Data-driven memory-dependent abstractions of dynamical systems via a Cantor-Kantorovich metric

Adrien Banse, Licio Romao, Alessandro Abate and Raphaël M. Jungers

***Abstract*— Abstractions of dynamical systems enable their verification and the design of feedback controllers using simpler, usually discrete, models. In this paper, we propose a data-driven abstraction mechanism based on a novel metric between Markov models. Our approach is based purely on observing output labels of the underlying dynamics, thus opening the road for a fully data-driven approach to construct abstractions. Another feature of the proposed approach is the use of memory to better represent the dynamics in a given region of the state space. We show through numerical examples the usefulness of the proposed methodology.**

***Index Terms*— Abstraction of dynamical systems, Markov models, Formal methods**

I. Introduction

The complexity of dynamical systems emerging from several industrial applications has dramatically increased in the past years, which raises additional challenges for their analysis and control [1]–[4]. *Abstraction techniques* provide a way to tame complexity in the verification and/or control design step by producing a (usually discrete) representation of the underlying dynamical system (see e.g. [5]–[10]). The resulting abstract models are then used to indirectly verify the concrete model, or to design feedback controllers by means of a procedure referred to as control refinement [11].

Two of the main limitations for creating abstractions of dynamical systems are the curse of dimensionality and the reliance on the mathematical model representing the dynamics. The first limitation results from the partitioning of the state space of the concrete model into a finite set of blocks. For a given accuracy, the number of finite states grows exponentially with the dimension of the initial system (see e.g. [12]). This leads to discrete representations that, in order to meet specific accuracy levels, require prohibitive computational resources. For this reason, model-based methods to partition the state space in a smarter fashion have appeared in the literature [13]. Amongst them, *memory-dependent* abstractions have been used in [14], [15] to mitigate this issue. The second limitation has attracted much attention in recent years in the control community: in many applications, the system to control is too complicated for a model-based analysis, or a given mechanistic model is not even available. Therefore, there is a growing need for *data-driven* control approaches, namely techniques that are purely based on the observation of data from the system, thus without resorting to the cumbersome and potentially error-prone work of building a model for the system [16]–[18].

Our new approach, presented in this paper, aims at mitigating both of these limitations, by enabling the construction of a Markov model purely based on output data and allowing for a smart and frugal refinement of the state-space partitions. To enrich the representation of our discrete model, we consider memory-dependent Markov models, and propose an algorithmic procedure to further enrich the model in specific regions of the state space. Previous work such as [19], [20] also combine memory-dependent and data-driven approaches, but does not profit from an adaptive approach and use deterministic abstractions. Other previous works, such as [6], [7], use stochastic models such as Interval Markov Decision Processes to circumvent the uncertainty on the concrete system, but do not leverage memory.

To enable a non-uniform partitioning of the state space of the dynamics, we develop a novel notion of *metric between Markov models* that relies on the theory of optimal transport. More precisely, we leverage the Kantorovich metric in order to evaluate the difference between two Markov models in terms of the probability distributions that they define on the output language. To define this optimal transport metric, we equip the space of words with the Cantor distance[1]. Kantorovich metrics for Markov models have already been studied in [21]–[24], but with different underlying metrics. The Cantor distance has been widely investigated in the field of symbolic dynamics [25]–[27], leading to interesting results about its topological structure. We also present a new algorithm that efficiently computes the proposed metric between Markov models.

This project has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program under grant agreement No 864017 (L2C). A. Banse is supported by the French Community of Belgium in the framework of a FNRS/FRIA grant and R. M. Jungers is a FNRS honorary Research Associate.

A. Banse and R. M. Jungers are with the ICTEAM, UCLouvain, Louvain-la-Neuve, Belgium (e-mails: {adrien.banse, raphael.jungers}@uclouvain.be). L. Romao is with the Department of Aeronautics & Astronautics, Stanford University, Stanford, USA (e-mail: licio@stanford.edu). A. Abate is with the Department of Computer Science, University of Oxford, Oxford, UK (e-mail: alessandro.abate@cs.ox.ac.uk).

[1]In order to ease the reading of this paper, we use the term *metric* for the Kantorovich metric and the term *distance* for the underlying Cantor distance.

Contributions[2] We develop a new framework to abstract dynamical systems. The proposed abstraction technique leverages memory to implicitly build a non-uniform partition of the state space, hence enabling the refinement of the abstraction in regions where the underlying dynamics has more intricate behaviours. A second contribution consists in proposing a new metric between Markov models based on optimal transport, and showing a recursive algorithm for its computation that improves upon a naïve linear programming formulation. With that, we also contribute to well established research efforts outside of the control community (see e.g. [21], [29]–[31]).

Outline The rest of this paper is organised as follows. In Section II, we provide theoretical background for dynamical systems and labeled Markov chains. In Section III, we introduce adaptive memory abstractions, and show that they satisfy behavioural inclusion (see [5]). We provide a method to choose a convenient partitioning given a metric between Markov chains, and prove that it always yields well-defined abstractions. In Section IV, we introduce a specific Cantor-Kantorovich metric between Markov chains. We prove that it is indeed a metric, and we give an algorithm to approximate it efficiently. In Section V, we illustrate our procedure with the Cantor-Kantorovich metric on a numerical example. We finally conclude in Section VI.

Notations In this work, $\mathbb{R}$ is the set of reals, $\mathbb{N}$ is the set of natural and $\mathbb{N}_{\geq 0}$ is the set of non-negative natural numbers. Given a finite alphabet A, a *word* $w \in A^n$ is denoted $w = a_1 \dots a_n$. Given $s, t \in \mathbb{N}_{\geq 0}$, a *timed word*, denoted $w_{[-s,t]}$, is a couple $w \in A^{s+t+1}$ and time interval $[-s, t]$. It is denoted $w_{[-s,t]} = a_{-s} \dots a_t$ with small abuse of notation. Given a set X, its *Kleene closure* is noted X^*, and the i-th *functional power* of a function f is noted f^i. In terms of computational complexity, we say that $f(n) = \mathcal{O}(g(n))$ if there exists n_0 and $c > 0$ such that, for all $n \geq n_0$, $|f(n)| \leq cg(n)$, and we say that $f(n) = \Omega(g(n))$ if there exists n_0 and $c > 0$ such that, for all $n \geq n_0$, $f(n) \geq cg(n)$.

II. Preliminaries

In this section, we formally define dynamical systems and labeled Markov chains. We also introduce a notion of *behaviour* for both models.

A. Dynamical systems

A *dynamical system* is the 4-tuple $\Sigma = (X, A, f, h)$ that defines the relation

$$x_{k+1} = f(x_k), \quad y_k = h(x_k), \tag{1}$$

where $X \subseteq \mathbb{R}^d$ is the *state space*, A is a finite alphabet called the *output space*, $f : X \to X$ is an invertible *transition function*, and $h : X \to A$ is the *output function*. The variables x_k are called *states*, and variables y_k are called *outputs* at time k.

[2]Preliminary results have been presented in [28], but without proofs. Besides, this work considers a more general model for abstracting dynamical systems, and presents a new, thorough numerical example.

For $s, t \in \mathbb{N}$, given a word $w_{[-s,t]} = a_{-s} \dots a_t$ with $a_i \in A$ for all $i = -s, \dots, t$, one can define a subset of the state space as follows

$$\left[w_{[-s,t]}\right] = \{x \in X : h(f^i(x)) = a_i \, \forall i = -s, \dots, t\}. \tag{2}$$

For example, the subset $\left[011_{[-1,1]}\right]$ corresponds to the set of states $x \in X$ such that $h(f^{-1}(x)) = 0$, $h(x) = 1$ and $h(f(x)) = 1$. A set of words

$$W = \left\{w^1_{[-s_1,t_1]}, \dots, w^k_{[-s_k,t_k]}\right\} \tag{3}$$

defines a *partition* of the state space if

$$\bigcup_{i=1,\dots,k} \left[w^i_{[-s_i,t_i]}\right] = X, \tag{4}$$

$$\forall i \neq j, \; \left[w^i_{[-s_i,t_i]}\right] \cap \left[w^j_{[-s_j,t_j]}\right] = \emptyset. \tag{5}$$

In this case, a subset $[w_{[-s,t]}]$ is called a *block*.

In this paper, we consider data-driven applications in which the initial state of a dynamical system is sampled. We assume that the state space X is endowed with a measure λ on the probability space $(X, \mathcal{B}(X), \lambda)$, where $\mathcal{B}(X)$ is the Borel σ-algebra generated by the topology of X. In other words, we consider that the system (1) is such that

$$x_0 \sim \lambda. \tag{6}$$

For clarity, for a given word $w_{[-s,t]}$, we call the quantity $\lambda\left(\left[w_{[-s,t]}\right]\right)$ the *probability* of observing the sequence $y_{-s} \dots y_t = w$, which we denote as $\mathbb{P}_\lambda(y_{-s} \dots y_t = w)$. Similarly, given two words $w_{[-s,t]}$ and $w'_{[-s',t']}$, we call the quantity

$$\lambda\left(f^{-1}\left(\left[w'_{[-s',t']}\right]\right) \cap \left[w_{[-s,t]}\right]\right) / \lambda\left(\left[w_{[-s,t]}\right]\right) \tag{7}$$

the *conditional probability* of observing $y_{-s'+1} \dots y_{t'+1} = w'$ knowing that $y_{-s} \dots y_t = w$, which we denote by $\mathbb{P}_\lambda(y_{-s'+1} \dots y_{t'+1} = w' | y_{-s} \dots y_t = w)$.

We now introduce the notion of finite *behaviour* for this class of dynamical systems. The behaviour of a dynamical system Σ with initial measure λ, denoted as $B(\Sigma, \lambda) \subseteq A^*$, is the set of finite sequences $w \in A^k$, for all $k \in \mathbb{N}$, such that $\mathbb{P}_\lambda(y_0 \dots y_{k-1} = w) > 0$. In other words, the behaviour of Σ contains the set of words that can be reached with positive probability.

B. Labeled Markov chains

A labeled Markov chain is defined as a 5-tuple $\Gamma = (S, A, \tau, \mu, l)$ where S is a finite set of *states*, A is a finite *alphabet*, τ is the *transition matrix* defined on $S \times S$, μ is an *initial measure* defined on S and $l : S \to A$ is a *labeling function*. For two states $s, s' \in S$, the entry $\tau_{s,s'}$ of the transition matrix is defined as

$$\tau_{s,s'} = \mathbb{P}(X_{k+1} = s' \mid X_k = s), \tag{8}$$

where $X_1, X_2, \dots$ is a sequence of random variables.

Consider the equivalence relation on S defined as $s \sim s'$ if and only if $l(s) = l(s')$. For any $a \in A$, the notion of *equivalence class* is therefore defined as

$$[\![a]\!] = \{s \in S : l(s) = a\}. \tag{9}$$

For any sequence of labels $w = a_1 \dots a_k$ of length k, this allows to define its probability induced by the Markov chain as

$$p^k(w) = \sum_{s_1 \in [\![a_1]\!]} \mu_{s_1} \sum_{s_2 \in [\![a_2]\!]} \tau_{s_1,s_2} \cdots \sum_{s_n \in [\![a_n]\!]} \tau_{s_{n-1},s_n}. \quad (10)$$

Remark 1: Using a similar algorithm as the *forward backward procedure* for hidden Markov models (see [32] and references therein), it is possible to compute $p^l(w)$ for all $w \in A^l$ for all $l = 1, \dots, k$ in $k|S|^2$ operations. □

Similarly, we define the finite *behaviour* of the labeled Markov chain Γ, noted $B(\Gamma) \subseteq A^*$, as the set of all finite sequences $w \in A^k$, for all $k \in \mathbb{N}$, such that $p^k(w) > 0$.

III. A PROCEDURE TO CONSTRUCT MEMORY-DEPENDENT ABSTRACTIONS

In this section, we first introduce the notion of *adaptive memory abstractions*. We then propose a data-driven procedure to construct tuneable adaptive memory abstractions.

A. Adaptive-memory abstractions

We introduce in Definition 1 the notion of *adaptive memory abstraction*, and we give an illustration in Figure 1.

Definition 1 (Adaptive memory abstraction): Given a dynamical system Σ, a measure λ and a set W that defines a partition, an *adaptive memory abstraction* of Σ is a labeled Markov chain $\Gamma_W = (S, A, \tau, \mu, l)$ composed as follows.

- The states correspond to the blocks of the partition, that is $S = W$
- For each node $w_{[-s,t]}$, the initial measure is defined as
$$\mu_{w_{[-s,t]}} = \mathbb{P}_\lambda \left(y_{-s} \dots y_t = w \right) \quad (11)$$
- For each two nodes $w_{[-s,t]}$ and $w'_{[-s',t']}$ the transition probability is defined as
$$\tau_{w_{[-s,t]}, w'_{[-s',t']}} = \mathbb{P}_\lambda \left(y_{-s'+1} \dots y_{t'+1} = w' \,|\, y_{-s} \dots y_t = w \right) \quad (12)$$
- For each node $w_{[-s,t]} = a_{-s} \dots a_t$, the labeling is defined as $l\left(w_{[-s,t]}\right) = a_0$ □

Since $x_0 \sim \lambda$, one needs to make sure that all block $\left[w_{[-s,t]}\right] \subseteq X$ can be captured by the sampling measure.

Assumption 1: For all $w_{[-s,t]} \in W$, $\mathbb{P}_\lambda(y_{-s} \dots y_t = w) > 0$.

Assumption 1 is a natural necessary condition in the context of data-driven methods. Informally, it states that the behaviour of the dynamical system can be reliably sampled. On a more technical level, it implies that the transition matrix of the corresponding abstraction is stochastic. For the sake of brevity, this result and its proof can be found in Appendix I.

Given a partition W, one can sample trajectories and compute the probabilities (11) and (12) in a Monte-Carlo fashion. For this reason, we say that these abstractions are *data-driven*. Finally, we say that these abstractions are *safe* because, for any partition W, all finite sequence output by the dynamical system Σ with initial measure λ can be simulated by Γ_W. This is formally written in the following proposition.

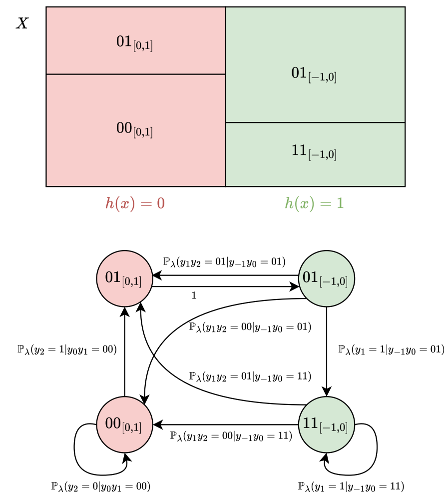


Fig. 1: Illustration of an adaptive memory abstraction Γ_W for a certain dynamical system Σ with measure λ. In this example, $W = \{01_{[0,1]}, 00_{[0,1]}, 01_{[-1,0]}, 11_{[-1,0]}\}$, as illustrated above. The corresponding abstraction Γ_W is given below, with all the possibly non-zero transition probabilities.

Proposition 1: Given a dynamical system Σ, a measure λ and its adaptive memory abstraction Γ_W, if Assumption 1 is satisfied, then it holds that $B(\Sigma, \lambda) \subseteq B(\Gamma_W)$.

Proof: It suffices to show that, for all $k \in \mathbb{N}$, for all words $w_{[0,k-1]} = b_0 \dots b_{k-1}$, if $\lambda([w_{[0,k-1]}]) > 0$ in the original system, then it holds that $p^k(w) > 0$ in the abstraction Γ_W. Since Assumption 1 is satisfied, the condition $p^k(w) > 0$ holds if there exists a sequence of states $s_0, \dots, s_{k-1} \in W$ such that $l(s_i) = b_i$ for all $i = 0, \dots, k-1$, and $\lambda([s_i] \cap f^{-1}([s_{i+1}])) > 0$ for all $i = 0, \dots, k-2$. We will therefore prove the latter.

First, it holds that

$$\left[w_{[0,k-1]}\right] = \bigcap_{i=0,\dots,k-1} f^{-i}([(b_i)_{[0,0]}]), \quad (13)$$

where $(b)_{[0,0]}$ denotes the word composed only of the letter b. Moreover, since W defines a partition, it holds that $[(b_i)_{[0,0]}] = \bigcup_{s_i \in [\![b_i]\!]} [s_i]$ for all $i = 0, \dots, k-1$. Therefore, one can write

$$\begin{aligned} \left[w_{[0,k-1]}\right] &= \bigcap_{i=0,\dots,k-1} f^{-i} \left(\bigcup_{s_i \in [\![b_i]\!]} [s_i] \right) \\ &= \bigcap_{i=0,\dots,k-1} \bigcup_{s_i \in [\![b_i]\!]} f^{-i}([s_i]). \end{aligned} \quad (14)$$

Now, since $\lambda([w_{[0,k-1]}]) > 0$, then there exists at least a sequence $s_0, \dots, s_{k-1} \in W$ such that $l(s_i) = b_i$ and such that $\lambda\left(\bigcap_{i=0,\dots,k-1} f^{-i}([s_i])\right) > 0$. Furthermore, one can say that the inequality above implies that

$$\lambda\left(f^{-i}([s_i]) \cap f^{-(i+1)}([s_{i+1}])\right) > 0 \quad (15)$$

for all $i = 0, \dots, k-2$. This is explained by the fact that, for any three sets A, B, C, if $\lambda(A \cap B \cap C) > 0$, then it holds that $\lambda(A \cap B) > 0$, $\lambda(A \cap C) > 0$ and $\lambda(B \cap C) > 0$. By invertibility of f, (15) implies that $\lambda\left([s_i] \cap f^{-1}([s_{i+1}])\right) > 0$ for all $i = 0, \dots, k-2$, and the proof is completed. ■

B. A data-driven procedure for abstractions

In the previous section, the choice of a convenient set W is not discussed. In this section, we present a data-driven procedure that tackles this problem. Our procedure, called REFINE, takes as input a dynamical system Σ, an initial measure λ, a *metric* D between labeled Markov chains, and a number of iterations N. It is descripted in Algorithm 1, and an illustration of this procedure is given in Figure 2.

Algorithm 1 REFINE$(\Sigma, \lambda, \mathsf{D}, N)$

$W \leftarrow \{(a)_{[0,0]}\}_{a \in A}$ — Start with a coarse partition
Sample trajectories from Σ
Find A_0 s.t. $\mathbb{P}_\lambda(y_0 = a) = 0 \ \forall a \in A_0$
$W \leftarrow W \setminus \{(a)_{[0,0]}\}_{a \in A_0}$
Construct Γ_W from samples of Σ
for $n = 1, \dots, N$ **do**
  for $i = 1, \dots, |W|$ **do**
    $W'_i \leftarrow W \setminus \left\{ w^i_{[0,t_i]} \right\}$ — Try to refine each block
    $W'_i \leftarrow W'_i \cup \left\{ (w^i a)_{[0,t_i+1]} \right\}_{a \in A}$
    Sample trajectories from Σ
    Find W_0 s.t. $\mathbb{P}_\lambda(y_0 \dots y_t = w) = 0 \ \forall w_{[0,t]} \in W_0$
    $W'_i \leftarrow W'_i \setminus W_0$
    Construct $\Gamma_{W'_i}$ from samples of S
    $\mathsf{D}_i \leftarrow \mathsf{D}(\Gamma_W, \Gamma_{W'_i})$
  end for
  $j = \arg\max_{i=1,\dots,|W|} \mathsf{D}_i$ — Greedy choice
  $W \leftarrow W'_j$
  $\Gamma_W \leftarrow \Gamma_{W'_j}$
end for
return W — Return a refined partition

The algorithmic procedure REFINE presented here depends on a general notion of metric between Markov models, rather than a particular one as in [28, Algorithm 2]. As it is based solely on trajectories of the original dynamics, we refer to REFINE as a data-driven abstraction algorithm. Besides, as formally stated in Appendix II, under the assumption that any output trace can be sampled with non-zero measure, REFINE leads to a valid partition of the state space of the original dynamics.

Remark 2: Many variants of the REFINE algorithm can be considered. For example, one could expand the memory in the past by expanding each word $w^i_{[-s_i,0]}$ into $\{(aw^i)_{[-(s_i+1),0]}\}_{a \in A}$. Also, inspired by algorithms from reinforcement learning (cf. TD learning scheme in [33]), one could compare the current model with the $|A|^n$ models possible after n steps instead of choosing between the possible $|A|$ different models, and take the one for which the distance is the largest. □

The output W from the REFINE algorithm depends on the chosen metric between the abstractions. In the following section, we introduce a novel metric between labeled Markov chains, and we discuss the interpretation of the corresponding REFINE output.

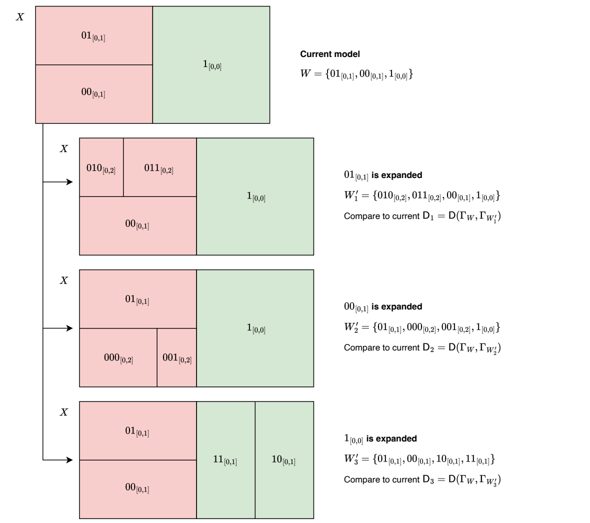


Fig. 2: Illustration of the REFINE algorithm for a system with $A = \{0, 1\}$. The current model contains three states $W = \left\{01_{[0,1]}, 00_{[0,1]}, 1_{[0,0]}\right\}$. Three sets W'_1, W'_2, W'_3 are then constructed, each time expanding one block of the partition into $|A| = 2$ sub-blocks. For each abstraction $\Gamma_{W'_i}$, the distance with the current one is computed: the current model is updated with the model for which the distance is the largest.

IV. THE CANTOR-KANTOROVICH METRIC

In this section, we introduce a new metric $\mathsf{CK}(\Gamma_1, \Gamma_2)$, named Cantor-Kantorovich metric, between two labeled Markov chains. This new metric was first introduced in [28], but without proofs. Here we prove that the metric is well-defined[3], then we present an algorithm to approximate it efficiently.

A. Definition

Given any two Markov chains $\Gamma_1 = (S_1, A, \tau_1, \mu_1, l_1)$ and $\Gamma_2 = (S_2, A, \tau_2, \mu_2, l_2)$ defined on the same set of outputs, for a fixed $k \in \mathbb{N}_{>0}$, let $p_1^k : A^k \to [0,1]$ and $p_2^k : A^k \to [0,1]$ be the probabilities induced by Γ_1 and Γ_2 such as defined in (10).

Definition 2 (Kantorovich metric): Let $\mathsf{D} : A^k \times A^k \to \mathbb{R}_{>0}$ be any metric between words of length k. The *Kantorovich metric* between p_1^k and p_2^k is defined as

$$\mathsf{K}_\mathsf{D}(p_1^k, p_2^k) = \min_{\pi^k \in \Pi(p_1^k, p_2^k)} \sum_{w_1, w_2 \in A^k} \mathsf{D}(w_1, w_2) \pi^k(w_1, w_2), \tag{16}$$

where $\Pi(p_1^k, p_2^k)$ is the set of all *couplings* of p_1^k and p_2^k, that is the set of all joint distribution $\pi^k : A^k \times A^k \to [0,1]$ such that the constraints

$$\pi^k(w_1, w_2) \geq 0 \quad \forall w_1, w_2 \in A^k, \tag{17}$$

$$\begin{aligned} \sum_{w_2 \in A^k} \pi^k(w_1, w_2) = p_1^k(w_1) \quad \forall w_1 \in A^k, \\ \sum_{w_1 \in A^k} \pi^k(w_1, w_2) = p_2^k(w_2) \quad \forall w_2 \in A^k \end{aligned} \tag{18}$$

[3] In the sense that it satisfies positivity, symmetry and triangle inequality.

hold. □

For a given k, the definition of the Kantorovich metric depends on an underlying distance D over the set of words of length k. We propose to endow the latter with the *Cantor distance*, defined as follows (see e.g. [27, Section 2.1]).

Definition 3 (Cantor distance): The *Cantor distance* between any two sequences $w_1 = a_1 \dots a_k$ and $w_2 = b_1 \dots b_k$ is defined as

$$\mathsf{C}(w_1, w_2) = \inf\{2^{-|c|} : c \text{ is a common prefix of } w_1 \text{ and } w_2\} \quad (19)$$

if $w_1 \neq w_2$, and $\mathsf{C}(w_1, w_2) = 0$ otherwise. □

Lemma 1 (See [27]): The Cantor distance satisfies the strong triangular inequality. That is, for any $w_1, w_2, w_3 \in A^k$,

$$\mathsf{C}(w_1, w_3) \leq \max\{\mathsf{C}(w_1, w_2), \mathsf{C}(w_2, w_3)\}. \quad (20)$$

We now define the Cantor-Kantorovich metric between two labeled Markov chains.

Definition 4 (Cantor-Kantorovich metric): Let Γ_1 and Γ_2 be two labeled Markov chains defined on the same set of labels A. Their *Cantor-Kantorovich metric* is defined as

$$\mathsf{CK}(\Gamma_1, \Gamma_2) = \lim_{k \to \infty} \mathsf{K}_\mathsf{C}(p_1^k, p_2^k), \quad (21)$$

where K_C is the Kantorovich metric with the Cantor distance as underlying distance, and where p_1^k and p_2^k are the probabilities respectively induced by Γ_1 and Γ_2. □

We prove in the next section that this metric is well-defined, in the sense that it satisfies positivity, symmetry and triangle inequality.

For a fixed length k, computing the Kantorovich metric $\mathsf{K}_\mathsf{C}(p_1^k, p_2^k)$ can be achieved by solving the linear program (16). The latter can be seen as an optimal transport problem where, at each $w \in A^k$, the "supplies" are given by $p_1^k(w)$ and the "demands" by $p_2^k(w)$. For each $w_1, w_2 \in A^k$, the cost of moving $\pi^k(w_1, w_2)$ of supply mass from w_1 to w_2 amounts to $\pi^k(w_1, w_2)\mathsf{C}(w_1, w_2)$. The specific choice of the Cantor distance as the underlying distance for the metric, allows to visualise this with a tree, as shown in Figure 3.

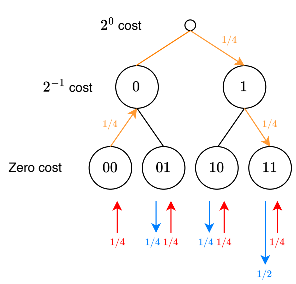


Fig. 3: Illustration of the Kantorovich metric linear program (16) for $A = \{0, 1\}$ and $k = 2$. The supplies are given in red, and the demands in blue. To solve this problem, one has to move $\pi^2(00, 11) = 1/4$ supply mass from 00 to 11. To do that, the mass has to travel up to the root. The corresponding Cantor distance is $\mathsf{C}(00, 11) = 2^0$. The total cost is therefore $\mathsf{K}_\mathsf{C}(p_1^2, p_2^2) = 1/4$.

Remark 3: A naïve approach to compute the metric is to use linear programming (LP) or combinatorial optimisation (CO) methods [34]. However, such methods are simply too costly to be used in practice. In the best case, the complexity is worse than quadratic, that is the number of operations is $\Omega(|A|^{2k})$. In the following section, we leverage the particular underlying Cantor distance to derive an algorithm that computes $\mathsf{K}_\mathsf{C}(p_1^k, p_2^k)$ that scales better. □

The Cantor-Kantorovich metric defines a metric space in which two labeled Markov chains are close if they have similar short-horizon behaviours. Indeed, the Cantor distance can be interpreted as a discount factor, and a large Cantor-Kantorovich metric means that the probabilities on sequences of labels differ close to the initial step of the random walks. The procedure $\textsc{Refine}(\Sigma, \lambda, \mathsf{CK}, N)$ therefore tends to choose the model with a different short-term behaviour.

B. A recursive algorithm for approximating the Cantor-Kantorovich metric

In this section, we state Theorem 1, a central recursive result for computing the Cantor-Kantorovich metric. This fact will be useful for two things. First, it implies that the Cantor-Kantorovich metric satisfies positivity, symmetry and triangle inequality, and that it can be approximated, as stated in Theorem 2. Second, it provides an efficient algorithm to approximate it.

We first state two lemmata that will be useful to prove Theorem 1. The proofs of the latter are moved in the appendices (see Appendix III and Appendix IV).

Lemma 2: For any $k \geq 1$, let π^k be the solution of (16). For all $w \in A^k$, it holds that $\pi^k(w, w) = \min\{p_1^k(w), p_2^k(w)\}$.

Lemma 3: For any $k \geq 2$, let π^k be the solution of (16). For all $w \in A^{k-1}$ such that $p_1^{k-1}(w) > p_2^{k-1}(w)$, then

$$\begin{aligned} \sum_{\substack{w' \in A^{k-1} \\ w' \neq w}} \sum_{a_1, a_2 \in A} \pi^k(wa_1, w'a_2) &= p_1^{k-1}(w) - p_2^{k-1}(w) \\ \sum_{\substack{w' \in A^{k-1} \\ w' \neq w}} \sum_{a_1, a_2 \in A} \pi^k(w'a_1, wa_2) &= 0. \end{aligned} \quad (22)$$

Else if $p_1^{k-1}(w) \leq p_2^{k-1}(w)$, then

$$\begin{aligned} \sum_{\substack{w' \in A^{k-1} \\ w' \neq w}} \sum_{a_1, a_2 \in A} \pi^k(wa_1, w'a_2) &= 0 \\ \sum_{\substack{w' \in A^{k-1} \\ w' \neq w}} \sum_{a_1, a_2 \in A} \pi^k(w'a_1, wa_2) &= p_2^{k-1}(w) - p_1^{k-1}(w). \end{aligned} \quad (23)$$

Theorem 1: For any $k \geq 1$, let π^k be the solution of (16). Then it holds that

$$\begin{aligned} \mathsf{K}_\mathsf{C}(p_1^{k+1}, p_2^{k+1}) &= \mathsf{K}_\mathsf{C}(p_1^k, p_2^k) \\ &+ 2^{-k} \sum_{w \in A^k} \left(r^k(w) - \sum_{a \in A} r^{k+1}(wa) \right), \end{aligned} \quad (24)$$

where $r^k(w) = \min\{p_1^k(w), p_2^k(w)\}$ for any $w \in A^k$.

Proof: For the sake of conciseness, we note $\mathsf{K}_\mathsf{C}^k = \mathsf{K}_\mathsf{C}(p_1^k, p_2^k)$. This proof is divided into two parts. First, we prove that the right-hand side of (24) is a lower bound for $\mathsf{K}_\mathsf{C}^{k+1}$. Second, we prove that it is also an upper bound.

First we note that $\mathsf{K}_\mathsf{C}^{k+1}$ is equal to

$$\begin{aligned}&\sum_{w_1,w_2\in A^k}\sum_{a_1,a_2\in A}\mathsf{C}(w_1a_1,w_2a_2)\pi^{k+1}(w_1a_1,w_2a_2)\\ =&\sum_{\substack{w_1,w_2\in A^k\\ w_1\neq w_2}}\mathsf{C}(w_1,w_2)\sum_{a_1,a_2\in A}\pi^{k+1}(w_1a_1,w_2a_2)\\ &+2^{-(k+1)}\sum_{w\in A^k}\sum_{\substack{a_1,a_2\in A\\ a_1\neq a_2}}\pi^{k+1}(wa_1,wa_2)\\ &:=C_1+C_2.\end{aligned} \tag{25}$$

In the first part of this proof, we show that the two following expressions hold:

$$C_1\geq \mathsf{K}_\mathsf{C}^k, \tag{26}$$

$$C_2=2^{-k}\sum_{w\in A^k}\left(r^k(w)-\sum_{a\in A}r^{k+1}(wa)\right). \tag{27}$$

In the first instance, we show that (26) holds. To do this, let $\mu^k : A^k\times A^k\to[0,1]$ be defined as $\mu^k(w_1,w_2)=\sum_{a_1,a_2\in A}\pi^{k+1}(w_1a_1,w_2a_2)$. We show that μ^k satisfies the constraints (17) and (18). Indeed, $\mu^k(w_1,w_2)\geq 0$, and

$$\begin{aligned}\sum_{w_2\in A^k}\mu(w_1,w_2)&=\sum_{w_2\in A^k}\sum_{a_1,a_2\in A}\pi^{k+1}(w_1a_1,w_2a_2)\\ &=\sum_{a_1}p_1^{k+1}(w_1a_1)=p_1^k(w_1),\end{aligned} \tag{28}$$

and similarly for the second condition in (18). This implies that μ^k is a coupling, thereby a feasible solution of (16). This yields

$$\mathsf{K}_\mathsf{C}^k\leq\sum_{a_1,a_2\in A}\mathsf{C}(w_1,w_2)\mu^k(w_1,w_2)=C_1. \tag{29}$$

As a second step, we show that (27) holds. More precisely, we show that, for all $w\in A^k$, the following holds:

$$\sum_{\substack{a_1,a_2\in A\\ a_1\neq a_2}}\pi^{k+1}(wa_1,wa_2)=r^k(w)-\sum_{a\in A}r^{k+1}(wa), \tag{30}$$

which implies that

$$C_2=2^{-(k+1)}\sum_{w\in A^k}\left[r^k(w)-\sum_{a\in A}r^{k+1}(wa)\right]. \tag{31}$$

We prove the claim. Assume without loss of generality that w is such that $p_1^k(w)>p_2^k(w)$, then

$$\begin{aligned}&\sum_{a_1,a_2\in A}\pi^{k+1}(wa_1,wa_2)\\ &=\sum_{a_1\in A}\sum_{w'\in A^k}\sum_{a_2\in A}\pi^{k+1}(wa_1,w'a_2)\\ &\qquad-\sum_{\substack{w'\in A^k\\ w'\neq w}}\sum_{a_2\in A}\pi^{k+1}(wa_1,w'a_2)\\ &=\sum_{a_1\in A}p_1^k(wa_1)-\sum_{\substack{w'\in A^k\\ w'\neq w}}\sum_{a_1,a_2\in A}\pi^{k+1}(wa_1,w'a_2)\end{aligned} \tag{32}$$

Following Lemma 3, this is equal to $p_1^k(w)-\big(p_1^k(w)-p_2^k(w)\big)=r^k(w)$. And the following holds:

$$\begin{aligned}&\sum_{\substack{a_1,a_2\in A\\ a_1\neq a_2}}\pi^{k+1}(wa_1,wa_2)\\ &=\sum_{\substack{a_1\neq a_2\in A\\ a_1\neq a_2}}\pi^{k+1}(wa_1,wa_2)-\sum_{a\in A}\pi^{k+1}(wa_1,wa_2)\\ &=r^k(w)-\sum_{a\in A}r^{k+1}(wa)\end{aligned} \tag{33}$$

by Lemma 2. This concludes that the right-hand side of (24) is a lower bound for $\mathsf{K}_\mathsf{C}^{k+1}$.

We now move to the second part of this proof. To provide an upper bound, we will show that we can construct a feasible $k+1$ solution feasible μ^{k+1} such that

$$\begin{aligned}&\sum_{w_1,w_2\in A^{k+1}}\mathsf{C}(w_1,w_2)\mu^{k+1}(w_1,w_2)\\ &\qquad=\mathsf{K}_\mathsf{C}^k+\sum_{w\in A^k}\left[r^k(w)-\sum_{a\in A}r^{k+1}(wa)\right].\end{aligned} \tag{34}$$

Consider π^k, an optimal solution at step k. We will construct μ^{k+1} in the following greedy way. Initialise μ^{k+1} with only zero elements, and for all $w\in A^k$, $a\in A$, we initialise $\delta(wa)=0$. We start by updating the blocks $\mu^{k+1}(w_1a_1,w_2a_2)$ where $w_1\neq w_2$. For all w such that $p_1^k(w)>p_2^k(w)$, for all $a\in A$ such that $p_1^{k+1}(wa)>p_2^{k+1}(wa)$, do the following.

1) Let $\tilde\delta(wa)=p_1^{k+1}(wa)-p_2^{k+1}(wa)$.
   If $\sum_{a'\neq a}\delta(wa')+\tilde\delta(wa)>p_1^k(w)-p_2^k(w)$, let $\delta(wa)=(p_1^k(w)-p_2^k(w))-\sum_{a'\neq a}\delta(wa')$.
   Else let $\delta(wa)=\tilde\delta(wa)$.
2) Find a $w'\neq w$ such that

$$\pi^k(w,w')>\sum_{a_1,a_2\in A}\mu^{k+1}(wa_1,w'a_2). \tag{35}$$

   Now, for any $a'\in A$, let

$$\begin{aligned}\psi(a')=&\,p_2^{k+1}(w'a')-p_1^{k+1}(w'a')\\ &-\sum_{\substack{w''\in A^k\\ w''\neq w'}}\sum_{a_1\in A}\mu^{k+1}(w''a_1,w'a')\end{aligned} \tag{36}$$

   Then, find $a'\in A$ such that $\psi(a')>0$.
   Now, if $\delta(wa)>\psi(a')$, then:
   - Update $\mu(wa,w'a')\leftarrow\psi(a')$
   - Update $\delta(wa)\leftarrow\delta(wa)-\psi(a')$
   - Return to 2.

   Else, update $\mu(wa,w'a')\leftarrow\delta(wa)$.

We claim that, in the procedure above, there always exists such a w' for a given wa. Otherwise, $\sum_{w'\neq w}\pi^k(w,w')<p_1^k(w)-p_2^k(w)$, which is impossible by Lemma 2. Also, we claim that there also always exists such a' for a given wa and

w'. Otherwise, for all $a' \in A$,

$$\begin{aligned}&\sum_{a'\in A}\sum_{\substack{w''\in A^k\\ w''\neq w'}}\sum_{a_1\in A}\mu^{k+1}(w''a_1, w'a')\\ &\quad=\sum_{a'\in A} p_2^{k+1}(w'a') - p_1^{k+1}(w'a'),\end{aligned} \tag{37}$$

which means by construction that

$$\begin{aligned}&\sum_{a'\in A}\sum_{\substack{w''\in A^k\\ w''\neq w'}}\sum_{a_1\in A}\pi^{k+1}(w''a_1, w'a')\\ &\quad>\sum_{a'\in A} p_2^{k+1}(w'a') - p_1^{k+1}(w'a'),\end{aligned} \tag{38}$$

which is $p_2^k(w) - p_1^k(w) > p_2^k(w) - p_1^k(w)$ by Lemma 3. Moreover, by construction we have that, for all $w \neq w'$,

$$\pi^k(w, w') = \sum_{a_1,a_2\in A^k}\mu^{k+1}(wa_1, w'a_2). \tag{39}$$

Now, we construct the diagonal blocks $\mu^{k+1}(wa_1, wa_2)$. For each $w \in A^k$ and $a \in A$, let

$$\begin{aligned}\tilde{p}_1^{k+1}(wa) &= p_1^{k+1}(wa) - \sum_{w'\neq w}\sum_{a\in A}\mu^{k+1}(wa, w'a'),\\ \tilde{p}_2^{k+1}(wa) &= p_2^{k+1}(wa) - \sum_{w'\neq w}\sum_{a\in A}\mu^{k+1}(w'a', wa).\end{aligned} \tag{40}$$

Now, for a given w, let us solve the following balanced optimal transport problem:

$$\begin{aligned}&\inf_{\mu^{k+1}} 2^{-k}\sum_{\substack{a_1,a_2\in A\\ a_1\neq a_2}}\mu^{k+1}(wa_1, wa_2)\\ &\text{s.t. } \forall a_1 \in A: \sum_{a_2}\mu^{k+1}(wa_1, wa_2) = \tilde{p}_1^{k+1}(wa_1),\\ &\quad\forall a_2 \in A: \sum_{a_1}\mu^{k+1}(wa_1, wa_2) = \tilde{p}_2^{k+1}(wa_2).\end{aligned} \tag{41}$$

Following the definition of $\tilde{p}$ and $\tilde{q}$, this is a balanced optimal transport whose trivial solution is given by

$$2^{-k}\left(r^k(w) - \sum_{a\in A} r^{k+1}(wa)\right). \tag{42}$$

Now we conclude the proof. By (39) and (40), μ^{k+1} is a coupling of p_1^{k+1} and p_2^{k+1}. Indeed it is positive, and for any $w_1 \in A^k$ and $a_1 \in A$,

$$\begin{aligned}&\sum_{w_2\in A^k}\sum_{a_2\in A}\mu^{k+1}(w_1a_1, w_2a_2)\\ &=\sum_{a_2\in A}\mu^{k+1}(w_1a_1, w_1a_2) + \sum_{\substack{w_2\in A^k\\ w_2\neq w_1}}\sum_{a_2\in A}\mu^{k+1}(w_1a_1, w_2a_2)\\ &=\tilde{p}_1^{k+1}(w_1a_1) + \left(p_1^{k+1}(w_1a_1) - \tilde{p}_1^{k+1}(w_1a_1)\right) = p_1^{k+1}(w_1a_1),\end{aligned} \tag{43}$$

and similarly for p_2^{k+1}. Finally,

$$\begin{aligned}&\sum_{w_1,w_2\in A^k}\sum_{a_1,a_2\in A}\mathsf{C}(w_1a_1, w_2a_2)\mu^{k+1}(w_1a_1, w_2a_2)\\ &=\sum_{w_1\neq w_2}\mathsf{C}(w_1, w_2)\sum_{a_1,a_2}\mu^{k+1}(w_1a_1, w_2a_2)\\ &+2^{-k}\sum_{w}\sum_{\substack{a_1,a_2\\ a_1\neq a_2}}\mu^{k+1}(wa_1, wa_2).\end{aligned} \tag{44}$$

By (39), the first term is K_C^k, and by (42), the second term is $\sum_{w\in A^k}\left[r^k(w) - \sum_{a\in A} r^{k+1}(wa)\right]$. This provides an upper bound on $\mathsf{K}_\mathsf{C}^{k+1}$, and the proof is completed. ■

Theorem 1 is central for proving Theorem 2, stated next. The latter first states that CK satisfies positivity, symmetry and triangle inequality. This also implies that the Cantor-Kantorovich metric can be approximated with $\mathsf{K}_\mathsf{C}(p_1^k, p_2^k)$ for a finite k.

Theorem 2: The function CK satisfies positivity, symmetry and triangle inequality. Moreover, for any $k \geq 1$,

$$0 \leq \mathsf{CK}(\Gamma_1, \Gamma_2) - \mathsf{K}_\mathsf{C}(p_1^k, p_2^k) \leq 2^{1-k}\sum_{w\in A^k} r^k(w). \tag{45}$$

Proof: For the sake of conciseness, let $\mathsf{K}_\mathsf{C}^k := \mathsf{K}_\mathsf{C}(p_1^k, p_2^k)$, and let $S_k := \sum_{w\in A^k} r^k(w)$. Proving that the metric is well-defined reduces to proving that the sequence $(\mathsf{K}_\mathsf{C}^k)_{k\geq 1}$ converges. Theorem 1 implies that

$$\mathsf{K}_\mathsf{C}^k = (1 - S_1) + \sum_{i=1}^{k-1} 2^{-i}(S_i - S_{i+1}). \tag{46}$$

Now let us focus on $S_i - S_{i+1}$. By the law of total probability, for all $i \geq 1$, we have that $0 \leq r^i(w) - \sum_{a\in A} r^{i+1}(wa) \leq r^i(w)$, and therefore, by summing over all the words $w \in A^i$, we have that

$$0 \leq S_i - S_{i+1} \leq S_i \leq 1. \tag{47}$$

One can see that the maximal value of (46) is attained when $S_1 = 0$, which implies that $S_i = 0$ for $i = 2, 3, \ldots$, and therefore the sequence $(\mathsf{K}_\mathsf{C}^k)_{k\geq 1}$ is upper bounded by 1. Moreover, one can assert that the sequence is also monotone, following (47). By the monotone convergence theorem, the limit therefore exists and is equal to $\mathsf{CK}(\Gamma_1, \Gamma_2) = \lim_{k\to\infty}\mathsf{K}_\mathsf{C}^k = \sup_{k\geq 1}\mathsf{K}_\mathsf{C}^k$.

Now, it remains to show that (45) holds. Inequalities (47) implies that

$$\begin{aligned}\mathsf{CK}(\Gamma_1, \Gamma_2) - \mathsf{K}_\mathsf{C}^p &= \sum_{i=p}^{\infty} 2^{-i}(S_i - S_{i+1})\\ &\leq \sum_{i=p}^{\infty} 2^{-i}S_i \leq S_p\sum_{i=p}^{\infty} 2^{-i} = 2^{1-p}S_p,\end{aligned} \tag{48}$$

which is the claim. ■

Remark 4: Following Theorem 2, one can compute a-priori the number of iterations needed to reach an accuracy $\varepsilon \in (0, 1)$. Indeed the right-hand side of (45) is upper bounded by 2^{1-k}, and therefore it suffices to take $k = \lceil\log_2(\varepsilon^{-1})\rceil + 1$ to guarantee an ε-accurate solution. □

Based on the above results, let us introduce Algorithm 3, which takes as input two labeled Markov chains Γ_1 and

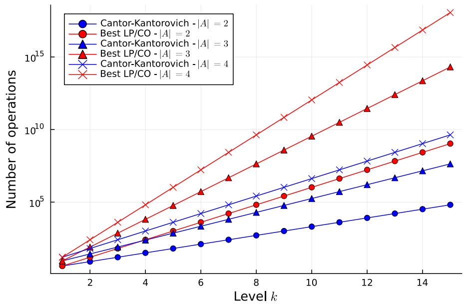


Fig. 4: Comparaison of the computational complexity of two methods to solve (16) for $|A| = 2, 3, 4$. In red, the best LP/CO method, and in blue our method as described in Algorithm 3.

Γ_2, and a desired accuracy ε, and computes the metric of interest. This algorithm relies on the recursive algorithm CK-REC, which is a natural implementation of Theorem 1 that is described in Algorithm 2.

Algorithm 2 CK-REC(ACC, l, w, k)

$r = \min\{p_1^l(w), p_2^l(w)\}$
if $r = 0$ **then**
  Stop
end if
if $l = k$ **then**
  ACC $\leftarrow$ ACC $+ 2^{1-k}r$
  Stop
end if
ACC $\leftarrow$ ACC $+ 2^{-k}r$
for $a \in A$ **do**
  CK-REC(ACC, $l+1$, wa, k)
end for

Algorithm 3 CANTOR-KANTOROVICH($\Gamma_1, \Gamma, \varepsilon$)

For $l = 1, \dots, k$, compute $p_1^l(w)$ and $p_2^l(w)$ See Remark 1
$k \leftarrow \lceil \log_2(\varepsilon^{-1}) \rceil + 1$ See Remark 4
ACC $\leftarrow 0$
for $a \in A$ **do**
  CK-REC(ACC, 1, a, k)
end for
return ACC

Algorithm 3 terminates in $2k|S|^2 + \mathcal{O}(|A|^{k+1})$ operations. Indeed, the first term is the number of operations needed to compute the probabilities $p_1^l(w)$ and $p_2^l(w)$ for $l = 1, \dots, k$ (see Remark 1), and the second term is the number of operations corresponding for a DFS[4] in a tree of $|A|^{k+1}$ nodes with a constant number of operations at each node. In order to illustrate the gain in complexity compared to classical LP or CO methods, we provide in Figure 4 the functions $|A|^{k+1}$ and $|A|^{2k}$ for $|A| = 2, 3, 4$ for $k = 1, \dots, 15$ (see Remark 3).

[4]Depth-First-Search, see e.g. [35] for an introduction.

V. APPLICATION: CK METRIC FOR ABSTRACTIONS

In this section, we apply the REFINE procedure described in Algorithm 1 with the Cantor-Kantorovich metric CK to abstract a given dynamical system.

A. Electron subject to Lorentz force

We are interested in the position of an electron subject to the Lorentz force (see e.g. [36]). The law is given by

$$m \begin{pmatrix} \dot{v}_1(t) \\ \dot{v}_2(t) \\ \dot{v}_3(t) \end{pmatrix} = q \begin{pmatrix} E_1 + B_3 v_2(t) - B_2 v_3(t) \\ E_2 + B_1 v_3(t) - B_3 v_1(t) \\ E_3 + B_2 v_1(t) - B_1 v_2(t) \end{pmatrix}, \tag{49}$$

where $v_1(t)$, $v_2(t)$ and $v_3(t)$ are respectively the x-axis, y-axis and z-axis components of the velocity of the electron in [m/s]. Every other constant and their unities are given in Table I.

Symbol	Physical quantity	Unit	Value
m	Mass	[kg]	9.1×10^{-31}
q	Charge	[C]	1.6×10^{-19}
E_1	x-axis electric field	[V/m]	-1.0×10^{-10}
E_2	y-axis electric field	[V/m]	5.0×10^{-11}
E_3	z-axis electric field	[V/m]	0
B_1	x-axis magnetic field	[T]	0
B_2	y-axis magnetic field	[T]	0
B_3	z-axis magnetic field	[T]	1.0×10^{-11}

TABLE I: Constants in the Lorentz force equation (49)

Since $B_1 = B_2 = E_3 = 0$, then $v_3(t) = 0$ and the dynamical system can be written on a 2D-plane. Moreover, we are interested in the position of the electron, rather than its velocity: hence, the dynamical equations become the description of a 4-dimensional affine dynamical system given by

$$\begin{pmatrix} \dot{p}_1(t) \\ \dot{p}_2(t) \\ \dot{v}_1(t) \\ \dot{v}_2(t) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & qB_3/m \\ 0 & 0 & -qB_3/m & 0 \end{pmatrix} \begin{pmatrix} p_1(t) \\ p_2(t) \\ v_1(t) \\ v_2(t) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ qE_1/m \\ qE_2/m \end{pmatrix}, \tag{50}$$

where $p_1(t)$ and $p_3(t)$ respectively denote the x-axis and y-axis components of the position of the electron.

We study this dynamical system in discrete time. To do that, we approximate the derivative with the explicit Euler scheme $\dot{x}(t) \approx (x_{t+1} - x_t)/h$, with $h = 0.1$, which gives $x_{t+1} = (I + hA)x_t + hb$, with $x_t = (p_{1,t}, p_{2,t}, v_{1,t}, v_{2,t}) \in \mathbb{R}^4$, $I \in \mathbb{R}^{4\times 4}$ the identity matrix, and $A \in \mathbb{R}^{4\times 4}$ and $b \in \mathbb{R}^4$ respectively the matrix and vectors of the continous-time affine system given in (50). We consider an obstacle defined by $O = [0.5, 1.5] \times [-0.5, 0.5] \times \mathbb{R} \times \mathbb{R}$. The output function h is defined as

$$h(x_t) = \begin{cases} 0 & \text{if } x_t \in O, \\ 1 & \text{else if } p_{1,t}, \geq 1.5 \\ 2 & \text{otherwise.} \end{cases} \tag{51}$$

Finally, we take the uniform measure on $[-1, 4] \times [-1, 1] \times [-1, 1] \times [-1, 1]$ for sampling the initial state, which defines λ. An illustration of the dynamical system Σ is given in Figure 5.

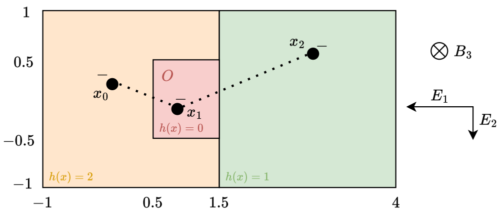


Fig. 5: Illustration of the Lorentz force dynamical system Σ as defined in Section V-A.

B. Abstraction-based analysis

In this section, we are interested in approximating the initial safe set measure

$$P_H = \lambda(\{x_0 \in \mathbb{R}^4 : x_t \in \mathbb{R}^4 \setminus O \, \forall t = 0, \dots, H\}), \quad (52)$$

where $H \in \mathbb{N}$ is a given horizon. We show that our procedure REFINE with the Cantor-Kantorovich metric CK yields better results than explicit grid-based approaches. In all of our numerical experiments, we approximated the Cantor-Kantorivich metric with the CANTOR-KANTOROVICH algorithm descripted in Algorithm 3 with $\varepsilon = 10^{-3}$.

We proceed as follows. We are given a labeled Markov chain $\Gamma = (S, A, \tau, \mu, l)$ abstracting the dynamical system where each state $s \in S$ corresponds to a block of a partition of the state space. If a block has a non-empty intersection with the obstacle O, we label the corresponding state as unsafe, and the other ones as safe. The set of safe states is denoted as S_{safe}. To approximate P_H, for a given level of confidence $\beta \in (0, 1)$, we identify the set of states from which the probability of a safe random walk is greater than $1 - \beta$, noted S_β and formally defined as the set of $s_0 \in S$ such that

$$\sum_{s_1 \in S_{\text{safe}}} \tau_{s_0, s_1} \cdots \sum_{s_H \in S_{\text{safe}}} \tau_{s_{H-1}, s_H} \geq 1 - \beta. \quad (53)$$

Then, for a given β, P_H is approximated as

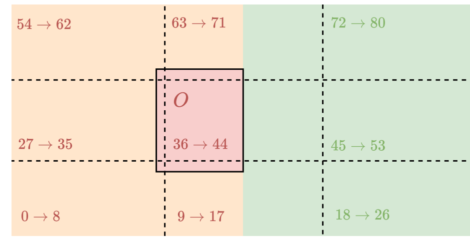
$$P_H \approx \sum_{s_0 \in S_\beta} \mu_{s_0}. \quad (54)$$

We compare two different approaches. In the first one, we uniformly grid the state space. If each dimension is divided into p parts, there will be p^4 blocks, noted $B_1, \dots, B_{p^4}$. From this partition, we compute a labeled Markov chain where the states correspond to the p^4 blocks with initial measure $\mu_{B_i} = \lambda(B_i)$, and where the transition probability from B_i to B_j is

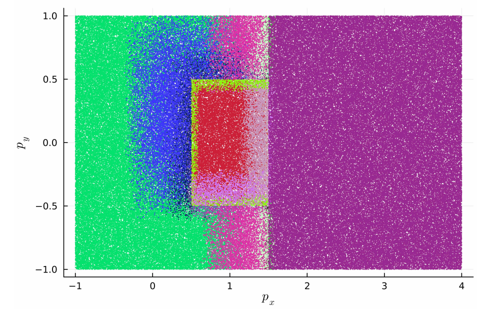
$$P_{B_i, B_j} = \frac{\lambda(f^{-1}(B_j) \cap B_i)}{\lambda(B_i)}. \quad (55)$$

We consider two labeled Markov chains constructed in this way, respectively with $p = 2$ (16 states) and $p = 3$ (81 states). The second approach is to compute the adaptive memory abstraction Γ_W with $W = \text{REFINE}(\Sigma, \lambda, \text{CK}, N)$. Again we study two cases, $N = 6$ (15 states) and $N = 14$ (31 states). An illustration of the two approaches can be found in Figure 6.

We compare the approximation of P_H with the four abstractions descripted above, and for different confidence levels. The approximation results and the true probability P_H can be found in Figure 7. We start by discussing the results of the abstractions generated by the REFINE algorithm (green and orange lines on Figure 7). When the asked level of confidence is too strict, for example $\beta = 0.01$, then one needs more states to well approximate P_H, as we can see on Figure 7a. On the other hand, when the level of confidence is too high, for example $\beta = 0.25$, then it is more likely to to overapproximate P_H, as one can see on Figure 7b for $H = 4$ with the 15 states abstraction. For this example, the confidence level is reasonable when it is equal to 0.05, as one can see on Figure 7c. In this case, the 31 states abstraction perfectly approximates P_H until $H = 8$. Now, one can see that the approximation is better using a the REFINE algorithm compared to the naïve approach. Indeed the former with 15 states always estimates better P_H than the latter with 81 states.



(a) Partition with classical approach.

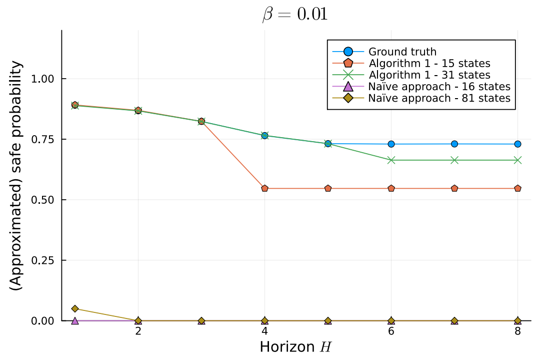


(b) Partition with Algorithm 1.

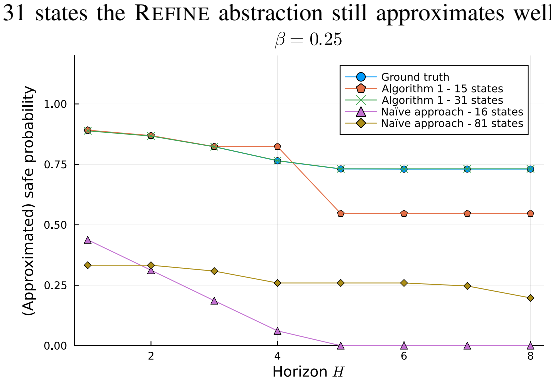
Fig. 6: Comparison between a classical partition and the partition computed with Algorithm 1 on a four dimensional example (here projected on (p_x, p_y)). Above: the classical approach, where each dimension is divided into three parts, which gives a total of $3^4 = 81$ blocks. Below: our adaptive memory abstraction with $N = 6$, which leads to 15 blocks in the partition. Since the algorithm does not explicitely compute a closed form expression for the blocks, we illustrate them by coloring samples. While there are less blocks here than the 81 blocks in the above partition, one sees that they are computed smartly, and the critical places in the system (close to the obstacles) benefit from a finer resolution than the large regions away from the obstacle.

VI. CONCLUSIONS

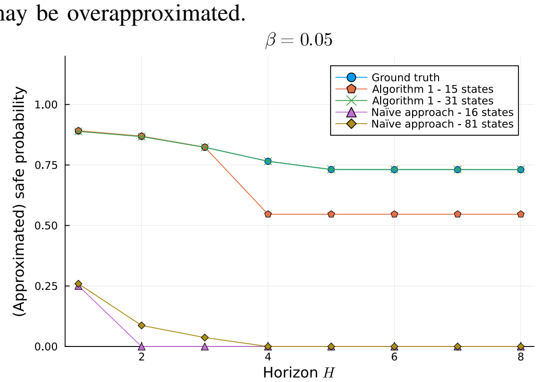
In this work, we have introduced Markov models encoding adaptive-memory schemes to abstract dynamical systems via samples. We have proved that the abstractions preserve safety properties of the abstracted systems. Along with this, given a notion of metric between Markov models, we have proposed a



(a) Confidence level $\beta = 0.01$. The confidence level is too low, but with 31 states the REFINE abstraction still approximates well P_H.



(b) Confidence level $\beta = 0.25$. The confidence level is too high, and P_H may be overapproximated.



(c) Confidence level $\beta = 0.05$. The confidence level is reasonable, and the REFINE abstractions yield good results with both 15 and 31 states.

Fig. 7: Approximation of P_H for $H = 0, \ldots, 8$ with different abstractions and for different confidence levels β.

generic and tuneable procedure to choose a convenient memory scheme. We have then introduced a Cantor-Kantorovich metric for Markov chains, proving that the latter satisfies positivity, symmetry and triangle inequality, and we have provided an algorithm to approximate it efficiently. We have finally performed an abstraction-based safety analysis on a example corresponding to a real-life dynamical system. We showed that our method yields better approximations than classical, grid-based abstraction approaches in the litterature.

For further work, we would like to test our method with other metrics between Markov models, and compare numerical results. For example, one could investigate metrics that encode control specifications, unlike our metric. On the other hand, we would like to improve the algorithm for computing the Cantor-Kantorovich metric.

APPENDIX I
CONSEQUENCE OF ASSUMPTION 1

Proposition 2: Given a dynamical system Σ, a measure λ and a set W that defines a partition, if Assumption 1 is satisfied, then the adaptive memory abstraction $\Gamma_W = (W, A, \tau, \mu, l)$ has a stochastic transition matrix τ.

Proof: By (12) and Assumption 1, it suffices to show that, for all $w_{[-s,t]} \in W$,

$$\sum_{w'_{[-s',t']} \in W} \lambda\left(f^{-1}\left(\left[w'_{[-s',t']}\right]\right) \cap \left[w_{[-s,t]}\right]\right) = \lambda\left(\left[w_{[-s,t]}\right]\right) \quad (56)$$

is satisfied. Since the sets $f^{-1}\left(\left[w_{[-s,t]}\right]\right)$ also form a partition of X, (56) holds by the law of total probability, and the proof is completed. ■

APPENDIX II
PROCEDURE REFINE YIELDS A PARTITION

Assumption 2: The dynamical system Σ and measure λ are such that, for all k, for all $w \in A^k$, if $\lambda([w]_{[0,k-1]}) = 0$, then $[w]_{[0,k-1]} = \emptyset$.

Proposition 3: Suppose that Assumption 2 holds. Let Σ be a dynamical system with initial measure λ, and let D be a metric between labeled Markov chains. Then, for all $N \in \mathbb{N}_{>0}$ the output of $\text{REFINE}(\Sigma, \lambda, \mathsf{D}, N)$ defines a partition of X and the matrix of the corresponding abstraction is stochastic.

Proof: By Proposition 2, we have to show that each current model Γ_W is such that W defines a partition of X and satisfies Assumption 1. We prove the claim by induction.

First, we prove that the set W such as defined after line 4 defines a partition and satisfies Assumption 1. The set $W = \{(a)_{[0,0]}\}_{a \in A}$ defines a partition by definition of the output function h. Now, the set $W \setminus \{(a)_{[0,0]}\}_{a \in A_0}$ satisfies Assumption 1 by construction. Moreover, it still defines a partition because the sets $[a_{[0,0]}]$ for $a \in A_0$ are all empty by Assumption 2.

We now assume that W defines a partition and satisfies Assumption 1. We prove that, for any $w^i_{[0,t_i]} \in W$, the set W'_i such as after 12 is a partition and satisfies Assumption 1. We first show that the expanded set $\tilde{W} = (W \setminus \{w^i_{[0,t_i]}\}) \cup \{(w^i a)_{[0,t_i+1]}\}_{a \in A}$ defines a partition. Condition 4 is satisfied because

$$\begin{aligned}
&\bigcup_{w_{[0,t]} \in \tilde{W}} \left[w_{[0,t]}\right] \\
&= \left(\bigcup_{w_{[0,t]} \in W \setminus \left\{w^i_{[0,t_i]}\right\}} \left[w_{[0,t]}\right]\right) \cup \left(\bigcup_{a \in A} \left[(w^i a)_{[0,t_i+1]}\right]\right) \\
&= \left(\bigcup_{w_{[0,t]} \in W \setminus \left\{w^i_{[0,t_i]}\right\}} \left[w_{[0,t]}\right]\right) \cup \left[w^i_{[0,t_i]}\right],
\end{aligned} \quad (57)$$

which is the whole state space X by the recursion assumption. And condition 5 is also satisfied because $\left[(w^i a)_{[0,t_i+1]}\right] \subseteq \left[w^i_{[0,t_i]}\right]$ for all $a \in A$. Now, with W_0 as defined in line 11, the $\tilde{W} \setminus W_0$ satisfies Assumption 1 by construction. Finally, it still defines a partition because the sets $[w_{[0,t]}]$ for $w_{[0,t]} \in W_0$ are empty by Assumption 2, and the proof is completed. ■

Appendix III
Proof of Lemma 2

We first prove that $\pi^k(w,w) \le \min\{p_1^k(w), p_2^k(w)\}$. Constraints (18) imply that

$$p_1^k(w) = \pi^k(w,w) + \sum_{\substack{w' \in A^k \\ w \neq w'}} \pi^k(w,w') \ge \pi^k(w,w),$$
$$p_2^k(w) = \pi^k(w,w) + \sum_{\substack{w' \in A^k \\ w \neq w'}} \pi^k(w',w) \ge \pi^k(w,w), \tag{58}$$

which implies $\pi^k(w,w) \le \min\{p_1^k(w), p_2^k(w)\}$. We now prove that $\pi^k(w,w) \ge \min\{p_1^k(w), p_2^k(w)\}$. By contradiction, let π^k be an optimal solution to (16) such that there exists $w \in A^k$ and $\varepsilon > 0$ with $\pi^k(w,w) = \min\{p_1^k(w), p_2^k(w)\} - \varepsilon$. Assume without loss of generality that $\min\{p_1^k(w), p_2^k(w)\} = p_1^k(w)$, then constraints (18) imply that

1) there exists $w' \neq w$ such that $\pi^k(w,w') = \varepsilon'$ for some $\varepsilon' \in (0, \varepsilon]$, and
2) there exists $w'' \neq w$ such that $\pi^k(w'',w) = \varepsilon''$ for some $\varepsilon'' \in (0, \varepsilon]$.

Let $\mathsf{K}_{\mathsf{C}}(p_1^k, p_2^k)$ denote the Kantorovich metric corresponding to such π^k. Now assume, again without loss of generality, that $\varepsilon' \le \varepsilon''$. Consider then $(\pi^k)'$ such that $(\pi^k)'(w_1,w_2) = \pi^k(w_1,w_2)$ for all $w_1, w_2 \in A^k$ except

1) $(\pi^k)'(w,w) = \pi^k(w,w) + \varepsilon'$,
2) $(\pi^k)'(w,w') = \pi^k(w,w') - \varepsilon'$,
3) $(\pi^k)'(w'',w') = \pi^k(w'',w') + \varepsilon'$,
4) $(\pi^k)'(w'',w) = \pi^k(w'',w) - \varepsilon'$.

The joint distribution $(\pi^k)'$ is feasible since it satisfies constraints (17) and (18). Now let $\mathsf{K}'_{\mathsf{C}}(p_1^k, p_2^k)$ denote the solution corresponding to such $(\pi^k)'$, it holds that

$$\mathsf{K}'_{\mathsf{C}}(p_1^k, p_2^k) = \mathsf{K}_{\mathsf{C}}(p_1^k, p_2^k) - \varepsilon'[\mathsf{C}(w,w') + \mathsf{C}(w',w'') - \mathsf{C}(w',w'')]. \tag{59}$$

By triangular inequality of C, we have that $\mathsf{K}'_{\mathsf{C}}(p_1^k, p_2^k) < \mathsf{K}_{\mathsf{C}}(p_1^k, p_2^k)$, which contradicts the fact that π^k is optimal, and the proof is completed. ■

Appendix IV
Proof of Lemma 3

For some $w \in A^{k-1}$, assume without loss of generality that $p_1^{k-1}(w) > p_2^{k-1}(w)$. First, by the constraints (18), it holds that

$$\sum_{\substack{w' \in A^{k-1} \\ w' \neq w}} \sum_{a_1, a_2 \in A} \pi^k(wa_1, w'a_2) \ge p_1^{k-1}(w) - p_2^{k-1}(w). \tag{60}$$

Now, we proceed similarly as for Lemma 2. Suppose by contradiction that $\sum_{\substack{w' \in A^{k-1} \\ w' \neq w}} \sum_{a_1,a_2 \in A} \pi^k(wa_1, w'a_2) > p_1^{k-1}(w) - p_2^{k-1}(w)$. Then there exists $w' \neq w \in A^{k-1}$, and $a_1, a_2 \in A$ such that $\pi^k(w'a_1, wa_2) = \varepsilon' > 0$. There also exists $w'' \in A^{k-1}$ such that $w'' \neq w$ and $w'' \neq w'$, and $a_3, a_4 \in A$ such that $\pi^k(wa_3, w''a_4) = \varepsilon'' > 0$. Assume w.l.o.g. that $\varepsilon' \le \varepsilon''$, and consider a solution $(\pi^k)'$ such that $(\pi^k)' = \pi^k$, except for

1) $(\pi^k)'(wa_3, w''a_4) = \pi^k(wa_3, w''a_4) - \varepsilon'$,
2) $(\pi^k)'(w'a_1, wa_2) = \pi^k(w'a_1, wa_2) - \varepsilon'$,
3) $(\pi^k)'(w'a_1, w''a_4) = \pi^k(w'a_1, w''a_4) + \varepsilon'$, and
4) $(\pi^k)'(wa_3, wa_2) = \pi^k(wa_3, wa_2) + \varepsilon'$.

The joint distribution $(\pi^k)'$ is feasible since it still satisfies the constraints (17) and (18). Note that, since the Cantor distance satisfies the strong triangular inequality (see Lemma 1),

$$\begin{aligned} \mathsf{C}(w'a_1, w''a_4) &\le \max\{\mathsf{C}(w'a_1, wa_2), \mathsf{C}(wa_2, w''a_4)\} \\ &= \max\{\mathsf{C}(w'a_1, wa_2), \mathsf{C}(wa_3, w''a_4)\}. \end{aligned} \tag{61}$$

Moreover, $\mathsf{C}(wa_3, wa_2) = 2^{-k}$. Now let $\mathsf{K}'_{\mathsf{C}}(p_1^k, p_2^k)$ denote the solution corresponding to such $(\pi^k)'$, we have that $\mathsf{K}_{\mathsf{C}}(p_1^k, p_2^k) - \mathsf{K}'_{\mathsf{C}}(p_1^k, p_2^k)$ is

$$\begin{aligned} \mathsf{K}_{\mathsf{C}}(p_1^k, p_2^k) - \mathsf{K}'_{\mathsf{C}}(p_1^k, p_2^k) &= -\varepsilon' \begin{bmatrix} + & \mathsf{C}(w'a_1, wa_2) \\ + & \mathsf{C}(wa_3, w''a_4) \\ - & \mathsf{C}(w'a_1, w''a_4) \\ - & 2^{-k} \end{bmatrix} \\ &\le -\varepsilon' \begin{bmatrix} + & \mathsf{C}(w'a_1, wa_2) + \mathsf{C}(wa_3, w''a_4) \\ - & \max\{\mathsf{C}(w'a_1, wa_2), \mathsf{C}(wa_3, w''a_4)\} \\ - & 2^{-k} \end{bmatrix} \\ &\le -\varepsilon'[2^{-(k-1)} - 2^{-k}] \\ &\le 0, \end{aligned} \tag{62}$$

which contradicts the fact that π^k is optimal. ■

References

[1] E. Lee, "The past, present and future of cyber-physical systems: A focus on models," *Sensors*, vol. 15, pp. 4837–4869, Feb. 2015.

[2] K.-D. Kim and P. R. Kumar, "Cyberphysical systems: A perspective at the centennial," *Proceedings of the IEEE*, vol. 100, pp. 1287–1308, May 2012.

[3] R. Alur, *Principles of Cyber-Physical Systems*. MIT Press, 2015.

[4] E. A. Lee and S. A. Seshia, *Introduction to Embedded Systems A Cyber-Physical Systems Approach*. MIT Press, 2016.

[5] P. Tabuada, *Verification and Control of Hybrid Systems: A Symbolic Approach*. Springer US, 2009.

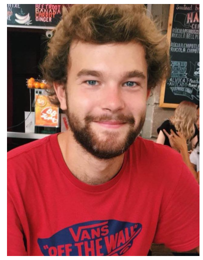
[6] T. Badings, L. Romao, A. Abate, and N. Jansen, "Probabilities are not enough: Formal controller synthesis for stochastic dynamical models with epistemic uncertainty," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 14701–14710, June 2023.

[7] T. Badings, L. Romao, A. Abate, D. Parker, H. Poonawala, M. Stoelinga, and N. Jansen, "Robust control for dynamical systems with non-Gaussian noise via formal abstractions," *JAIR*, vol. 76, pp. 341–391, 2023.

[8] A. Banse, L. Romao, A. Abate, and R. Jungers, "Data-driven memory-dependent abstractions of dynamical systems," in *Proceedings of The 5th Annual Learning for Dynamics and Control Conference* (N. Matni, M. Morari, and G. J. Pappas, eds.), vol. 211 of *Proceedings of Machine Learning Research*, pp. 891–902, PMLR, 15–16 Jun 2023.

[9] A. Lavaei, S. Soudjani, A. Abate, and M. Zamani, "Automated verification and synthesis of stochastic hybrid systems: A survey," *Automatica*, vol. 146, p. 110617, Dec. 2022.

[10] A. van der Schaft, "Equivalence of dynamical systems by bisimulation," *IEEE Transactions on Automatic Control*, vol. 49, pp. 2160–2172, Dec. 2004.
[11] S. Haesaert, S. Soudjani, and A. Abate, "Verification of general Markov decision processes by approximate similarity relations and policy refinement," *SIAM Journal on Control and Optimisation*, vol. 55, no. 4, pp. 2333–2367, 2017.
[12] G. Reissig and M. Rungger, "Feedback refinement relations for symbolic controller synthesis," in *53rd IEEE Conference on Decision and Control*, IEEE, Dec. 2014.
[13] S. E. Z. Soudjani and A. Abate, "Adaptive and sequential gridding for abstraction and verification of stochastic processes," *SIAM Journal on Applied Dynamical Systems*, vol. 12, no. 2, pp. 921–956, 2012.
[14] A.-K. Schmuck and J. Raisch, "Asynchronous l-complete approximations," *Systems and Control Letters*, vol. 73, pp. 67–75, Nov. 2014.
[15] R. Majumdar, N. Ozay, and A.-K. Schmuck, "On abstraction-based controller design with output feedback," in *Proceedings of the 23rd International Conference on Hybrid Systems: Computation and Control*, HSCC 20, ACM, Apr. 2020.
[16] C. De Persis and P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness," *IEEE Transactions on Automatic Control*, vol. 65, pp. 909–924, Mar. 2020.
[17] Z. Wang and R. M. Jungers, "A data-driven method for computing polyhedral invariant sets of black-box switched linear systems," *IEEE Control Systems Letters*, vol. 5, pp. 1843–1848, Nov. 2021.
[18] J. Berberich, J. Kohler, M. A. Muller, and F. Allgower, "Data-driven model predictive control with stability and robustness guarantees," *IEEE Transactions on Automatic Control*, vol. 66, pp. 1702–1717, Apr. 2021.
[19] R. Coppola, A. Peruffo, and M. Mazo, "Data-driven abstractions for verification of linear systems," *IEEE Control Systems Letters*, pp. 1–1, 2023.
[20] R. Coppola, A. Peruffo, and M. Mazo, "Data-driven abstractions for control systems," 2024.
[21] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden, "Metrics for labelled Markov processes," *Theoretical Computer Science*, vol. 318, pp. 323–354, June 2004.
[22] F. van Breugel, B. Sharma, and J. Worrell, *Approximating a Behavioural Pseudometric Without Discount for Probabilistic Systems*, pp. 123–137. Springer Berlin Heidelberg, 2007.
[23] N. Madras and D. Sezer, "Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances," *Bernoulli*, vol. 16, Aug. 2010.
[24] D. Rudolf and N. Schweizer, "Perturbation theory for Markov chains via Wasserstein distance," *Bernoulli*, vol. 24, Nov. 2018.
[25] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Nov. 1995.
[26] N. P. Fogg, S. bastien Ferenczi, C. Mauduit, and A. Siegel, *Substitutions in Dynamics, Arithmetics and Combinatorics*. Springer-Verlag Berlin Heidelberg, 2002.
[27] G. Rozenberg and A. Salomaa, *Handbook of Formal Languages*. Springer, 2002.
[28] A. Banse, L. Romao, A. Abate, and R. M. Jungers, "Data-driven abstractions via adaptive refinements and a Kantorovich metric," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, IEEE, Dec. 2023.
[29] J. Garca, . Viss, and F. Fernndez, "A taxonomy for similarity metrics between Markov decision processes," *Machine Learning*, vol. 111, pp. 4217–4247, Oct. 2022.
[30] T. Chen and S. Kiefer, "On the total variation distance of labelled Markov chains," in *Proceedings of the Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, CSL-LICS 14, ACM, July 2014.
[31] Z. Rached, F. Alajaji, and L. Campbell, "The KullbackLeibler divergence rate between Markov sources," *IEEE Transactions on Information Theory*, vol. 50, pp. 917–921, May 2004.
[32] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
[33] R. S. Sutton and A. G. Barto, *Reinforcement Learning An Introduction*. A Bradford Book, 2018.
[34] A. M. Oberman and Y. Ruan, "An efficient linear programming method for optimal transportation," 2015.
[35] R. Sedgewick, *Algorithms*. Addison-Wesley, 2011.
[36] I. Grant and W. Phillips, *Electromagnetism*. Manchester Physics Series, Wiley, 2013.



Adrien Banse received his M.Sc. degree in Mathematical Engineering from UCLouvain, Belgium, in 2022. How is now a Ph.D. candidate at the Department of Mathematical Engineering in the Institute of Information and Communication Technologies (INMA/ICTEAM) at UCLouvain. His research interests lie are at the intersection of Computer Science and Applied Mathematics, including data-driven formal methods and analysis of dynamical systems. He is a FRIA/FNRS fellow.

Licio Romao received the B.Sc. degree in electrical engineering from the Federal University of Campina Grande (UFCG), Campina Grande, Brazil, in 2014, the M.Sc. degree in electrical engineering from the University of Campinas (UNICAMP), Campinas, Brazil, in 2017, and the D.Phil. (Ph.D.) degree in engineering science from the University of Oxford, Oxford, U.K., in 2021. He is currently a Postdoctoral Researcher with the Department of Aeronautics and Astronautics, Stanford University, USA, and has previously worked as a postdoctoral researcher with the Department of Computer Science, University of Oxford. His research focuses on the design of feedback controllers for stochastic dynamical systems. It leverages tools from formal methods, optimisation, statistics, stochastic calculus, and control theory. He is a recipient of the 2021 IET Control and Automation Doctoral Dissertation prize.

Alessandro Abate is Professor of Verification and Control in the Department of Computer Science at the University of Oxford. Earlier, he did research at Stanford University and at SRI International, and was an Assistant Professor at the Delft Center for Systems and Control, TU Delft. He received MS and PhD from the University of Padua and UC Berkeley. His research interests lie on the formal verification and control of stochastic hybrid systems, and in their applications to cyber-physical systems, particularly involving safety criticality and energy. He blends in techniques from probability theory and logic, machine learning and AI, towards assured autonomy.

Raphaël M. Jungers is a Professor at UCLouvain, Belgium. His main interests lie in the fields of Computer Science, Graph Theory, Optimization and Control. He received a Ph.D. in Mathematical Engineering from UCLouvain (2008), and a M.Sc. in Applied Mathematics, both from the Ecole Centrale Paris, (2004), and from UCLouvain (2005). He has held various invited positions, at the Université Libre de Bruxelles (2008-2009), at the Laboratory for Information and Decision Systems of the Massachusetts Institute of Technology (2009-2010), at the University of LAquila (2011, 2013, 2016), and at the University of California Los Angeles (2016-2017). He is a FNRS, BAEF, and Fulbright fellow. He has been an Associate Editor for the IEEE CSS Conference Editorial Board, and the journals NAHS, Systems and Control Letters, and IEEE Transactions on Automatic Control. He was the recipient of the IBM Belgium 2009 award and a finalist of the ERCIM Cor Baayen award 2011. He was the co-recipient of the SICON best paper award 2013-2014.