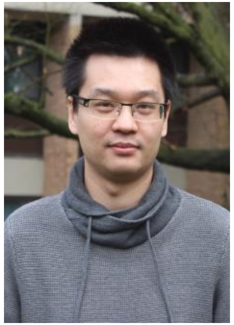
Data-driven abstractions

Oxymoron or panacea?

Raphael Jungers

ICTEAM, UCLouvain
On leave in Oxford U.

Joint work with



Zheming Wang
(Zhejiang Univ.)

Guillaume O. Berger
(UCBoulder)

Licio Romao
(Oxford)

Alessandro Abate
(Oxford)

Julien calbert

Adrien Banse

Data-driven control

Data-driven control

Highly complex systems
(black-box, third-party, remote, etc.)

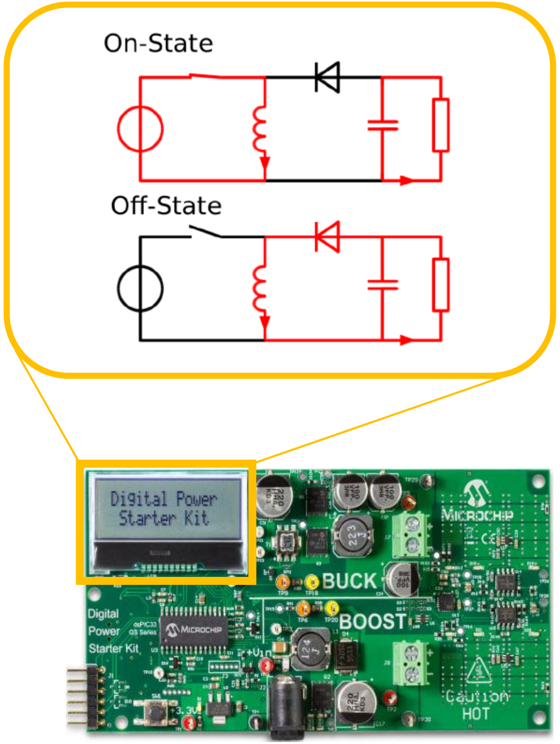
Lot of data
(sensors, simulations, automated tests, etc.)

⇒ **Data-driven analysis**

Data-driven control

Proprietary systems

Black-box cyber-physical systems



Human-in-the-loop

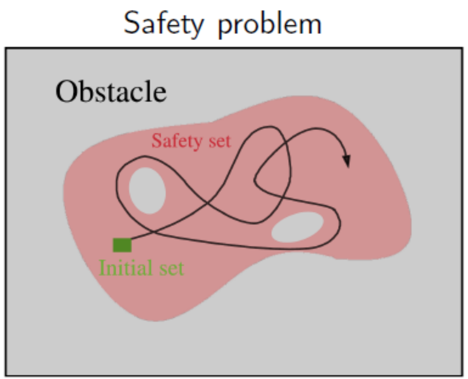
Networked multi-agent systems

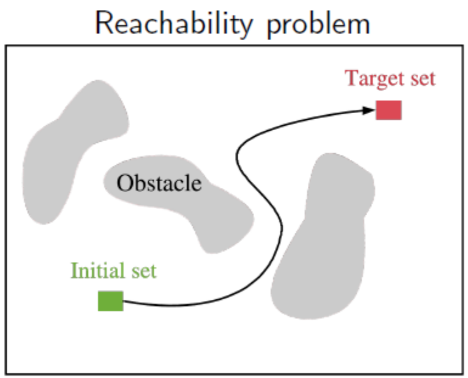
Opinion dynamics (flocking, consensus, etc.) over **hidden** switching networks

Cooperation of open multi-agent systems

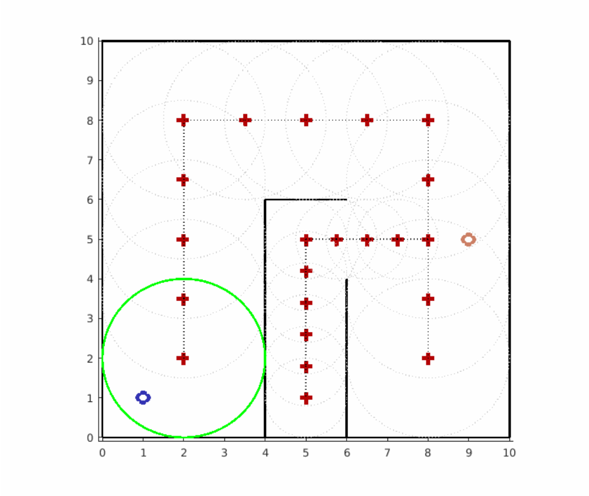
…

Formal methods





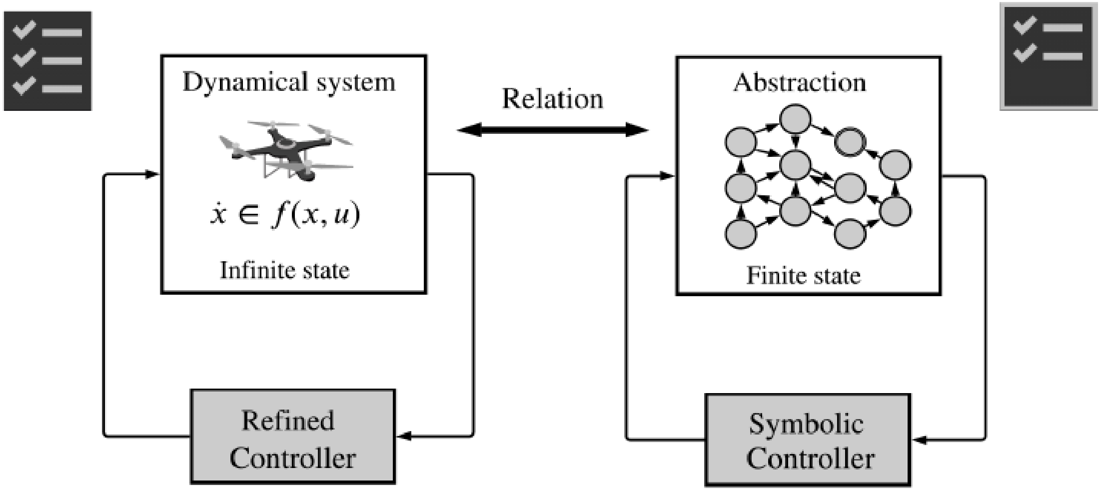
'Reactive synthesis'



General safety-critical optimal control problems

Symbolic control

"*Given a […] system and some desired property, one extracts a finite, discrete system while preserving all properties of interest*"
[Alur et al., 2000]



Outline

- **Introduction**

Data-Driven abstractions: opportunities and challenges

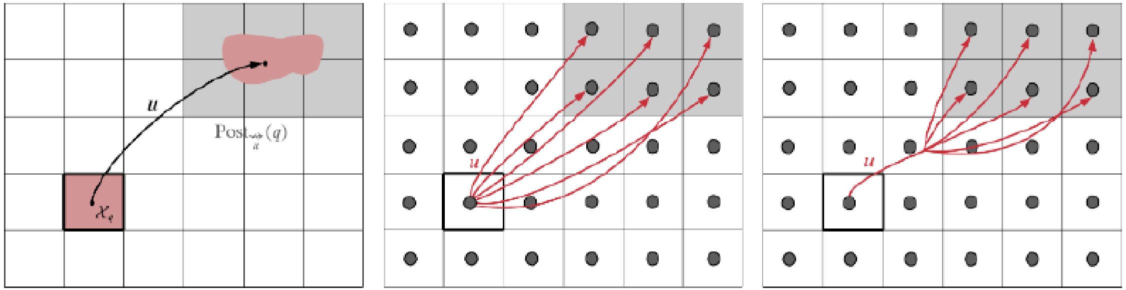
- Existing approaches

- The markovian issue and Markov abstractions

- Conclusions and perspectives

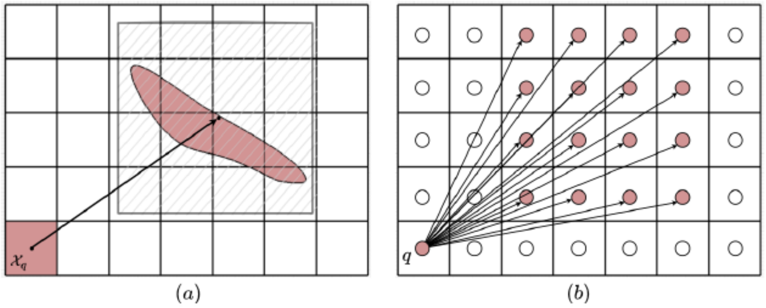
Symbolic control, Abstraction (ABCD), formal methods

“Given a […] system and some desired property, one extracts a finite, discrete system while preserving all properties of interest”
[Alur et al., 2000]

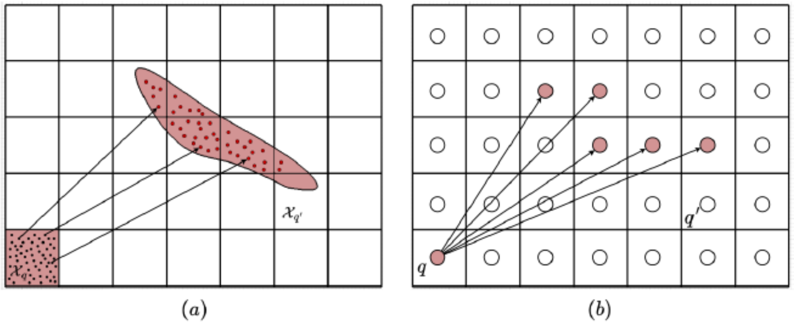


Data driven abstraction is an opportunity

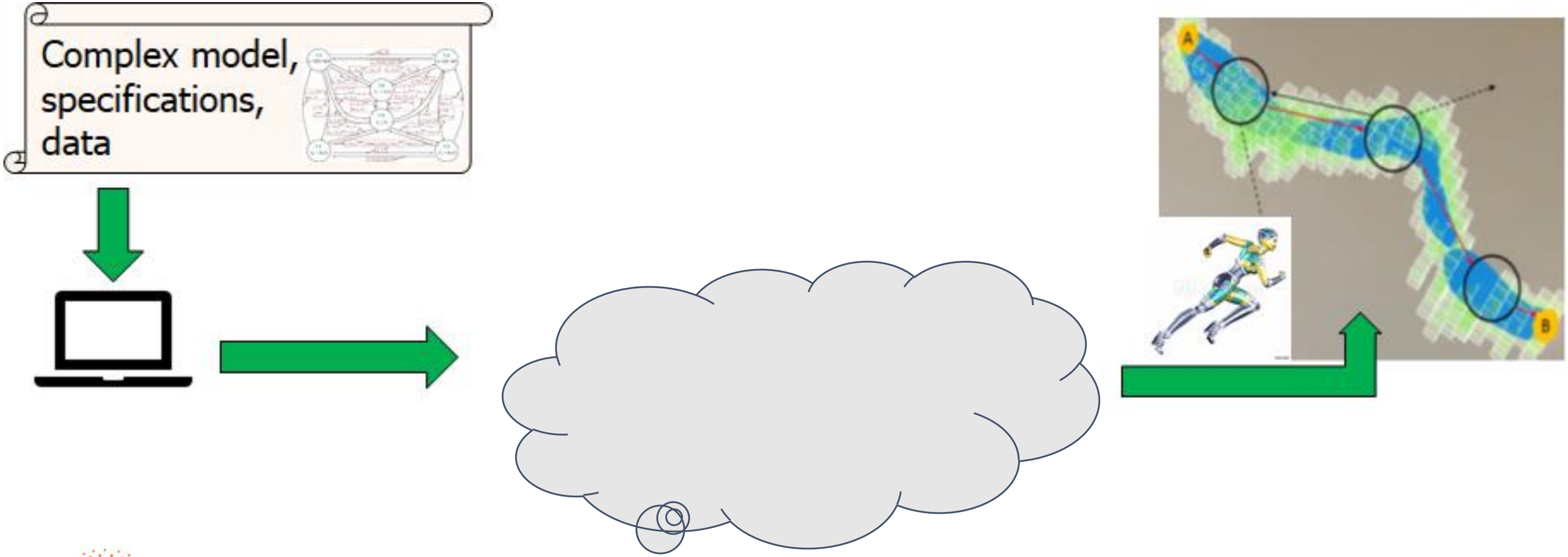
Model-based
growth bound



Data-driven: only put
what you observe!



Our long-term goal



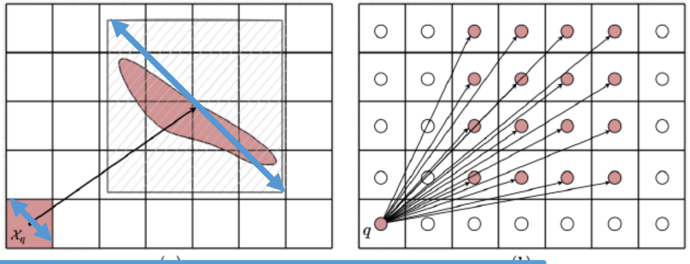
Outline

- Introduction
- **Existing approaches: Incremental stability and the Scenario approach**
- The markovian issue and Markov abstractions
- Conclusions and perspectives

The (in)famous Lipschitz assumption

Aka delta-ISS, contraction, differential positivity, …

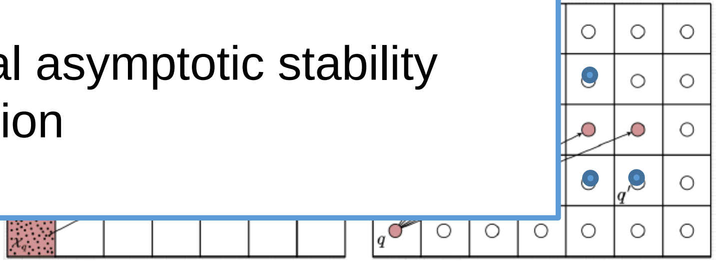
[**Pola, Girard, Tabuada**, Approximately bisimilar symbolic models for nonlinear control systems, *Automatica* 2008.]
[**Rungger, Zamani.** SCOTS: A tool for the synthesis of symbolic controllers. HSCC 2016]



- **Assumption:** the dynamics function is Lipschitz-continuous with L<1
- ➔ Stability of the **error-dynamics**

Arguably a strong assumption. Indeed:

- **Theorem:** delta-ISS on a compact ➔ global asymptotic stability
- Requires potentially infinitesimal discretization



(a) (b)

Lipschitz systems

[**Sadraddini, Belta.** Formal guarantees in data-driven model identification and control synthesis. HSCC 2018]

Positive systems

[**Makdesi, Girard, Fribourg.** Data-driven abstraction of monotone systems. L4DC PMLR 2021]

Scenario approach for data-driven formal methods

Switched systems stability

[**Kenanian, Balkan, Jungers, Tabuada** Automatica 2018]

incremental -ISS

[**Kazemi, Majumdar, Salamati, Soudjani, Wooding.** Data-driven abstraction-based control synthesis. Arxiv 2022]

[**Lavaei, Frazzoli.** Data-driven synthesis of symbolic abstractions with guaranteed confidence. IEEE Control Systems Letters, 2022.]

behavioural relation

[**Coppola, Peruffo, Mazo.** Data-driven Abstractions for Verification of Deterministic Systems. Arxiv 2022]

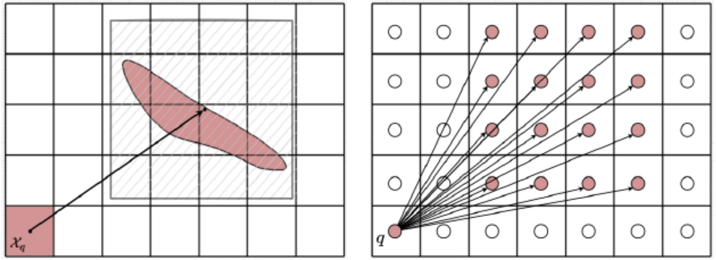
optimal control

[**Badings, Abate, Jansen, Parker, Poonawala, Stoelinga.** Sampling-based robust control of autonomous systems with non-Gaussian noise. AAAI 2022]

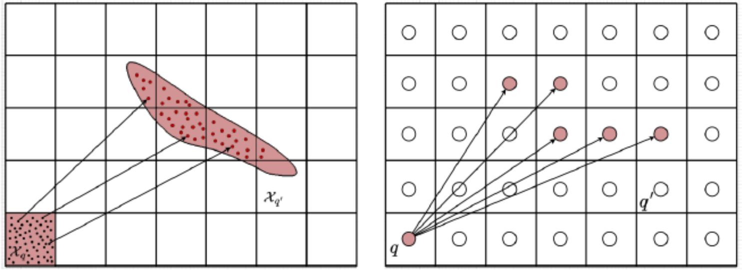
formal verification

[**Nejati, Lavaei, Jagtap, Soudjani, Zamani**. Formal Verification of Unknown Discrete- and Continuous-Time Systems: A Data-Driven Approach. IEEE TAC 2023.]

[**Devonport, Arcak.** Data-driven reachable set computation using adaptive Gaussian process classification and Monte Carlo methods. ACC 2020.]



(a) (b)



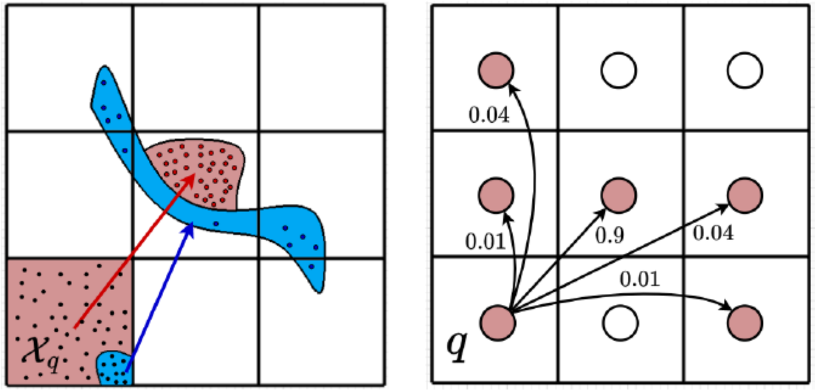
(a) (b)

Outline

- Introduction
- Existing approaches: Scenario approach and Incremental stability
- **The markovian issue and Markov abstractions**
- Conclusions and perspectives

An ML-inspired approach

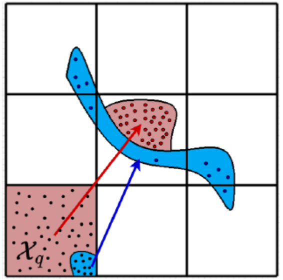
- Focus on the **behaviour** of the system
- Given a deterministic system, let us use **probabilities** to represent our uncertainty
- **Goal:** find the places where the 'non-lipschitz behaviour' matters most, and refine there
- In effect, we are trading **epistemic uncertainty vs. Aleatoric uncertainty**



Epistemic vs. Aleatoric uncertainty

Uncertainty about reality

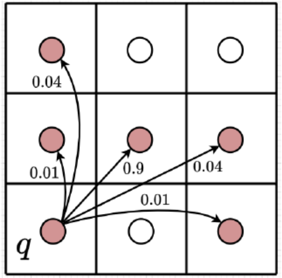
(Bayesian view)

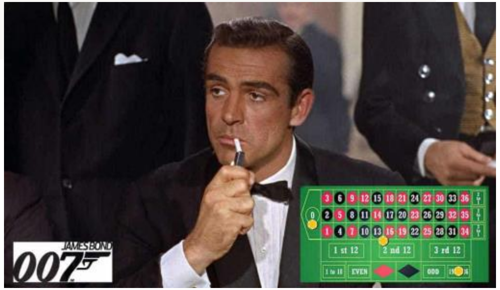


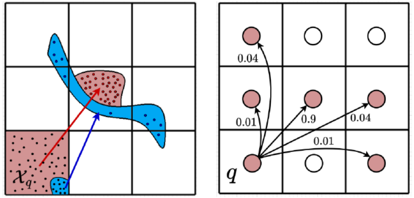
An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom

Intrinsic uncertainty

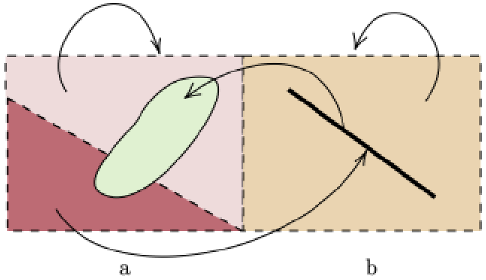
(frequentist view)

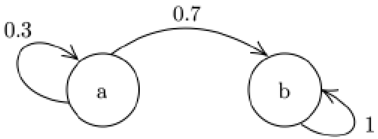






The Markovian issue

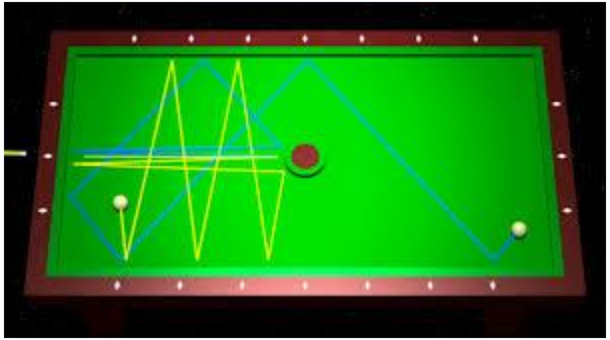




Memory-1: aba is impossible

➔Missing trajectory

Memory and dynamics



Block-shifts, Debruijn graphs

[**Morse, Hedlund.** **Symbolic dynamics**. *American Journal of Mathematics*, *1938*]

> *The methods used in the study of recurrence and transitivity frequently combine classical differential analysis with a more abstract symbolic analysis*

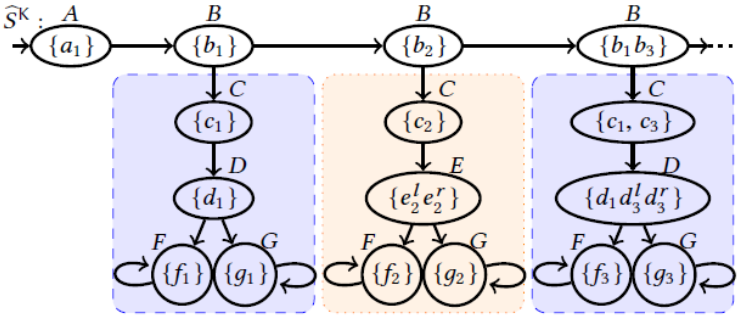
Multiple Lyapunov functions

$$\forall k \in \mathbb{N},\ x(k) \neq 0 \Rightarrow V(x(k)) > 0 \text{ and } V(x(k+p)) < V(x(k)).$$

[**Bliman, Ferrari-Trecate**. Stability analysis of discrete-time switched systems through Lyapunov functions with **nonminimal state**. *ADHS 2003]*

(also Dullerud, Daafouz,…)

L-complete abstractions



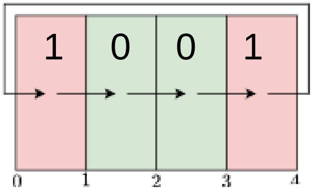
[**Moor, Raisch, O'Young**. Supervisory control of hybrid systems via l-complete approximations. WODES 1998]

[**Majumdar, Ozay, Schmuck**. On abstraction-based controller design with output feedback. HSCC 2020]

Memory lifts to alleviate the Markovian issue

Algorithm:

- Assume perfect sampling of each cell (no chance constrained error)
- Build the hierarchy of corresponding markov chains
- The markov chain 'converges' towards the true system



The Memory lift has eliminated spurious trajectories

The *behaviour* is the set of output traces

Main theorem [Banse, Romao, Abate, Jungers, L4DC 2023]

If the system is observable, F is continuous, and X is compact
Then there exists a sequence ε_ℓ such that $\lim_{\ell\to\infty} \varepsilon_\ell = 0$ for which $\mathcal{B}(S(\varepsilon_\ell)) = \mathcal{B}(\Sigma_\ell)$,with $S(\varepsilon_\ell)$ a perturbed system of the form

$$x_{k+1} = F(x_k + w_k), y_k = H(x_k), \| w_k \| \leq \varepsilon_\ell$$

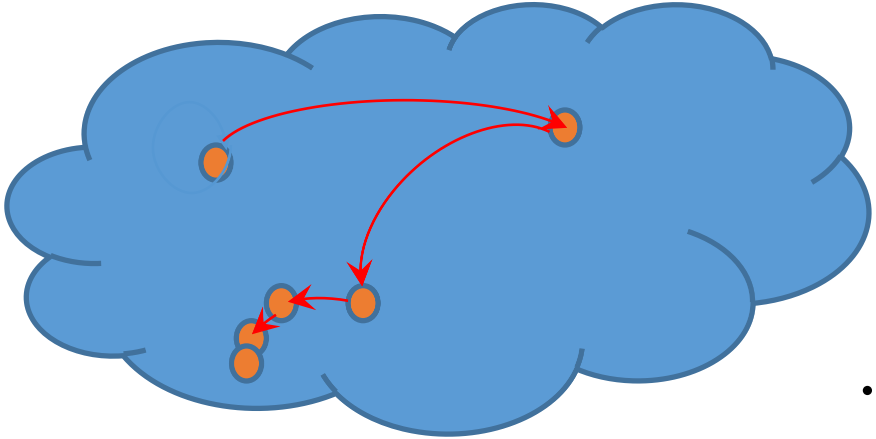
Approximating the behaviour with Markov Chains

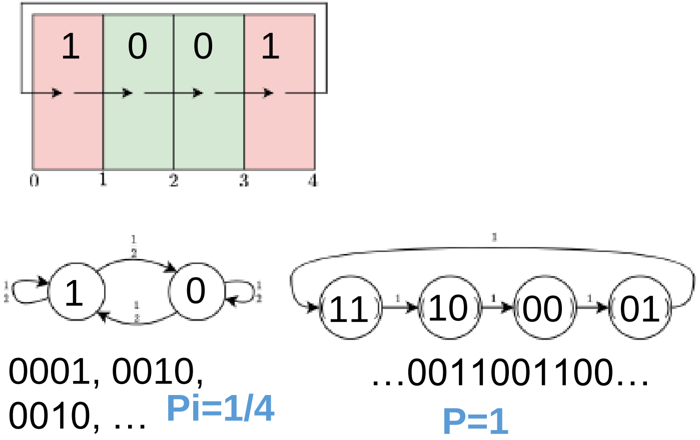
And a new distance on Markov Chains

Memory lifts to alleviate the Markovian issue:

A practical algorithm

- The markov chain ‘converges’ towards the true system
- ➔ need for a **metric** between Markov Chains

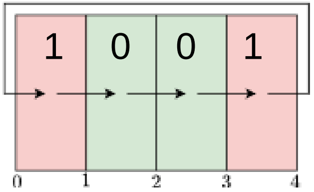


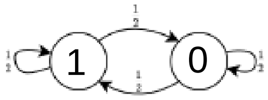


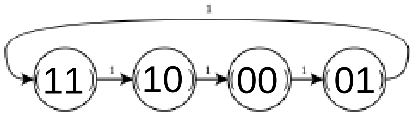
- Our metric must reflect the difference in terms of **behaviour** (probability distribution on the output signal (traces))

The Cantor-Kantorovich metric

- For a fixed length (here, 5), the behaviour is a **probability distribution on a set of output signals**
- **I.The Kantorovich metric**: a well-known (and appropriate) way to measure a distance between probability distributions
- Aka earth-mover distance, Wasserstein distance, optimal transport…







0001, 0010, 0010, …

…0011001100…

P(w_i)=1/32

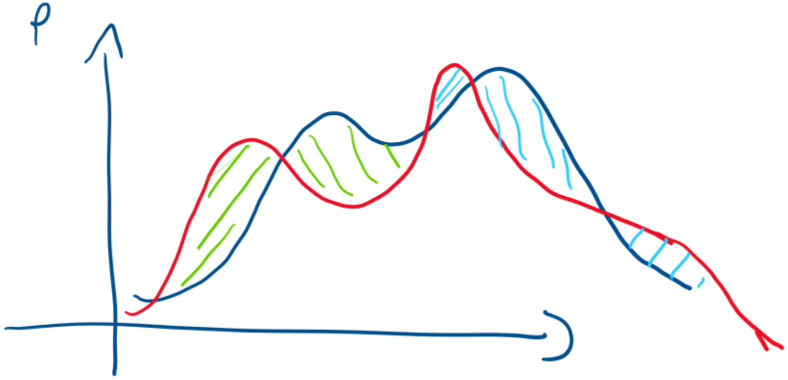
00000, 00001, 00011,…

00110…

P(w)=1

The Cantor-Kantorovich metric

- For a fixed length (here, 5), the behaviour is a **probability distribution on a set of output signals**



- **I.The Kantorovich metric**: a well-known (and appropriate) way to measure a distance between probability distributions
- Aka earth-mover distance, Wasserstein distance, optimal transport…

Nice **duality results for efficient computation** (Farkas'lemma)

P(w_i)=1/32

00000, 00001, 00011,…

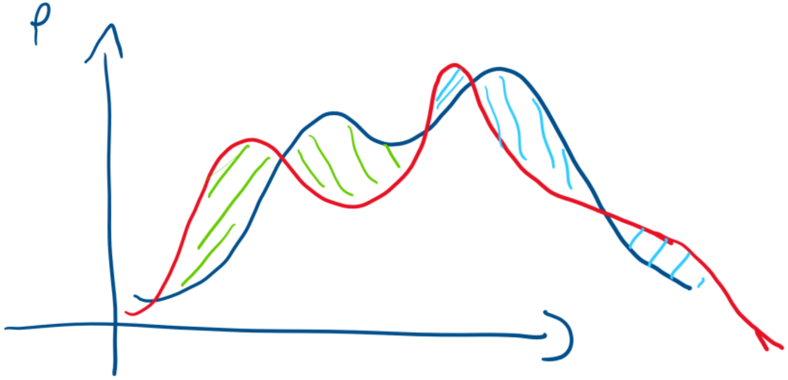
00110…

P(w)=1

$$K(p_1^n, p_2^n) = \min_{\pi^n \in \Pi(p_1^n, p_2^n)} \sum_{w_1, w_2 \in \mathcal{A}^n} d_B(w_1, w_2)\pi^n(w_1, w_2)$$

The Cantor-Kantorovich metric

- For a fixed length (here, 5), the behaviour is a **probability distribution on a set of output signals**



- **I.The Kantorovich metric**: a well-known (and appropriate) way to measure a distance between probability distributions
- Aka earth-mover distance, Wasserstein distance, optimal transport…
- It relies on a notion of **distance within the probability space**

$$K(p_1^n, p_2^n) = \min_{\pi^n \in \Pi(p_1^n, p_2^n)} \sum_{w_1, w_2 \in \mathcal{A}^n} \boxed{d_B(w_1, w_2)} \pi^n(w_1, w_2)$$

The Cantor-Kantorovitch metric

II. The Cantor distance

A simple distance between words:

Definition:
D_B(w_1,w_2):= ½^{# of the first different letter}

Example: w1= 01011   W2=01000   ➔   d_B(w1,w2) = 1/8

The cantor distance is an **ultrametric** (strong triangular inequality)

$$d_B(w_1, w_3) \leq \max\{d_B(w_1, w_2), d_B(w_2, w_3)\}$$

The Cantor-Kantorovitch metric

II. The Cantor distance

A simple distance between words:

Definition:
D_B(w_1,w_2):= ½^{# of the first different letter}

Example: w1= 01011   W2=01000   ➔   d_B(w1,w2) = 1/8

The cantor distance is an **ultrametric** (strong triangular inequality)

$$d_B(w_1, w_3) \leq \max\{d_B(w_1, w_2), d_B(w_2, w_3)\}$$

The Cantor-Kantorovitch metric

Finally, we define the **C-K metric between two Markov chains** as the limit of the Kantorovitch metric for large horizons n (i.e. the length of the traces)

$$d(\Sigma_1, \Sigma_2) = \lim_{n \to \infty} K(p_1^n, p_2^n)$$

Theorem: the metric is well defined. Moreover,

$$0 \leq d(\Sigma_1, \Sigma_2) - K(p_1^n, p_2^n) \leq 2^{-n}$$

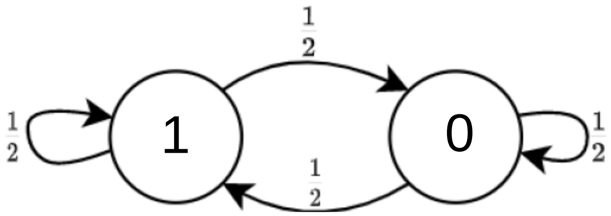
Theorem: the metric can be computed efficiently

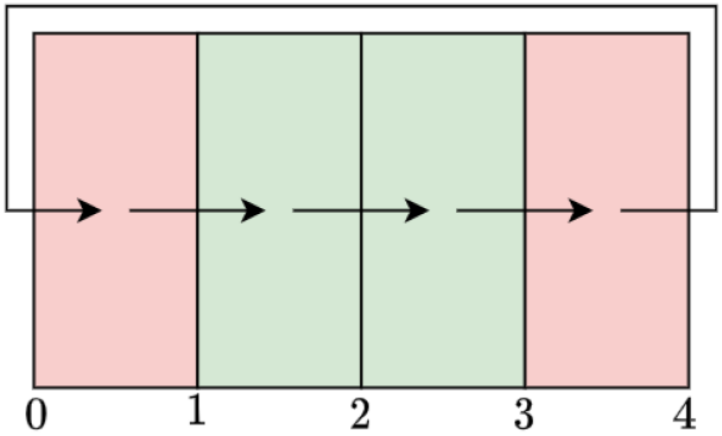
[Banse, Romao, Abate, Jungers, Arxiv 2023]

Opportunistic smart refinement of Markov models

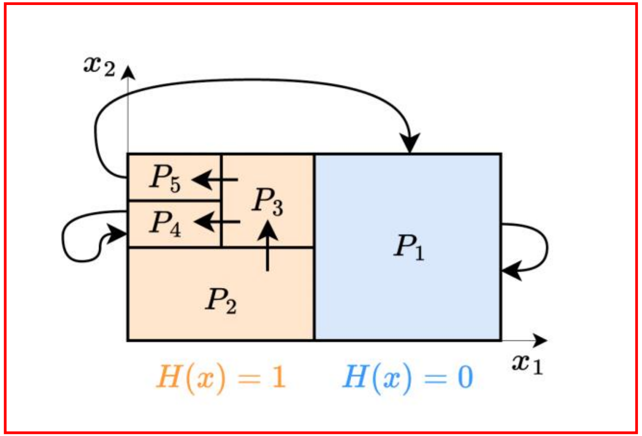
Algorithm

- **Build the elementary** Markov Model (1 node=1 output)
- At step N, build N Markov models by **prolonging separately each node**
- Keep the model **maximizing the C-K metric wrt previous** model
- **Iterate** until convergence





A deterministic toy example



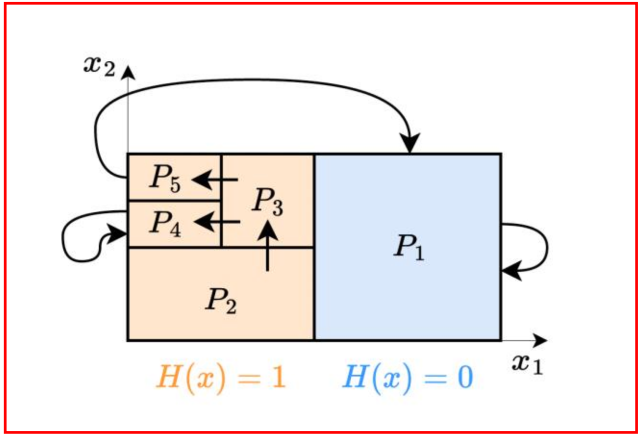
In our setting (variants are possible), we assume that we have an **oracle that returns the probability** of a queried transition
(e.g. P(110→1100))

$$F(x) = \begin{cases} x & \text{if } x \in P_1 \cup P_4, \\ (x_1/2 + 1/2, x_2 + 1/2) & \text{if } x \in P_2, \\ (x_1 - 1/2, x_2) & \text{if } x \in P_3, \\ (2x_1 + 1, 4x_2 - 3/4) & \text{else}, \end{cases}$$

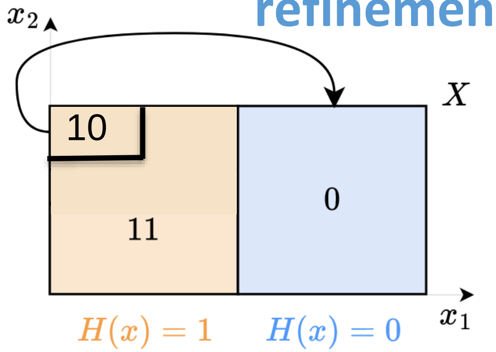
A deterministic toy example

At initial step, the algorithm **starts** with the partition defined by the **outputs**

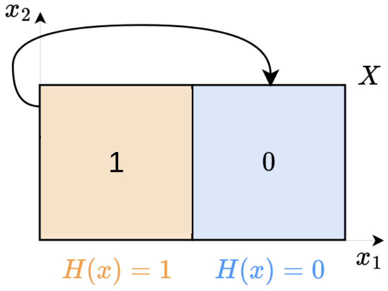
A deterministic toy example

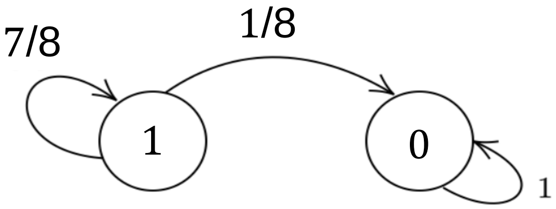


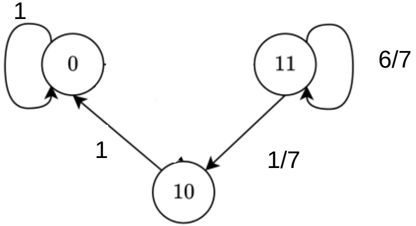
We now **separately explore the refinement** of each of our cells



Refine '1'

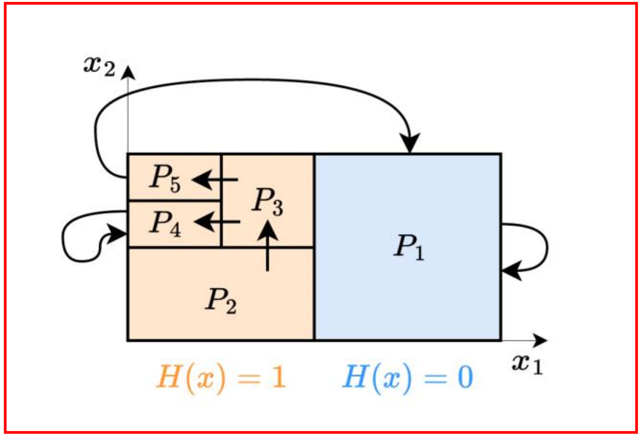






A deterministic toy example

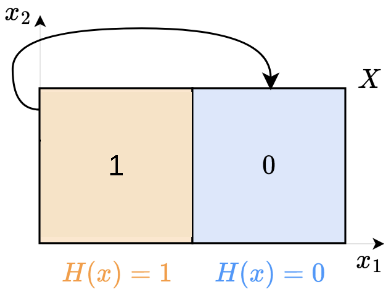
A deterministic toy example

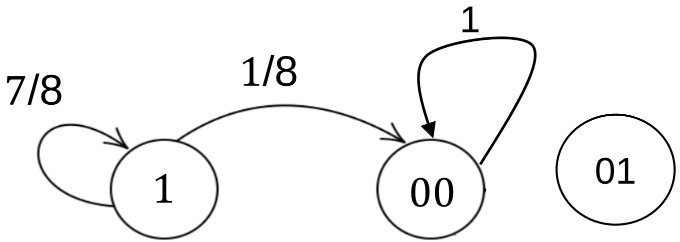


At initial step, the algorithm **starts** with the partition defined by the **outputs**

We now **separately explore the refinement** of each of our cells

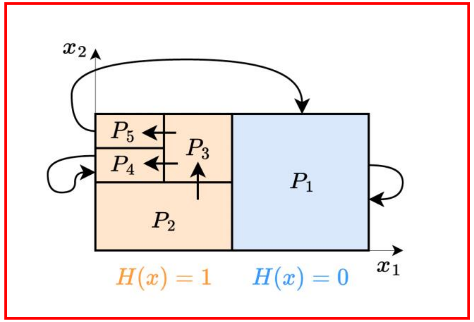
Refine ‘0’



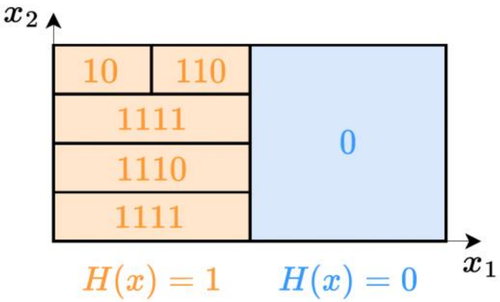


A deterministic toy example

A deterministic toy example



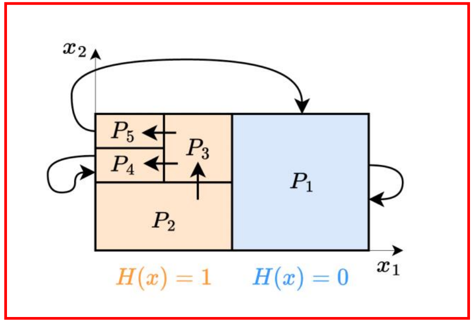
	$\mathcal{W}$	$d\left(\Sigma_{\mathcal{W}}, \Sigma_{\mathcal{W}'_j}\right)$
$k = 0$	$\{0, \mathbf{1}\}$	0.0015
$k = 1$	$\{0, 10, \mathbf{11}\}$	0.0059
$k = 2$	$\{0, 10, 110, \mathbf{111}\}$	0.0039
$k = 3$	$\{0, 10, 110, 1110, 1111\}$	-



A deterministic toy example

Solve an optimal control problem on successive partitions

Iteration	Controller (10)	Expected reward (9)
$k=0$	$K(x)=\begin{cases}0 & \text{if } x\in[0]_S\\ 0 & \text{if } x\in[1]_S\end{cases}$	**14.4784**
$k=1$	$K(x)=\begin{cases}0 & \text{if } x\in[0]_S\\ 0 & \text{if } x\in[10]_S\\ 1/4 & \text{if } x\in[11]_S\end{cases}$	**18.8726**
$k=2$	$K(x)=\begin{cases}0 & \text{if } x\in[0]_S\\ 0 & \text{if } x\in[10]_S\\ 0 & \text{if } x\in[110]_S\\ 1/4 & \text{if } x\in[111]_S\end{cases}$	**19.0311**
$k=3$	$K(x)=\begin{cases}0 & \text{if } x\in[0]_S\\ 0 & \text{if } x\in[10]_S\\ 0 & \text{if } x\in[110]_S\\ 1/2 & \text{if } x\in[1110]_S\\ 1/4 & \text{if } x\in[1111]_S\end{cases}$	**19.1022**



$$x'_k = x_k + \binom{0}{1} u_k (mod 1), x_{k+1} = F(x'_k), u_k = K(x_k) \in \left\{0, \frac{1}{4}, \frac{1}{2}\right\}.$$

$$\max_K \mathbb{E}_{x_1 \sim \mu(X)} \sum_k \gamma^k r(x_k)$$

where $r(x) = 1$ if $H(x) = 0$, and $r(x) = 0$ otherwise.

Tame wilder behaviours

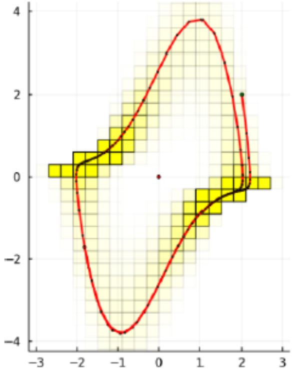
Two-dimensional system: the Van der Pol oscillator

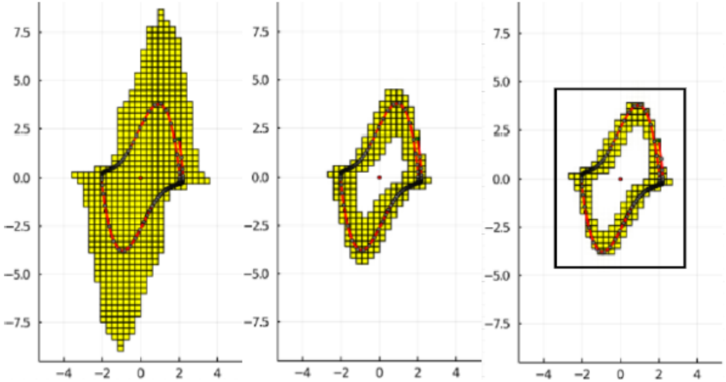
We use the entropy of the markov model as a proxy to estimate our epistemic uncertainty wrt a given discretization

➔ **Opportunistic discretization based on the entropy of the Markov Chain**

$$\begin{cases} \dot{x} = y \\ \dot{y} = \mu(1 - x^2)y - x \end{cases}$$

with $\mu > 0$.





(a) (b) (c)

$$C(\boldsymbol{\pi}) = -\sum_{q \in \mathcal{Q}} \pi_q \log_2(\pi_q)$$

Outline

- Introduction
- Existing approaches: Scenario approach and Incremental stability
- The markovian issue and Markov abstractions
- **Conclusions and perspectives**

Conclusions

- **Memory** is a way to alleviate conservativeness of data-driven abstractions. Focuses on the **behaviour rather than state-space accuracy**
- We (provide a framework for) **iteratively refine the abstraction** in an opportunistic way ➔ **Nonstandard abstractions**
- **Technical contributions**
  - Convergence theorem (with no Lipschitz assumption)
  - Introduce a new 'C-K metric' on Markov Chain
  - Efficient computation algorithm for the 'C-K metric'
  - Proof of concept for smart/nonstandard abstractions
- Many **open problems** and directions

Thanks!

Questions?

We're hiring! (raphael.jungers@uclouvain.be)