Explainable and Scalable Planning with Probabilistic Temporal Logic Specifications

University of Texas at Austin

PI: Ufuk Topcu

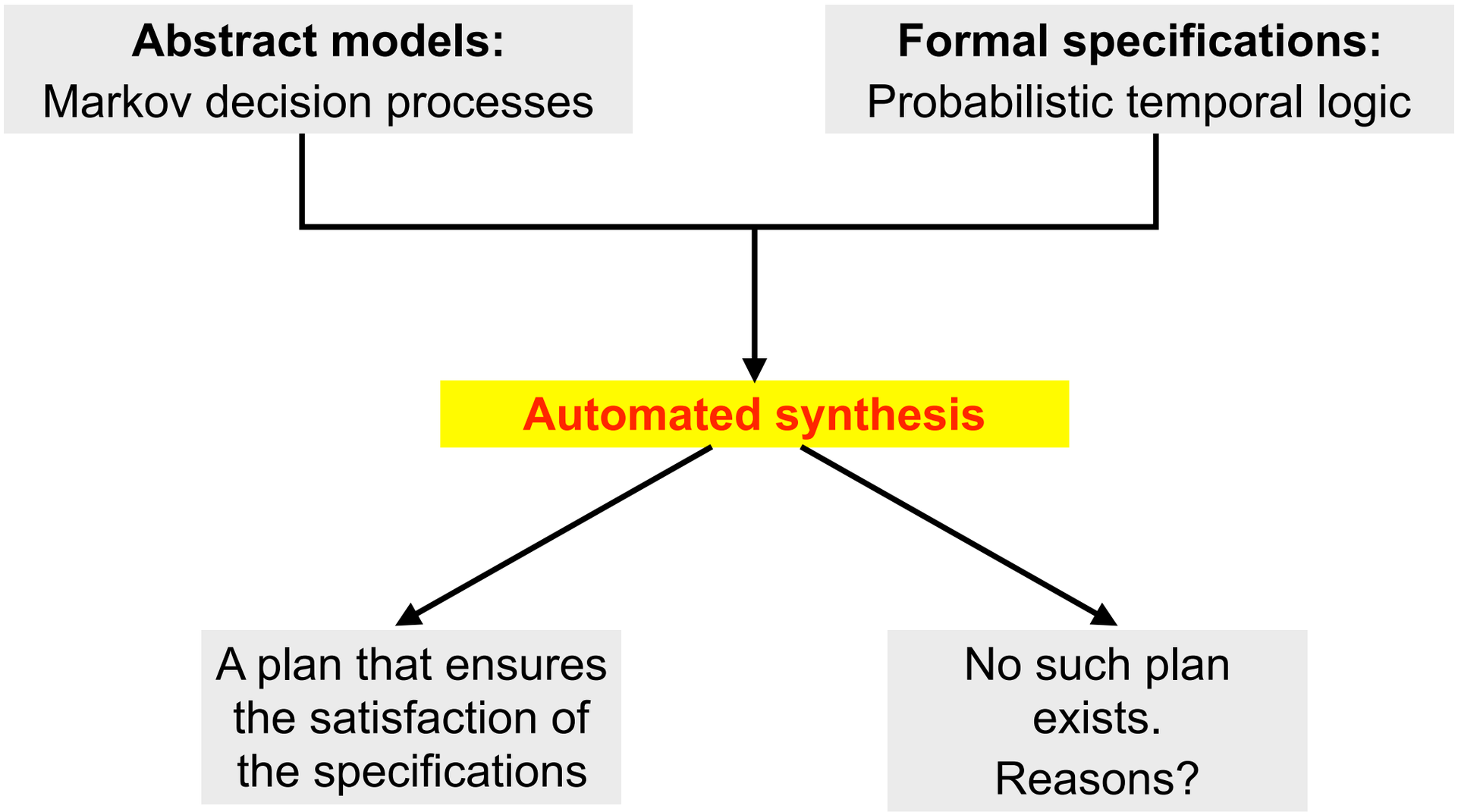
Postdoc: Nils Jansen

Graduate students: Murat Cubuktepe, Mahsa Ghasemi

u-t-autonomous.info

Overview of the approach

(formally specify and automatically synthesize plans)



Outline and main contributions

(and publications in the first year)

Overview of planning in uncertain Markov decision processes

Planning in parametric Markov decision processes subject to temporal logic specifications (mainly Thrust I)

- Convex-optimization-based sequential methods with convergence guarantees
- Orders of magnitude more scalable compared to conventional methods

"Sequential Convex Programming for the Efficient Verification of Parametric MDPs" appeared in TACAS 2017.

Explainable feedback from planning in Markov decision processes (mainly Thrust III)

- Structured counterexamples in Markov decision processes
- Minimal and sound explanations in natural-like languages

"Counterexamples for Robotic Planning Explained in Structured Natural Language" submitted to ICRA 2018.

Plans for the next year

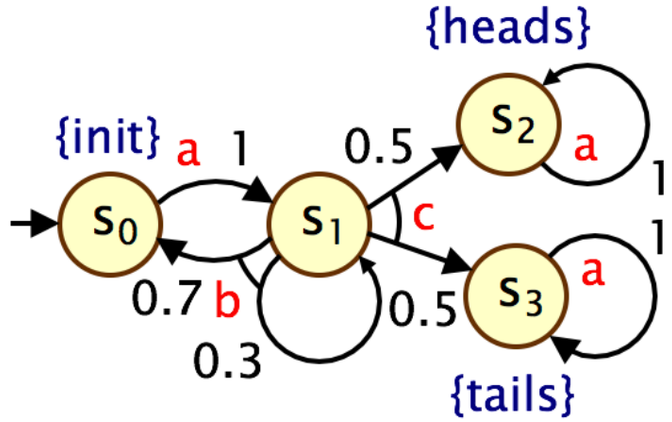
Markov decision processes (MDPs)

Many systems and processes are inherently probabilistic

- Unreliable behavior of components (and people)
- Unpredictable evolution of events
- Communication losses

An **MDP** M is a tuple M = (S, A, P, s_0, AP, L), where

- S is a finite set of states,
- A is a finite set of action,
- P:S×A×S→[0,1] is the transition probability function,
- s_0 is the initial state,
- AP is a finite set of atomic propositions, and
- L : S → 2^{AP} is a labeling function.



Specifying behavior with temporal logic

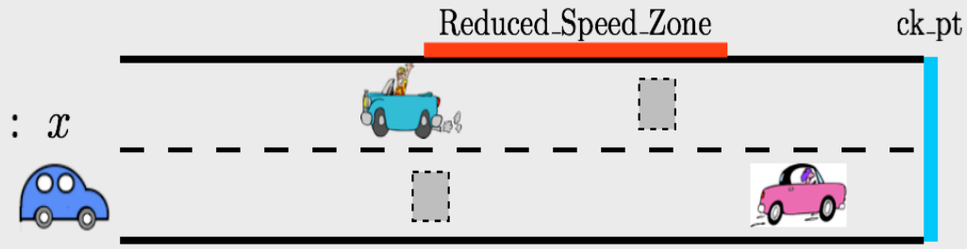
Linear Temporal Logic (LTL) = **Propositional Logic** + **Temporal Operators**

$\wedge$ (and)
$\vee$ (or)
$\rightarrow$ (implies)
$\neg$ (not)

$\Diamond$ (eventually)
$\Box$ (always)
$\mathcal{U}$ (until)

- Reason about infinite sequences $\sigma = s_0 s_1 s_2 \ldots$ of states
- Many different dialects of temporal logic (with probabilistic and epistemic modalities)
- Specify safe, allowable, required, or desired behavior of system and/or environment.

Traffic rules:
- No collision $\Box\,(\mathrm{dist}(x, \mathrm{Obs}) \geq X_{\mathrm{safe}} \wedge \mathrm{dist}(x, \mathrm{Loc}(\mathrm{Veh})) \geq X_{\mathrm{safe}})$
- Obey speed limits $\Box\,((x \in \mathrm{Reduced\_Speed\_Zone}) \rightarrow (v \leq v_{\mathrm{reduced}}))$
- Stay in travel lane unless blocked
- Intersection precedence & merging, stop line, passing,...



Goals:
- Eventually visit the check point $\Diamond(x = \mathrm{ck\_pt})$
- Every time check point is reached, eventually come to start

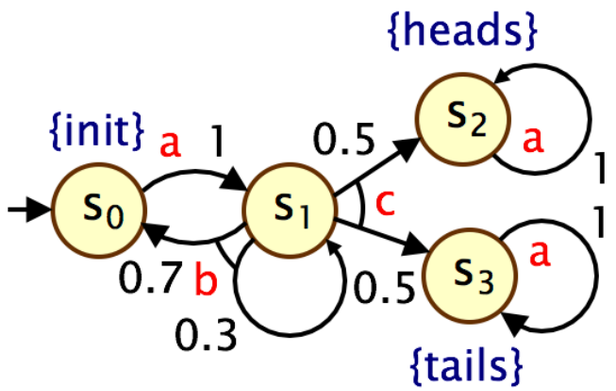
$$\Box((x = \mathrm{ck\_pt}) \rightarrow \Diamond(x = \mathrm{start}))$$

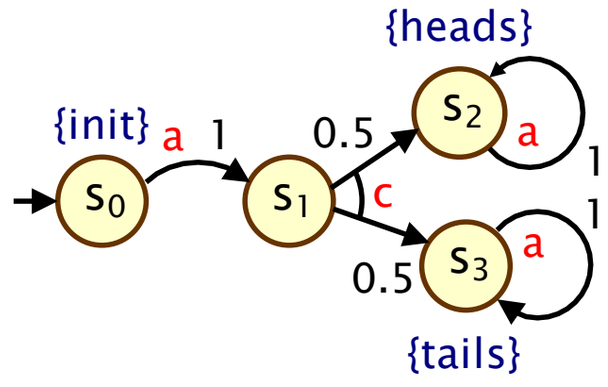
Environment assumptions:
- Each intersection is clear infinitely often $\Box\Diamond(\mathrm{Intersection} = \mathrm{empty})$
- Limited sensing range, detect obstacles before too late,...

Probabilistic satisfaction of temporal logic specifications

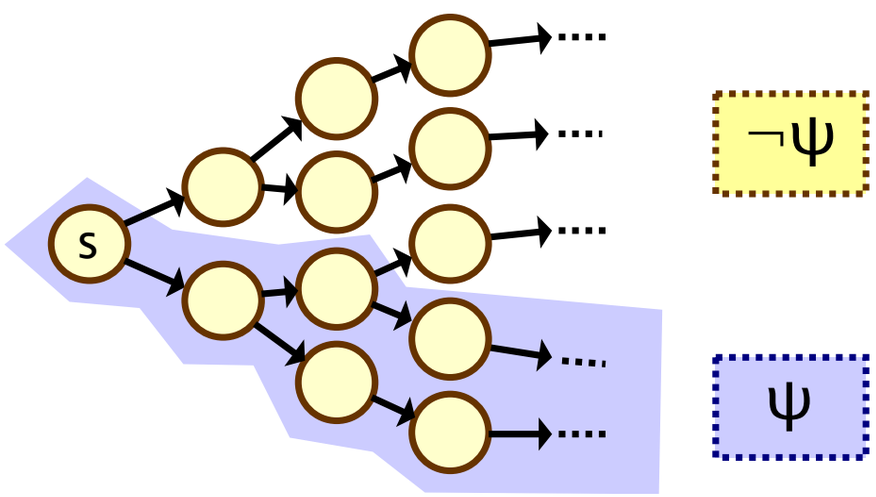
Policy: π: S → A (other names: plan, scheduler, strategy,…)

MDP





A policy induces a Markov chain



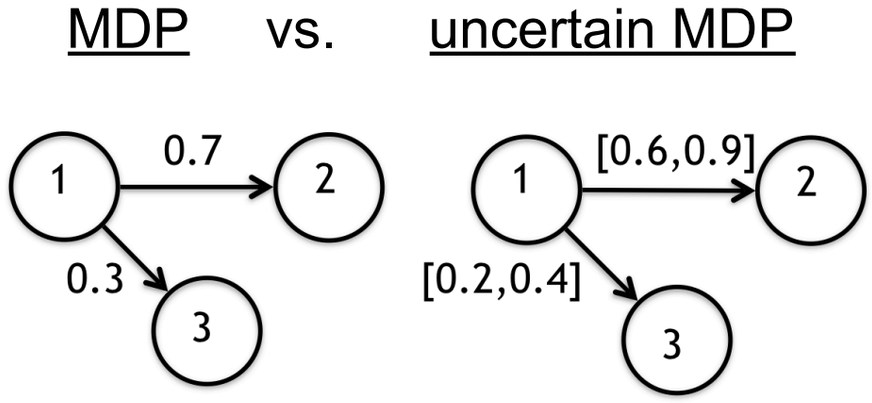
$\mathrm{Prob}(s, \psi)$

mass of executions that satisfy the specification

Some interesting challenges in planning with MDPs

(related to planning for human space missions)

Probabilities are hard to obtain precisely

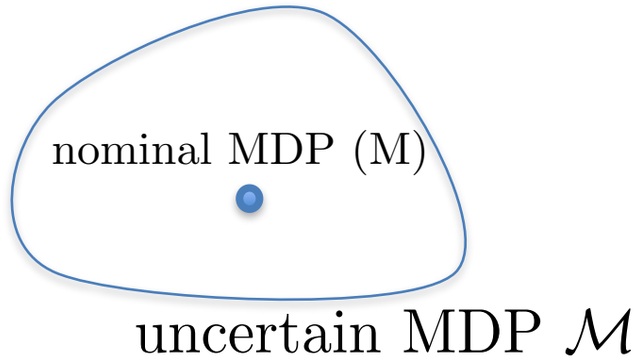


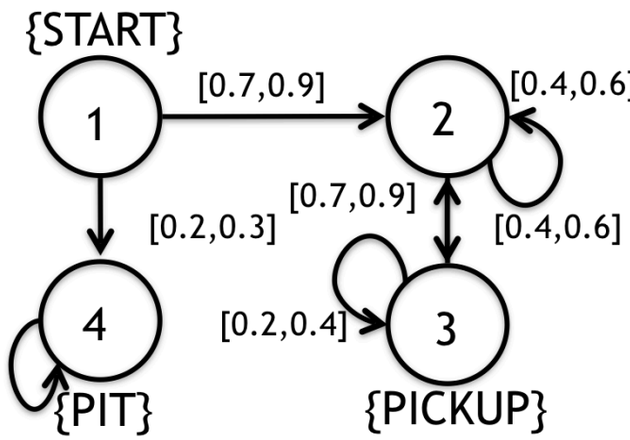
Need to scale problems with hundreds of tasks

Need planning artifacts explainable to the crew and planners

Robust policies in uncertain MDPs

- **Given:**
  - Uncertain MDP with initial state s_0
  - Temporal logic specification ϕ





Task:
Repeatedly PICKUP and always avoid PIT

- **Problem**: Compute a policy π* that maximizes the worst-case (over all viable transition functions) probability of satisfying ϕ:

$$\pi^* = \arg \max_{\pi} \min_{M \in \mathcal{M}} \text{Prob}^{\pi, M}(s_0 \models \phi)$$

Solution overview

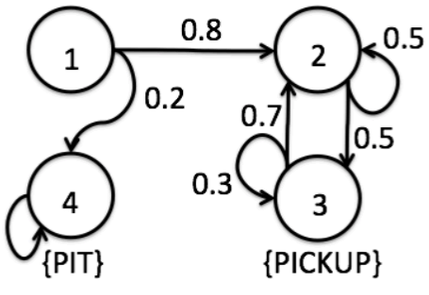
1. Specification ϕ → deterministic Rabin automaton A_ϕ
2. Create product MDP $M_p = M \times A_\phi$
3. Compute winning set in M_p
4. Compute control policy to maximize probability of reaching winning set (dynamic programming)
5. Project policy back to the original MDP M

Set from which ϕ is satisfied with certainty.

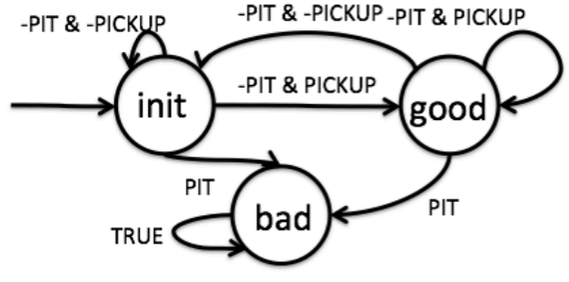
ε-approximate solution to the fixed-point operation

$$TV(s) = \max_{a \in A(s)} \left[r(s,a) + \min_{p \in \mathcal{P}_s^a} p^T V \right]$$

$M_p =$



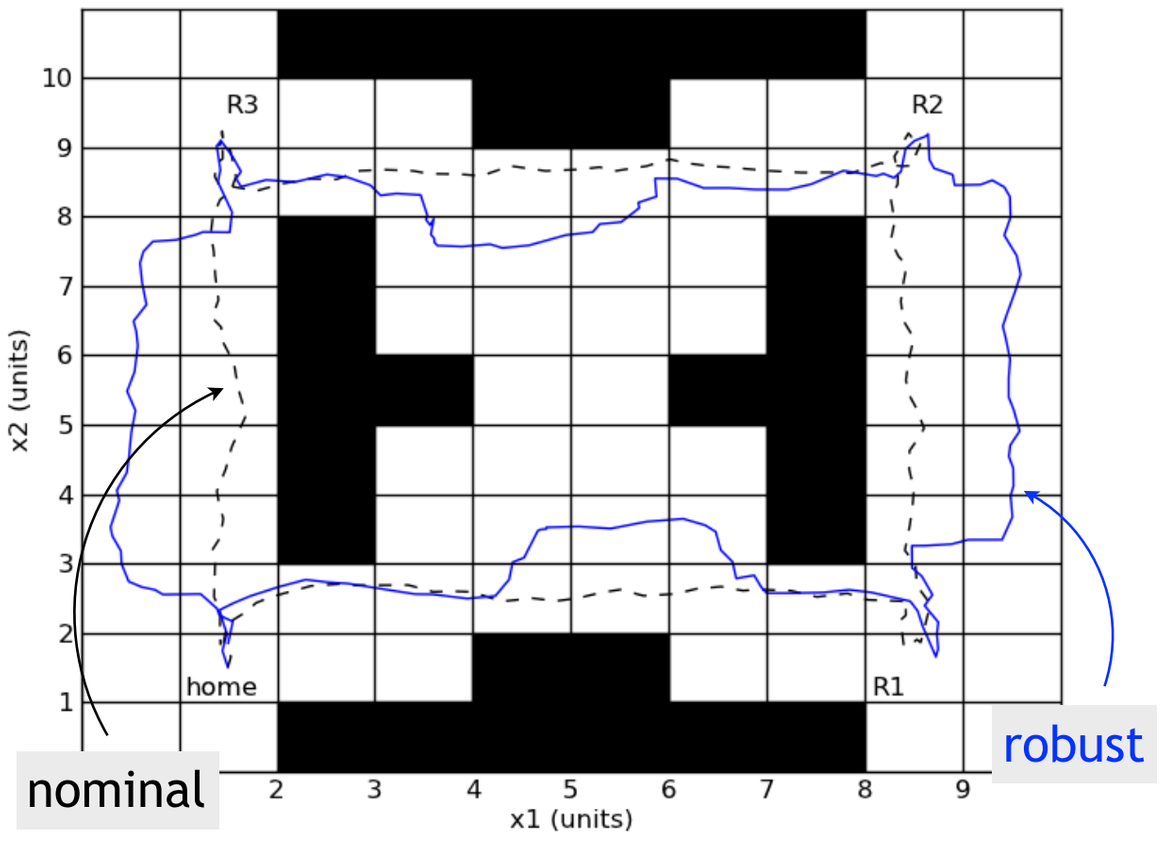
X

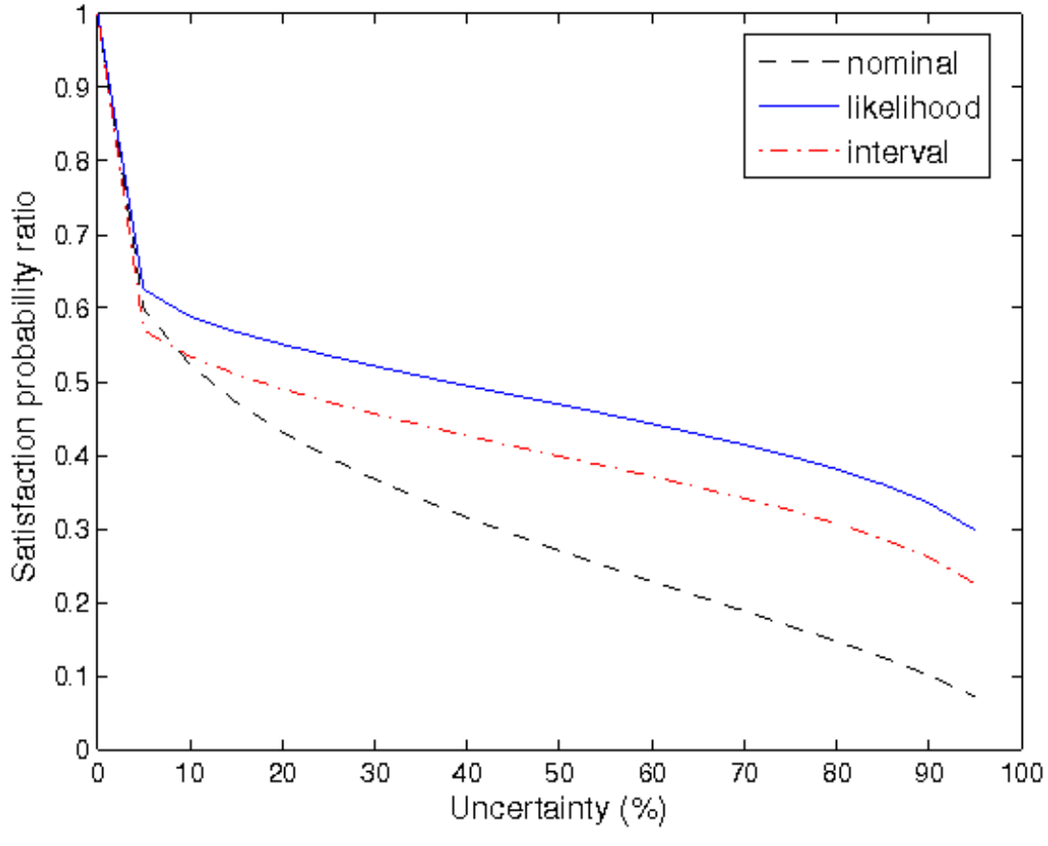


“Repeatedly PICKUP and always avoid PIT”

An example: Accounting for uncertainties in probabilities matters

Informal task: Start + end at HOME. Avoid OBSTACLES. Visit R1, R2, R3.

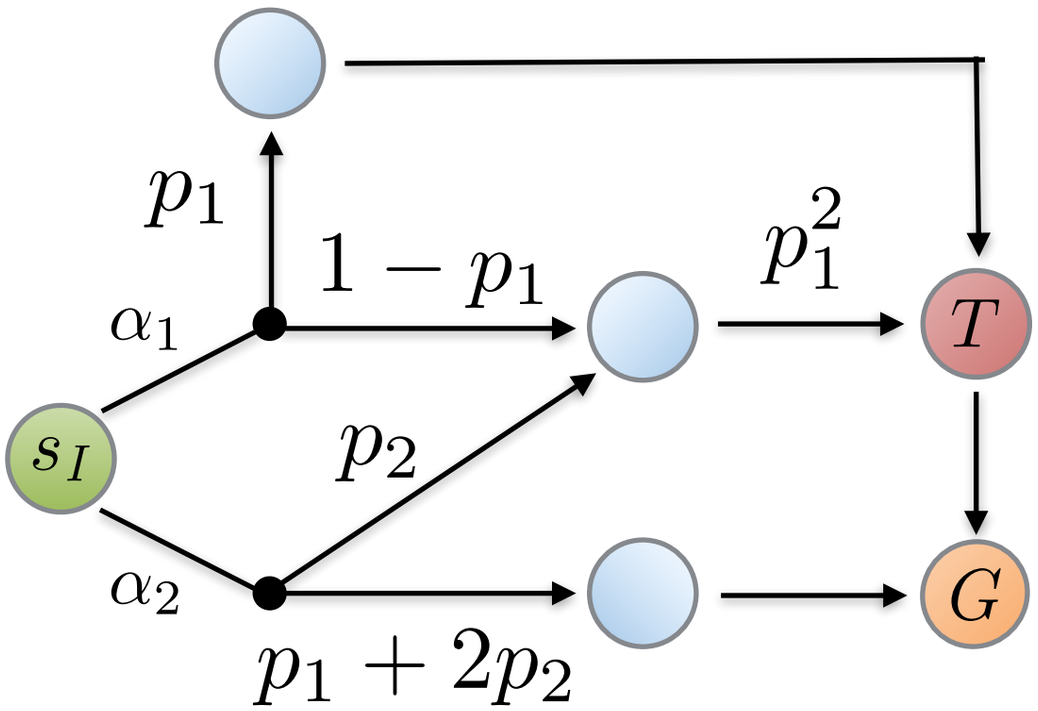




Computation time:

- nominal (0.47 sec)
- robust (5.7 sec)

Parameter synthesis in parametric MDPs (pMDPs)



Safety specification

$$\varphi = \mathbb{P}_{\leq\lambda}(\Diamond T),\ T \subseteq S$$

Performance specification

$$\psi = \mathbb{E}_{\leq\kappa}(\Diamond G),\ G \subseteq S$$

Objective function $f : V \to \mathbb{R}$

Parameters $p_1, p_2, \ldots, p_n \in V$

Given pMDP $\mathcal{M}$, find a well-defined valuation of parameters and a scheduler $\sigma \in Sched^{\mathcal{M}}$ such that

$$\mathcal{M}^\sigma \models \varphi \wedge \psi$$

and value for objective function $f : V \to \mathbb{R}$ is minimal.

An application: model repair

Say: Concrete MDP, minimal probability to reach T is 0.25

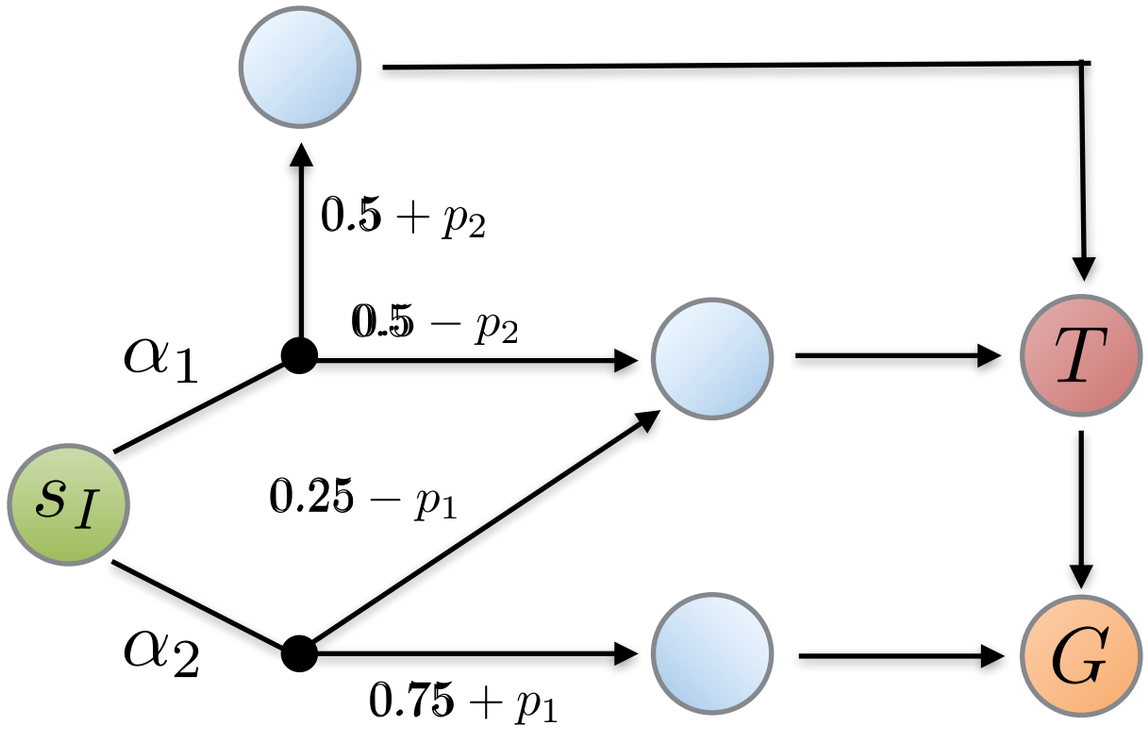
Repair the MDP such that the minimal probability is 0.2

Approach:

Introduce parameters to adjust transition probabilities

Find parameter valuation such that cost is minimal

Cost function: $p_1^2 + p_2^2$



a parametric MDP

Conventional solution based on nonlinear programming

$$\begin{aligned} &\text{minimize} && f \\ &\text{subject to} \end{aligned}$$

objective function over parameters

$$\begin{aligned} p_{s_I} &\leq \lambda \\ c_{s_I} &\leq \kappa \end{aligned}$$

safety and performance specifications

$$\forall s \in S. \quad \sum_{\alpha \in Act(s)} \sigma^{s,\alpha} = 1$$

$$\forall s \in S \ \ \forall \alpha \in Act(s). \quad \sum_{s' \in S} \mathcal{P}(s,\alpha,s') = 1$$

well-defined schedulers and parameter instantiations

$$\forall s \in T. \quad p_s = 1$$

$$\forall s \in S \setminus T. \quad p_s = \sum_{\alpha \in Act(s)} \sigma^{s,\alpha} \cdot \sum_{s' \in S} \mathcal{P}(s,\alpha,s') \cdot p_{s'}$$

safety probability computation

$$\forall s \in G. \quad c_s = 0$$

$$\forall s \in S \setminus G. \quad c_s = \sum_{\alpha \in Act(s)} \sigma^{s,\alpha} \cdot \left(c(s,\alpha) + \sum_{s' \in S} \mathcal{P}(s,\alpha,s') \cdot c_{s'} \right)$$

expected performance computation

Problem variables:

Randomized scheduler:

$\{\sigma^{s,\alpha} \mid s \in S, \alpha \in Act(s)\}$

Probability of reaching T:

$\{p_s \mid s \in S\}$

Expected cost of reaching G:

$\{c_s \mid s \in S\}$

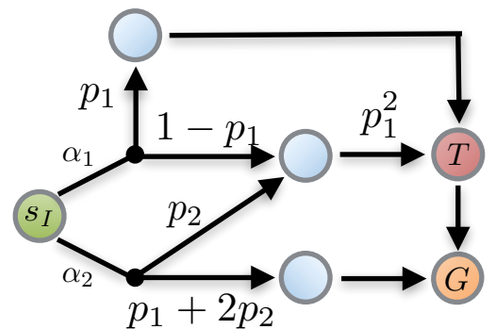
A useful observation

non-negative-valued variable

signomials

non-negative-valued variable

$$p_s = \sum_{\alpha \in Act(s)} \sigma^{s,\alpha} \cdot \sum_{s' \in S} \mathcal{P}(s, \alpha, s') \cdot p_{s'}$$



$$f = \sum_{k=1}^{K} c_k \cdot x_1^{a_{1k}} \cdots x_n^{a_{nk}}$$

strictly positive

real

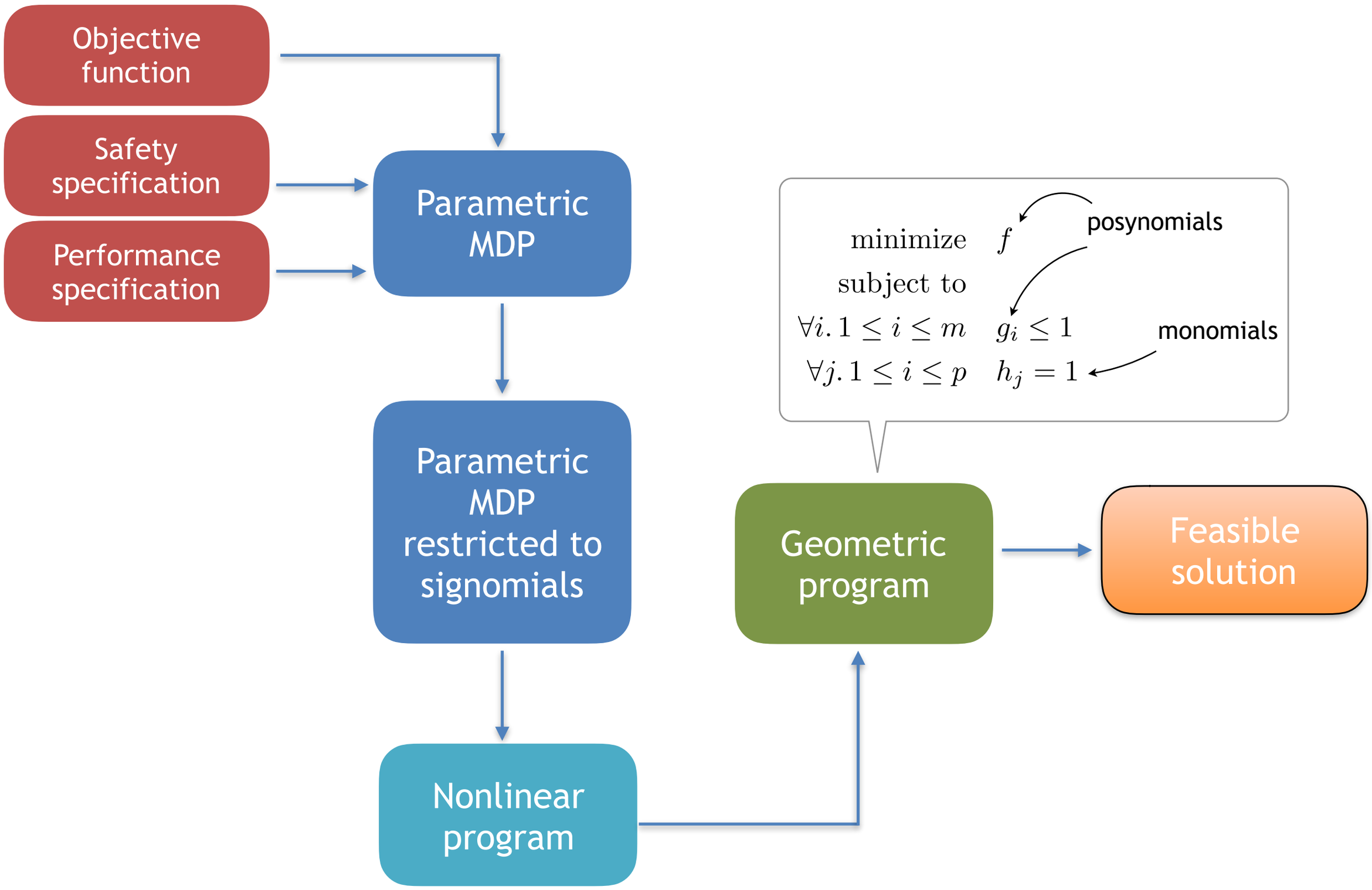
monomial

> 0: **posynomial**

no restriction: **signomial**

Question: Can we somehow exploit this structure and solve the parameter synthesis problem as a convex optimization problem (maybe bunch of them)?

Workflow


Objective function
Safety specification
Performance specification
Parametric MDP
Parametric MDP restricted to signomials
Nonlinear program
Geometric program
Feasible solution
minimize f
subject to
$\forall i. 1 \leq i \leq m$ $g_i \leq 1$
$\forall j. 1 \leq i \leq p$ $h_j = 1$
posynomials
monomials

Convexification

$$p_s = \sum_{\alpha \in Act(s)} \sigma^{s,\alpha} \cdot \sum_{s' \in S} \mathcal{P}(s, \alpha, s') \cdot p_{s'}$$

relaxation

upper bound on actual probability

$$p_s \geq \sum_{\alpha \in Act(s)} \sigma^{s,\alpha} \cdot \sum_{s' \in S} \mathcal{P}(s, \alpha, s') \cdot p_{s'}$$

division transformation

still signomials (not geometric program yet)

$$\frac{\sum_{\alpha \in Act(s)} \sigma^{s,\alpha} \cdot \sum_{s' \in S} \mathcal{P}(s, \alpha, s') \cdot p_{s'}}{p_s} \leq 1$$

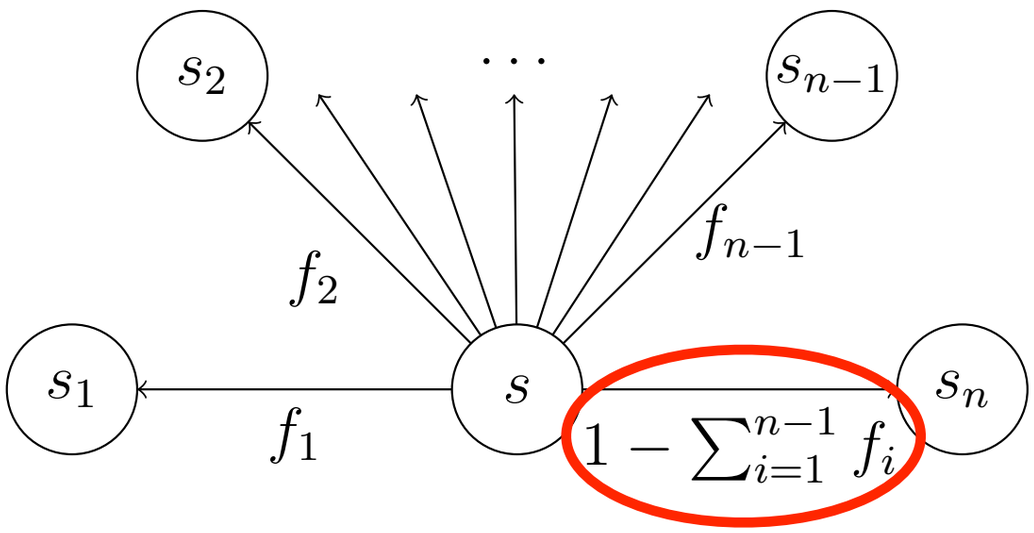
Convexification

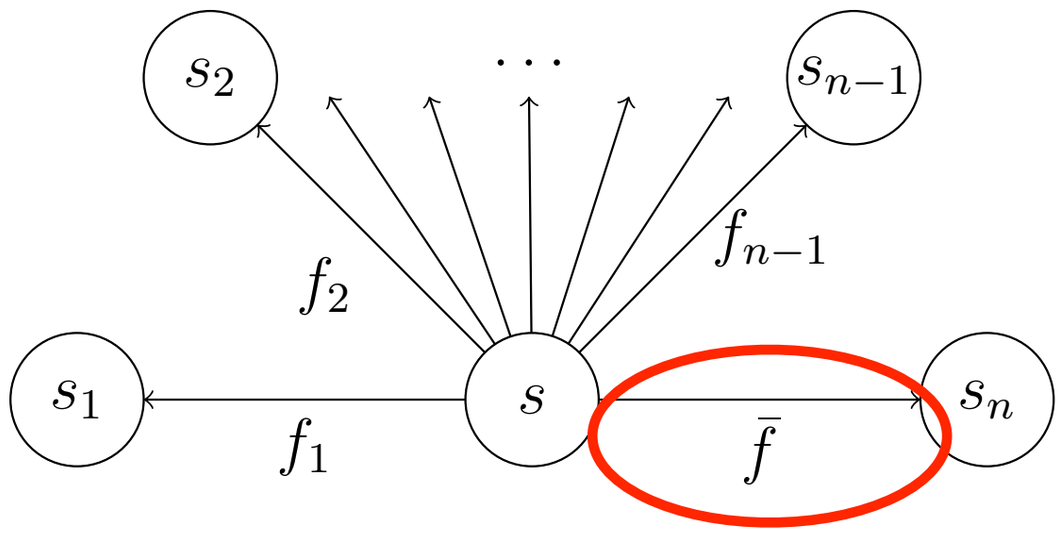
$$\mathcal{P}(s,\alpha,\bar{s}) = 1 - \sum_{s' \in S \setminus \{\bar{s}\}} \mathcal{P}(s,\alpha,s') \Longrightarrow \bar{\mathcal{P}}(s,\alpha,\bar{s}) = \bar{p}_{s,\alpha,\bar{s}} \in L$$

signomial

posynomial

lifting variable





Geometric program (with relaxation tightening)

$$\text{minimize} \quad \sum_{p \in V} \frac{1}{p} + \sum_{\bar{p} \in L} \frac{1}{\bar{p}} + \sum_{s \in S, \alpha \in Act(s)} \frac{1}{\sigma_{s,\alpha}}$$

regularization

$$\text{subject to}$$

$$\frac{p_{s_I}}{\lambda} \leq 1$$

$$\frac{c_{s_I}}{\kappa} \leq 1$$

$$\forall s \in S. \quad \sum_{\alpha \in Act(s)} \sigma^{s,\alpha} \leq 1$$

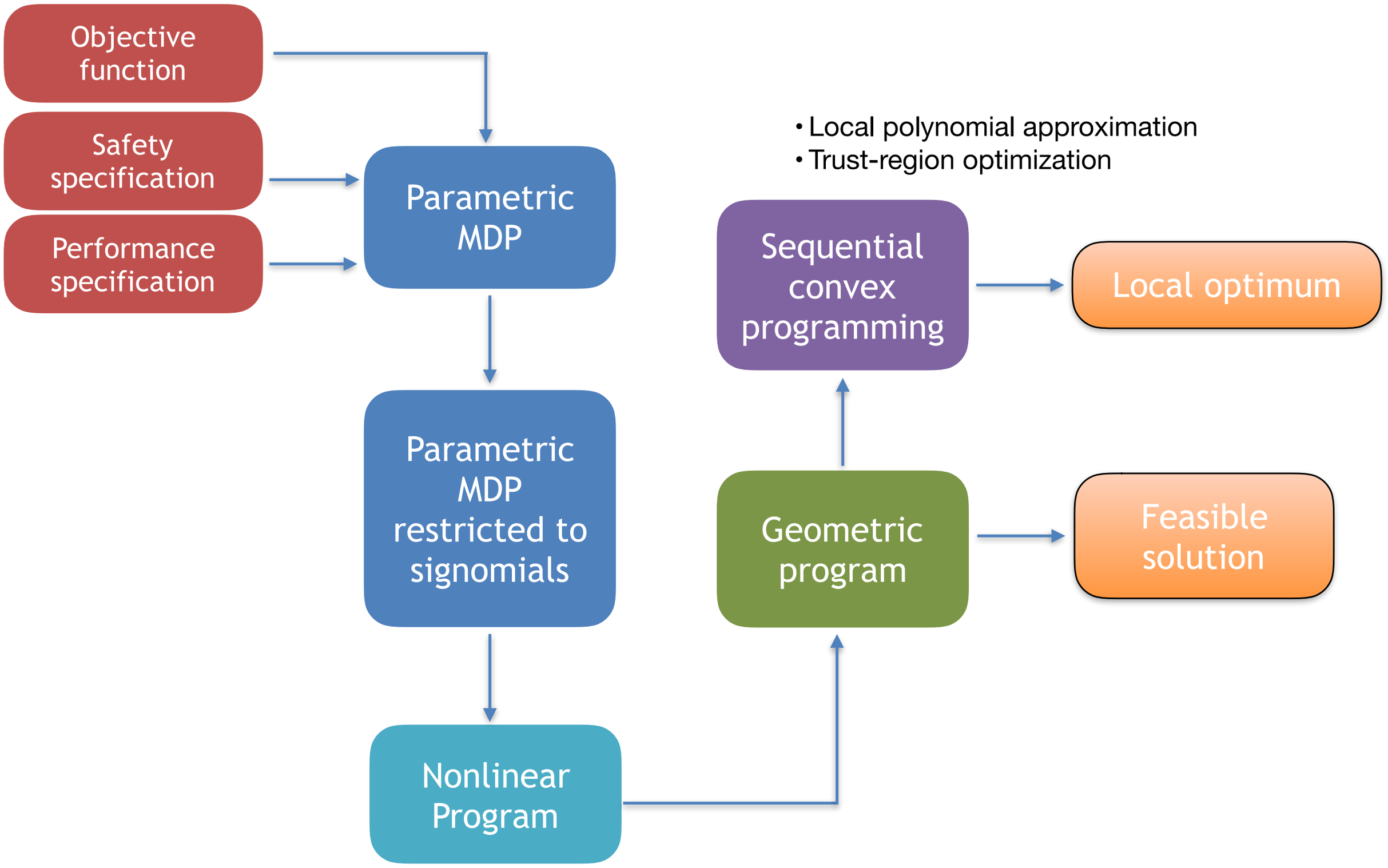
$$\forall s \in S \, \forall \alpha \in Act(s). \quad \sum_{s' \in S} \mathcal{P}(s, \alpha, s') \leq 1$$

$$\forall s \in S \setminus T. \quad \frac{\sum_{\alpha \in Act(s)} \sigma^{s,\alpha} \cdot \sum_{s' \in S} \mathcal{P}(s, \alpha, s') \cdot p_{s'}}{p_s} \leq 1$$

$$\forall s \in S \setminus G. \quad \frac{\sum_{\alpha \in Act(s)} \sigma^{s,\alpha} \cdot \left(c(s, \alpha) + \sum_{s' \in S} \mathcal{P}(s, \alpha, s') \cdot c_{s'} \right)}{c_s} \leq 1$$

Theorem: The solution to the geometric program gives a well-defined scheduler and parameter instantiation. But it may be sub-optimal.

Workflow


Objective function
Safety specification
Performance specification
Parametric MDP
Parametric MDP restricted to signomials
Nonlinear Program
• Local polynomial approximation
• Trust-region optimization
Sequential convex programming
Local optimum
Geometric program
Feasible solution

Numerical experiments

Alternative tools for optimization TO even in the smallest instances

Benchmark	#states	#par	specs	proposed method MOSEK (s)		only feasibility Z3
BRP (pMC)	5382	2	EC, $\mathbb{P}$, $*$	23.17	(6.48)	–
	112646	2	EC, $\mathbb{P}$, $*$	3541.59	(463.74)	–
	112646	4	EC, $\mathbb{P}$, $*$	4173.33	(568.79)	–
	5382	2	EC, $\mathbb{P}$	3.61		904.11
	112646	2	EC, $\mathbb{P}$	479.08		*TO*
NAND (pMC)	4122	2	EC, $\mathbb{P}$, $*$	14.67	(2.51)	–
	35122	2	EC, $\mathbb{P}$, $*$	1182.41	(95.19)	–
	4122	2	EC, $\mathbb{P}$	1.25		1.14
	35122	2	EC, $\mathbb{P}$	106.40		11.49
BRP (pMDP)	5466	2	EC, $\mathbb{P}$, $*$	31.04	(8.11)	–
	112846	2	EC, $\mathbb{P}$, $*$	4319.16	(512.20)	–
	5466	2	EC, $\mathbb{P}$	4.93		1174.20
	112846	2	EC, $\mathbb{P}$	711.50		*TO*
CONS (pMDP)	4112	2	EC, $\mathbb{P}$, $*$	102.93	(1.14)	–
	65552	2	EC, $\mathbb{P}$, $*$	*TO*		–
	4112	2	EC, $\mathbb{P}$	6.13		*TO*
	65552	2	EC, $\mathbb{P}$	1361.96		*TO*

Outline and main contributions

(and publications in the first year)

Overview of planning in uncertain Markov decision processes

Planning in parametric Markov decision processes subject to temporal logic specifications (mainly Thrust I)

- Convex-optimization-based sequential methods with convergence guarantees
- Orders of magnitude more scalable compared to conventional methods

"Sequential Convex Programming for the Efficient Verification of Parametric MDPs" appeared in TACAS 2017.

Explainable feedback from planning in Markov decision processes (mainly Thrust III)

- Structured counterexamples in Markov decision processes
- Minimal and sound explanations in natural-like languages

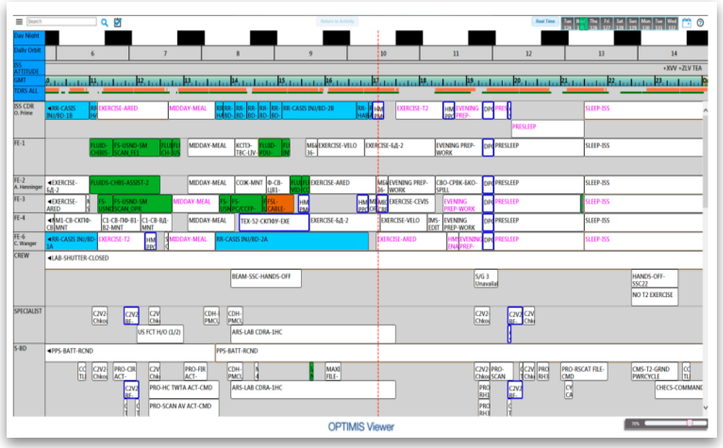
"Counterexamples for Robotic Planning Explained in Structured Natural Language" submitted to ICRA 2018.

Plans for the next year

Explainable planning artifacts

Human operator or planner

user interface



Automated mission planning

Can the human operator understand the planning artifacts?

- Plans
- Counterexamples

Compute counterexamples that can be understood by “humans”

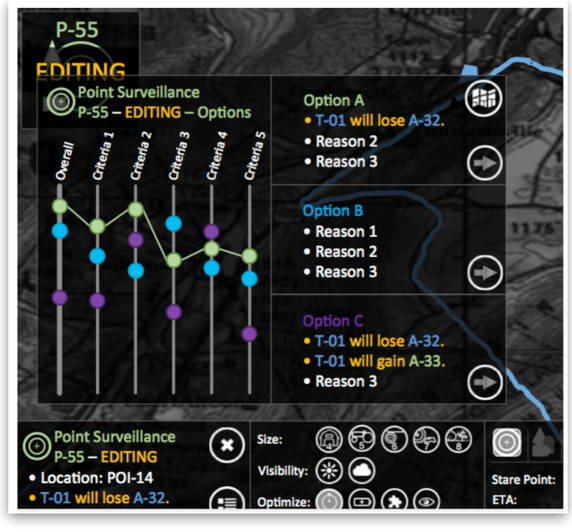
- Uses the same alphabet and grammar with humans
- Respects the limitations (expressivity, bandwidth, etc.) of the interface

Structured Counterexamples

Process algebra statements to create missions out of “plays”:

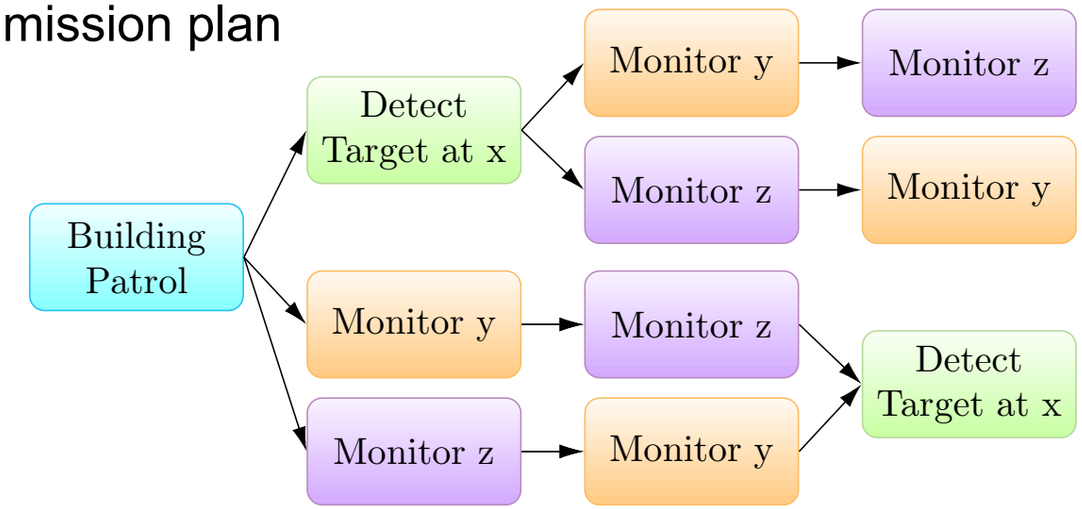
“Building Patrol” · (“Detect Target at x” + (“Monitor y” || Monitor z”))

sequential alternative interleaving

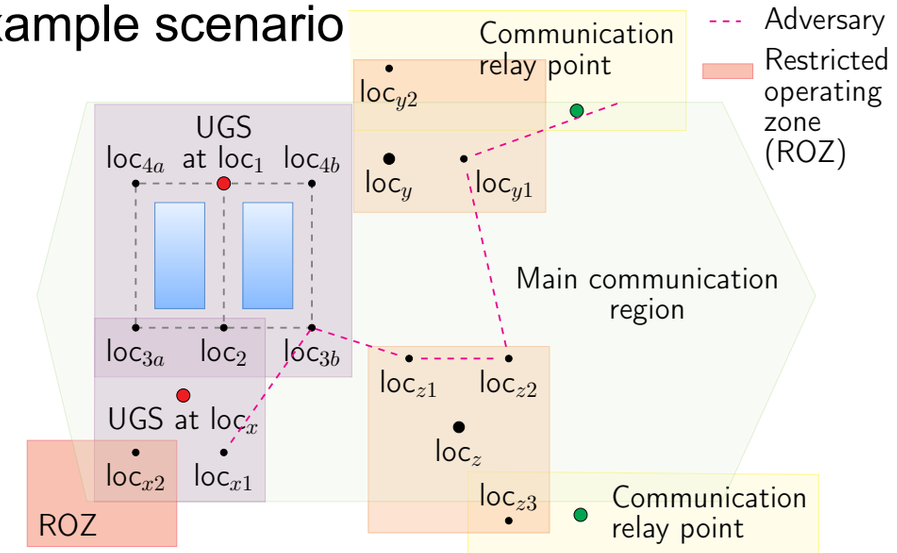


A UAV operator interface (Air Force)

A mission plan



An example scenario

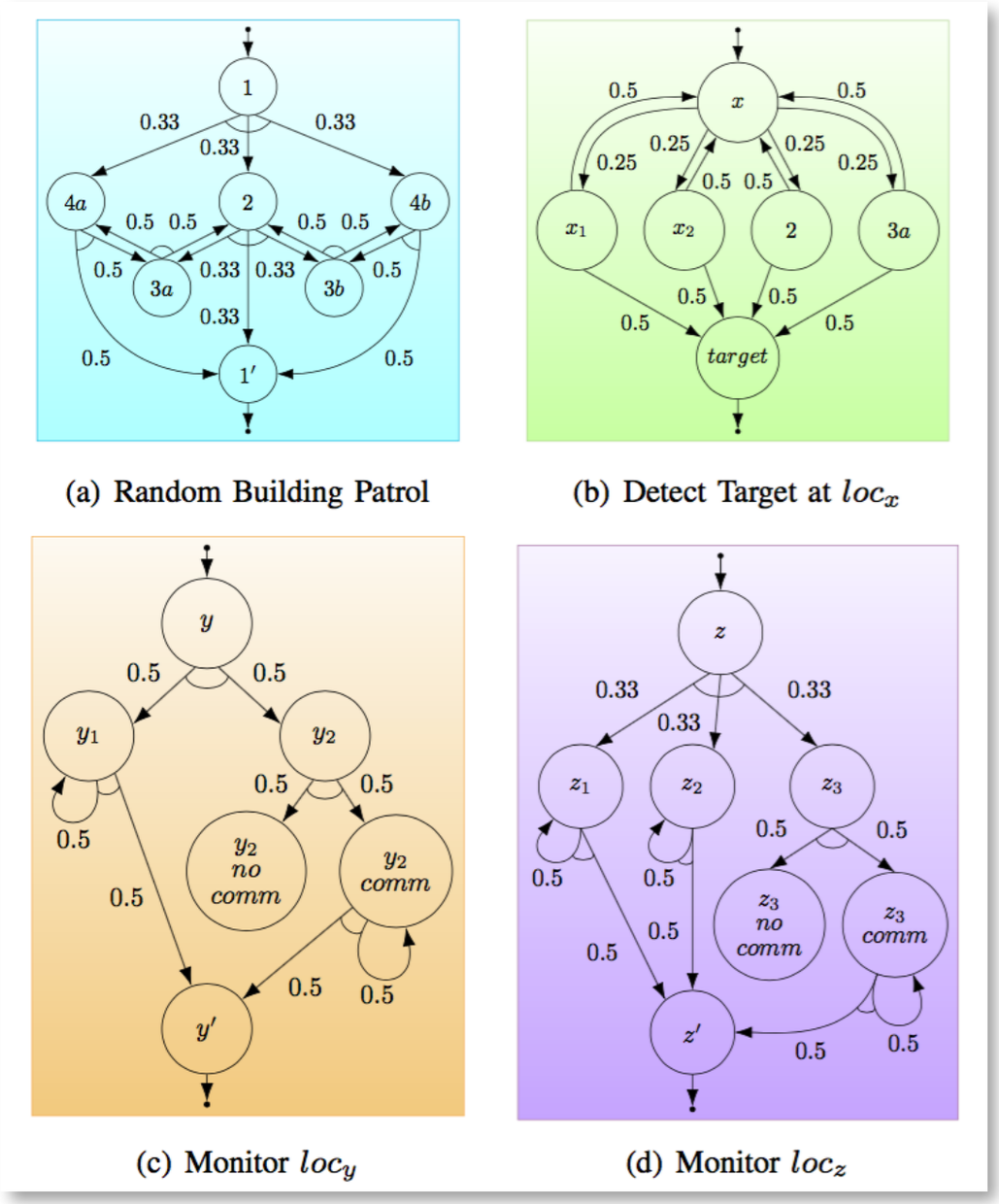


Example structures:

- Minimal number of plays
- Temporal or logical relations at the play level

(An) Abstraction of Plays

- Model each play as a discrete-time Markov chain with special entrance and exit conditions, where probabilistic distributions are used to represent uncertainties in system behavior
- Compose plays into a Markov decision process (MDP), where the nondeterminism is introduced through the alternative and interleaving operators



Problem statement: Find a subsystem of the MDP that violates the probabilistic specifications and involves a minimal number of plays

Counterexamples with minimal number of plays

as a mixed integer linear program

$$\text{minimize} \sum_{1 \le i \le n} \omega_i$$

Binary variables indicate if a state partition is included in the counterexample

$$\text{such that}$$

$$p_{\overline{s}} > \lambda$$

The probabilistic reachability property is violated

T: set of target states

$$\forall s \in T, \text{ for } s \in \Omega_i : \; p_s = \omega_i$$

X: set of exit states

$$\forall s \in S \setminus T, \text{ for } s \in \Omega_i : \; p_s \le \omega_i$$

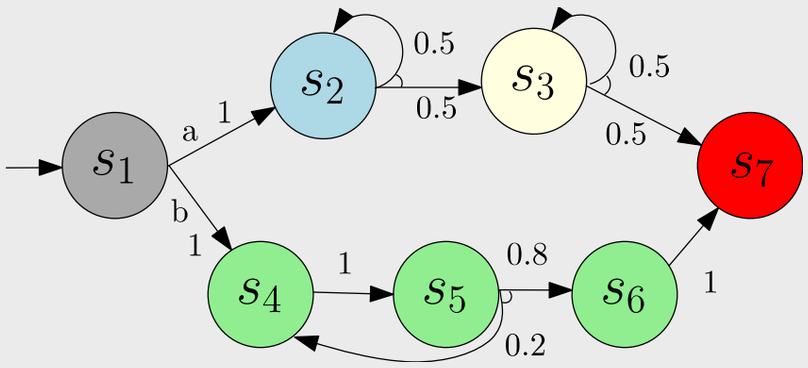
$$\forall s \in S \setminus (T \cup X) : \; p_s \le \sum_{s' \in \mathsf{succ}(s,\tau)} P(s,\tau,s') \cdot p_{s'}$$

only one action is chosen at exit states

$$\forall s \in X \setminus T, a \in \alpha : \; p_s \le (1 - \theta_{s,a}) + \sum_{s' \in \mathsf{succ}(s,a)} P(s,a,s') \cdot p_{s'}$$

$$\forall s \in X, \text{ for } s \in \Omega_i : \; \sum_{a \in \alpha} \theta_{s,a} = \omega_i$$

Intuition: encoding MDP transition probabilities



Property: $\mathcal{P}_{\le 0.1}(\mathsf{F}\ s_7)$

Structured counterexample:

$\text{grey}(s_1) \rightarrow \text{green}(s_4, s_5, s_6) \rightarrow \text{red}(s_7)$

Sample results on the UAV example

The probability of entering ROZ should be smaller than λ

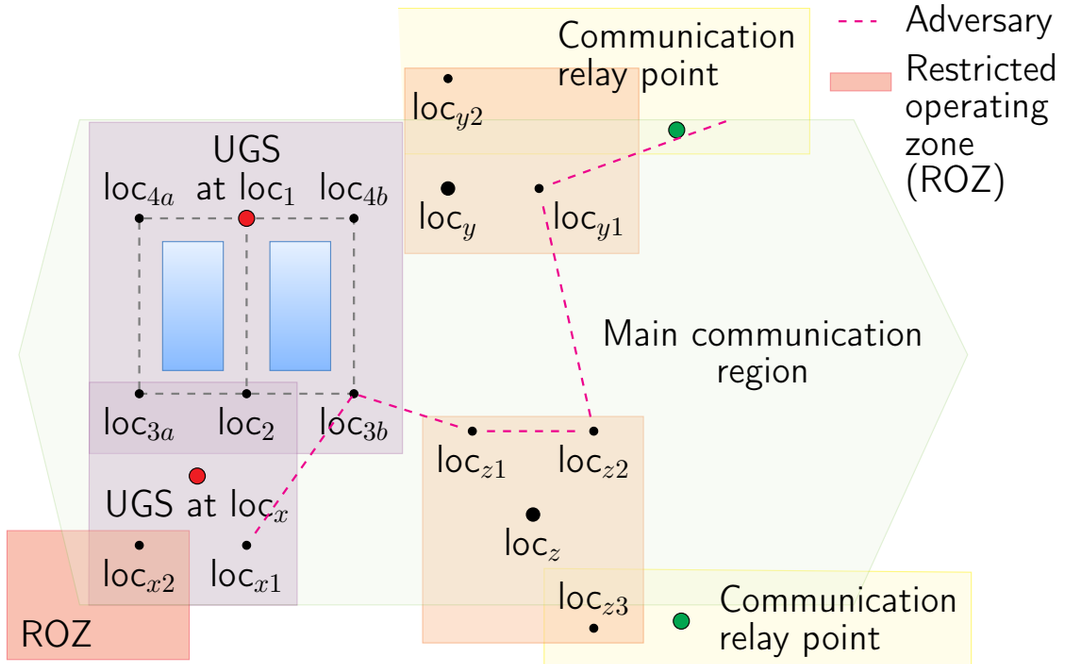
- λ=0.4, “Building Patrol” → “Detect Target at x”
- λ=0.5, no counterexample

The probability of losing communication should < λ

- λ=0.2, “Building Patrol” → “Monitor y”
- λ=0.3, “Building Patrol” → “Monitor y” → “Monitor z”
- λ=0.4, no counterexample

The probability of being detected by adversary should < λ

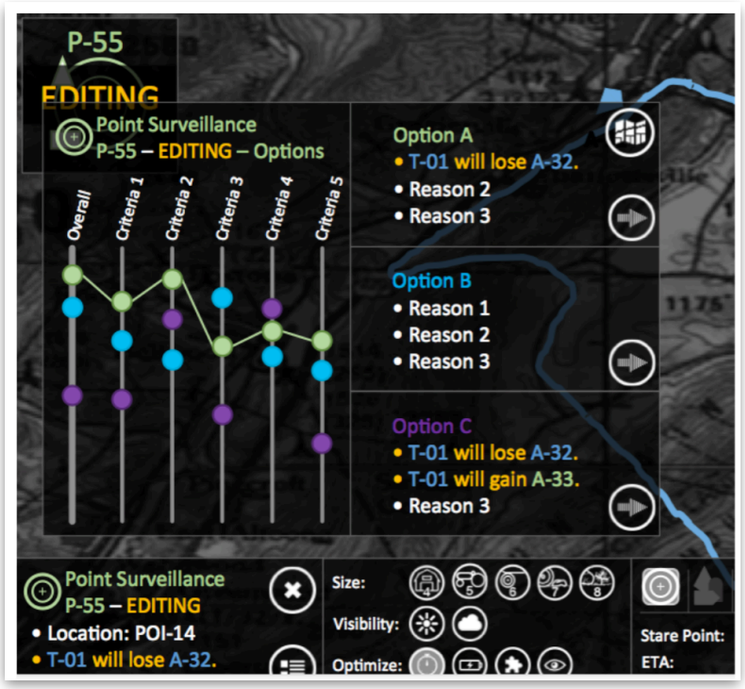
- λ=0.2, “Building Patrol”
- λ=0.4, “Building Patrol” → “Monitor y”
- λ=0.8, “Building Patrol” → “Monitor z” → “Monitor y”
- λ=0.9, no counterexample



Counterexamples with explanations in structured natural language

(structure: The robot ⟨action⟩ when ⟨proposition⟩.)

Image source: AFRL

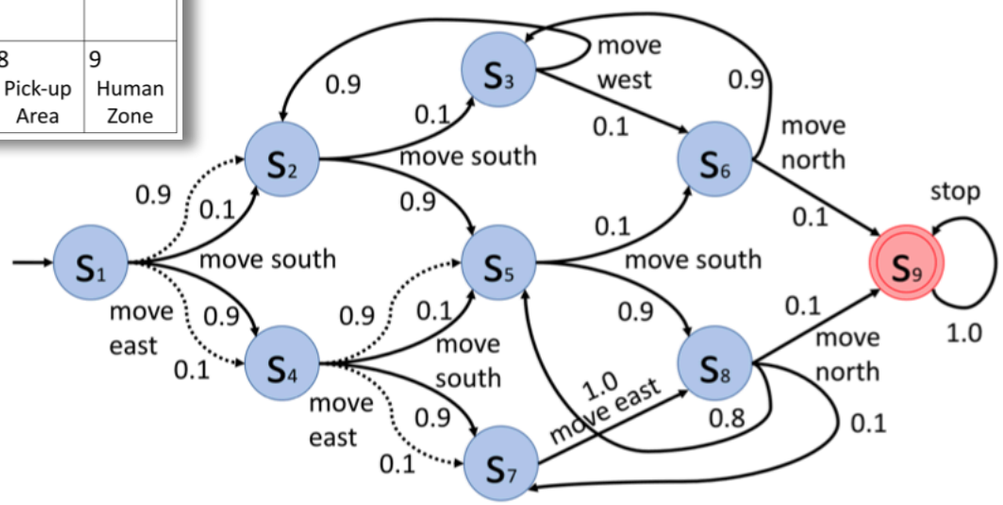


Counterexamples with explanations in structured natural language

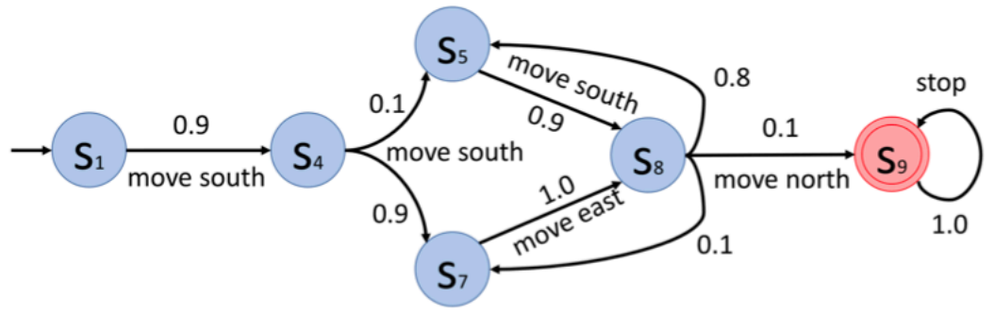
(structure: The robot ⟨action⟩ when ⟨proposition⟩.)

1 Charging Station	2	3 Delivery Area
4	5	6
7	8 Pick-up Area	9 Human Zone

An MDP for a planning problem



A counterexample



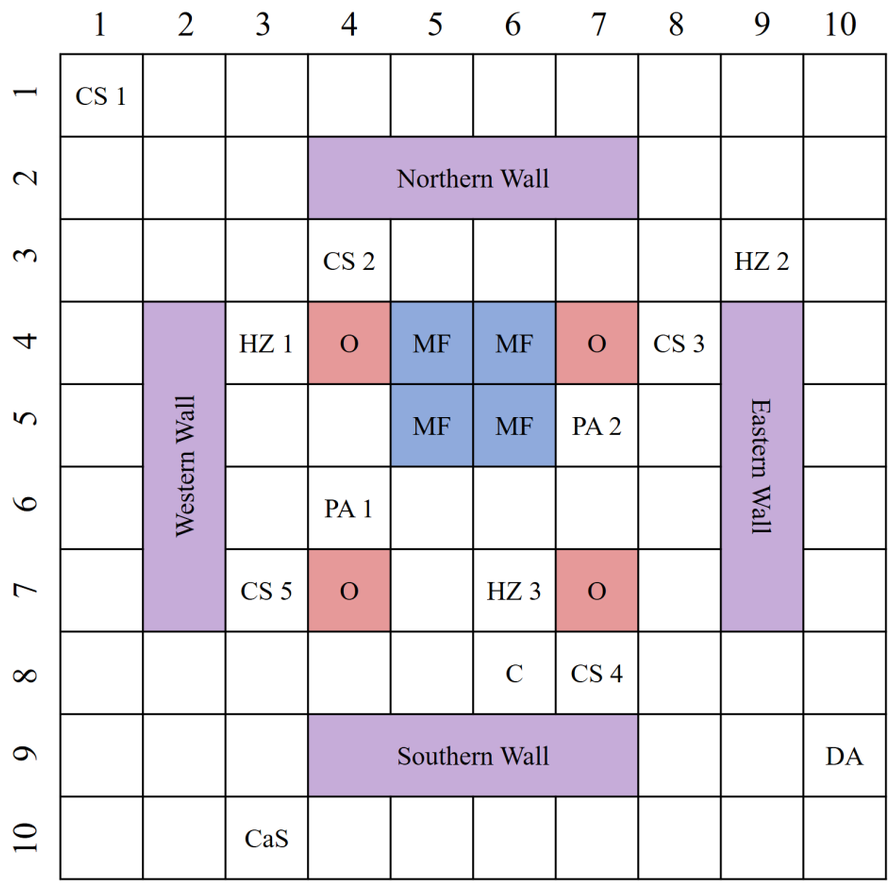
There exists an explanation with 4 sentences.

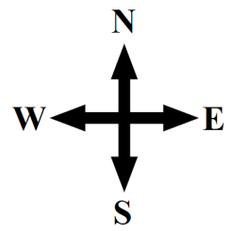
An explanation in structured natural language with 6 sentences

(S1) The robot moves south when in charging station.
(S2) The robot moves south when south of charging station.
(S3) The robot moves south when north of pick-up area.
(S4) The robot moves east when west of pick-up area.
(S5) The robot moves north when in pick-up area.
(S6) The robot stops when in human zone.

Sample results on explainable counterexamples

Planning for a warehouse robot





Legends

Obstacle	O
Charging Station	CS
Checkpoint	C
Pick-up Area	PA
Delivery Area	DA
Human Zone	HZ
Calibration Station	CaS
Magnetic Field	MF

			alternative method		proposed method		
N	# States	# Transitions	# States	Time (s)	# States	# Sentences	Time (s)
10	100	208	9	0.11	9	3	1.39
20	400	788	19	**24.86**	39	3	**4.43**
30	900	1,768	–	**time-out**	29	3	**1.54**
40	1,600	3,148	–	**time-out**	79	3	**17.45**
50	2,500	4,928	–	**time-out**	99	3	**32.33**

Accomplishment, critique and plans

Accomplishments: Significant progress in Thrusts I and II in year 1

- Scalability in planning in uncertain and parametric Markov decision processes with probabilistic temporal logic specifications
- Minimal and sound natural-language-like explanations of core failure reasons in plans in stochastic environments

Critique:

- Progress in demonstrations on a case study slower than expected
- Delayed the work on compositional synthesis in Thrust I

Recently obtained sample constraints

Delay due to unexpected progress in Thrust I based on alternative approaches

Plans for year 2 (and 3):

- Thrust I: Further scalability from compositional algorithms. The results from year 1 are particularly suitable for distributed optimization.
- Thrust II: Interpretability through sparsity in model ”repair”.
- Thrust III: Extend the explanations to richer natural-like language structures. Explanations for plans (in addition to counterexamples).
- Validation: Incorporate the representative constraints and preferences recently obtained from NASA into case studies in every thrust.

Application to ISS-like mission planning

Heterogeneous constraints

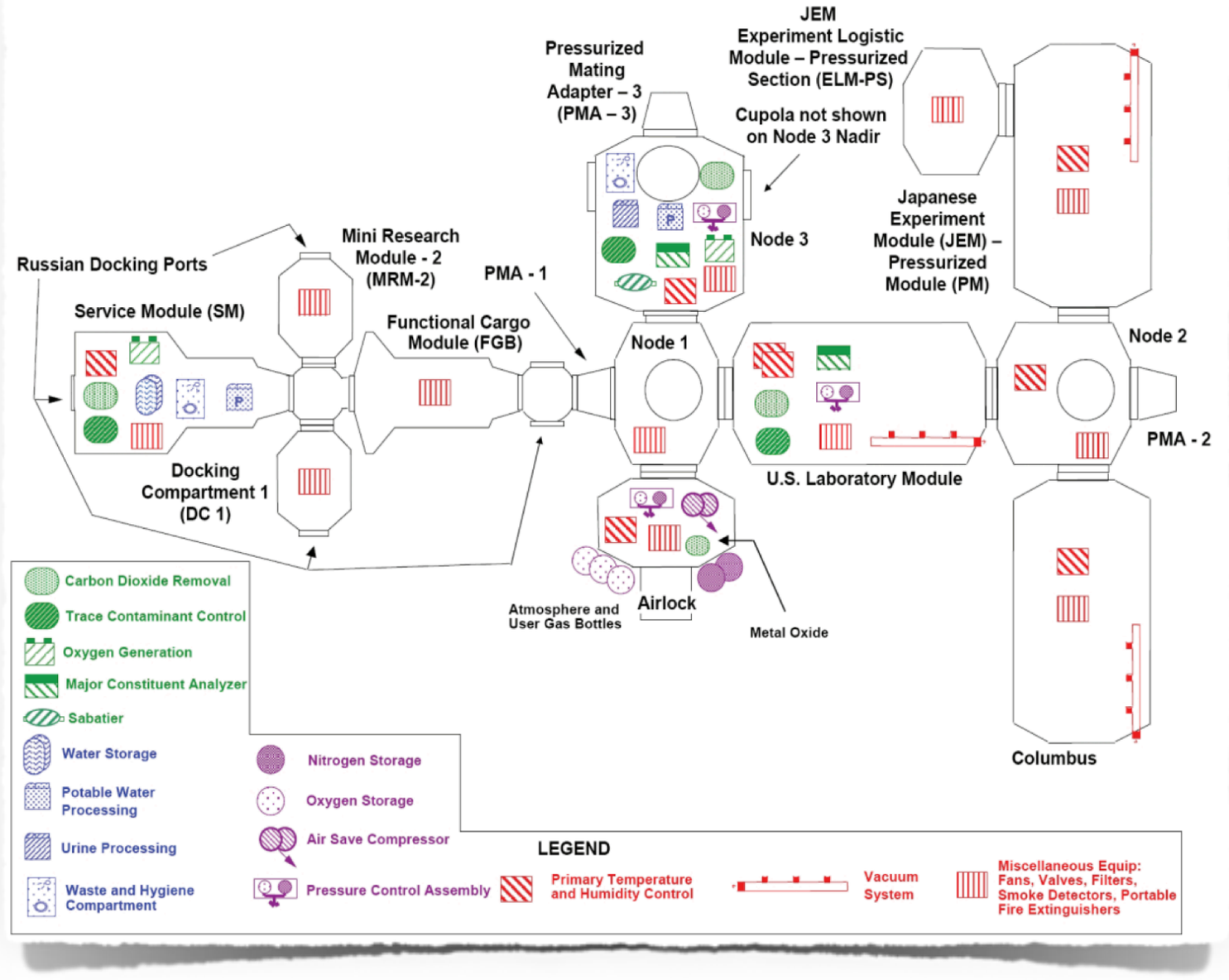
- Hard (flight rules) + soft (crew preferences)
- MaxSAT-based formulations may be suitable
- Connections to stochastic environments are unexplored

Co-existence of stochasticity and nondeterminism

- To incorporate worst-case uncertainties due to lack of probabilistic knowledge

Resource (time, energy, etc.) constraints

- Identify the right (i.e., minimal yet useful) level of fidelity



Additional required information

Relation to other funded research

- The PI has other projects in the general area of temporal-logic-constrained control protocol synthesis.
- On the other hand, the main problems and most of the approaches in the current project are unique.
- The PI leverages the communality in the basics by exposing multiple students to the problems in the NASA-funded effort.

Interactions with NASA

- Visit and seminar at NASA Ames.
- Collaboration with Masahiro Ono from JPL.

Major activities and milestones

- Expect to follow the timeline from the original proposal with the major research components as outlined in the previous slide.
- Visit(s) and seminar at NASA Centers
- Submission of papers for publication: Upcoming “windows of submission” in Jan-Mar 2018
- Presentation of the results at conferences and workshops