

An Information Theory of Finite Abstractions and their Fundamental Scalability Limits

Giannis Delimpaltadakis and Gabriel Gleizer

Abstract—Finite abstractions are discrete models of dynamical systems, such that the set of abstraction trajectories contains all system trajectories. There is a consensus that abstractions suffer from a scalability bottleneck: to obtain a sufficient “accuracy” (how closely the abstraction models the system), the required abstraction size often becomes very large, rendering computation infeasible. This is accentuated for complex, high-dimensional systems, due to the curse of dimensionality. Yet, after decades of research, there are no formal results on the size-accuracy tradeoff of abstractions. Here, we derive a statistical, quantitative theory of the size-accuracy tradeoff of abstractions of autonomous deterministic systems and uncover fundamental limits on their scalability, through *rate-distortion theory*—the information theory of lossy compression. Abstractions are viewed as *encoder-decoder* pairs, encoding trajectories of dynamical systems. *Rate* measures abstraction size, while *distortion* describes accuracy, defined as the spatial average deviation between abstract trajectories and system ones. We obtain a fundamental lower bound on the minimum achievable abstraction distortion, given the system dynamics and the abstraction size; and vice-versa a lower bound on the minimum required size, for given distortion. The bound depends on the complexity of the dynamics, through trajectory entropy and a constant depending on the trajectory manifold’s geometry. We demonstrate its tightness on some dynamical systems. Finally, we showcase how this new theory enables constructing minimal abstractions, optimizing the size-accuracy tradeoff, through an example on a chaotic system.

I. INTRODUCTION

Modern engineering systems are becoming more complex and must meet intricate specifications in safety-critical situations. For instance, a self-driving car must follow traffic rules, avoid collisions, and optimize speed and fuel consumption. Due to the complexity of these systems, traditional analytic methods for verification and control are intractable. For over two decades, to address verification and control of complex dynamics and objectives, *abstraction*-based methods have flourished [1], [2]. Given a dynamical system, these methods construct a finite system—the abstraction—, arising from partitioning the state (and control) space of the original

system, such that all trajectories of the system are contained, in a formal sense, in the set of abstraction trajectories. Employing this property, one may solve an intractable verification or control problem for the original system over the finite abstraction, with formal guarantees of correctness. Over the years, research on abstractions has spanned deterministic systems [3]–[5], stochastic systems [6]–[8], and data-driven scenarios [9]–[12].

Despite their immense success, there is a consensus that abstractions suffer from a scalability bottleneck: to obtain a sufficient “accuracy” (how closely the abstraction represents the system), the required abstraction size often becomes very large, rendering computation infeasible. This is accentuated for complex, high-dimensional systems, due to the curse of dimensionality induced by partitioning. However, after considerable interest on abstractions in the past decades, there are still no formal results concerning their size-accuracy tradeoff.

Contributions

We derive a statistical, quantitative theory of the size-accuracy tradeoff of abstractions of autonomous deterministic systems and uncover fundamental limits on their scalability. To that end, we establish connections with *rate-distortion theory* – the information theory of lossy compression [13, Chapter 10]. The key observation for the theory is that abstractions are information-theoretic *encoder-decoder* pairs, encoding trajectories of dynamical systems, in a higher-dimensional ambient space. *Rate* measures abstraction size, while *distortion* is defined as the spatial average deviation between abstract trajectories and system ones, thus capturing the average accuracy of an abstraction. Then, building on recent developments in rate-distortion theory for generalized measurable sets [14], [15], we derive fundamental limits of abstractions’ size-accuracy tradeoff: a fundamental lower bound on the minimum achievable abstraction distortion, for given system dynamics and abstraction size; conversely, we also obtain a lower bound on the minimum abstraction size, for given distortion. The fundamental lower bound depends on the complexity of the dynamics, through generalized entropy, which we show how to compute. We demonstrate the tightness of the bound on certain dynamical systems. Finally, we showcase how the developed theory can be employed to construct optimal abstractions, in terms of the size-accuracy tradeoff, through an example on a chaotic system, and we provide a discussion towards a general procedure for constructing optimal abstractions.

Related work

Through decades of research, there has been considerable effort to construct scalable abstractions. Indicatively, [16]–

Giannis Delimpaltadakis is with the Robust and Intelligent Autonomous Systems lab, Italian Institute of Artificial Intelligence (AI4I), Turin, Italy. Gabriel Gleizer is with the Delft Center for Systems and Control, Mechanical Engineering, Delft University of Technology. Emails: ioannis.delimpaltadakis@ai4i.it, g.gleizer@tudelft.nl.

This research is partially supported by the project “Chaotic sampling for secure and sustainable networks of control systems” with file number 21937 of the research programme VENI AES 2024 which is (partly) financed by the Dutch Research Council (NWO) under the grant <https://doi.org/10.61686/WZNAX74774>.

[20] adapt the partition's resolution depending on the *local uncertainty* a given state-space region induces to the abstraction. Further, [21] constructs multi-resolution abstractions, employing feedback-refinement relations. The work [22] employs optimal control, such that the generated trajectories result in smaller abstraction cells and only a portion of the state space needs to be partitioned. Although the above methods result in more scalable abstractions, they neither provide quantitative results on the size-accuracy tradeoff, nor optimize some metric describing it. Another approach to derive more accurate abstractions is introducing memory [23], [24], based on sequences of outputs. In [25], it is shown that the size of such memory-based abstractions increases exponentially with the sequence length for deterministic chaotic systems. Apart from adaptive-partitioning techniques, compositional methods [5], [26] decompose the system to smaller ones, that are abstracted more efficiently. However, they do not address scalability issues of abstracting each subsystem. Further, it is also worth mentioning [27], which, for a particular class of stochastic abstractions, demonstrates that partitioning the control space may be unnecessary.

The connection between information theory and symbolic dynamics is well-known [28]; listing the whole literature on the topic is impossible. Worth mentioning is the work in [29], which employs rate-distortion theory to characterize complexity of dynamical systems and their relationship with so-called *shifts*¹. Nonetheless, [29] does not consider the deviation between a shift and a dynamical system, but rather focuses on asymptotic results (arbitrarily large partition size, steady-state trajectories) and the qualitative question of if a system can be embedded into a shift. Thus, it does not provide a quantitative theory of the size-accuracy tradeoff. Finally, the works [30]–[32] employ rate-distortion theory, to compress models that are already discrete and do not focus on abstracting continuous dynamics with formal guarantees.

II. PRELIMINARIES

A. Measure spaces, Hausdorff measure, generalized entropy

We first recall some notions related to information theory over general measurable spaces, based on [14], [15]. When we say “dimension” of a set, we refer to its intrinsic dimension; e.g. the dimension of a curve in $\mathbb{R}^n$ is 1. We denote the m -dimensional Hausdorff measure² by $\mathcal{H}^m$. Denote the restriction of $\mathcal{H}^m$ to the compact set $\mathcal{K}$ by $\mathcal{H}^m_{\mathcal{K}}$. Consider a measure μ over the measure space $(\mathcal{X}, \Sigma_{\mathcal{X}}, \nu)$, where $\Sigma_{\mathcal{X}}$ is the Borel σ -algebra of $\mathcal{X} \subseteq \mathbb{R}^n$. When μ is absolutely continuous w.r.t. ν (denoted by $\mu \ll \nu$), we denote the Radon–Nikodym derivative by $\frac{d\mu}{d\nu}$. When $\nu = \mathcal{H}^m_{\mathcal{X}}$ (assuming $\mathcal{X}$ is m -dimensional) and μ is a probability measure, then $\frac{d\mu}{d\mathcal{H}^m_{\mathcal{X}}}$ is the probability distribution associated to μ . Absolute continuity $\mu \ll \mathcal{H}^m_{\mathcal{X}}$ suggests that μ is not concentrated in arbitrarily small balls in $\mathcal{X}$. We denote the volume of the unit

ball in $\mathbb{R}^n$ as $v_n := \frac{\pi^{n/2}}{\Gamma(n/2+1)}$, where the Gamma function $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$.

Let $\mathcal{X} \subset \mathbb{R}^n$ be a finite union of bounded, m -dimensional, C^1 -manifolds. Denote by $c_{\mathcal{X}} \in (0, \infty]$ the smallest constant such that

$$\mathcal{H}^m_{\mathcal{X}}(B(\hat{x}, \delta) \cap \mathcal{X}) \leq c_{\mathcal{X}} \delta^m, \quad \text{for all } \hat{x} \in \mathbb{R}^n \text{ and } \delta > 0, \quad (1)$$

where $B(\hat{x}, \delta) := \{x \in \mathbb{R}^n : \|x - \hat{x}\| < \delta\}$. The constant $c_{\mathcal{X}}$ is finite in many cases of interest, including compact manifolds (see [14, Lemma 1]).

Consider a random variable x , distributed over the measure space $(\mathcal{X}, \Sigma_{\mathcal{X}}, \mathcal{H}^m_{\mathcal{X}})$ with probability measure μ_x . The generalized entropy of x (w.r.t the Hausdorff measure) is

$$h(x) := -\mathbb{E}_x \left[\log \left(\frac{d\mu_x}{d\mathcal{H}^m_{\mathcal{X}}} \right) \right], \quad (2)$$

where $\mathbb{E}_x[\cdot]$ denotes the expectation operator w.r.t. the random variable x . The generalized entropy is the extension of the classical Shannon entropy to continuous spaces, and is a measure of uncertainty or complexity of a random variable. For a measure space $(\mathcal{X}, \Sigma_{\mathcal{X}}, \mathcal{H}^m_{\mathcal{X}})$ with $0 < \mathcal{H}^m_{\mathcal{X}}(\mathcal{X}) < \infty$: a) $h(x)$ is maximized for the uniform distribution $\mu_x = \mathcal{H}^m_{\mathcal{X}}/\mathcal{H}^m_{\mathcal{X}}(\mathcal{X})$, b) $h(x)$ is bounded above, when $\mu_x \ll \mathcal{H}^m_{\mathcal{X}}$. Finally, employing a similar generalization as in (2), let us denote the *generalized Rényi entropy* with parameter $a \in \mathbb{R}_+ \setminus \{1\}$ by $h_a(x) := \frac{1}{1-a} \log \mathbb{E}_x \left[\left(\frac{d\mu_x}{d\mathcal{H}^m_{\mathcal{X}}} \right)^{a-1} \right]$.

The example below shows how the above apply to computing the entropy of a dynamical system's trajectories.

Example II.1 (Entropy of trajectories of the doubling map). Consider the dynamical system $x^+ = f(x)$, where the doubling map $f : [0, 1] \rightarrow [0, 1] : x \mapsto 2x \bmod 1$. Consider the set of 3-length trajectories of the system $\mathcal{B} := \{(x_0, f(x_0), f(f(x_0))) : x_0 \in [0, 1]\} \subseteq [0, 1]^3$. Notice that $\mathcal{B}$ is the union of 4 straight-line segments:

$$\begin{aligned} \mathcal{B} = & \{(x_0, 2x_0, 4x_0) : x_0 \in [0, .25]\} \cup \\ & \{(x_0, 2x_0, 4x_0 - 1) : x_0 \in [.25, .5]\} \cup \\ & \{(x_0, 2x_0 - 1, 4x_0 - 2) : x_0 \in [.5, .75]\} \cup \\ & \{(x_0, 2x_0 - 1, 4x_0 - 3) : x_0 \in [.75, 1]\}. \end{aligned}$$

Consider random initial conditions $x_0 \sim U[0, 1]$, where $U[0, 1]$ is the uniform distribution over $[0, 1]$. It is known that $U[0, 1]$ is invariant under the doubling map. Thus, the random variable $\xi(x_0) = (x_0, f(x_0), f(f(x_0))) \in \mathcal{B}$, i.e. the trajectories, is uniformly distributed over $\mathcal{B}$; i.e., the probability measure $\mu_{\xi} = \mathcal{H}^1_{\mathcal{B}}/\mathcal{H}^1_{\mathcal{B}}(\mathcal{B})$. Its generalized entropy

$$\begin{aligned} h(\xi) &= -\mathbb{E}_{\xi} \left[\log \left(\frac{d\mu_{\xi}}{d\mathcal{H}^1_{\mathcal{B}}} \right) \right] = -\int_{\mathcal{B}} \log \left(\frac{d\mu_{\xi}}{d\mathcal{H}^1_{\mathcal{B}}} \right) d\mu_{\xi} \\ &= -\int_{\mathcal{B}} \log \left(\frac{1}{\mathcal{H}^1_{\mathcal{B}}(\mathcal{B})} \right) d\mu_{\xi} \\ &= \log(\sqrt{21}) \int_{\mathcal{B}} d\mu_{\xi} = \log(\sqrt{21}), \end{aligned}$$

where we have used that the total length of the line segments is $\mathcal{H}^1_{\mathcal{B}}(\mathcal{B}) = \sqrt{21}$, and that $\int_{\mathcal{B}} d\mu_{\xi} = 1$.

¹A class of discrete systems. Abstractions can be cast as shifts.

²The Hausdorff measure is a generalization of the Lebesgue measure, generalizing the notions of curve length or surface area. E.g., $\mathcal{H}^1(C) = 2\pi$, where C is the unit circle embedded in $\mathbb{R}^n$.

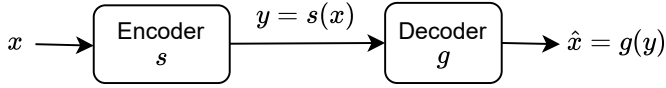


Fig. 1: The typical source coding setting.

B. Rate-distortion theory on measurable spaces

A typical setting in information theory is source coding, see Fig. 1. A source emits a *message* $x \in \mathcal{X} \subseteq \mathbb{R}^n$, which is a random variable over the measure space $(\mathcal{X}, \Sigma_{\mathcal{X}}, \mathcal{H}_{\mathcal{X}}^m)$, with associated probability measure μ_x , where $\mathcal{X}$ is assumed to be m -dimensional. The *encoder* $s : \mathcal{X} \rightarrow \mathcal{Y}$, where $\mathcal{Y}$ is finite, outputs the coded message $s(x) = y \in \mathcal{Y}$. Finally, the *decoder* $g : \mathcal{Y} \rightarrow \hat{\mathcal{X}}$, upon receiving y , decodes it into $g(y) = \hat{x}$. Compression takes place by encoding the continuous message x into a low-dimensional, finite coded message y . The encoder cardinality $|\mathcal{Y}|$ determines the compression, and the *rate* is defined as $\log|\mathcal{Y}|$. A *distortion* function $d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathbb{R}_+$ measures the deviation of $\hat{x}$ from the original message x . A typical distortion function, when $\hat{\mathcal{X}} = \mathbb{R}^n$, is the squared error $d(x, \hat{x}) = \|x - \hat{x}\|^2$.³

Of particular interest is the fundamental limit of the rate-distortion tradeoff, i.e. the following quantity:

$$D(R) := \inf_{s, g} \mathbb{E}_x[d(x, \hat{x}) \mid s, g] \\ \text{s.t. } s : \mathcal{X} \rightarrow \mathcal{Y}, g : \mathcal{Y} \rightarrow \hat{\mathcal{X}}, \\ \log|\mathcal{Y}| \leq R, y = s(x), \hat{x} = g(y),$$

where the expectation is taken w.r.t. the random variable x . In words, $D(R)$ is the minimum achievable average distortion, for a given rate threshold R . The function $D(R)$ has an inverse, $R(D)$, which is the minimum rate, for a given maximum expected distortion threshold D . The following result provides a fundamental lower bound on $D(R)$.

Theorem II.1 (Generalized Shannon lower bound [15, Thm. 3.1, simplified]). *Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a finite union of compact, m -dimensional, C^1 -manifolds, and $\mu_x \ll \mathcal{H}_{\mathcal{X}}^m$. Assume that $\hat{\mathcal{X}} \subseteq \mathbb{R}^n$ and that $(\hat{\mathcal{X}}, \Sigma_{\hat{\mathcal{X}}})$ is measurable. Consider the Euclidean distortion $d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathbb{R}_+ : (x, \hat{x}) \mapsto \|x - \hat{x}\|^2$. Then*

$$D(R) \geq D_*(R) := \frac{m}{2} \left(\frac{e^{-R+h(x)-m/2}}{c_{\mathcal{X}} \Gamma(1+m/2)} \right)^{2/m}. \quad (3)$$

Proof Sketch. This is [15, Thm. 3.1] specialized for finite unions of compact C^1 -manifolds and Euclidean distortion. $\square$

C. Transition systems

Definition II.2 (Transition system). *A transition system S is a tuple $S = (\mathcal{X}, \xrightarrow{S})$, where $\mathcal{X}$ is the state space and $\xrightarrow{S} \subseteq \mathcal{X} \times \mathcal{X}$ is a transition relation.*

A transition system $S = (\mathcal{X}, \xrightarrow{S})$ is *deterministic* if, for any $x \in \mathcal{X}$, there exists at most one $x' \in \mathcal{X}$, such that $(x, x') \in \xrightarrow{S}$. Given a transition system $S = (\mathcal{X}, \xrightarrow{S})$, its l -length behavior

³Here, we present a simplified setting of source coding, where the encoder-decoder is deterministic, and the source emits a single message. For the general theory, see [13], [15].

$\mathcal{B}_l^S$ is defined as $\mathcal{B}_l^S := \left\{ \xi : \xi = \{x_i\}_{i=0}^{l-1}, (x_i, x_{i+1}) \in \xrightarrow{S}, i = 0, 1, \dots, l-2 \right\}$. That is, the l -length behavior is the set of l -long trajectories. Notice that $\mathcal{B}_l^S \subseteq \mathcal{X}^l$.

III. ABSTRACTIONS AND THE CURSE OF DIMENSIONALITY

A. Finite abstractions of dynamical systems

Throughout this work, we consider deterministic dynamical systems $x^+ = f(x)$, with $f : \mathcal{X} \rightarrow \mathcal{X}$. Dynamical systems obtain the transition-system representation $S = (\mathcal{X}, \xrightarrow{S})$, where $\xrightarrow{S} := \{(x, y) : y = f(x), x \in \mathcal{X}\}$. We make the following assumption.

Assumption 1 (The state space). *The set $\mathcal{X} \subseteq \mathbb{R}^n$ is n -dimensional, connected and compact.*

Under Assumption 1, $\mathcal{B}_l^S$ is an n -dimensional subset of $\mathcal{X}^l \subseteq \mathbb{R}^{nl}$. We now introduce *abstractions* of dynamical systems.

Definition III.1 (Measurable Partition). *Given a set $\mathcal{X}$, a finite collection of measurable, disjoint sets $\mathcal{Y} = \{Y_i\}$, such that $\bigcup_i Y_i = \mathcal{X}$, is a measurable partition of $\mathcal{X}$.*

Definition III.2 (Abstraction). *Given a dynamical system with transition-system representation $S = (\mathcal{X}, \xrightarrow{S})$ and a measurable partition $\mathcal{Y} = \{Y_i\}$ of $\mathcal{X}$, a transition system $A = (\mathcal{Y}, \xrightarrow{A})$ is an abstraction of S if, for any $x, x' \in \mathcal{X}$ and $Y, Y' \in \mathcal{Y}$, such that $x \in Y$ and $x' \in Y'$, we have $(x, x') \in \xrightarrow{S} \implies (Y, Y') \in \xrightarrow{A}$.*

We consider abstractions based on state-space partitions for ease of exposition, but our results apply to more general notions, such as *simulation-based abstractions* ([1, Def. 9.2] specialized to autonomous systems), and *finite covers of the system's trajectory space*. More detail is given in Remark 5.

Although the dynamical system S is deterministic, the abstraction A is generally non-deterministic. With a slight abuse of formality, we often treat trajectories $\{\omega_i\}_{i=0}^{l-1}$ of the abstraction (with $\omega_i \in \mathcal{Y}$) as subsets of $\mathcal{X}^l$, that is $\{\omega_i\}_{i=0}^{l-1} \equiv \omega_0 \times \omega_1 \times \dots \times \omega_{l-1} \subseteq \mathcal{X}^l$. We will also make use of the set of l -long reachable sets of the abstraction A , that is $\mathcal{R}_l^A := \{\Omega \in 2^{\mathcal{X}^l} : \Omega = \bigcup_{\omega \in \mathcal{B}_l^A, \omega_0=Y} \omega, Y \in \mathcal{Y}\}$. Observe

that the cardinality $|\mathcal{R}_l^A| = |\mathcal{Y}|$.

Theorem III.3 (Behavioral inclusion [1, Theorem 4.18, simplified]). *For a dynamical system $x^+ = f(x)$ with transition system representation $S = (\mathcal{X}, \xrightarrow{S})$, a partition $\mathcal{Y}$ of $\mathcal{X}$ and an abstraction $A = (\mathcal{Y}, \xrightarrow{A})$ of S , the following holds for any l : for any $\xi \in \mathcal{B}_l^S$, there exists $\omega \in \mathcal{B}_l^A$ such that $\xi \in \omega$.*

An immediate result of Thm. III.3 is the following:

Corollary III.4. *For a dynamical system $x^+ = f(x)$ with transition system representation $S = (\mathcal{X}, \xrightarrow{S})$, a partition $\mathcal{Y}$ of $\mathcal{X}$ and an abstraction $A = (\mathcal{Y}, \xrightarrow{A})$ of S , the set $\mathcal{R}_l^A$ is a finite cover of $\mathcal{B}_l^S$.*

Proof. Follows directly from Thm. III.3 and the fact that $|\mathcal{R}_l^A| = |\mathcal{Y}|$. $\square$

This observation is instrumental in this work. Through behavioral inclusion, abstractions encode information about the infinite, continuous system behavior $\mathcal{B}_S^l$ into the finite set $\mathcal{R}_l^A$. While this enables computational methods to verification problems for dynamical systems, it also generally entails information loss, as the following section explains.

B. Abstraction-based verification and information loss

In typical verification problems, we are given a set of initial conditions $\Xi_0 \subseteq \mathcal{X}$ for the system S and we have to check if the corresponding set of system trajectories $\Xi = \{\xi \in \mathcal{B}_l^S : \xi_0 \in \Xi_0\}$ satisfies a given property. For example, in the case of safety, we have to check if $\Xi \cap \mathcal{U}_l = \emptyset$, where $\mathcal{U}_l := \{\xi \in \mathcal{X}^l : \exists i \in \{0, \dots, l-1\}, \xi_i \in \mathcal{U}\}$ and $\mathcal{U} \subseteq \mathcal{X}$ is an unsafe set. Computing the exact reachable set Ξ is generally impossible. Abstractions A address this problem by computing the corresponding set of abstract state trajectories $\Omega_A = \bigcup_{\omega \in \mathcal{B}_l^A, \omega_0 \in \Xi_0 \neq \emptyset} \omega$, which is tractable, as the abstraction

is finite. Notice that, by Thm. III.3, we have $\Xi \subseteq \Omega_A$. Hence, e.g. for safety verification if $\Omega_A \cap \mathcal{U}_l = \emptyset$, then one may safely deduce that the system is safe, as $\Omega_A \cap \mathcal{U}_l = \emptyset \implies \Xi \cap \mathcal{U}_l = \emptyset$.

As abstractions group system states $x \in \mathcal{X}$ in sets $Y \subseteq \mathcal{X}$, information loss is inevitable. In general, the partition $\mathcal{Y}$ needs to have a relatively high resolution, to recover a meaningful verification answer. E.g., in the extreme case of $|\mathcal{Y}| = 1$, for any set of initial conditions $\Xi_0 \subseteq \mathcal{X}$, the abstraction returns $\Omega_A = \mathcal{X}^l$, i.e. the whole ambient space of l -length trajectories. As such, for small $|\mathcal{Y}|$, the abstraction A does not accurately represent the system S . On the other hand, for large $|\mathcal{Y}|$, where the abstraction is more accurate, the computations on the abstraction become heavier – even intractable. Thus, there is a trade-off between abstraction accuracy and partition size $|\mathcal{Y}|$. In what follows, we provide a statistical, quantitative theory of this size-accuracy tradeoff and provide bounds on it, based on rate-distortion theory.

IV. INFORMATION-THEORETIC FRAMEWORK FOR FINITE ABSTRACTIONS

In what follows, consider the dynamical system $x^+ = f(x)$, with $x \in \mathcal{X} \subseteq \mathbb{R}^n$, under Assumption 1. The dynamical system admits the transition system representation $S = (\mathcal{X}, \xrightarrow{S})$. Towards deriving a statistical quantification on the size-accuracy tradeoff of abstractions, we impose a probability distribution $p_{\xi_0} : \mathcal{X} \rightarrow \mathbb{R}_+$ on the system's initial conditions. Verification, then, becomes: sampling an initial condition $\xi_0 \in \mathcal{X}$, with $\xi_0 \sim p_{\xi_0}$, and afterward employing the abstraction to give a verification answer.⁴

Let us show how an abstraction A can be viewed as an encoder-decoder pair of system trajectories $\xi \in \mathcal{B}_l^S$. For the following, we refer the reader to Figure 2. The system (source) samples an initial condition $\xi_0 \sim p_{\xi_0}$ and generates

the trajectory $\xi = (\xi_0, \dots, \xi_{l-1}) \in \mathcal{B}_l^S$ (the message). The encoder $s_A : \mathcal{B}_l^S \rightarrow \mathcal{Y}$ looks at the initial condition ξ_0 and returns the corresponding abstract initial condition:⁵

$$s_A(\xi) := Y, \text{ s.t. } \xi_0 \in Y \quad (4)$$

The decoder g_A , upon receiving the initial condition $s_A(\xi)$, outputs the set of all abstract state trajectories corresponding to $s_A(\xi)$. That is, for the decoder we have $g_A : \mathcal{Y} \rightarrow \mathcal{R}_l^A$ with

$$g_A(y) := \bigcup_{\omega \in \mathcal{B}_l^A, \omega_0 = y} \omega. \quad (5)$$

The rate, determined by the encoder's size, is $\log(|\mathcal{Y}|)$. Indeed, notice that *the abstraction encodes the system's trajectories $\mathcal{B}_l^S$ into exactly $|\mathcal{Y}|$ outcomes, that is $\{g_A(z) : z \in \mathcal{Y}\} = \mathcal{R}_l^A$.*

To capture the accuracy of the abstraction, and compare the message ξ and output $\Omega_A = g_A(s_A(\xi))$, we employ a distortion function $d : \mathcal{B}_l^S \times 2^{\mathcal{X}^l} \rightarrow \mathbb{R}^+$ defined by

$$d(\xi, \Omega_A) := \sup_{\xi' \in \Omega_A} \frac{1}{l} \|\xi - \xi'\|^2. \quad (6)$$

In words, $d(\xi, \Omega_A)$ is the worst possible distortion between system trajectories ξ and abstract trajectories Ω_A , averaged over the time horizon l . This is in-line with abstraction-based verification, where the worst-case outcome is considered.

Let us now explain what “expected (or average) distortion”, for a given abstraction A , means in the context of verification. The expected distortion $\mathbb{E}_{\xi_0}[d(\xi, \Omega_A) \mid A]$ is taken w.r.t. the initial-condition distribution p_{ξ_0} . Thus, for N verification problems with $N \rightarrow \infty$, where the initial condition $\xi_0 \sim p_{\xi_0}$, $\mathbb{E}_{\xi_0}[d(\xi, \Omega_A) \mid A]$ is the average distortion. As d measures the distance between system trajectories and abstract state trajectories, the expected distortion $\mathbb{E}_{\xi_0}[d(\xi, \Omega_A) \mid A]$ is thus the *spatial*, statistical average of the deviation between system trajectories and abstract state trajectories, over initial conditions in $\mathcal{X}$ with distribution p_{ξ_0} .

Remark 1 (Initial-condition distribution). *The distribution p_{ξ_0} weighs how much each initial condition $\xi_0 \in \mathcal{X}$ (and thus each trajectory $\xi \in \mathcal{B}_l^S$) contributes to the average distortion $\mathbb{E}_{\xi_0}[d(\xi, \Omega_A) \mid A]$. Arguably, a suitable choice for p_{ξ_0} is the uniform distribution, as when constructing an abstraction, the initial condition is unknown and all initial conditions are considered equally likely; however, if the system is known to obey a specific distribution p_{ξ_0} , we may use this one.*

Finally, the optimal abstraction size-accuracy tradeoff is captured by the following rate-distortion quantity:

$$\begin{aligned} D_{abs}(R) &:= \inf_A \mathbb{E}_{\xi_0}[d(\xi, \Omega_A) \mid A] \\ &\text{s.t. } A \text{ is an abstraction of } S, \\ &\text{(4), (5) hold,} \\ &\log|\mathcal{Y}| \leq R, \Omega_A = g_A(s_A(\xi)). \end{aligned}$$

⁴More precisely, ξ_0 is a random variable over $(\mathcal{X}, \Sigma_{\mathcal{X}}, \mathcal{L}_{\mathcal{X}}^n)$, where $\mathcal{L}_{\mathcal{X}}^n$ is the n -dimensional Lebesgue measure, with probability measure μ_{ξ_0} such that $\frac{d\mu_{\xi_0}}{d\mathcal{L}_{\mathcal{X}}^n} = p_{\xi_0}$.

⁵As explained in Section III, for verification, the only information given to the abstraction is the initial condition ξ_0 . Nonetheless, for given dynamics f , ξ_0 and ξ are equivalent and carry the exact same information, as $\xi_0 \mapsto \xi$ is bijective. Thus using ξ as the encoder input is equivalent to using ξ_0 . We adopt ξ to regard $\mathcal{B}_l^S$ as the message space, which is convenient since the distortion (6) is defined over trajectories (see also Remark 3).

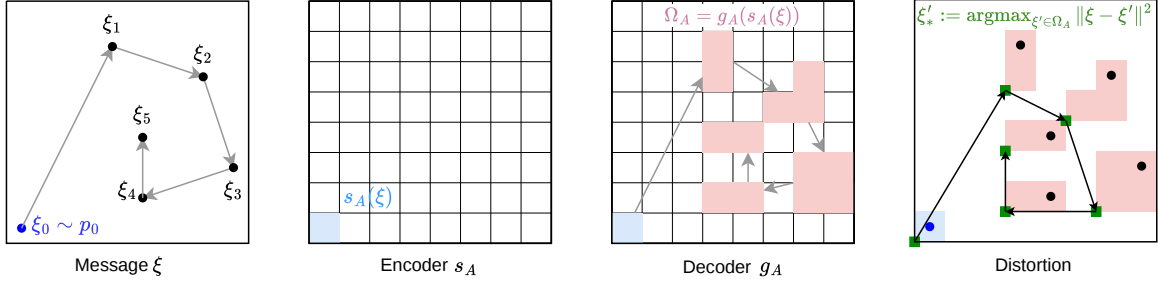


Fig. 2: Abstractions as a source coding scheme. From the left: 1) a sample trajectory ξ of system S with its initial state ξ_0 highlighted in blue; 2) the state-space partition $\mathcal{Y}$, and the corresponding abstract initial condition in cyan; 3) the set of abstract trajectories Ω_A , in red; 4) the specific trajectory $\xi'_* \in \Omega_A$ that deviates the most from the true system trajectory ξ .

That is, the minimum average deviation of abstract state trajectories and system trajectories, over all abstractions with a given upper-bound e^R on partition size. Likewise, we also consider the inverse $R_{abs}(D)$, which is the (log of the) minimum partition size for a given upper threshold D on the average deviation of abstract state trajectories and system ones.

Remark 2 (Statistics of abstractions' accuracy and size). *The proposed theory does not aim at providing (probabilistic) guarantees on the correctness of abstractions. These are a-priori provided by Definition III.2, through behavioral inclusion or related properties. Instead, the theory developed here provides (guarantees on the) statistical quantification of abstractions' accuracy and size.*

Remark 3 (The message space is $\mathcal{B}_l^S$). *Even though we have reduced everything thus far to the initial condition distribution p_{ξ_0} , the message space is $\mathcal{B}_l^S$, i.e. the system trajectories. Indeed, although the expectation $\mathbb{E}[d(\xi, \Omega_A) \mid A]$ can be taken either w.r.t. $\xi_0 \sim p_{\xi_0}$ or w.r.t. the random variable $\xi \in \mathcal{B}_l^S$ (as $\xi_0 \mapsto \xi$ is bijective), the distortion d considers the whole $\xi \in \mathcal{B}_l^S$. As such, in the coming section, to derive bounds on $D_{abs}(R)$ and $R_{abs}(D)$, employing the theory from Section II-B, we reason about the random variable $\xi \in \mathcal{B}_l^S$ and its probability measure μ_ξ over $(\mathcal{B}_l^S, \Sigma_{\mathcal{B}_l^S}, \mathcal{H}_{\mathcal{B}_l^S}^n)$, which is solely determined by the initial condition distribution p_{ξ_0} and the system dynamics f . Hence, we take expectations $\mathbb{E}_\xi$ and $\mathbb{E}_{\xi_0}$ interchangeably.*

V. RATE-DISTORTION THEORY AND A FUNDAMENTAL LIMIT FOR ABSTRACTIONS

A. A fundamental limit on abstracting dynamical systems

Having modeled the statistics of abstraction-based verification as a source coding problem, we now proceed to probing the fundamental limits of abstractions' size-accuracy tradeoff, by providing lower bounds on $R_{abs}(D)$ and $D_{abs}(R)$.

Note that abstractions, given the message, output sets and the associated distortion (6) is set-based. This is in contrast to typical encoder-decoder pairs considered in Thm. II.1, which output points and the distortion function is the Euclidean distance. Thus, the results from Section II-B do not straightforwardly apply, to derive bounds on D_{abs} and R_{abs} . In what follows, we derive said bounds, both employing Thm. II.1 and quantifying the aforementioned distortion disparity. This

enables a rate-distortion theory for abstractions. First, we present an intermediate, purely geometric result, providing a lower bound on the average distortion of a given abstraction.

Proposition V.1 (Abstraction vs. encoder distortion). *Consider a dynamical system $x^+ = f(x)$ with transition system representation $S = (\mathcal{X}, \xrightarrow{S})$. Let Assumption 1 hold. Let $\xi \in \mathcal{B}_l^S$ be a trajectory of S , with $\xi_0 \sim p_{\xi_0} : \mathcal{X} \rightarrow \mathbb{R}_+$. Consider a measurable partition $\mathcal{Y}$ of $\mathcal{X}$ and an associated abstraction A , and let $\Omega_A = g_A(s_A(\xi))$, where s, g are given by (4) and (5). Consider an encoder-decoder pair (s_{q_A}, g_{q_A}) , where $s_{q_A}(\xi) = g_A(s_A(\xi))$ and $g_{q_A}(z) = x_c(z)$, where $x_c(z) := \arg \min_y \sup_{y' \in z} \|y - y'\|^2$ is the Chebyshev center of the set z . Denote the Chebyshev radius of z , by $r_c(z) := \min_y \sup_{y' \in z} \|y - y'\|$. Let $\xi_{q_A} = g_{q_A}(s_{q_A}(\xi))$. The following lower holds for the average distortion of the abstraction:*

$$\mathbb{E}_{\xi_0}[d(\xi, \Omega_A)] \geq \frac{1}{l} \mathbb{E}_{\xi_0}[\|\xi - \xi_{q_A}\|^2] + \frac{1}{l} \mathbb{E}_{\xi_0}[r_c^2(\Omega_A)], \quad (7)$$

where d is the distortion function in (6).

Prop. V.1 suggests that the average distortion of an abstraction is lower bounded by the expected distortion of a particular encoder-decoder pair (the one outputting the Chebyshev centers of the abstraction's outputs) plus a term depending on the size of the abstraction's outputs. Employing Prop. V.1, in Theorem V.2 below, we derive a fundamental lower bound on $D_{abs}(R)$, by lower-bounding each of the two terms in the right-hand side of (7) separately, over all abstractions with the same rate. The first term in the right-hand side of (7) can be lower bounded as in Thm. II.1, being the expected distortion of an encoder-decoder pair with the same rate as the abstraction. To bound the second term, we observe that the abstraction's outputs Ω_A form the nl -dimensional cover $\mathcal{R}_l^A$ of $\mathcal{B}_l^S$, and the cover's size is equal to the abstraction's size; the bound is then obtained by lower-bounding over all possible nl -dimensional covers of $\mathcal{B}_l^S$, using geometric measure theory (see Lemma VIII.1). For an illustrative example of the above, see Fig. 3.

Theorem V.2 (Shannon lower bound for abstractions). *Consider a dynamical system $x^+ = f(x)$ with transition-system representation $S = (\mathcal{X}, \xrightarrow{S})$, and let Assumption 1 hold. Let $\xi \in \mathcal{B}_l^S$ be a trajectory of S , with $\xi_0 \sim p_{\xi_0} : \mathcal{X} \rightarrow \mathbb{R}_+$. Assume that:*

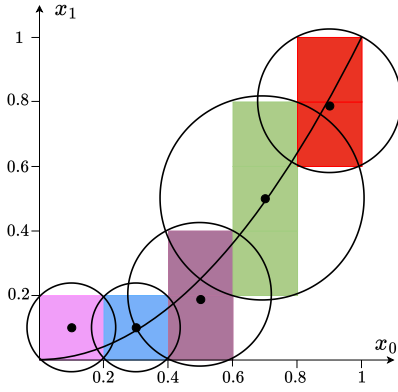


Fig. 3: Consider the dynamical system $x^+ = x^2$ with state-space $\mathcal{X} = [0, 1]$. The parabola depicts the set of trajectories $\mathcal{B}_2^S$, embedded in $[0, 1]^2$. Consider an abstraction A with associated partition sets $Y_i = [0.2(i-1), 0.2i]$ for $i = 1, \dots, 4$ and $Y_5 = [0.8, 1]$. The abstraction transitions, thus, are $\xrightarrow{A} = \{(Y_1, Y_1), (Y_2, Y_1), (Y_3, Y_1), (Y_3, Y_2), (Y_4, Y_2), (Y_4, Y_3), (Y_4, Y_4), (Y_5, Y_4), (Y_5, Y_5)\}$. The colored rectangles represent the abstraction outputs Ω_A , depending on the initial condition ξ_0 . For example, if $\xi_0 \in Y_4$, then the output Ω_A is the green rectangle $\{(x_0, x_1) : x_0 \in Y_4, x_1 \in Y_2 \cup Y_3 \cup Y_4\}$. Observe that the abstraction's outputs form a 2-dimensional cover of $\mathcal{B}_2^S$, as per Cor. III.4. The dots represent the Chebyshev centers for each different output Ω_A , and the circles are the corresponding Chebyshev balls. E.g., when $\xi_0 \in Y_4$, we have $x_c(\Omega_A) = (.7, .5)$ and $r_c(\Omega_A) = \sqrt{0.1}$. The abstraction's average distortion is lower bounded as in (7).

- 1) $\mathcal{B}_l^S$ is a finite union of bounded, n -dimensional C^1 -manifolds,
- 2) $\mu_\xi \ll \mathcal{H}_{\mathcal{B}_l^S}^n$, with $p_\xi := \frac{d\mu_\xi}{d\mathcal{H}_{\mathcal{B}_l^S}^n}$.

The average distortion of any abstraction A with partition size $|\mathcal{Y}| \leq e^R$, where $R > 0$, is lower bounded as follows

$$D_{abs}(R) \geq \frac{n}{2l} \left(\frac{e^{-R+h(\xi)-n/2}}{c_{\mathcal{B}_l^S} \Gamma(1+n/2)} \right)^{2/n} + \frac{1}{l} c_{\mathcal{B}_l^S}^{-2/n} \sup_{s \in (1, \infty]} e^{\frac{2}{n}(-\frac{s}{s-1}R+h_s(\xi))}, \quad (8)$$

where $c_{\mathcal{B}_l^S}$ is defined by (1).

Notice that a valid lower bound on $D_{abs}(R)$ is obtained for any value of $s \in (1, \infty]$ in the right-hand side of (8); the sup over s provides the tightest bound. In the numerical examples in Section VI, we compute the bound for multiple values of s . Further, one may recover a lower bound on $R_{abs}(D)$ numerically, by fixing D in the left-hand side of (8) and solving numerically for R (as the right-hand side is a decreasing function of R , this is trivially computed by, e.g., bisection methods). Nonetheless, by specifically picking $s = \infty$, one may analytically invert the bound (8) to obtain a lower bound on $R_{abs}(D)$, as the following corollary shows.

Corollary V.3 (to Thm. V.2). *Under the assumptions of Thm. V.2, the minimum abstraction size, for an upper threshold*

D on average abstraction distortion, is lower bounded as:

$$R_{abs}(D) \geq \frac{n}{2} \log \left(\frac{n}{2l c_{\mathcal{B}_l^S}^{2/n} \Gamma(1+n/2)^{2/n}} + \frac{1}{l c_{\mathcal{B}_l^S}^{2/n}} e^{\frac{2}{n}h_\infty(\xi)} \right) - \frac{n}{2} \log(D). \quad (9)$$

Proof Sketch. Pick $s = \infty$ in (8), fix the right-hand side equal to D , and solve for R . $\square$

Hence, Thm V.2 and Cor. V.3 provide fundamental limits on the size-accuracy tradeoff, or the scalability, of abstractions, for given dynamics $x^+ = f(x)$.

Remark 4 (On the assumptions of Thm. V.2). *The first assumption of Thm. V.2, requiring $\mathcal{B}_l^S$ to be a union of smooth manifolds, is satisfied whenever the dynamics f is piecewise continuously-differentiable. The second assumption is satisfied whenever f is piecewise continuous and the initial condition distribution p_{ξ_0} is such that $\mu_{\xi_0} \ll \mathcal{L}^n$, where $\mathcal{L}^n$ the Lebesgue measure in $\mathbb{R}^n$.*

Remark 5 (Simulation-based abstractions and trajectory covers). *Our results apply to abstractions based on ε -approximate simulation relations [1, Def. 9.2], with one minor modification. In such abstractions, the state space is a cover of $\mathcal{X}$, and each initial condition ξ_0 may be related to multiple $Y_i \in \mathcal{Y}$. For verification, though, any such initial condition Y_i can be picked, and the resulting reachable set Ω_A is guaranteed to contain ξ (by the simulation-relation definition). Thus, when there are multiple $Y_i \in \mathcal{Y}$ such that $\xi_0 \in Y_i$, the encoder in (4) picks only one arbitrarily and deterministically. The reader can verify that all results apply directly, and the resulting bounds are the same. More importantly, our results apply to any finite cover $\mathcal{R}_l^A$ of the system's behavior $\mathcal{B}_l^S$. As simulation-based abstractions induce finite covers $\mathcal{R}_l^A$ of $\mathcal{B}_l^S$ (Cor. III.4), trajectory covers generalize abstractions. Research on such a generalization is left for future work.*

In the coming section, we provide: a) closed-form expressions on $h(\xi)$, $h_s(\xi)$ and $c_{\mathcal{B}_l^S}$, for certain classes of dynamics $x^+ = f(x)$, and b) an interpretation on how the complexity of the dynamics and the time-horizon l affect the fundamental lower bound (8) in Thm. V.2. Afterwards, in Section V-C, we show how the curse of dimensionality arises, through a relaxed version of the bounds of Thm. V.2 and Cor. V.3.

B. Interpretation and calculus for Theorem V.2

Before we proceed with the interpretation of Thm. V.2, let us show how one may compute $h(\xi)$, $h_s(\xi)$ and $c_{\mathcal{B}_l^S}$, which are required to compute the lower bound (8). Let us define the function $b_l : \mathcal{X} \rightarrow \mathcal{B}_l^S$ by

$$b_l(x) := [x^T \quad f(x)^T \quad f(f(x))^T \quad \dots \quad f^{(l-1)T}(x)]^T, \quad (10)$$

which maps an initial state into its l -long trajectory.

Proposition V.4 (Computing $h(\xi)$ and $h_s(\xi)$). *Consider a system $x^+ = f(x)$ with $f : \mathcal{X} \rightarrow \mathcal{X}$ measurable and piecewise*

Lipschitz⁶ and differentiable. Let Assumption 1 hold, and $\xi_0 \sim p_{\xi_0} : \mathcal{X} \rightarrow \mathbb{R}_+$. The following expressions hold

$$h(\xi) = h(\xi_0) + \frac{1}{2} \int_{\mathcal{X}} p_{\xi_0}(x) \log \det(J_{b_l}(x)^T J_{b_l}(x)) dx, \quad (11)$$

$$h_s(\xi) = \frac{1}{1-s} \log \int_{\mathbb{R}^n} \frac{p_{\xi_0}(x)^s}{\det(J_{b_l}(x)^T J_{b_l}(x))^{\frac{s-1}{2}}} dx, \quad s > 1, \quad (12)$$

$$h_\infty(\xi) = \text{ess inf}_{\mathcal{X}} \left(\frac{1}{2} \log \det(J_{b_l}(x)^T J_{b_l}(x)) - \log p_{\xi_0}(x) \right), \quad (13)$$

where J_{b_l} denotes the Jacobian matrix of b_l . Moreover, $h(\xi) \geq h(\xi_0)$ and $h_s(\xi) \geq h_s(\xi_0)$ for any $s \in [1, \infty)$.

Proposition V.5 (Computing $c_{\mathcal{B}_l^S}$). Consider a system $x^+ = f(x)$, with $f : \mathcal{X} \rightarrow \mathcal{X}$ differentiable a.e., and let Assumption 1 hold. The following facts on $c_{\mathcal{B}_l^S}$ hold:

- 1) $c_{\mathcal{B}_l^S} \leq v_n$, if f is affine;
- 2) $c_{\mathcal{B}_l^S} \leq M^l v_n$, if f is piecewise affine with M modes;
- 3) $c_{\mathcal{B}_l^S} \leq v_n \left(\sum_{i=0}^{l-1} L^{2i} \right)^{n/2}$, if f is Lipschitz continuous with constant L .

Remark 6 ($c_{\mathcal{B}_l^S}$ at high rates). As the partition size $|\mathcal{Y}|$ grows large, the Chebyshev balls of the abstraction outputs (c.f. Prop. V.1 and Lemma VIII.1) become small. Hence, in the case of smooth f , their intersection with the manifold $\mathcal{B}_l^S$ approaches the case of an affine system, with $c_{\mathcal{B}_l^S} \leq v_n$. Similarly, for piecewise affine f and sufficiently small balls – at least $|\mathcal{Y}| \geq M^l$ –, these can be chosen to intersect with at most one piece each. Thus, in the high-rate regime, to reduce conservatism of the bound, one may use the approximation $c_{\mathcal{B}_l^S} \approx v_n$. We demonstrate this in the numerical examples in Section VI.

We proceed to discussing Thm. V.2. First, inspecting (8), systems with more complex dynamics lead to bigger abstraction distortion, for fixed abstraction size, since the right-hand side is increasing w.r.t. $h(\xi)$ and the Rényi entropy $h_s(\xi)$; equivalently, more complex systems require bigger abstraction size for the same distortion.

Regarding the effect of the time-horizon l on the bound (8), we have to inspect the effect that l has on $h(\xi)$ and $h_s(\xi)$. Let us first demonstrate that, for the “simple” dynamics of exponentially stable systems, the abstraction distortion can be made arbitrarily small for $l \rightarrow \infty$, for sufficiently large abstraction size.

Example V.1 (Exponentially stable systems). Consider a system $x^+ = f(x)$ whose origin is exponentially stable on a given compact set in $\mathbb{R}^n$. Then, there is a Lyapunov function $V : \mathbb{R}^n \mapsto \mathbb{R}_+$ satisfying $V(x) \geq \frac{1}{r^2} \|x\|^2$ for a given $r > 0$ and, for all x s.t. $V(x) \leq 1$ (w.l.o.g.), $V(f(x)) \leq aV(x)$, with $a \in [0, 1)$. This allows us to create an abstraction A with the associated partition $Y_i = \{x \in \mathbb{R}^n \mid a^i < V(x) \leq a^{i-1}\}$ for $i = 1, \dots, N$, and $Y_{N+1} = \{x \in \mathbb{R}^n \mid V(x) \leq a^N\}$; and transitions $Y_i \xrightarrow{A} Y_j$ if and only if $j > i$ or $j = i = N + 1$. The abstraction encapsulates the fact that, after N steps or

⁶That is, $\mathcal{X}$ is a finite union $\cup_i \mathcal{X}_i$ of Lebesgue-measurable sets such that the restriction of f to each $\mathcal{X}_i$ is Lipschitz. This condition may be relaxed to f approximately Lipschitz, see [33, Thm. 3.1.8, Sec. 3.2.1], which also implies approximate differentiability.

less, all trajectories reach the sublevel set $V(x) \leq a^N$. We get that $2r$ and $2\sqrt{a^N}r$ are overapproximations of the diameters of Y_i , $i \leq N$, and Y_{N+1} , respectively. Then, recalling the distortion (6), for any trajectory ξ of the system, $l > N$,

$$d(\xi, \Omega_A) \leq \frac{1}{l} (4Nr^2 + 4(l - N)a^N r^2) \underset{l \rightarrow \infty}{=} 4a^N r^2,$$

which can be made arbitrarily small by suitable choice of N .

Further, as the following example shows, for Schur LTI systems, the bound (8) converges to 0, for $l \rightarrow \infty$, which demonstrates the bound’s tightness.

Example V.2 (Schur LTI systems). For an LTI system $x^+ = Ax$, where the matrix A is Schur-stable (its spectral radius is less than 1), we have $c_{\mathcal{B}_l^S} \leq v_n$ and

$$\lim_{l \rightarrow \infty} J_{b_l}(x)^T J_{b_l}(x) = \sum_{i=0}^{\infty} (A^i)^T A^i = W,$$

where W is the discrete observability Gramian [34, Thm. 6.D1] for the pair (A, I) ; i.e., W is the solution to the Lyapunov equation $W - A^T W A = I$. Thus, both $h(\xi)$, $h_s(\xi)$ are finite, for $l \rightarrow \infty$, and the bound (8) converges to 0.

Conversely, the example below shows that, even for marginally stable systems, the bound may not vanish with $l \rightarrow \infty$.

Example V.3 (Marginally stable LTI system). For the simple system $x^+ = x$, we have that $\det(J_{b_l}(x)^T J_{b_l}(x)) = \det(lI) = l^n$ and, by Prop. V.4, $h(\xi) = h(x_0) + \frac{n \log l}{2}$ and

$$\begin{aligned} h_s(\xi) &= \frac{1}{1-s} \log \left(\frac{1}{l^{n(s-1)/2}} \int_{\mathbb{R}^n} p_0^s dx \right) \\ &= -\frac{n(s-1) \log l}{2(1-s)} + \frac{1}{1-s} \log \int_{\mathbb{R}^n} p_0^s dx \\ &= \frac{n \log l}{2} + h_s(\xi_0). \end{aligned}$$

Replaced in the distortion bound (8), the l in the denominator is canceled out, indicating a positive lower bound for any l . This independence on l is expected, as abstracting $x^+ = x$ is the same as encoding the initial condition ξ_0 .

C. A relaxed lower bound and the curse of dimensionality

We now derive relaxed lower bounds that do not depend on the dynamics, and thus do not require computing trajectory entropies, as the following corollary to Thm. V.2 shows.

Corollary V.6 (Relaxed lower bound for abstractions). Consider a dynamical system $x^+ = f(x)$ under the assumptions of Thm. V.2, with $\xi_0 \sim p_{\xi_0} : \mathcal{X} \rightarrow \mathbb{R}_+$. Assume that f is either (piecewise) affine or Lipschitz continuous with constant L . The average distortion of any abstraction A with partition size $|\mathcal{Y}| \leq e^R$, where $R > 0$, is lower bounded as follows:

$$\begin{aligned} D_{abs}(R) &\geq K(l) \left(\frac{n}{2} \left(\frac{e^{-R+h(\xi_0)-n/2}}{\Gamma(1+n/2)} \right)^{2/n} \right. \\ &\quad \left. + \sup_{s \in (1, \infty]} e^{\frac{2}{n}(-\frac{s}{s-1}R + h_s(\xi_0))} \right), \end{aligned} \quad (14)$$

where

- 1) $K(l) = \frac{1-L^2}{lv_n^{2/n}}$, if f is Lipschitz with $L < 1$;
- 2) $K(l) = \frac{1}{l^2 v_n^{2/n}}$, if f is Lipschitz with $L = 1$;
- 3) $K(l) = \frac{L^2-1}{lL^{2l} v_n^{2/n}}$ if f is Lipschitz with $L > 1$;
- 4) $K(l) = \frac{1}{lM^{2l/n} v_n^{2/n}}$, if f is piecewise affine with M pieces.

Moreover, picking $s = \infty$ yields:

$$R_{abs}(D) \geq \frac{n}{2} \log \left(K(l) \frac{n}{2} \frac{e^{\frac{2}{n}h(\xi_0)-1}}{\Gamma(1+n/2)^{2/n}} + K(l)e^{\frac{2}{n}h(\xi_0)} \right) - \frac{n}{2} \log(D). \quad (15)$$

While more conservative than Thm. V.2, Cor. V.6 confirms abstractions' curse of dimensionality: for any system and finite l , the partition size e^R *fundamentally increases exponentially* with n to achieve a prescribed distortion (by (15)).

VI. NUMERICAL EXAMPLES

A. The doubling map

Consider the doubling map from Example II.1. For any trajectory length l , its behavior $\mathcal{B}_l^S$ is composed by 2^{l-1} line segments in $\mathbb{R}^l$ described by $(x_0, 2x_0, 4x_0, \dots, 2^{l-1}x_0) \bmod 1$, uniformly distributed with $p_\xi(\xi) = 1/\sqrt{1+4+\dots+4^{l-1}} = \sqrt{3/(4^l-1)}$, giving $h(\xi) = h_s(\xi) = \frac{1}{2}(\log(4^l-1) - \log 3)$ for all $s \in (1, \infty]$. Using Prop. V.5 for piecewise affine systems, we obtain $c_{\mathcal{B}_l^S} \leq \sqrt{(4^l-1)/3}v_n$, enabling us to compute the lower bound in Theorem V.2. In light of Remark 6, we also determine the high-rate lower bound by picking $c_{\mathcal{B}_l^S} = v_1 = 2$. The lower-bound curves can be seen in Fig. 4 (left) for $s = 2$ and $s = \infty$. It is apparent that the tightest bound is obtained with $s = \infty$ and $c_{\mathcal{B}_l^S} = v_1$. In this case, the bound is consistently half of that of the optimal distortion $D_{abs}(R)$, which is remarkably close. As a comparison, the standard Shannon lower bound is 1.42 times smaller than the optimal quantizer distortion of a uniform random variable in $\mathbb{R}^1$, in the standard source-coding setting with Euclidean distortion.

Let us explain how we were able to compute the actual optimal achievable abstraction distortion. Following the reasoning in Section V, we first build an optimal cover for $\mathcal{B}_l^S$ (afterwards, we show that this optimal cover admits a distortion that is equal to that of a specific abstraction with the same rate, and thus its rate-distortion curve is optimal, among all abstractions). For a given l , consider $R = \log(k2^{l-1})$, where k is an arbitrary natural number. Since all segments are equiprobable and congruent, and probability is uniform among them, the optimal cover of $\mathcal{B}_l^S$ is obtained by cutting each of the 2^{l-1} segments in k equal pieces.⁷ The cover is then the collection $M = \{C_i\}_{i=1}^{2^{l-1}k}$ of all these pieces—see Fig. 4 for an example with $l = 3$ and $k = 2$. Now let $C_\xi \in M$ denote the piece including the trajectory ξ . The expected error between ξ and the Chebyshev center of C_ξ is $\mathbb{E}[\|\xi - x_c(C_\xi)\|^2] = \frac{1-4^{-l}}{9k^2}$, which is obtained by computing the squared length of each segment, $L^2 = 1/(2^{l-1}k)^2(1+2^2+\dots+(2^{l-1})^2) = (4^l-1)/(3k^24^{l-1}) =$

$4(1-4^{-l})/3k^2$, followed by using the variance of the uniform distribution, giving $L^2/12$. Then, for the cover's set-based distortion, we have $d(\xi, C_\xi) = \frac{1}{l} \max_{\xi' \in C_\xi} \|\xi - \xi'\|^2 = \frac{1}{l}(\|\xi - x_c(C_\xi)\| + L/2)^2$, since C_ξ is a straight-line segment. We, thus have $\mathbb{E}[d(\xi, C_\xi)] = \frac{1}{l} \mathbb{E}[(\|\xi - x_c(C_\xi)\| + L/2)^2] = 7L^2/(12l)$, where we used that $X := \|\xi - x_c(C_\xi)\| + L/2$ is a uniform random variable in $[L/2, L]$, and thus $\mathbb{E}[X^2] = \text{Var}(X) + \mathbb{E}[X]^2 = L^2/48 + 9L^2/16 = 7L^2/12$. As M is the optimal cover, we have $D_{cover}(R) = \mathbb{E}[d(\xi, C_\xi)] = \frac{7}{36l}(4^l-1)e^{-2R}$, where we used $R = \log(k2^{l-1})$, and $D_{cover}(R)$ is the optimal distortion among all nl -dimensional covers of $\mathcal{B}_l^S$.

Finally, we show that, for $R = \log(k2^{l-1})$, $l = 1, 2, \dots$,

$$D_{abs}(R) = D_{cover}(R) = \frac{7}{36l}(4^l-1)e^{-2R}.$$

Notice that, in general, $D_{abs}(R) \geq D_{cover}(R)$, as behaviors of abstractions are covers. However, the optimal cover built above admits an abstraction A that gives the same distortion, and thus we have $D_{abs}(R) = D_{cover}(R)$. First, let $\mathcal{Y}$ be the uniform grid of $[0, 1]$ with segments of length $1/k2^{l-1}$. Each trajectory of the abstraction is a sequence of segments of lengths $1/k2^{l-1}, 1/k2^{l-2}, \dots, 1/k$; thus, for any given trajectory ξ , the abstraction output $\Omega_A(\xi)$ is a box in $\mathbb{R}^l$ containing any related trajectory ξ (see Fig. 4). In fact, ξ lies on one of the diagonals of the box $\Omega_A(\xi)$, and this diagonal is precisely C_ξ . We thus have $\arg \sup_{\xi' \in \Omega_A(\xi)} \|\xi - \xi'\|^2 = \arg \sup_{\xi' \in C_\xi} \|\xi - \xi'\|^2$, and hence $d(\xi, \Omega_A(\xi)) = d(\xi, C_\xi)$. Consequently, the abstraction has the same distortion as the optimal cover.

B. A 3D nonlinear system and uniform-grid abstractions

Consider the nonlinear system $f: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ where

$$f(x) = \begin{bmatrix} 0.9x_1 + 0.1 \sin x_2 \\ 2x_2^3 - x_2 \\ 0.9x_3 + 0.1x_1x_2, \end{bmatrix}$$

and $\mathcal{X} = [-1, 1]^3$, which is forward invariant under f . This system has multiple equilibria, hence the origin is not stable in $\mathcal{X}$. For each N in $\{10, 20, 50, 100\}$, we build abstractions A_N by using uniform partitioning of $\mathcal{X}$ with grids of size $N \times N \times N$ and determining the transition map using interval arithmetic. Then, we compute the distortion lower bound from Theorem V.2 using Prop. V.4 and Prop. V.5, case 3.⁸ Furthermore, lower bounds were also computed by picking $c_{\mathcal{B}_l^S} = v_3$, in light of Remark 6. The resulting distortion lower bound curves can be seen in Fig. 5. In this case, as the abstraction we construct is not necessarily the optimal one, its expected distortion is generally 100x higher than the fundamental lower bound. Still, this demonstrates the validity of the lower bound, even in cases with complex, nonlinear dynamics; even more importantly, it indicates how conservative standard abstractions with uniform grids might be.

VII. CONCLUSION AND FUTURE RESEARCH

We have developed a statistical, quantitative theory on the size-accuracy tradeoff of finite abstractions of dynamical

⁷For uniform distributions, with convex and symmetric distortion functions as the one in (6), optimal covers (or *quantizers*) are uniform [35].

⁸The entropies were computed using Monte-Carlo integration with 10000 samples, while Jacobians and the Lipschitz constant were determined using automatic differentiation.

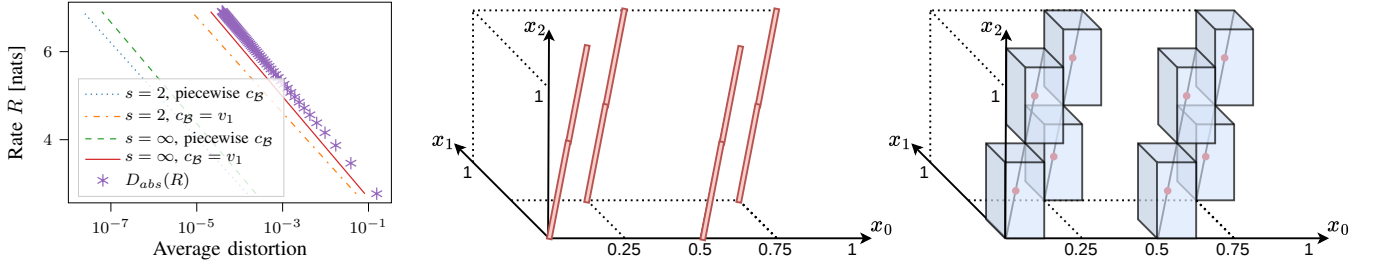


Fig. 4: Section VI-A: Comparison between $D_{abs}(R)$ and the fundamental lower bound from Theorem V.2, for $l = 5$ (left). Optimal cover of $\mathcal{B}_3^S$ with $k = 2$ (center), and abstract reachable sets (right, blue boxes). On the right, the lines inside the boxes represent $\mathcal{B}_3^S$, with their Chebyshev centers marked in red. The maximal distance between any point in $\mathcal{B}_3^S$ and its corresponding blue box is obtained at one of the edges intersecting with the trajectory.

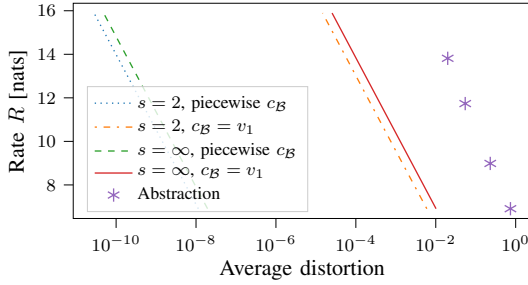


Fig. 5: Section VI-B: Comparison between average distortion ($l = 5$) from the constructed abstractions and the fundamental lower bound from Theorem V.2.

systems. Through this theory, we have uncovered fundamental limits on their scalability: given the dynamics, we have obtained a fundamental bound on the achievable abstraction accuracy, for a given abstraction size; and, conversely, a lower bound on the abstraction size, for a prescribed accuracy. To that end, we have established connections with rate-distortion theory. From an information-theoretic perspective, we have developed rate-distortion theory for the class of encoder-decoder pairs that abstractions constitute: set-based, with set-based distortion. This novel theory quantifies scalability limits of abstractions, and provides insights on how the complexity of the dynamics to be abstracted dictates these limits.

Importantly, the developed theory may be employed to construct minimal abstractions, harnessing their full scalability potential. This work made clear that, to construct minimal abstractions, one has to encode system trajectories, through covers in a high-dimensional ambient space. This is demonstrated in Section VI-A, where we construct a minimal abstraction of the doubling-map dynamics. Future research will thus focus on the general problem of constructing minimal abstractions.

Further, extending our theory to control-oriented abstractions, e.g. alternating simulations [1, Def. 9.6] is a topic of future research. A main challenge is incorporating inputs in the information-theoretic framework. Although one may cast inputs the same way as states, and proceed with extending our theory, this might yield conservative results. A more sensible approach might be capturing the causal relationship between control and trajectories, through causal rate-distortion theoretic concepts, e.g. non-anticipative rate-distortion functions [36].

VIII. TECHNICAL RESULTS AND PROOFS

Proof of Prop. V.1. For any given $\xi \in \mathcal{B}_l^S$, we will prove that

$$d(\xi, \Omega_A) \geq \frac{1}{l} \|\xi - \xi_{q_A}\|^2 + \frac{1}{l} r_c^2(\Omega_A),$$

where note that $\xi \in \Omega_A$ (by behavioral inclusion, c.f. Thm. III.3) and $\xi_{q_A} = x_c(\Omega_A)$. Then, the proof is complete by applying the expectation operator to the above inequality.

Notice that, since $\Omega_A \subseteq \mathcal{X}^l$ is bounded, $\overline{\Omega_A}$ is compact, and taking the closure does not change suprema/infima taken over Ω_A or the Chebyshev radius, that is $\sup_{y \in \Omega_A} \|y - x'\| = \max_{y \in \overline{\Omega_A}} \|y - x'\|$, for any x' . Hence we equivalently work here with $\overline{\Omega_A}$, for which maxima are attained. Define $w(x') = \max_{y \in \overline{\Omega_A}} \|y - x'\|^2$. The function w is convex, being the pointwise maximum of the convex quadratic maps $x' \mapsto \|y - x'\|^2$. We have $x_c(\Omega_A) = \arg \min_{x' \in \mathbb{R}^{nl}} w(x')$ and $r_c^2(\Omega_A) = w(x_c(\Omega_A)) = \max_{y \in \overline{\Omega_A}} \|y - x_c(\Omega_A)\|^2$.

Define the set of maximizers

$$M := \{y \in \overline{\Omega_A} : \|y - x_c(\Omega_A)\| = r_c(\Omega_A)\}.$$

The subdifferential of w at $x_c(\Omega_A)$ is $\partial w(x_c(\Omega_A)) = 2 \text{conv}\{x_c(\Omega_A) - y : y \in M\}$, where conv denotes the convex hull operator. Since $x_c(\Omega_A)$ minimizes w , the optimality condition $0 \in \partial w(x_c(\Omega_A))$ gives $0 \in \text{conv}\{x_c(\Omega_A) - y : y \in M\}$. Hence there exist finitely many points $y_1, \dots, y_m \in M$ and coefficients $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$, such that

$$\sum_{i=1}^m \lambda_i (y_i - x_c(\Omega_A)) = 0. \quad (16)$$

Now, fix $\xi \in \mathcal{B}_l^S$ and let $\xi_* \in \arg \max_{y \in \overline{\Omega_A}} \|y - \xi\|^2$. By definition of ξ_* , for every $y_i \in M$ we have $\|y_i - \xi\|^2 \leq \|\xi_* - \xi\|^2$. Taking the convex combination with the λ_i gives

$$\sum_{i=1}^m \lambda_i (\|y_i - \xi\|^2 - \|\xi_* - \xi\|^2) \leq 0.$$

Since $\|y_i - \xi\|^2 = \|y_i - x_c(\Omega_A)\|^2 + \|x_c(\Omega_A) - \xi\|^2 + 2(y_i - x_c(\Omega_A))^T(x_c(\Omega_A) - \xi)$ and $\|y_i - x_c(\Omega_A)\|^2 = r_c(\Omega_A)^2$, for the above inequality we have

$$r_c(\Omega_A)^2 + \|x_c(\Omega_A) - \xi\|^2 - \|\xi_* - \xi\|^2 \leq 0,$$

where the cross term has vanished, using (16). Finally, using (6),

$$r_c(\Omega_A)^2 + \|\xi_{q_A} - \xi\|^2 \leq \|\xi_* - \xi\|^2 = l \cdot d(\xi, \Omega_A).$$

□

Towards proving Thm. V.2, we use the following lemma.

Lemma VIII.1. *Let $M \subset \mathbb{R}^n$ be a finite union of bounded, m -dimensional C^1 -manifolds. Let X be a random variable in M with probability measure $\mu_X \ll \mathcal{H}_M^m$ and density $p = \frac{d\mu_X}{d\mathcal{H}_M^m}$. Consider a finite measurable cover $\mathcal{R} := \{R_i\}_{i=1}^N$ of M , and a measurable function $w : M \rightarrow \mathcal{R}$, which is such that $x \in w(x)$ for all $x \in M$. Then:*

$$\mathbb{E}_X[r_c(w(X))^2] \geq c_M^{-2/m} \sup_{s \in (1, \infty]} e^{\frac{2}{m} h_s(X)} N^{-\frac{2}{m(1-1/s)}}, \quad (17)$$

where c_M is defined by (1), and $r_c(R_i)$ denotes the Chebyshev radius of R_i .

Proof. Define $S_i := \{x \in \mathcal{M} : w(x) = R_i\} = w^{-1}(\{R_i\})$. Since w is measurable and single-valued, the sets S_i are measurable and pairwise disjoint. Moreover, since $w(x) \in \mathcal{R}$ for all $x \in \mathcal{M}$, they cover $\mathcal{M}$. Thus, $\{S_i\}_{i=1}^N$ forms a measurable partition of $\mathcal{M}$. Furthermore, since $x \in w(x)$, we have $S_i \subseteq R_i$ for every i . Let

$$p_i := \mu_X(S_i) = \int_{S_i} p d\mathcal{H}^m,$$

so that $\sum_{i=1}^N p_i = 1$, and let $r_i := r_c(S_i)$ whenever $S_i \neq \emptyset$ (set $r_i := 0$ otherwise). Since $S_i \subseteq R_i$, $r_i \leq r_c(R_i)$. Thus,

$$\mathbb{E}_X[r_c(w(X))^2] = \sum_{i=1}^N p_i r_c(R_i)^2 \geq \sum_{i=1}^N p_i r_i^2.$$

Hence, it suffices to lower bound $\sum_{i=1}^N p_i r_i^2$.

For a given i , by definition, $S_i \subset \overline{B}(c_i, r_i) \cap M$ for some Chebyshev center c_i . For any $s > 1$, we have:

$$\begin{aligned} p_i &\leq \int_{\overline{B}(c_i, r_i) \cap M} p d\mathcal{H}^m \\ &= \int_M p \mathbf{1}_{\overline{B}(c_i, r_i) \cap M} d\mathcal{H}^m \\ &\leq \left(\int_M p^s d\mathcal{H}^m \right)^{1/s} \left(\int_M (\mathbf{1}_{\overline{B}(c_i, r_i) \cap M})^{\frac{s}{s-1}} d\mathcal{H}^m \right)^{1-1/s} \\ &\leq \|p\|_{L^s(M)} (\mathcal{H}^m(\overline{B}(c_i, r_i) \cap M))^{1-1/s} \\ &\leq \|p\|_{L^s(M)} (c_M r_i^m)^{1-1/s}, \end{aligned}$$

where $\|p\|_{L^s(M)} := \left(\int_M p^s d\mathcal{H}^m \right)^{1/s}$, in the third step we used Hölder's inequality, and in the final step we used the inequality (1).⁹ Defining $K_s := \|p\|_{L^s(M)} c_M^{1-1/s}$, from the inequality above we have:

$$r_i \geq \left(\frac{p_i}{K_s} \right)^{\frac{1}{m(1-1/s)}} \implies r_i^2 \geq \left(\frac{p_i}{K_s} \right)^\alpha,$$

where $\alpha := 2/(m(1-1/s)) = 2s/(m(s-1))$. Multiplying by p_i gives

$$p_i r_i^2 \geq K_s^{-\alpha} p_i^{1+\alpha} \implies \sum_{i=1}^N p_i r_i^2 \geq K_s^{-\alpha} \sum_{i=1}^N p_i^{1+\alpha}. \quad (18)$$

⁹Although (1) is defined for open balls, it also holds for closed balls, since $\overline{B}(c_i, r_i) \subset B(c_i, r_i + \varepsilon)$ for every $\varepsilon > 0$, and one may let $\varepsilon \downarrow 0$.

Our job now is to find a lower bound to $\sum_{i=1}^N p_i^{1+\alpha}$ over discrete probabilities p_i . First, notice that $\alpha > 0$ since $s > 1$. Therefore, the map $t \mapsto t^{1+\alpha}$ is convex in $t \in [0, +\infty)$. Thus, by Jensen's inequality,

$$\sum_{i=1}^N p_i^{1+\alpha} \geq N \left(\frac{1}{N} \sum_{i=1}^N p_i \right)^{1+\alpha} = N \left(\frac{1}{N} \right)^{1+\alpha} = N^{-\alpha}.$$

Substituting in (18) gives

$$\sum_{i=1}^N p_i r_i^2 \geq (K_s N)^{-\alpha} = \left(\int_M p^s d\mathcal{H}^m \right)^{-\alpha/s} N^{-\alpha} c_M^{-2/m}.$$

Now, by definition of the Rényi entropy, we get

$$h_s(x) = \frac{1}{1-s} \log \mathbb{E}[p(x)^{s-1}] = \frac{1}{1-s} \log \int \left(\frac{d\mu_x}{d\mathcal{H}_X^m} \right)^{s-1} d\mu_x,$$

which, by the properties of the Radon–Nikodym derivative, gives

$$\begin{aligned} h_s(x) &= \frac{1}{1-s} \log \int \left(\frac{d\mu_x}{d\mathcal{H}_X^m} \right)^s d\mathcal{H}_X^m \\ &= \frac{1}{1-s} \frac{s}{\alpha} \log \int p^s d\mathcal{H}_X^m \\ &= -\frac{s}{\alpha(1-s)} \log \left(\int p^s d\mathcal{H}_X^m \right)^{-\alpha/s}. \end{aligned}$$

And, using the fact that $\alpha(1-s) = -2s/m$, we get

$$\begin{aligned} h_s(x) &= \frac{m}{2} \log \left(\int p^s d\mathcal{H}_X^m \right)^{-\alpha/s} \\ \iff \left(\int p^s d\mathcal{H}_X^m \right)^{-\alpha/s} &= e^{\frac{2}{m} h_s(x)}. \end{aligned}$$

Therefore, for any $s > 1$, we have

$$\mathbb{E}_X[r_c(w(X))^2] \geq c_M^{-2/m} e^{\frac{2}{m} h_s(X)} N^{-\frac{2}{m(1-1/s)}}.$$

Since this holds for every $s \in (1, \infty)$, and $h_s(X) \rightarrow h_\infty(X)$ as $s \rightarrow \infty$, taking the supremum over $s \in (1, \infty]$ yields (17). □

Proof of Thm. V.2. We make use of Prop. V.1. Take (7) and minimize both sides over all possible partitions $\mathcal{Y}$ with size $|\mathcal{Y}| \leq e^R$ and associated abstractions A . We have

$$D_{abs}(R) \geq \inf_{A, |\mathcal{Y}| \leq e^R} \frac{1}{l} \mathbb{E}_\xi[\|\xi - \xi_{q_A}\|^2] + \frac{1}{l} \mathbb{E}_\xi[r_c^2(\Omega_A)],$$

where recall that $\mathbb{E}_{\xi_0}[\cdot] = \mathbb{E}_\xi[\cdot]$, and that for a given abstraction A with corresponding encoder-decoder pair s_A, g_A , we have $\xi_{q_A} = g_{q_A}(s_{q_A}(\xi))$ with $s_{q_A}(\xi) = g_A(s_A(\xi))$ and $g_{q_A}(z) = x_c(z)$, where $x_c(z)$ is the Chebyshev center of the set z ; and $r_c(\Omega_A)$ is the Chebyshev radius of Ω_A . Notice that (s_{q_A}, g_{q_A}) is an encoder-decoder pair with rate R (the same as the abstraction) and ξ_{q_A} is its output. Hence, the first term in the left-hand side of the above inequality can be lower bounded by employing Thm. II.1, to obtain:

$$D_{abs}(R) \geq \frac{n}{2l} \left(\frac{e^{-R+h(\xi)-n/2}}{c_{\mathcal{B}_l^s} \Gamma(1+n/2)} \right)^{2/n} + \frac{1}{l} \inf_{A, |\mathcal{Y}| \leq e^R} \mathbb{E}_\xi[r_c^2(\Omega_A)].$$

To bound the second term, we employ Lemma VIII.1. Notice that the abstraction's outputs $\Omega_A \in \mathcal{R}_l^A$ are elements of the nl -dimensional cover $\mathcal{R}_l^A$ of $\mathcal{B}_l^S$ (which is n -dimensional) with cardinality $|\mathcal{R}_l^A| = |\mathcal{Y}| \leq e^R$. Thus, the term $\inf_{A, |\mathcal{Y}| \leq e^R} \mathbb{E}[r_c^2(\Omega_A)]$ can be lower bounded as in (17), where we replace M by $\mathcal{B}_l^S$, $\mathcal{R}_l^A$ by $\mathcal{R}$, m by n , N by $|\mathcal{Y}|$, μ_X by μ_ξ , $w(X)$ by $g_A(s_A(\xi)) =: \Omega_A$, and take $|\mathcal{Y}| \leq e^R$, to obtain:

$$\frac{1}{l} \inf_{A, |\mathcal{Y}| \leq e^R} \mathbb{E}_\xi[r_c^2(\Omega_A)] \geq \frac{1}{l} c_{\mathcal{B}_l^S}^{-2/n} \sup_{s \in (1, \infty]} e^{\frac{2}{n}(-\frac{s}{s-1}R + h_s(\xi))}$$

□

Proof of Prop. V.4. Fix any measurable subset $\mathcal{A} \subseteq \mathcal{X}$. Because $\mu_{\xi_0}(\mathcal{A}) = \mu_\xi(b_l(\mathcal{A}))$, the definitions of p_ξ and p_{ξ_0} imply that

$$\int_{b_l(\mathcal{A})} p_\xi(y) d\mathcal{H}_{\mathcal{B}_l^S}^n(y) = \int_{\mathcal{A}} p_{\xi_0}(x) d\mathcal{L}^n(x).$$

But also, since the first n components of $b_l(x)$ equal x itself, we have that b_l is injective, allowing the use of the area formula [33, Thm. 3.2.5] to get

$$\int_{b_l(\mathcal{A})} p_\xi(y) d\mathcal{H}_{\mathcal{B}_l^S}^n(y) = \int_{\mathcal{A}} p_\xi(b_l(x)) \sqrt{\det(J_{b_l}(x)^T J_{b_l}(x))} d\mathcal{L}^n$$

implying that, for almost all $x \in \mathcal{X}$,

$$p_\xi(b_l(x)) = \frac{p_{\xi_0}(x)}{\sqrt{\det(J_{b_l}(x)^T J_{b_l}(x))}}. \quad (19)$$

The expression above is well-defined everywhere as $J_{b_l}^T J_{b_l} = I + J_f^T J_f + \dots \succeq I$, implying that $\det(J_{b_l}^T J_{b_l}) \geq 1$.

Then, (2) becomes

$$\begin{aligned} h(\xi) &= - \int_{\mathbb{R}^n} p_{\xi_0}(x) \log(p_{\xi_0}(x)) d\mathcal{L}^n \\ &\quad + \frac{1}{2} \int_{\mathbb{R}^n} \log \det(J_{b_l}(x)^T J_{b_l}(x)) p_{\xi_0}(x) d\mathcal{L}^n. \end{aligned}$$

Likewise, the area formula gives

$$\begin{aligned} h_s(\xi) &= \frac{1}{1-s} \log \int_{\mathcal{B}_l^S} p_\xi^s d\mathcal{H}^n \\ &= \frac{1}{1-s} \log \int_{\mathbb{R}^n} p_\xi(b_l(x))^s \det(J_{b_l}(x)^T J_{b_l}(x))^{\frac{1}{2}} d\mathcal{L}^n, \end{aligned}$$

and, applying (19) gives (12). In the case of $s = \infty$, we have

$$\begin{aligned} h_s(\xi) &= \frac{s}{1-s} \log \left(\int_{\mathbb{R}^n} \frac{p_{\xi_0}(x)^s}{\det(J_{b_l}(x)^T J_{b_l}(x))^{\frac{s-1}{2}}} dx \right)^{1/s} \\ &\stackrel{s \rightarrow \infty}{=} - \log \operatorname{ess\,sup} \frac{p_{\xi_0}(x)}{\sqrt{\det(J_{b_l}(x)^T J_{b_l}(x))}} \\ &= \operatorname{ess\,inf} \log \frac{\sqrt{\det(J_{b_l}(x)^T J_{b_l}(x))}}{p_{\xi_0}(x)}, \end{aligned}$$

which gives (13) by the properties of log.

Finally, since $\det(J_{b_l}^T J_{b_l}) \geq 1$, we establish that $h(\xi) \geq h(\xi_0)$ and $h_s(\xi) \geq h_s(\xi_0)$ for any $s > 1$. □

Lemma VIII.2. Let $X \subset \mathbb{R}^N$, and $w : X \rightarrow \mathbb{R}^N$, $N \geq n$, be an L -Lipschitz function satisfying $w(x) = [x^T \ g(x)]^T$ for some function g . Then for every $y \in \mathbb{R}^N$ and $\delta > 0$,

$$\mathcal{H}^n(w(X) \cap B(y, \delta)) \leq L^n v_n \delta^n.$$

Proof. Fix $y \in \mathbb{R}^N$ and $\delta > 0$ and define $Z := w(X) \cap B(y, \delta)$ and its pre-image $E := w^{-1}(Z) \subset \mathbb{R}^N$. Let $P : \mathbb{R}^N \rightarrow \mathbb{R}^n$ denote the Euclidean projection onto the first n coordinates. Because $w(x) = [x^T \ g(x)]^T$, we have that $P(w(x)) = x$ for all $x \in X$. Thus, for any $x \in E$, since $w(x) \in B(y, \delta)$, we get $\|x - P(y)\| = \|P(w(x)) - P(y)\| \leq \|w(x) - y\| < \delta$, where the first inequality is because orthogonal projections are nonexpansive. Hence, $E \subseteq B(P(y), \delta)$, and consequently $\mathcal{H}^n(E) \leq v_n \delta^n$. Since w is L -Lipschitz, by fundamental properties of the Hausdorff measure [33, Sec. 2.10.11],

$$\mathcal{H}^n(Z) = \mathcal{H}^n(w(E)) \leq L^n \mathcal{H}^n(E) \leq L^n v_n \delta^n.$$

□

Proof of Prop. V.5. We again use the function $b_l : \mathcal{X} \rightarrow \mathcal{B}_l^S$, defined by (10). Since by assumption $\mathcal{X}$ is full dimensional in $\mathbb{R}^n$, the tightest value for $c_{\mathcal{X}} = c_{\mathbb{R}^n} = v_n$. Now we look at each case.

Case (1) follows trivially by the observation that $\mathcal{B}_l^S$ is an n -dimensional affine subset of $\mathbb{R}^{nl}$, and that the intersection of an nl -ball of radius r and a plane of dimension n is a ball of dimension n and radius $\leq r$. Hence, $\mathcal{H}_{\mathcal{B}_l^S}(B(z, \delta) \cap \mathcal{B}_l^S) \leq v_n \delta^n$, for all $z \in \mathbb{R}^{nl}$.

Case (2): If f is piecewise affine, so is $\mathcal{B}_l^S$, which has at most M^l disjoint pieces. Denote by Z_i each such piece of $\mathcal{B}_l^S$, which is a bounded, connected n -dimensional subset of some affine subspace of $\mathbb{R}^{nl}$. Thus, $\mathcal{B}_l^S = \bigcup_{i=1}^N Z_i$, with $N \leq M^l$. Then, for all $z \in \mathbb{R}^{nl}$ and $\delta > 0$,

$$\mathcal{H}^n\left(\bigcup_i Z_i \cap B(z, \delta)\right) = \sum_{i=1}^N \mathcal{H}^n(Z_i \cap B(z, \delta)) \leq M^l v_n \delta^n,$$

where in the last inequality we used case (1) and $N \leq M^l$.

Case (3): It is easy to see that b_l satisfies the conditions of Lemma VIII.2 with Lipschitz constant $(\sum_{i=0}^{l-1} L^{2i})^{1/2}$. Hence the result comes from applying Lemma VIII.2. □

Proof of Cor. V.6. We start by applying the property that $h(\xi) \geq h(\xi_0)$ and $h_s(\xi) \geq h_s(\xi_0)$ from Prop. V.4 in (8), Thm. V.2. The fact that the RHS of (8) is monotonically increasing with $h(\xi)$ and $h_s(\xi)$ gives

$$\begin{aligned} D_{abs}(R) &\geq \frac{n}{2l} \left(\frac{e^{-R+h(\xi_0)-n/2}}{c_{\mathcal{B}_l^S} \Gamma(1+n/2)} \right)^{2/n} \\ &\quad + \frac{1}{l} c_{\mathcal{B}_l^S}^{-2/n} \sup_{s \in (1, \infty]} e^{\frac{2}{n}(-\frac{s}{s-1}R + h_s(\xi_0))}. \end{aligned} \quad (20)$$

Define $\tilde{K}(l) := \frac{1}{l} (\frac{1}{c_{\mathcal{B}_l^S}})^{2/n}$. Applying Prop. V.5 case (2) to $\tilde{K}(l)$ gives, for the piecewise affine case with M pieces,

$$\tilde{K}(l) \geq K(l) = (l M^{2l/n} v_n^{2/n})^{-1},$$

which is case (4) of the corollary. For the remaining cases, we apply Prop. V.5, case (3), and use the fact that $\sum_{i=0}^{k-1} a^i = k$ for $a = 1$, and otherwise $\sum_{i=0}^{k-1} a^i = (a^k - 1)/(a - 1) \leq \max(1, a^k)/|a - 1|$. Thus, we bound $\tilde{K}(l)$ accordingly via

- 1) $\tilde{K}(l) \geq \frac{1-L^2}{l v_n^{2/n}}$, if f is Lipschitz with $L < 1$;
- 2) $\tilde{K}(l) \geq \frac{1}{l^2 v_n^{2/n}}$, if f is Lipschitz with $L = 1$;
- 3) $\tilde{K}(l) \geq \frac{L^2-1}{l L^{2l} v_n^{2/n}}$ if f is Lipschitz with $L > 1$.

The final expression (14) is then obtained by applying the respective bounds on $\tilde{K}(l)$ to (20), which is then inverted for the $s = \infty$ case to arrive at (15). $\square$

REFERENCES

- [1] P. Tabuada, *Verification and control of hybrid systems: a symbolic approach*. Springer Science & Business Media, 2009.
- [2] A. Lavaei, S. Soudjani, A. Abate, and M. Zamani, “Automated verification and synthesis of stochastic hybrid systems: A survey,” *Automatica*, vol. 146, p. 110617, 2022.
- [3] A. Girard, G. Pola, and P. Tabuada, “Approximately bisimilar symbolic models for incrementally stable switched systems,” *IEEE Transactions on Automatic Control*, vol. 55, no. 1, pp. 116–126, 2009.
- [4] M. Rungger and M. Zamani, “SCOTS: A tool for the synthesis of symbolic controllers,” in *Proceedings of the 19th international conference on hybrid systems: Computation and control*, 2016, pp. 99–104.
- [5] K. Mallik, A.-K. Schmuck, S. Soudjani, and R. Majumdar, “Compositional synthesis of finite-state abstractions,” *IEEE Transactions on Automatic Control*, vol. 64, no. 6, pp. 2629–2636, 2018.
- [6] M. Zamani, P. M. Esfahani, R. Majumdar, A. Abate, and J. Lygeros, “Symbolic control of stochastic systems via approximately bisimilar finite abstractions,” *IEEE Transactions on Automatic Control*, vol. 59, no. 12, pp. 3135–3150, 2014.
- [7] M. Lahijanian, S. B. Andersson, and C. Belta, “Formal verification and synthesis for discrete-time stochastic systems,” *IEEE Transactions on Automatic Control*, vol. 60, no. 8, pp. 2031–2045, 2015.
- [8] A. Abate, J.-P. Katoen, J. Lygeros, and M. Prandini, “Approximate model checking of stochastic hybrid systems,” *European Journal of Control*, vol. 16, no. 6, pp. 624–641, 2010.
- [9] R. Coppola, A. Peruffo, and M. Mazo, “Data-driven abstractions for verification of linear systems,” *IEEE Control Systems Letters*, vol. 7, pp. 2737–2742, 2023.
- [10] T. Badings, L. Romao, A. Abate, D. Parker, H. A. Poonawala, M. Stoelinga, and N. Jansen, “Robust control for dynamical systems with non-gaussian noise via formal abstractions,” *Journal of Artificial Intelligence Research*, vol. 76, pp. 341–391, 2023.
- [11] A. Devonport, A. Saoud, and M. Arcak, “Symbolic abstractions from data: A PAC learning approach,” in *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 2021, pp. 599–604.
- [12] M. Kazemi, R. Majumdar, M. Salamati, S. Soudjani, and B. Wooding, “Data-driven abstraction-based control synthesis,” *Nonlinear Analysis: Hybrid Systems*, vol. 52, p. 101467, 2024.
- [13] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.
- [14] E. Riegler, H. Bölcskei, and G. Koliander, “Rate-distortion theory for general sets and measures,” in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 101–105.
- [15] E. Riegler, G. Koliander, and H. Bölcskei, “Lossy compression of general random variables,” *Information and Inference: A Journal of the IMA*, vol. 12, no. 3, pp. 1759–1829, 2023.
- [16] S. Esmail Zadeh Soudjani and A. Abate, “Adaptive and sequential gridding procedures for the abstraction and verification of stochastic processes,” *SIAM Journal on Applied Dynamical Systems*, vol. 12, no. 2, pp. 921–956, 2013.
- [17] S. Adams, M. Lahijanian, and L. Laurenti, “Formal control synthesis for stochastic neural network dynamic models,” *IEEE Control Systems Letters*, vol. 6, pp. 2858–2863, 2022.
- [18] Y. Tazaki and J.-i. Imura, “Discrete-state abstractions of nonlinear systems using multi-resolution quantizer,” in *International Workshop on Hybrid Systems: Computation and Control*. Springer, 2009, pp. 351–365.
- [19] A. Weber, M. Rungger, and G. Reissig, “Optimized state space grids for abstractions,” *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 5816–5821, 2016.
- [20] A. Saoud and A. Girard, “Multirate symbolic models for incrementally stable switched systems,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 9278–9284, 2017.
- [21] K. Hsu, R. Majumdar, K. Mallik, and A.-K. Schmuck, “Multi-layered abstraction-based controller synthesis for continuous-time systems,” in *Proceedings of the 21st International Conference on Hybrid Systems: Computation and Control (part of CPS Week)*, 2018, pp. 120–129.
- [22] J. Calbert, L. N. Egidio, and R. M. Jungers, “Smart abstraction based on iterative cover and non-uniform cells,” *IEEE Control Systems Letters*, vol. 8, pp. 2301–2306, 2024.
- [23] A.-K. Schmuck and J. Raisch, “Asynchronous l-complete approximations,” *Systems & Control Letters*, vol. 73, pp. 67–75, 2014.
- [24] A. Banse, G. Delimpaltadakis, L. Laurenti, M. Mazo Jr, and R. M. Jungers, “Memory-dependent abstractions of stochastic systems through the lens of transfer operators,” in *Proceedings of the 28th ACM International Conference on Hybrid Systems: Computation and Control*, 2025, pp. 1–12.
- [25] G. A. Gleizer and M. Mazo Jr, “Chaos and order in event-triggered control,” *IEEE Transactions on Automatic Control*, vol. 68, no. 11, pp. 6541–6556, 2023.
- [26] A. Lavaei, S. Soudjani, and M. Zamani, “Compositional abstraction of large-scale stochastic systems: A relaxed dissipativity approach,” *Nonlinear Analysis: Hybrid Systems*, vol. 36, p. 100880, 2020.
- [27] G. Delimpaltadakis, M. Lahijanian, M. Mazo Jr, and L. Laurenti, “Interval markov decision processes with continuous action-spaces,” in *Proceedings of the 26th ACM International Conference on Hybrid Systems: Computation and Control*, 2023, pp. 1–10.
- [28] D. Lind and B. Marcus, *An introduction to symbolic dynamics and coding*. Cambridge university press, 2021.
- [29] E. Lindenstrauss and M. Tsukamoto, “From rate distortion theory to metric mean dimension: variational principle,” *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3590–3609, 2018.
- [30] D. Abel, D. Arumugam, K. Asadi, Y. Jinnai, M. L. Littman, and L. L. Wong, “State abstraction as compression in apprenticeship learning,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 3134–3142.
- [31] O. Biza, R. Platt, J.-W. van de Meent, and L. L. Wong, “Learning discrete state abstractions with deep variational inference,” *arXiv preprint arXiv:2003.04300*, 2020.
- [32] D. T. Larsson, D. Maity, and P. Tsiotras, “A generalized information-theoretic framework for the emergence of hierarchical abstractions in resource-limited systems,” *Entropy*, vol. 24, no. 6, p. 809, 2022.
- [33] H. Federer, *Geometric measure theory*, ser. Grundlehren Math. Wiss. Springer, Cham, 1969, vol. 153.
- [34] C.-T. Chen, *Linear System Theory and Design (3rd Edition)*. Oxford University Press, 1999.
- [35] R. Gray and D. Neuhoff, “Quantization,” *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2325–2383, 1998.
- [36] C. D. Charalambous, P. A. Stavrou, and N. U. Ahmed, “Nonanticipative rate distortion function and relations to filtering theory,” *IEEE Transactions on Automatic Control*, vol. 59, no. 4, pp. 937–952, 2013.