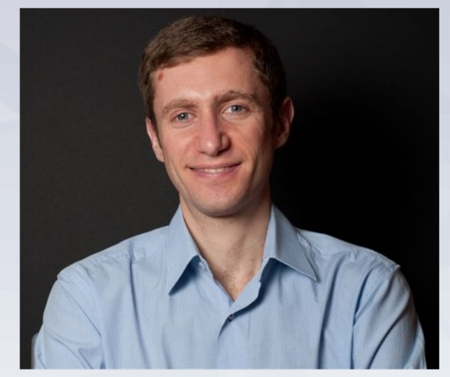
Bayes-Adaptive Search for Data-Efficient Verification of Parametric MDPs

Viraj B. Wijesuriya[1], Alessandro Abate[2]

viraj.wijesuriya@cs.ox.ac.uk[1], alessandro.abate@cs.ox.ac.uk[2]

Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford, OX1 3QD



1. Introduction

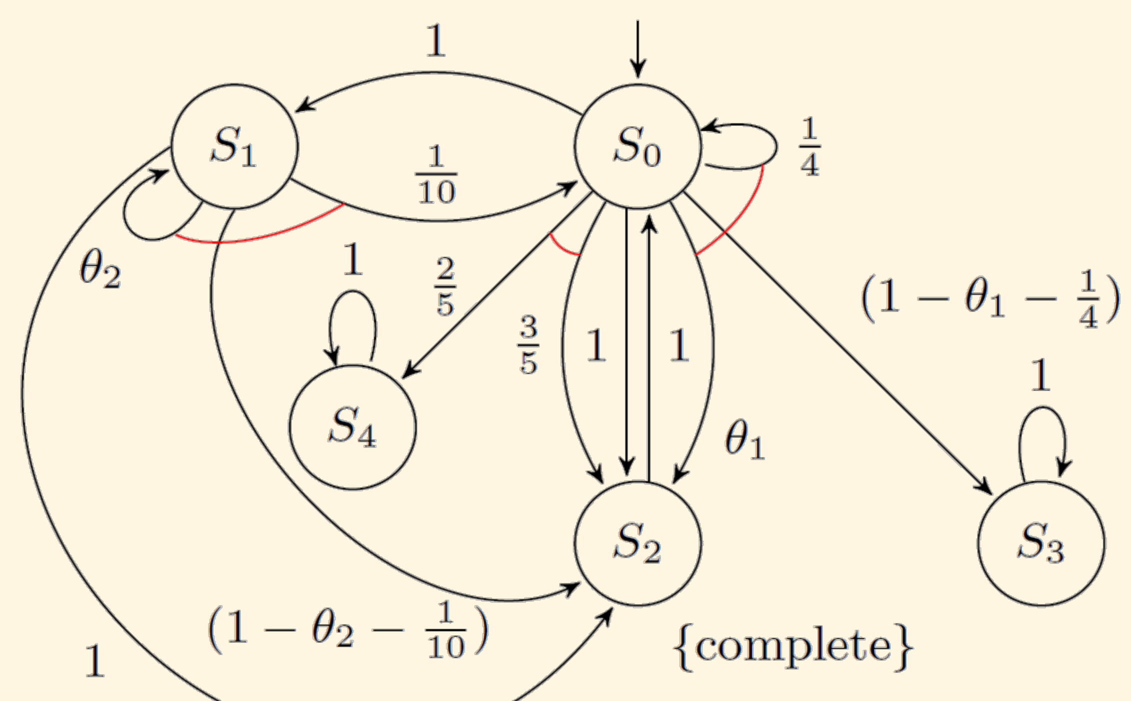
Problem: Formal verification relies on full access to accurate models. Such models are often hard to obtain for systems encompassing partially understood behaviours

Our goal: Verification of unbounded-time PCTL properties on partially unknown systems with actions

Our approach: Incorporate available information captured by a parameterised model with the active collection of limited data from the system

Parametric Markov Decision Process (pMDP)

$$\mathbf{M}_\Theta = (S, \mathrm{Act}, \mathbb{T}_\theta, \iota_{init}, \mathrm{AP}, L, \Theta)$$



2. Bayes-Adaptive Reinforcement Learning

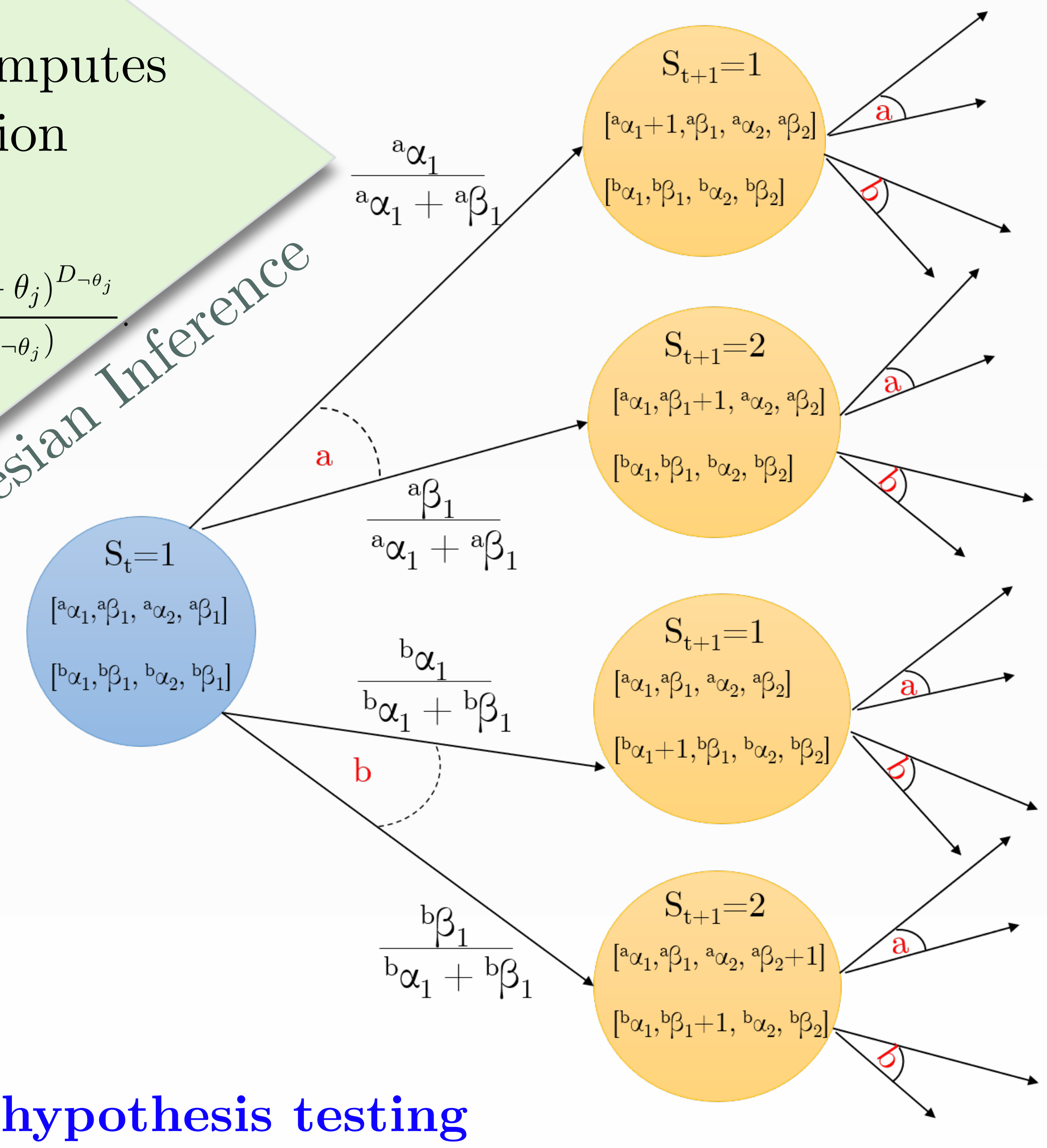
- Propose a method for selecting a strategy that computes the most accurate confidence for a finite amount of data

- Synthesise strategies on joint space of states and beliefs to optimise expected return $\hat{s}_t = p[\mathcal{P} \mid h_t]$

- Immediate confidence gain $\mathbb{G}_{t+1}$ after choosing action a_t at state s_t: $\mathbb{G}_{t+1} = |\mathcal{C}_{t+1} - \mathcal{C}_t|$

- We use an approximate bilinear Q-function:

$$Q(s, h, a; \beta) = \sigma(\mathbf{y}(h)^\top \beta \mathbf{x}(s, a))$$



Parameter Synthesis (1)

φ a desired PCTL system property

Use **parameter synthesis**[3] to induce models satisfying the property

$$\Theta_\varphi = \{\theta \in \Theta : M(\theta) \models \varphi\}$$

Strategy Synthesis (2)

Owing to [1], optimal strategy is memory-dependent

- Use Information State Search: **Bayes-Adaptive MDPs**

- Linear Temporal Difference Search [5] with Linear Value Function Approximation

Bayesian Inference (3)

Collect data from the underlying system

Bayesian inference computes a posterior distribution over parameters

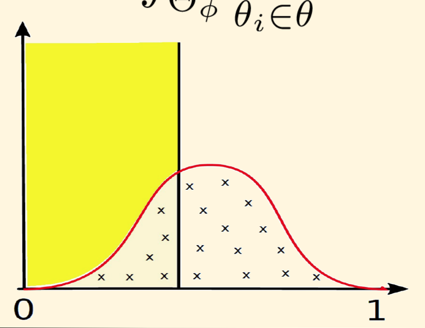
$$p(\theta_j \mid D) = \frac{\mathbb{P}(D \mid \theta_j)p(\theta_j)}{\mathbb{P}(D)} = \frac{p(\theta_j)\theta_j^{D_{\theta_j}}(1-\theta_j)^{D_{\neg\theta_j}}}{\mathbb{P}(D_{\theta_j}, D_{\neg\theta_j})}$$

(4) (5)

Bayesian Confidence

Integrate the posterior over Θ_φ to obtain the confidence

$$\mathcal{C} = \mathbb{P}(\mathbf{S} \models \phi \mid D) = \int_{\Theta_\phi} \prod_{\theta_i \in \theta} p(\theta_i \mid D_{\theta_i, \neg\theta_i}) d\theta,$$



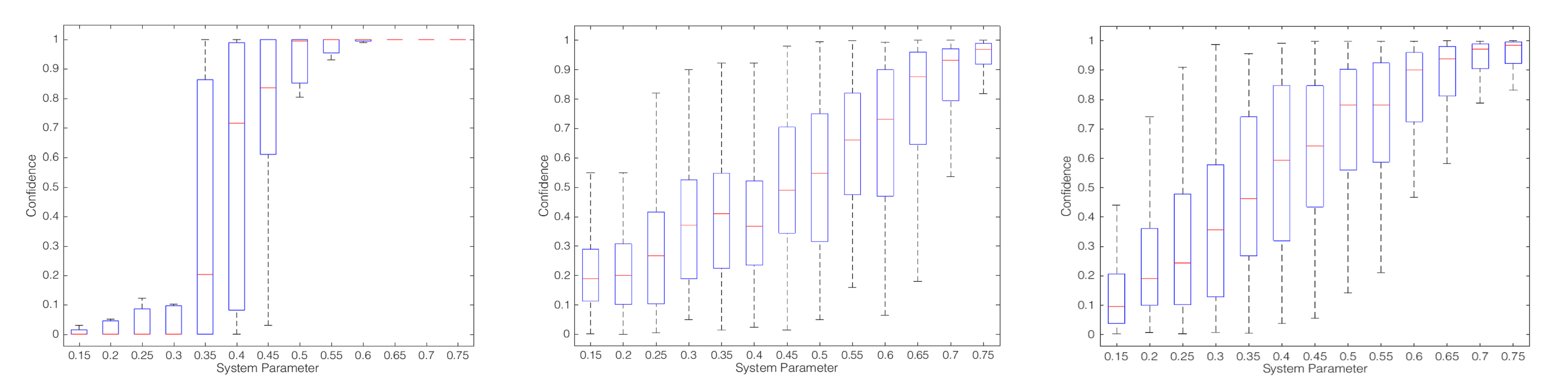
3. Results

Evaluate against:

- strategy that randomly selects actions at each state
- randomly selected memoryless strategy

Our synthesised strategy:

- is the Bayes-Optimal strategy
- improves the accuracy of the confidence calculation
- uses lower amounts of data to verify the property
- robust to limited amount of data with insufficient coverage



Synthesised strategy — Random static strategy — No strategy

4. Future Work

- Integration with **Bayesian hypothesis testing**
- Extension to **non-linearly parameterised pMDPs**
- Use alternative parameter synthesis tools to *allow* **more parameters**
- Achieve **faster runtimes**

[1] Gretton, C., Price, D., Thiebaux, S.: Implementation and comparison of solution methods for decision processes with non-Markovian rewards. In: UAI. pp. 289-296.
[2] Polgreen, E., Wijesuriya, V.B., Haesaert, S., Abate, A.: Data-efficient Bayesian verification of Parametric Markov Chains. In: QEST 2016.
[3] Kwiatkowska, M.Z., Norman, G., Parker, D.: PRISM 4.0: Verification of probabilistic real-time systems. In: CAV. Lecture Notes in Computer Science, Springer (2011).
[4] Polgreen, E., Wijesuriya, V.B., Haesaert, S., Abate, A.: Automated Experiment Design for Efficient Verification of Parametric Markov Decision Processes. In: QEST 2017.
[5] Silver, D., Sutton, R.S., Müller, M., : Temporal-difference search in computer Go. In Machine Learning. May 2012, Volume 87, Issue 2, pp. 183–219.

University of Oxford — Department of Computer Science — Oxford Control and Verification