Physics Letters A 380 (2016) 998–1002

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

Exact complexity: The spectral decomposition of intrinsic computation

James P. Crutchfield [a,*], Christopher J. Ellison [b], Paul M. Riechers [a]

[a] Complexity Sciences Center and Department of Physics, University of California at Davis, One Shields Avenue, Davis, CA 95616, United States
[b] Center for Complexity and Collective Computation, University of Wisconsin-Madison, Madison, WI 53706, United States

ARTICLE INFO

Article history:
Received 10 July 2015
Received in revised form 26 December 2015
Accepted 7 January 2016
Available online 11 January 2016
Communicated by C.R. Doering

Keywords:
Excess entropy
Statistical complexity
Projection operator
Resolvent
Entropy rate
Predictable information

ABSTRACT

We give exact formulae for a wide family of complexity measures that capture the organization of hidden nonlinear processes. The spectral decomposition of operator-valued functions leads to closed-form expressions involving the full eigenvalue spectrum of the mixed-state presentation of a process's ϵ-machine causal-state dynamic. Measures include correlation functions, power spectra, past-future mutual information, transient and synchronization informations, and many others. As a result, a direct and complete analysis of intrinsic computation is now available for the temporal organization of finitary hidden Markov models and nonlinear dynamical systems with generating partitions and for the spatial organization in one-dimensional systems, including spin systems, cellular automata, and complex materials via chaotic crystallography.

The emergence of organization in physical, engineered, and social systems is a fascinating and now, after half a century of active research, widely appreciated phenomenon [1–5]. Success in extending the long list of instances of emergent organization, however, is not equivalent to understanding what organization itself is. How do we say objectively that new organization has appeared? How do we measure quantitatively how organized a system has become?

Computational mechanics' answer to these questions is that a system's organization is captured in how it stores and processes information—how it computes [6]. *Intrinsic computation* was introduced two decades ago to analyze the inherent information processing in complex systems [7]: How much history does a system remember? In what architecture is that information stored? And, how does the system use it to generate future behavior?

Computational mechanics, though, is part of a long historical trajectory focused on developing a physics of information [8–10]. That nonlinear systems actively process information goes back to Kolmogorov [11], who adapted Shannon's communication theory [12] to measure the information production rate of chaotic dynamical systems. In this spirit, today computational mechanics is routinely used to determine physical and intrinsic computational properties in single-molecule dynamics [13], in complex materials [14], and even in the formation of social structure [15], to mention several recent examples.

Thus, measures of complexity are important to quantifying how organized nonlinear systems are: their randomness and their structure. Moreover, we now know that randomness and structure are intimately intertwined. One cannot be properly defined or even practically measured without the other [16, and references therein].

Measuring complexity has been a challenge: Until recently, in understanding the varieties of organization to be captured; still practically, in terms of estimating metrics from experimental data. One major reason for these challenges is that systems with emergent properties are hidden: We do not have direct access to their internal, often high-dimensional state space; we do not know a priori what the emergent patterns are. Thus, we must "reconstruct" their state space and dynamics [17–20]. Even then, when successful, reconstruction does not lead easily or directly to measures of structural complexity and intrinsic computation [7]. It gives access to what is hidden, but does not say what the mechanisms are nor how they work.

Our view of the various kinds of complexity and their measures, though, has become markedly clearer of late. There is a natural semantics of complexity in which each measure answers a specific question about a system's organization. For example:

- How random is a process? Its *entropy rate* h_μ [11].
- How much state information must be stored for optimal prediction? Its *statistical complexity* C_μ [7].

* Corresponding author.
E-mail addresses: chaos@ucdavis.edu (J.P. Crutchfield), cellison@wisc.edu (C.J. Ellison), pmriechers@ucdavis.edu (P.M. Riechers).

http://dx.doi.org/10.1016/j.physleta.2016.01.008

- How much of the future can be predicted? Its *past–future mutual information* or *excess entropy* **E** [16].
- How much information must an observer extract to know a process's hidden states? Its *transient information* **T** and *synchronization information* **S** [16].
- How much of the generated information (h_μ) affects future behavior? Its *bound information* b_μ [21].
- What's forgotten? Its *ephemeral information* ρ_μ [21].

And there are other useful measures ranging from degrees of irreversibility to quantifying model redundancy; see, for example, Ref. [22] and the proceedings in Refs. [23,24].

Unfortunately, except in the simplest cases where expressions are known for several, to date typically measures of intrinsic computation require extensive numerical simulation and estimation. Here we answer this challenge, providing exact expressions for a process's measures in terms of its ϵ-machine. In particular, we show that the spectral decomposition of this hidden dynamic leads to closed-form expressions for complexity measures. In this way, the remaining task in analyzing intrinsic computation reduces to mathematically constructing or reliably estimating a system's ϵ-machine in the first place.

Background

Our main object of study is a process $\mathcal{P}$, by which we mean the rather prosaic listing of all of a system's behaviors or realizations $\{\ldots x_{-2}, x_{-1}, x_0, x_1, \ldots\}$ and their probabilities: $\Pr(\ldots X_{-2}, X_{-1}, X_0, X_1, \ldots)$. We assume the process is stationary and ergodic and the measurement values range over a finite alphabet: $x \in \mathcal{A}$. This class describes a wide range of processes from statistical mechanical systems in equilibrium and in nonequilibrium steady states to nonlinear dynamical systems in discrete and continuous time on their attracting invariant sets.

Following Shannon and Kolmogorov, information theory gives a natural measure of a process's randomness as the uncertainty in measurement blocks: $H(L) = H[X_{0:L}]$, where H is the Shannon–Boltzmann entropy of the distribution governing the block $X_{0:L} = X_0, X_1, \ldots, X_{L-1}$. We monitor the *block entropy growth*—the average uncertainty in the next measurement X_{L-1} conditioned on knowing the preceding block $X_{0:L-1}$:

$$\begin{aligned} h_\mu(L) &= H(L) - H(L-1) \\ &= H[X_{L-1}|X_{0:L-1}] \\ &= -\left\langle \sum_{x_{L-1}\in\mathcal{A}} p(x_{L-1}) \log_2 p(x_{L-1}) \right\rangle_{\Pr(X_{0:L-1})} , \end{aligned} \quad (1)$$

where $p(x_{L-1}) = \Pr(x_{L-1}|x_{0:L-1})$. And when the limit exists, we say the process generates information at the *entropy rate*: $h_\mu = \lim_{L\to\infty} h_\mu(L)$.

Measurements, though, only indirectly reflect a system's internal organization. Computational mechanics extracts that hidden organization via the process's ϵ-machine [6], consisting of a set of recurrent *causal states* $\boldsymbol{\mathcal{S}} = \{\sigma^0, \sigma^1, \sigma^2, \ldots\}$ and transition dynamic $\{T^{(x)} : T^{(x)}_{i,j} = \Pr(x, \sigma^j|\sigma^i)\}_{x\in\mathcal{A}}$. Each causal state represents a collection of "equivalent" histories—equivalent in the sense that each history belonging to an equivalence class yields the same prediction over futures. The ϵ-machine is a system's unique, minimal-size, optimal predictor from which two key complexity measures can be directly calculated.

The entropy rate follows immediately from the ϵ-machine as the causal-state averaged transition uncertainty:

$$h_\mu = -\sum_{\sigma\in\boldsymbol{\mathcal{S}}} \Pr(\sigma) \sum_{x\in\mathcal{A}} \Pr(x|\sigma) \log_2 \Pr(x|\sigma) . \quad (2)$$

Here, the causal-state distribution $\Pr(\mathcal{S})$ is the stationary distribution $\langle\pi| = \langle\pi|T$ of the internal Markov chain governed by the row-stochastic matrix $T = \sum_{x\in\mathcal{A}} T^{(x)}$. The conditional probabilities $\Pr(x|\sigma)$ are the associated transition components in the labeled matrices $T^{(x)}_{\sigma,\sigma'}$. Note that the next state σ' is uniquely determined by knowing the current state σ and the measurement value x—a key property called *unifilarity*.

The amount of historical information the process stores also follows immediately: the *statistical complexity*, the Shannon–Boltzmann entropy of the causal-state distribution:

$$C_\mu = -\sum_{\sigma\in\boldsymbol{\mathcal{S}}} \Pr(\sigma) \log_2 \Pr(\sigma) . \quad (3)$$

In this way, the ϵ-machine allows one to directly determine two important properties of a system's intrinsic computation: its information generation and its storage. Since it depends only on block entropies, however, h_μ can be calculated via other presentations; though not as efficiently. For example, h_μ can be determined from Eq. (2) using any unifilar predictor, which necessarily is always larger than the ϵ-machine. Only recently was a (rather more complicated) closed-form expression discovered for the excess entropy **E** using a representation closely related to the ϵ-machine [22]. Details aside, no analogous closed-form expressions for the other complexity measures are known, including and especially those for finite-L blocks, such as $h_\mu(L)$.

Mixed-state presentation

To develop these, we shift to consider how an observer represents its knowledge of a hidden system's current state and then introduce a spectral analysis of that representation. For our uses here, the observer has a correct model in the sense that it reproduces $\mathcal{P}$ exactly. (Any model that does we call a *presentation* of the process. There may be many.) Using this, the observer tracks a process's evolution using a distribution over the hidden states called a *mixed state* $\eta \equiv \left(\Pr(\sigma^0), \Pr(\sigma^1), \Pr(\sigma^2), \ldots\right)$. The associated random variable is denoted $\mathcal{R}$. The question is how does an observer update its knowledge (η) of the internal states as it makes measurements—$x_0, x_1, \ldots$?

If a system is in mixed state η, then the probability of seeing measurement x is: $\Pr(X = x|\mathcal{R} = \eta) = \langle\eta|T^{(x)}|\mathbf{1}\rangle$, where $\langle\eta|$ is the mixed state as a row vector and $|\mathbf{1}\rangle$ is the column vector of all 1s. This extends to measurement sequences $w = x_0x_1\ldots x_{L-1}$, so that if, for example, the process is in statistical equilibrium, $\Pr(w) = \langle\pi|T^{(w)}|\mathbf{1}\rangle = \langle\pi|T^{(x_0)}T^{(x_1)}\cdots T^{(x_{L-1})}|\mathbf{1}\rangle$. The mixed-state evolution induced by measurement sequence w is: $\langle\eta_{t+L}| = \langle\eta_t|T^{(w)}/\langle\eta_t|T^{(w)}|\mathbf{1}\rangle$. The set $\boldsymbol{\mathcal{R}}$ of mixed states that we use here are those induced by all allowed words $w \in \mathcal{A}^*$ from initial mixed state $\eta_0 = \pi$. For each mixed state η_{t+1} induced by symbol $x \in \mathcal{A}$, the mixed-state-to-state transition probability is: $\Pr(\eta_{t+1}, x|\eta_t) = \Pr(x|\eta_t)$. And so, by construction, using mixed states gives a unifilar presentation. We denote the associated set of transition matrices $\{W^{(x)}\}$. They and the mixed states $\boldsymbol{\mathcal{R}}$ define a process's *mixed-state presentation* (MSP), which describes how an observer's knowledge of the hidden process updates via measurements. The row-stochastic matrix $W = \sum_{x\in\mathcal{A}} W^{(x)}$ governs the evolution of the probability distribution over allowed mixed states.

The use of mixed states is originally due to Blackwell [25], who expressed the entropy rate h_μ as an integral of a (then uncomputable) measure over the mixed-state space $\boldsymbol{\mathcal{R}}$. Although we focus here on the finite mixed-state case for simplicity, it is instructive to see in the general case the complicatedness revealed in a process using the mixed-state presentation: e.g., Figs. 17(a)–(c) of Ref. [26]. The Supplementary Materials give the detailed calculations for examples in the finite case.