Bergische Universität Wuppertal

Fakultät für Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational Mathematics (IMACM)

Preprint BUW-IMACM 26/15

F. Kasolis

Entropy-Guided Sampling in Tolerance Spaces

July 6, 2026

http://www.imacm.uni-wuppertal.de

Entropy-Guided Sampling in Tolerance Spaces

Fotios Kasolis

Department of Mathematics and Informatics, University of Wuppertal,
Gaußstraße 20, Wuppertal, 42119, Germany.

Contributing authors: kasolis@uni-wuppertal.de;

Building upon the notions of tolerance relations, entropy, and filtrations, we develop a unified framework for entropy-guided sampling in tolerance spaces, which provides a general setting for quantifying information and constructing entropy-guided sampling algorithms. A fundamental advantage of tolerance relations is that they make explicit the features with respect to which information is measured. Tolerance entropy quantifies the informational content induced by these features, entropy-guided sampling eliminates information redundancy, and filtrations provide a natural foundation for multiscale analysis. The presentation establishes connections with nonlinear dynamics, recurrence quantification analysis, and principal curves, while accommodating both generic and application-specific notions of information. The central thesis is that information is encoded by filtrations of tolerance relations, quantified through the density of their tolerance classes, and essentially preserved under entropy-guided sampling.

Keywords: tolerance spaces; tolerance entropy; entropy-guided sampling; filtrations; nonlinear dynamics

1 Introduction

The rapid growth of scientific computing and data-driven technologies has led to an unprecedented increase in the amount of sequential data generated by physical, biological, and engineered systems. Such data arise naturally from measured or simulated dynamical systems across virtually all scientific disciplines. Since storing and processing every observed state is often computationally prohibitive, considerable effort has been devoted to sparse representations, dimensionality reduction, compression, model reduction, and sampling methods that seek compact representations while preserving a notion of information relevant to a given application [1–8].

For sequences generated by dynamical systems, however, information is intrinsically linked to the relations between observed states rather than to the states

themselves. Consequently, the amount of information retained by a reduced representation depends fundamentally on the notion of indistinguishability used to compare observations. Classical entropy and entropy-rate notions quantify information through partitions or symbolic descriptions, providing fundamental measures of uncertainty and dynamical complexity [9–14]. Their practical computation, however, often relies on asymptotic constructions that are difficult to exploit for finite, large-scale, or online datasets.

Maximum entropy state sampling was originally introduced as an entropy-based framework for selecting representative states in reduced-order modeling and has subsequently been applied to nonlinear simulations, entropy-based basis generation, and surrogate modelling [1, 2, 15–21]. These developments demonstrated that entropy-guided sampling can substantially reduce computational complexity while preserving the essential information contained in dynamical orbits. At the same time, these observations highlight the need for a general notion of information capable of accommodating application-dependent notions of indistinguishability between observations.

The central premise of this work is that information cannot be defined independently of the criterion used to regard two observations as indistinguishable. Building upon this observation, we introduce a unified framework for entropy-guided sampling on tolerance spaces. The framework is founded on tolerance relations, which provide a flexible mathematical language for encoding multiscale and application-specific notions of indistinguishability, and on tolerance entropy, which quantifies the informational content induced by overlapping tolerance classes. In this way, information is measured directly with respect to the features of interest, naturally establishing connections between geometry, recurrence, and information. This viewpoint also relates naturally to nonlinear dynamics, recurrence quantification analysis, and manifold learning, where neighborhood relations determine the underlying structure of the data [3, 22–27]. The resulting formalism provides a systematic methodology for constructing both generic and application-specific tolerance relations, with particular emphasis on nonlinear dynamical systems.

The remainder of this paper is organized as follows. Section 2 develops the mathematical foundations of tolerance spaces, tolerance entropy, entropy-guided sampling, and filtrations of tolerance relations. Section 3 specializes the framework to metric tolerance relations and presents applications to nonlinear dynamical systems and geometry. Section 4 extends the framework to non-metric notions of proximity, illustrating its flexibility and applicability to problem-dependent notions of information. Finally, Section 5 illustartes the usability of the framework through numerical experiments.

2 Tolerance entropy sampling

Tolerance entropy sampling (TES) provides a unified framework for extracting representative subsequences from observed orbits by exploiting the relational geometry induced by tolerance relations. Unlike conventional sampling strategies based on temporal spacing, heuristic error indicators, or probabilistic methods, TES is founded on a sequential tolerance entropy monotonicity condition. Consequently, the framework is

independent of the particular notion of indistinguishability and unifies a broad family of sampling methodologies, with different choices of the underlying tolerance relation inducing different notions of indistinguishability, thereby leading to dynamically relevant sampling strategies.

2.1 Tolerance spaces

Let Ω be the state space of a dynamical system. A *tolerance relation* on Ω is a reflexive and symmetric binary relation $T \subseteq \Omega \times \Omega$, with (Ω, T) referred to as a *tolerance space*. The notation $x \sim_T y$ is used whenever $(x, y) \in T$, in which case we say that x is T-related to y. A tolerance relation induces tolerance classes of the form

$$[x]_T = \{y \in \Omega : x \sim_T y\}.$$

Unlike equivalence relations, tolerance relations are not transitive, and hence, the associated tolerance classes form a covering,

$$\Omega/T = \{[x]_T : x \in \Omega\},$$

rather than a partition of Ω. This distinction enables approximate indistinguishability and makes tolerance relations particularly suitable for numerical simulations and experimental data. Throughout this review, we consider families of ε-parameterized tolerance relations, where $\varepsilon > 0$ denotes a prescribed tolerance. For instance, *metric tolerance relations* of the form

$$M_\varepsilon = \{(x, y) \in \Omega \times \Omega : d(x, y) \leq \varepsilon\},$$

where $d : \Omega \times \Omega \to [0, \infty)$ is a metric on Ω encode metric proximity, while *energetic tolerance relations* of the form

$$E_\varepsilon = \{(x, y) \in \Omega \times \Omega : |K(x, f(x)) - K(y, f(y))| \leq \varepsilon\},$$

where $f : \Omega \to \Omega$ is a time-shift map and $K : \Omega \times \Omega \to \mathbb{R}$ an energy functional, encode information regarding energetically related states. Finally, the *linear tolerance relation* defined by

$$L_\varepsilon = \{(x, y) \in \Omega \times \Omega : a(x, y) \leq \varepsilon\},$$

where $a : \Omega \times \Omega \to [0, 1]$ denotes the alignment defect, measures the deviation from exact parallelism.

The preceding examples illustrate a common construction principle. Starting from an equivalence relation encoding exact indistinguishability, introduce a tolerance parameter $\varepsilon > 0$ and discard the transitivity property to obtain a family $\{T_\varepsilon : \varepsilon > 0\}$ of ε-parameterized tolerance relations. The resulting relations retain a notion of approximate indistinguishability, while encoding uncertainty, and finite observational resolution. Moreover, the resulting family naturally forms a filtration,

$$0 < \varepsilon_1 \leq \varepsilon_2 \Rightarrow T_{\varepsilon_1} \subseteq T_{\varepsilon_2},$$

thereby providing a multiscale representation of the underlying dynamics. This general construction underlies the TES framework, while the particular choice of tolerance relation determines the notion of information used during the sampling strategy.

2.2 Tolerance entropy

Similarly to partition entropy, the information content of a state is quantified through the size of its tolerance class. Given a tolerance space (Ω, T) and a finite orbit X in Ω, the *local tolerance density* of $x \in X$ is defined by

$$\rho_T(x) = \frac{|[x]_T|}{|X|} \in [1/|X|, 1],$$

where T is restricted to X and $|\cdot|$ denotes set cardinality. The density ρ_T measures the proportion of observed states that are indistinguishable from x with respect to the tolerance relation. Averaging the local densities over the entire orbit gives the *average tolerance density*

$$\overline{\rho}_T(X) = \frac{1}{|X|} \sum_{x \in X} \rho_T(x) = \frac{1}{|X|^2} \sum_{x \in X} |[x]_T|.$$

Definition 2.1 (Tolerance entropy)**.** *Let (Ω, T) be a tolerance space and X a finite orbit in Ω. The* tolerance entropy *of X with respect to the tolerance relation T is defined by*

$$h_T(X) = -\log \overline{\rho}_T(X).$$

Tolerance entropy quantifies the average geometric scale of the tolerance classes induced by the associated relation. Large values indicate that, on average, states possess relatively small tolerance classes and therefore carry greater relational information, whereas small values correspond to states that are relationally redundant.

Proposition 2.1 (Control of averaged Rényi entropies)**.** *Let (Ω, T) be a tolerance space, X a finite orbit in Ω, and $\alpha > 1$. Then*

$$\frac{1}{\alpha - 1} h_T(X) \leq h^{(\alpha)}_{\rho_T}(X) \leq \frac{\alpha}{\alpha - 1} h_T(X),$$

where

$$h^{(\alpha)}_{\rho_T}(X) = \frac{1}{1 - \alpha} \log \overline{\rho^{\alpha}_T}(X)$$

denotes the averaged Rényi-type relational entropy of order α.

Proof Since $0 < \rho_T \leq 1$ and $\alpha > 1$, we have $\rho^{\alpha}_T \leq \rho_T$. Averaging over X yields

$$\overline{\rho^{\alpha}_T}(X) \leq \overline{\rho_T}(X).$$

Moreover, Jensen's inequality implies

$$\overline{\rho^{\alpha}_T}(X) \geq \overline{\rho_T}(X)^{\alpha}.$$

Taking logarithms and using the fact that $1-\alpha<0$ reverses the inequalities and completes the proof. □

The averaged second-order Rényi-type entropy is of particular importance, since it identifies precisely the variance contribution omitted by the tolerance entropy; more precisely,

$$h_{\rho_T}^{(2)}(X) = 2h_T(X) - \log\left(1 + \frac{\mathrm{Var}(\rho_T)}{\overline{\rho}_T^2}\right),$$

and hence, for fixed tolerance entropy, the maximum value $2h_T(X)$ is obtained whenever the variance of the local tolerance density vanishes.

2.3 Evolution of the tolerance entropy

Let (Ω, T) be a tolerance space and $X_n = (x_1, \ldots, x_n)$ the first n states of an observed orbit. Denote by

$$A_n = (\chi_T(x_i, x_j))_{i,j=1}^n \in \{0,1\}^{n\times n},$$

the associated relation matrix, where χ_T is the characteristic function with values $\chi_T(x_i, x_j) = 1$ whenever $x_i \sim_T x_j$ and zero otherwise. The matrix A_n generalizes several binary matrices that arise in data analysis. In particular, for the metric tolerance relation, A_n coincides with the state recurrence matrix employed in recurrence quantification analysis, while such affinity matrices also appear in machine learning, including graph-based clustering, spectral methods, manifold learning, and nearest-neighbor methods. The TES formalism unifies these constructions by allowing arbitrary tolerance relations and by exploiting the resulting relation matrices to derive sequential entropy-based sampling criteria.

Using the notation $\rho_n = \overline{\rho}(X_n)$ and $h_n = h_T(X_n)$, the instantaneous average tolerance density and the associated tolerance entropy can be written in the form

$$\rho_n = \frac{\|A_n\|_{\mathrm{F}}^2}{n^2}, \qquad h_n = -\log \rho_n,$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. When a new state x_{n+1} is observed, the relation matrix is updated according to

$$A_{n+1} = \begin{pmatrix} A_n & a_n \\ a_n^\top & 1 \end{pmatrix},$$

where a_n records the tolerance relations between x_{n+1} and the previously observed states.

Proposition 2.2 (Tolerance density evolution)**.** *Let $N_n = 1^\top a_n$ be the number of previously observed states that are T-related to x_{n+1}. Then, the average tolerance density satisfies the one-step recursion*

$$(n+1)^2\rho_{n+1} = n^2\rho_n + 1 + 2N_n.$$

Consequently, for every $n \geq 1$,

$$\rho_n = \frac{1}{n} + \frac{2}{n^2} \sum_{k=1}^{n-1} N_k.$$

Proof Since A_n is binary, we have

$$\|A_{n+1}\|_{\mathrm{F}}^2 = \|A_n\|_{\mathrm{F}}^2 + 1 + 2N_n.$$

The unit contribution corresponds to the newly introduced reflexive pair, $x_{n+1} \sim_T x_{n+1}$, while the term $2N_n$ accounts for the newly created symmetric T-related pairs involving x_{n+1}. Substituting $\|A_n\|_{\mathrm{F}}^2 = n^2\rho_n$ gives

$$(n+1)^2\rho_{n+1} = n^2\rho_n + 1 + 2N_n.$$

Summing this identity from 1 to $n-1$ yields the telescoping formula

$$n^2\rho_n = 1 + \sum_{k=1}^{n-1}(1 + 2N_k) = n + 2\sum_{k=1}^{n-1} N_k,$$

from which the claim follows directly. □

As a direct consequence of Proposition 2.2, the tolerance entropy converges under mild assumptions.

Proposition 2.3 (Tolerance entropy convergence)**.** *If the quadratic Cesàro condition*

$$\frac{1}{n^2}\sum_{k=1}^{n-1} N_k \to N_\infty \qquad \text{as } n \to \infty, \qquad N_\infty \in [0, 1/2],$$

holds, then $\rho_n \to 2N_\infty$. *If, in addition,* $N_\infty > 0$, *continuity of the logarithm yields*

$$h_n = -\log \rho_n \to -\log(2N_\infty).$$

Thus, under the Cesàro assumption, the asymptotic behavior of the tolerance entropy is completely determined by the average tolerance density. For the derivation of the local entropy-guided sampling rule in the next section, it is convenient to rewrite the update in the equivalent local form

$$\rho_{n+1} - \rho_n = \frac{\delta_n - (2n+1)\rho_n}{(n+1)^2}, \qquad \rho_1 = 1,$$

which shows that the evolution of the average tolerance density is determined by the scalar increment $\delta_n = 1 + 2N_n$ together with the current value ρ_n. This recursion provides an efficient online update of the tolerance entropy without recomputing the relation matrix from scratch.

2.4 Tolerance entropy sampling

The TES problem can be formulated as a sequential entropy-monotone subsequence extraction problem. Given an observed orbit $X_n = (x_1, \dots, x_n)$ in a tolerance space (Ω, T), the objective is to extract a subsequence

$$Y_m = (y_1, \dots, y_m) \preceq X_n$$

with minimal cardinality m, while enforcing a positive tolerance entropy production rate. Hence, the TES optimization problem is

$$\min\{m : Y_m \preceq X_n\} \quad \text{subject to} \quad h_{k+1}(Y_m) > h_k(Y_m) \quad \forall k \in \{1, 2, \dots, m-1\}$$

or, equivalently, since $h_k(Y_m) = -\log \rho_k(Y_m)$,

$$\min\{m : Y_m \preceq X_n\} \quad \text{subject to} \quad \rho_{k+1}(Y_m) < \rho_k(Y_m) \quad \forall k \in \{1, 2, \dots, m-1\}$$

Using the scalar density evolution, the local admissibility condition becomes

$$\rho_{k+1} < \rho_k \Leftrightarrow \delta_k < (2k+1)\rho_k.$$

Thus, a candidate state is retained precisely when its observation decreases the average tolerance density, or equivalently, when it increases the tolerance entropy. This yields the local *TES novelty criterion*

$$\delta_k < (2k+1)\rho_k.$$

Theorem 2.1 (Fundamental properties of TES)**.** *Let $Y_m = (y_1, \dots, y_m) \preceq X_n$ be the subsequence produced by the TES novelty criterion. Then, Y_m is tolerance-separated. In particular, for all $i \in \{1, \dots, m\}$,*

$$[y_i]_T \cap Y_m = \{y_i\},$$

and the relation matrix restricted to Y_m satisfies

$$A_m = I_m.$$

Consequently, $\overline{\rho}_T(Y_m) = 1/m$, $h_T(Y_m) = \log m$. Moreover, every discarded state is T-related to at least one previously retained state.

Proof Let $Y_k = (y_1, \dots, y_k)$ denote the retained subsequence after k accepted states, and let x be a candidate state. If x is retained, then the TES novelty criterion gives

$$\delta_k < (2k+1)\rho_k.$$

Since, by induction, the retained states are tolerance-separated, we have $\rho_k = 1/k$. Therefore,

$$1 + 2N_k < \frac{2k+1}{k} = 2 + \frac{1}{k}.$$

Since N_k is a nonnegative integer, this implies

$$N_k = 0.$$

Hence the newly retained state is not T-related to any previously retained state. By induction, all retained states are pairwise non-T-related, and therefore

$$[y_i]_T \cap Y_m = \{y_i\}, \qquad A_m = I_m.$$

It follows that

$$\overline{\rho}_T(Y_m) = \frac{\|A_m\|_{\mathrm{F}}^2}{m^2} = \frac{m}{m^2} = \frac{1}{m},$$

and hence

$$h_T(Y_m) = -\log \overline{\rho}_T(Y_m) = \log m.$$

It remains to consider a discarded state. A candidate is discarded precisely when the TES novelty criterion fails, so that

$$\delta_k \geq (2k+1)\rho_k.$$

Again using $\rho_k = 1/k$, this gives

$$1 + 2N_k \geq 2 + \frac{1}{k}.$$

Since N_k is an integer, we obtain $N_k \geq 1$, and hence, every discarded state is T-related to at least one previously retained state. □

Theorem 2.1 shows that TES simultaneously produces a tolerance covering and a maximally tolerance-separated subsequence. In particular,

$$X_n = [y_1]_T \cup \cdots \cup [y_m]_T,$$

so that every observed state is represented by the tolerance class of at least one retained state. Consequently, TES compresses the observed orbit without leaving any state uncovered. Moreover, the restriction of T to the sampled subsequence satisfies $A_m = I_m$, and therefore $\rho_m = 1/m$, $h_m = \log m$. Thus, among all tolerance-separated subsequences of cardinality m, the TES representation attains the maximum possible tolerance entropy.

Remark 2.1. *For ordered data, such as dynamical orbits or general sequences, the locality of the TES novelty criterion does not imply a greedy sampling procedure, since the ordering constraint introduces global dependencies between successive selections.*

The TES novelty condition immediately gives rise to two realizations of TES. In the *offline* setting, the entire observed orbit is assumed to be available, allowing the complete relation matrix to be constructed before sampling. The novelty condition is then evaluated by traversing the columns of the relation matrix and progressively selecting representative states. This implementation is summarized in Algorithm 1. In contrast, the *online* setting assumes that states become available sequentially. Rather than storing the entire relation matrix, only the T-related pairs between the newly observed state and the previously retained states are evaluated. Since the novelty condition depends solely on the current average tolerance density and the newly introduced T-related pairs, TES can be implemented recursively without requiring access to future

Algorithm 1 Offline Tolerance Entropy Sampling (Offline TES)

Require: Tolerance relation matrix $A \in \{0,1\}^{n\times n}$
Ensure: Logical index vector $I \in \{0,1\}^n$

$I \leftarrow (1,0,\ldots,0)$
$\rho \leftarrow 1$
$k \leftarrow 1$
for $c = 1,\ldots,n-1$ **do**
    $\delta \leftarrow 1 + 2\cdot 1^\top A(I, c+1)$
    if $\delta - (2k+1)\rho < 0$ **then**
        $I_{c+1} \leftarrow 1$
        $\rho \leftarrow (k^2\rho + \delta)/(k+1)^2$
        $k \leftarrow k+1$
    end if
end for
return I

Algorithm 2 Online Tolerance Entropy Sampling (Online TES)

Require: Stream of states $(x_1, x_2, \ldots)$, tolerance relation T
Ensure: Sampled subsequence Y

$Y \leftarrow (x_1)$
$\rho \leftarrow 1$
$k \leftarrow 1$
for each new state x **do**
    $\delta \leftarrow 1 + 2\sum_{y\in Y} \chi_T(y,x)$
    if $\delta - (2k+1)\rho < 0$ **then**
        $Y \leftarrow (Y, x)$
        $\rho \leftarrow (k^2\rho + \delta)/(k+1)^2$
        $k \leftarrow k+1$
    end if
end for
return Y

observations. Consequently, the online Algorithm 2 has constant memory requirements aside from the extracted subsequence and its associated tolerance information.

The preceding developments reveal four fundamental features of TES. First, the sampling procedure is sequential and therefore admits both offline and online implementations. Second, the retained states preserve a tolerance covering of the original orbit. Third, the framework is independent of the particular choice of tolerance relation, allowing metric, energetic, linear, and other notions of indistinguishability to be treated within a common framework. Finally, under mild assumptions, the tolerance entropy converges asymptotically according to the average tolerance frequency induced by the chosen tolerance relation.

2.5 General fidelity horizon

The scalar density evolution also provides a natural link between the convergence of the average tolerance density and the asymptotic behaviour of the new T-related pairs. Indeed, rearranging

$$\rho_{n+1} - \rho_n = \frac{\delta_n - (2n+1)\rho_n}{(n+1)^2},$$

gives

$$\frac{\delta_n}{2n+1} = \rho_n + \frac{(n+1)^2}{2n+1}(\rho_{n+1} - \rho_n).$$

Therefore, convergence of the average tolerance density alone does not necessarily imply convergence of the average $\delta_n/(2n+1)$, since the latter also depends on the rate at which consecutive average densities approach one another. Nevertheless, since $(n+1)^2/(2n+1) \sim n/2$, if

$$\rho_n \to \rho_\infty \qquad \text{and} \qquad n(\rho_{n+1} - \rho_n) \to 0 \qquad \text{as } n \to \infty,$$

it follows that

$$\frac{\delta_n}{2n+1} \to \rho_\infty.$$

Consequently, the fidelity horizon needs to be stated under certain additional asymptotic regularity assumptions.

Theorem 2.2 (Fidelity horizon)**.** *Assume that there exist $n_0 \geq 1$, $\rho_\infty \in (0,1)$, and $\beta \geq 0$ such that, for all $n \geq n_0$,*

$$\left| \frac{\delta_n}{2n+1} - \rho_\infty \right| \leq \frac{\beta}{2n+1}.$$

Then every future state x_{n_0+t} is T-related to at least one state of the initial observed block X_{n_0} provided

$$t < \frac{\rho_\infty}{1-\rho_\infty} n_0 + \frac{1}{2} - \frac{\beta}{2(1-\rho_\infty)}.$$

Proof A future state x_{n_0+t} is T-related to at least one state of the initial block X_{n_0} if and only if

$$\sum_{i=1}^{n_0} \chi_T(x_i, x_{n_0+t}) \geq 1.$$

At time $n_0 + t - 1$, the quantity δ_{n_0+t-1} counts the new reflexive pair and all symmetric T-related pairs between x_{n_0+t} and the previously observed states. Since the $t-1$ states between x_{n_0+1} and x_{n_0+t-1} can contribute at most $2(t-1)$ non-reflexive symmetric entries, a sufficient condition forcing at least one relation with the initial block X_{n_0} is

$$\delta_{n_0+t-1} > 2t - 1.$$

By the assumed regularity estimate,

$$\delta_{n_0+t-1} \geq (2(n_0+t-1)+1)\rho_\infty - \beta = (2n_0+2t-1)\rho_\infty - \beta.$$

Therefore it is sufficient that

$$(2n_0+2t-1)\rho_\infty - \beta > 2t-1.$$

Solving this inequality for t and using $\rho_\infty \in (0,1)$ yields the claimed horizon. □

Every future state within the horizon remains T-related to the sampled subsequence, and hence, the extracted subsequence continues to provide a faithful representation of the underlying dynamics.

Remark 2.2. *In the context of TES, this fidelity horizon inequality quantifies the predictive validity of the TES sample.*

In practice, the limiting density ρ_∞ is not known a priori. Thus, the fidelity inequality must be used through a data-driven stabilization criterion. Let $\gamma > 0$ be a prescribed tolerance. Suppose that, after an initial observation length n_0, the average tolerance density has stabilized in the sense that

$$|\rho_n - \widehat{\rho}_{n_0}| \leq \gamma, \qquad n \geq n_0,$$

where $\widehat{\rho}_{n_0}$ is an empirical stabilized value, typically estimated as $\widehat{\rho}_{n_0} = \rho_{n_0}$ or as a moving-window average of the tail densities. Assume moreover that the normalized number of new T-related pairs satisfies the empirical regularity bound

$$\left| \frac{\delta_n}{2n+1} - \widehat{\rho}_{n_0} \right| \leq \gamma + \frac{\beta}{2n+1}, \qquad n \geq n_0.$$

If

$$\varrho_{n_0} = \widehat{\rho}_{n_0} - \gamma > 0,$$

then every future state x_{n_0+t} is T-related to at least one state of the observed block X_{n_0}, provided

$$t < \frac{\varrho_{n_0}}{1-\varrho_{n_0}} n_0 + \frac{1}{2} - \frac{\beta}{2(1-\varrho_{n_0})}.$$

Thus, the unknown asymptotic density ρ_∞ may be replaced by a conservative empirical lower estimate ϱ_{n_0} obtained from the observed density evolution.

2.6 Filtrations and multiscale complexity

The TES naturally extends from a single tolerance relation to filtrations of tolerance relations indexed by a tolerance parameter $\varepsilon \in (0,\infty)$. Examples include metric filtrations indexed by the tolerance radius, energetic filtrations indexed by admissible energy variations, and linear filtrations indexed by the alignment defect. Let

$$\{T_\varepsilon : \varepsilon \in (0,\infty)\}$$

be a filtration of tolerance relations. Applying TES to each relation in the filtration induces a corresponding filtration of informative representations

$$\{Y_\varepsilon : \varepsilon \in (0, \infty)\},$$

where Y_ε denotes the TES extracted subsequence associated with T_ε. Consequently, TES should be viewed not merely as a sampling algorithm, but as a multiscale representation of the observed orbit. Each sample Y_ε satisfies the fundamental TES properties established previously. In particular, the extracted subsequence forms a tolerance covering and is maximally tolerance-separated. The cardinalities of the extracted subsequences therefore define the *TES cardinality function*

$$m_X : (0, \infty) \to [1, |X|], \qquad m_X(\varepsilon) = |Y_\varepsilon|,$$

meaning that larger tolerances identify more states as being mutually related, and hence, within the same tolerance class. Consequently,

$$\varepsilon_1 \leq \varepsilon_2 \Rightarrow m_X(\varepsilon_1) \geq m_X(\varepsilon_2),$$

so that m_X is monotone decreasing along the tolerance filtration. Equivalently, the tolerance entropy of the extracted subsequence satisfies

$$h_{T_\varepsilon}(Y_\varepsilon) = \log m_X(\varepsilon),$$

and is therefore likewise monotone decreasing with respect to the filtration parameter. The TES cardinality function captures the multiscale complexity of the tolerance covering. Small tolerances produce finely resolved subsequences containing many representative states, whereas large tolerances produce progressively coarser descriptions of the observed dynamics. The discontinuities of m_X correspond to tolerance values at which the minimal representative set decreases because previously distinct representative states become mutually tolerance related. Large discontinuities indicate tolerance scales at which many representatives become simultaneously redundant, thereby revealing characteristic transitions in the multiscale organization of the observed orbit.

The TES cardinality function m_X measures the number of representative states required to cover the observed orbit at tolerance level ε. Consequently, it provides a natural notion of multiscale complexity associated with the tolerance filtration. Many multiscale systems exhibit asymptotic power-law scaling of covering statistics. Motivated by this observation, assume that m_X is regularly varying at the origin with index $-\Phi$, that is,

$$\lim_{\varepsilon \downarrow 0} \frac{m_X(\lambda\varepsilon)}{m_X(\varepsilon)} = \lambda^{-\Phi} \qquad \forall \lambda > 0.$$

By Karamata's representation theorem,

$$m_X(\varepsilon) = \varepsilon^{-\Phi} L(\varepsilon),$$

where L is slowly varying. In the common situation where $L(\varepsilon)$ converges to a positive constant C, this reduces to the familiar asymptotic power law

$$m_X(\varepsilon) \sim C\varepsilon^{-\Phi}.$$

This scaling hypothesis underlies many practical algorithms for estimating box-counting, covering, and correlation dimensions, where power-law growth of neighborhood or covering statistics reveals the effective complexity of the underlying system. In the present setting, however, the exponent is determined by the filtration of tolerance relations and therefore remains meaningful beyond metric or geometric constructions. Under the above asymptotic power-law hypothesis, the exponent Φ becomes an intrinsic descriptor of the multiscale complexity induced by the tolerance filtration. Since

$$\log m_X(\varepsilon) = -\Phi \log \varepsilon + \log C + o(1),$$

it is natural to define the corresponding logarithmic scaling exponent.

Definition 2.2 (TES index)**.** *Let $\{T_\varepsilon : \varepsilon > 0\}$ be a filtration of tolerance relations with associated TES cardinality function m_X. Whenever the limit exists, the tolerance index is defined by*

$$\Phi = -\lim_{\varepsilon \downarrow 0} \frac{d \log m_X(\varepsilon)}{d \log \varepsilon}.$$

Equivalently, if $m_X(\varepsilon) \sim C\varepsilon^{-\Phi}$ for some constant $C > 0$, then Φ is precisely the corresponding power-law exponent.

In practice, the TES index may be estimated as the least-squares slope of the log-log plot of $m_X(\varepsilon)$ versus ε over an appropriate scaling regime. Once the empirical scaling law of $m_X(\varepsilon)$ has been identified, it becomes possible to quantify not only the overall multiscale complexity through the tolerance index Φ, but also the characteristic tolerance scales at which representative states disappear from the TES representation. Since $m_X(\varepsilon)$ is monotone decreasing, it naturally describes the number of representatives surviving as the tolerance increases. After normalization,

$$M_X(\varepsilon) = \frac{m_X(\varepsilon) - m_X(\varepsilon_{\max})}{m_X(\varepsilon_{\min}) - m_X(\varepsilon_{\max})},$$

may therefore be interpreted as a survival function over the filtration parameter. If $m_X(\varepsilon_{\min}) = |X|$ and $m_X(\varepsilon_{\max}) = 1$, then under the power-law approximation,

$$M_X(\varepsilon) = \frac{(\varepsilon_{\min}/\varepsilon)^\Phi - (\varepsilon_{\min}/\varepsilon_{\max})^\Phi}{1 - (\varepsilon_{\min}/\varepsilon_{\max})^\Phi}.$$

Consequently, $F_X(\varepsilon) = 1 - M_X(\varepsilon)$ represents the cumulative fraction of representative states removed by TES up to tolerance ε. Differentiating yields the corresponding

probability density on the tolerance scale,

$$p_X(\varepsilon) = -\frac{d}{d\varepsilon} M_X(\varepsilon) = \frac{\Phi \varepsilon_{\min}^{\Phi}}{1 - (\varepsilon_{\min}/\varepsilon_{\max})^{\Phi}} \varepsilon^{-\Phi-1}.$$

The associated expected reduction scale is then

$$\mathbb{E}(\varepsilon) = \int_{\varepsilon_{\min}}^{\varepsilon_{\max}} \varepsilon \, p_X(\varepsilon) \, d\varepsilon.$$

For $\Phi \neq 1$, this evaluates to

$$\mathbb{E}(\varepsilon) = \frac{\Phi \varepsilon_{\min}^{\Phi}}{1 - (\varepsilon_{\min}/\varepsilon_{\max})^{\Phi}} \frac{\varepsilon_{\min}^{1-\Phi} - \varepsilon_{\max}^{1-\Phi}}{\Phi - 1},$$

whereas for $\Phi = 1$,

$$\mathbb{E}(\varepsilon) = \frac{\varepsilon_{\min}}{1 - \varepsilon_{\min}/\varepsilon_{\max}} \log\left(\frac{\varepsilon_{\max}}{\varepsilon_{\min}}\right).$$

Remark 2.3. *Unlike classical estimators based on pairwise metric distances, the TES index is derived from the minimum number of representative states required to cover the observed orbit over a prescribed interval of tolerance scales. Consequently, the estimator simultaneously quantifies the relational covering complexity and the intrinsic compressibility of the orbit relative to the underlying filtration of tolerance relations.*

3 Metric realization and connections

We now consider the important special case in which indistinguishability is induced by a metric. Let (Ω, d) be a metric space and recall the metric tolerance relation

$$M_\varepsilon = \{(x, y) \in \Omega \times \Omega : d(x, y) \leq \varepsilon\}.$$

The resulting metric tolerance entropy sampling (MTES) framework is a specialization of TES and therefore inherits all structural properties established for TES, including tolerance coverings, maximal tolerance separation, and entropy monotonicity. Its additional importance is that the metric setting connects the abstract tolerance formalism with several classical constructions in nonlinear dynamics, inference, and data analysis.

3.1 Connections to existing methods

For the MTES instance, the average tolerance density coincides with the empirical Grassberger-Procaccia correlation integral [**?**]. Hence, the associated tolerance entropy is the logarithmic information associated with the correlation sum, while under the power law assumption, the correlation dimension is recovered as the small-scale logarithmic scaling of the metric tolerance entropy. This connection also clarifies

the relation with entropy estimation in dynamical systems. The Grassberger-Procaccia method uses recurrence counts to estimate geometric and entropic invariants, including quantities related to Kolmogorov-Sinai entropy. MTES uses the same recurrence information differently. Instead of estimating a dynamical invariant, it first constructs a sparse orbit representation. Thus, average tolerance density measures statistical recurrence, whereas the MTES cardinality function measures relational covering complexity. The same viewpoint connects MTES with dimension estimation. The TES index Φ is estimated by the negative slope of the log-log plot of m_X against ε. Unlike classical dimension estimators, however, Φ should not be interpreted as a dimension in general. It is a sampling complexity index associated with the underlying tolerance filtration.

From the viewpoint of machine learning, MTES is related to neighbourhood-based learning, clustering, manifold learning, kernel methods, and inference. In all these settings, the basic object is not exact equality but similarity at a prescribed scale. MTES adds a sequential and information-theoretic principle; it selects a new representative only when the new state is not sufficiently explained by the previously selected tolerance covering. Thus, MTES may be regarded as an online geometric inference procedure that constructs a sparse geometric representation of the observed orbit.

3.2 Interpretation in terms of stability

The behavior of MTES is closely tied to the stability properties of the underlying dynamical system. Roughly, stable recurrent dynamics results in large average tolerance density, low tolerance entropy, and slow growth of the representative sample. Conversely, unstable dynamics continually generates geometrically novel states, decreasing the average tolerance density and increasing both the tolerance entropy and the sampling complexity.

Exponential stability has particularly strong implications for MTES. For simplicity, consider a finite orbit $X = (x_1, \ldots, x_n)$ in a metric space (Ω, d), and assume that x_* is exponentially stable, that is, there exist constants $C \geq 1$, $\lambda > 0$ such that

$$d(x_n, x_*) \leq Ce^{-\lambda n} d(x_1, x_*).$$

If an orbit converges exponentially to an equilibrium, then after the finite entrance time

$$\tau_\varepsilon \leq \frac{1}{\lambda} \log \frac{2Cd(x_1, x_*)}{\varepsilon},$$

the tail of the trajectory is contained in a single metric tolerance class. Consequently, $\rho_n \to 1$, $h_n \to 0$, and the MTES subsequence becomes finite. In particular,

$$m_X(\varepsilon) \leq 1 + \tau_\varepsilon,$$

meaning that for exponentially stable systems, MTES extracts only the transient dynamics, with logarithmic sampling complexity in the tolerance scale.

Lyapunov stability guarantees that nearby trajectories remain uniformly close for all future times, thereby preserving sufficiently large metric tolerance classes. If the orbit is contained in a compact Lyapunov-stable invariant set, then for every tolerance

$\varepsilon > 0$ the sampling complexity is uniformly bounded by the covering number of the invariant set,

$$m_X(\varepsilon) \leq N(K, \varepsilon).$$

If, in addition, the orbit is recurrent, then the average tolerance density admits a positive asymptotic limit $\rho_n \to \rho_\infty$, and consequently the tolerance entropy converges, $h_n \to -\log \rho_\infty$. Thus, Lyapunov stability controls the long-term sampling complexity, while recurrence determines the asymptotic tolerance entropy. More precisely, for recurrent Lyapunov-stable dynamics with effective dimension α, the covering-number bound $m_X(\varepsilon) \lesssim C\varepsilon^{-\alpha}$ is obtained. Thus recurrent Lyapunov stability yields finite sampling complexity at each fixed tolerance scale, while the dependence on the tolerance is governed by the metric size of the recurrent set.

For systems exhibiting sensitive dependence on initial conditions, nearby trajectories separate exponentially at a rate determined by the dominant Lyapunov exponent $\lambda_{\max} > 0$. Consequently, two initially indistinguishable states become distinguishable at tolerance ε after approximately the Lyapunov time

$$\tau_\varepsilon = \frac{1}{\lambda_{\max}} \log \frac{\varepsilon}{\delta},$$

where δ denotes their initial separation. Thus each MTES representative can cover at most τ_ε consecutive iterates, yielding the transient lower bound

$$m_X(n, \varepsilon) \gtrsim \frac{n}{\tau_\varepsilon} = \frac{\lambda_{\max} n}{\log(\varepsilon/\delta)}.$$

Hence the dominant Lyapunov exponent directly controls the transient sampling complexity, in the sense that larger exponents reduce the tolerance lifetime, generate geometrically novel states more rapidly, and require more representatives to preserve the dynamics at the prescribed resolution.

3.3 Interpretation in terms of geometry

Let $Y = (y_1, \ldots, y_m)$ denote the MTES subsequence at prescribed tolerance ε. Since the selected states inherit the temporal order of the original orbit, they define the polygonal MTES curve

$$\Gamma_\varepsilon = \bigcup_{i=1}^{m-1} [y_i, y_{i+1}].$$

Unlike many manifold learning and clustering methods, MTES preserves the chronological order of the observed orbit, thereby retaining dynamical information together with its geometric structure. This polygonal curve provides a finite-scale geometric skeleton of the observed orbit. Large values of ε produce coarse curves that suppress small-scale fluctuations, whereas smaller values of ε preserve finer features of the observed orbit. Thus, the tolerance acts as a geometric resolution parameter.

A connection between the MTES curve and principal curves is therefore understood at a fixed tolerance scale, with the limit $\varepsilon \downarrow 0$ recovering the original trajectory

and removing the smoothing effect. At a fixed tolerance, however, the MTES polygonal curve provides a principal-curve type approximation, that is, it passes through representative states, follows the dominant geometric trend of the orbit, and filters fluctuations below the tolerance scale. In this sense, MTES gives a constructive, sequential, and tolerance-controlled analogue of a principal curve.

Given the underlying vector space structure, a class averaged variant, AMTES, strengthens this connection. After the MTES representatives have been selected, define the class of a representative y_i by

$$[y_i]_{T_\varepsilon} = \{x_j \in X_n : \ d(x_j, y_i) \leq \varepsilon\}.$$

The averaged representative state is then

$$\bar{y}_i = \frac{1}{|[y_i]_{T_\varepsilon}|} \sum_{x_j \in [y_i]_{T_\varepsilon}} x_j,$$

while the AMTES curve is defined by

$$\bar{\Gamma}_\varepsilon = \bigcup_{i=1}^{m-1} [\bar{y}_i, \bar{y}_{i+1}].$$

Unlike the ordinary MTES curve, $\bar{\Gamma}_\varepsilon$ is not constrained to pass through the states of the MTES extracted subsequence. It is a centroidal geometric surrogate of the orbit and therefore behaves more like a principal curve. Indeed, classical principal curves satisfy the self-consistency condition

$$\gamma(t) = \mathbb{E}(X \mid \Pi(X) = \gamma(t)),$$

where Π denotes projection onto the curve. The AMTES representative $\bar{y}_i$ is an empirical conditional mean over the tolerance neighbourhood $[y_i]_{T_\varepsilon}$. Thus AMTES provides a finite-sample, tolerance-local analogue of this self-consistency condition.

The distinction between MTES and AMTES is essential for practical applications. MTES produces a sparse subsequence and is best suited for compression, sampling, and dynamics-preserving reduction. Averaged MTES produces a smoothed geometric surrogate and is better suited for denoising, principal-curve-type reconstruction, and statistical representation of an orbit. Both are finite-scale constructions, with the tolerance ε determining the degree of geometric smoothing.

In practice, the averaged variant should be implemented with care, see Algorithm 3. The MTES selection step should be performed using selected anchor states, whereas the centroids should be updated only as geometric averages. If the centroids themselves are used for future tolerance entropy tests, the algorithm no longer coincides with MTES, because the tolerance classes move during the sampling process. Thus, one should distinguish between anchor states z_i, which determine the MTES tolerance cover, and averaged representatives $\bar{y}_i$, which define the smoothed geometric curve.

Algorithm 3 Online Averaged Metric Tolerance Entropy Sampling (Online AMTES)

Require: Stream of states $(x_1, x_2, \ldots)$, tolerance ε
Ensure: Anchor and averaged states $Z = (z_1, \ldots, z_k)$, $\bar{Y} = (\bar{y}_1, \ldots, \bar{y}_k)$

$Z \leftarrow (z_1) = (x_1)$
$\bar{Y} \leftarrow (\bar{y}_1) = (x_1)$
$w_1 \leftarrow 1$
$\rho \leftarrow 1$
$k \leftarrow 1$
for each new state x **do**
    $J \leftarrow \{\, i : \ d(x, z_i) \leq \varepsilon \,\}$
    $\delta \leftarrow 1 + 2|J|$
    if $\delta - (2k+1)\rho < 0$ **then**
        $k \leftarrow k + 1$
        $z_k \leftarrow x$
        $\bar{y}_k \leftarrow x$
        $w_k \leftarrow 1$
        $\rho \leftarrow [(k-1)^2\rho + \delta]/k^2$
    else
        for all $i \in J$ **do**
            $\bar{y}_i \leftarrow (w_i \bar{y}_i + x)/(w_i + 1)$
            $w_i \leftarrow w_i + 1$
        end for
    end if
end for
return $Z, \bar{Y}$

4 Beyond metric tolerance relations

Although metric-induced tolerance relations provide the most direct connection with classical notions from dynamical systems and geometry, they are by no means the only meaningful choice. One of the principal advantages of the tolerance-space framework is that the notion of indistinguishability is not fixed a priori, but can instead be adapted to the features of interest in a given application. Consequently, tolerance relations may be constructed from geometric, energetic, probabilistic, dynamical, or other application-specific quantities, leading to different notions of information and entropy-guided sampling. For systems possessing an intrinsic energetic structure, such as port-Hamiltonian systems, it is often more natural to compare states through their energetic behaviour than through their geometric proximity, while in other applications the relevant information may be encoded in the local dynamics or in the frequency spectrum rather than the state itself. The tolerance space formalism accommodates all such choices within a common mathematical framework. As an illustration, we first introduce transition energy tolerance relations, which quantify information through the energy associated with one-step transitions rather than through metric state proximity, before discussing more general strategies for constructing application-specific tolerance relations.

4.1 Transition energy tolerance spaces

Besides geometric proximity, tolerance may also be defined through the energetic properties of the local dynamics. Let $f : \Omega \to \Omega$ denote the time-shift map, and suppose that Ω is a vector space equipped with a symmetric positive definite matrix $M = M^\top \succ 0$. Defining the one-step transition map by

$$v : \Omega \to \Omega, \qquad v(x) = f(x) - x,$$

the corresponding *transition energy* is

$$K(x) = \frac{1}{2} v(x)^\top M v(x).$$

This quantity is the squared transition norm induced by the inner product defined by M. Consequently, the *scalar transition energy tolerance relation*

$$x \sim_{K_\varepsilon} y \Leftrightarrow |K(x) - K(y)| \leq \varepsilon$$

groups states whose local evolution possesses comparable intrinsic transition energy. Unlike metric proximity tolerance, scalar transition energy tolerance compares the magnitude of the local evolution, making it naturally suited to systems whose dynamics are more meaningfully characterized by transition activity than spatial proximity.

A related, but distinct, energetic construction compares the transition vectors themselves through the quadratic form

$$E(x, y) = \frac{1}{2}\big(v(x) - v(y)\big)^\top M \big(v(x) - v(y)\big).$$

The corresponding *vector transition energy tolerance relation* is defined by

$$x \sim_{E_\varepsilon} y \Leftrightarrow E(x, y) \leq \varepsilon.$$

This relation identifies states whose local transition vectors are similar in both magnitude and direction. Consequently, it preserves considerably more information about the local dynamics than the scalar transition energy tolerance and is particularly useful when constructing reduced models that retain local vector-field behavior. Since M is symmetric positive definite, the Rayleigh quotient satisfies

$$\lambda_{\min}(M)\|v\|^2 \leq v^\top M v \leq \lambda_{\max}(M)\|v\|^2, \qquad v \in \Omega,$$

and therefore

$$E(x, y) \leq \varepsilon \Rightarrow \|v(x) - v(y)\| \leq \sqrt{\frac{2\varepsilon}{\lambda_{\min}(M)}}.$$

Thus, vector transition energy tolerance controls the metric proximity between local state transitions. If, in addition, the transition map is locally bi-Lipschitz, so that there exists constant $c > 0$ such that

$$\|x - y\| \leq c\|v(x) - v(y)\|,$$

then

$$E(x, y) \leq \varepsilon \Rightarrow \|x - y\| \leq c\sqrt{\frac{2\varepsilon}{\lambda_{\min}(M)}}.$$

Hence, the corresponding tolerance spaces satisfy

$$E_\varepsilon \subseteq M_{c\sqrt{2\varepsilon/\lambda_{\min}(M)}}.$$

Therefore, whenever the transition map is locally invertible, vector transition energy tolerance implies metric tolerance.

Scalar transition energy tolerance relations are therefore well suited for identifying slow, metastable, or quasi-stationary regimes. Vector transition energy tolerance further distinguishes the direction of the local evolution, allowing TES to preserve both the magnitude and orientation of the dynamics. Together, these constructions illustrate how the tolerance space framework naturally accommodates application-specific notions of indistinguishability while leaving the underlying TES algorithm unchanged.

4.2 Kernel-induced tolerance relations

The tolerance-space framework naturally extends beyond metric-induced indistinguishability by allowing tolerance relations to be constructed from positive semidefinite kernels. Consequently, TES establishes a direct connection with the broad class of kernel methods, where similarity is encoded through a kernel rather than a metric. Classical examples include Gaussian, polynomial, Laplacian, diffusion, graph, and Fourier kernels, all of which induce corresponding notions of tolerance.

Suppose that $k : \Omega \times \Omega \to \mathbb{R}$ is a normalized positive semidefinite kernel satisfying $k(x, x) = 1$, $x \in \Omega$. A natural kernel discrepancy is

$$D_k(x, y) = 1 - k(x, y),$$

which vanishes whenever two states are maximally indistinguishable. The corresponding kernel-induced tolerance relation is defined by

$$x \sim_{k_\varepsilon} y \Leftrightarrow D_k(x, y) \leq \varepsilon.$$

Since

$$D_k(x, y) = \frac{1}{2}\|\phi(x) - \phi(y)\|^2_{\mathcal{H}},$$

where $\phi : \Omega \to \mathcal{H}$ is the feature map associated with the reproducing kernel Hilbert space, kernel-induced tolerance compares states according to their separation in feature space rather than in the original state space. Consequently, different kernels

induce different notions of indistinguishability while the TES algorithm itself remains unchanged. This construction illustrates a general principle underlying the tolerance space framework. Application-specific information is encoded in the tolerance relation rather than in the entropy-guided sampling algorithm. Improvements in kernel design therefore translate directly into new notions of information and new sampling strategies without modifying the underlying TES procedure.

The relationship with kernel methods is conceptual rather than algorithmic. Classical kernel methods use the kernel matrix to construct feature representations, classifiers, regression models, or low-dimensional embeddings, typically through spectral decompositions of the associated kernel operator. In contrast, TES does not analyze the kernel spectrum. Instead, the kernel serves only for inducing a tolerance relation, after which entropy-guided sampling operates directly on the resulting tolerance classes. Thus, kernel methods and TES are complementary rather than competing frameworks. The former exploits kernels to construct representations and learning algorithms, whereas the latter uses kernels solely to define the notion of indistinguishability on which information quantification and entropy-guided sampling are based.

More generally, any symmetric notion of indistinguishability naturally induces a tolerance relation and therefore a corresponding tolerance space. Consequently, the construction of application-specific tolerance relations amounts to specifying the notion of information relevant to the problem at hand, while the entropy-guided sampling algorithm itself remains unchanged. This separation between the definition of information and the algorithm used to eliminate informational redundancy is one of the principal conceptual advantages of the tolerance space framework.

5 Numerical illustrations

We conclude with two numerical experiments illustrating complementary aspects of the proposed tolerance space framework and TES. Rather than serving as performance benchmarks, these examples demonstrate how different choices of tolerance relations produce informative representations adapted to the geometry and dynamics of the underlying system. We first consider a noisy contracting spiral, whose simple geometry allows the sampling mechanism and its averaged variants to be visualized directly. We then investigate a chaotic forced Duffing oscillator, illustrating how different notions of dynamical similarity reveal complementary geometric and energetic features of the same nonlinear trajectory while leaving the underlying TES algorithm unchanged.

5.1 A noisy contracting spiral

We first consider a simple two-dimensional contracting spiral observed in the presence of additive measurement noise. The deterministic orbit is defined parametrically by

$$r(t) = e^{-\alpha t}, \qquad x(t) = r(t)\cos t, \qquad y(t) = r(t)\sin t,$$

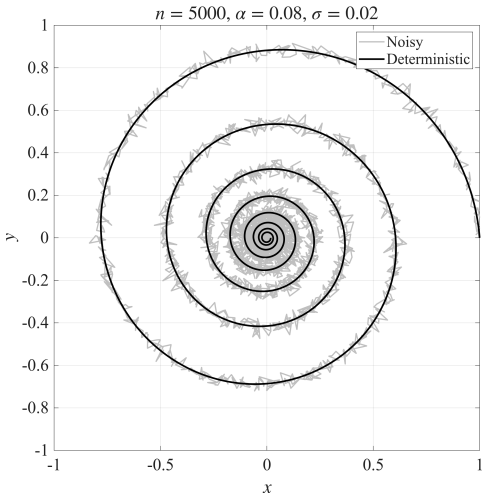

Fig. 1 Noisy observations of a contracting spiral together with the underlying deterministic orbit.

where $\alpha > 0$ denotes the contraction rate. The observed orbit is obtained by adding independent Gaussian perturbations,

$$\tilde{x}(t) = x(t) + \sigma\xi_1(t), \qquad \tilde{y}(t) = y(t) + \sigma\xi_2(t),$$

where ξ_1 and ξ_2 are independent standard normal random variables and $\sigma > 0$ is the noise intensity. The resulting dataset therefore consists of noisy observations of an asymptotically stable spiral rather than the exact deterministic orbit (Figure 1).

From the observed orbit we construct the normalized Euclidean distance matrix

$$D_{ij} = \frac{\|x_i - x_j\|_2}{\max\limits_{p,q} \|x_p - x_q\|_2},$$

whose entries lie in the interval $[0, 1]$. The corresponding metric tolerance relation is defined by

$$x_i \sim_\varepsilon x_j \Leftrightarrow D_{ij} \leq \varepsilon, \qquad \varepsilon \in [0, 1].$$

Applying the offline TES algorithm to this tolerance relation produces the MTES representation, whose representatives are selected directly from the observed orbit (Figure 2, left).

To illustrate the averaging procedure introduced in Section 3, we also compute the corresponding AMTES representation. After the MTES representatives have been determined, each representative is replaced by the arithmetic mean of all observations belonging to its associated tolerance class. Consequently, the AMTES representation no longer consists of points from the original orbit, but instead provides averaged representatives of these classes. Since random perturbations within each class tend to cancel, AMTES naturally suppresses observational noise while preserving the large-scale geometry of the spiral, thereby providing a smoother approximation of the underlying deterministic orbit (Figure 2, right).

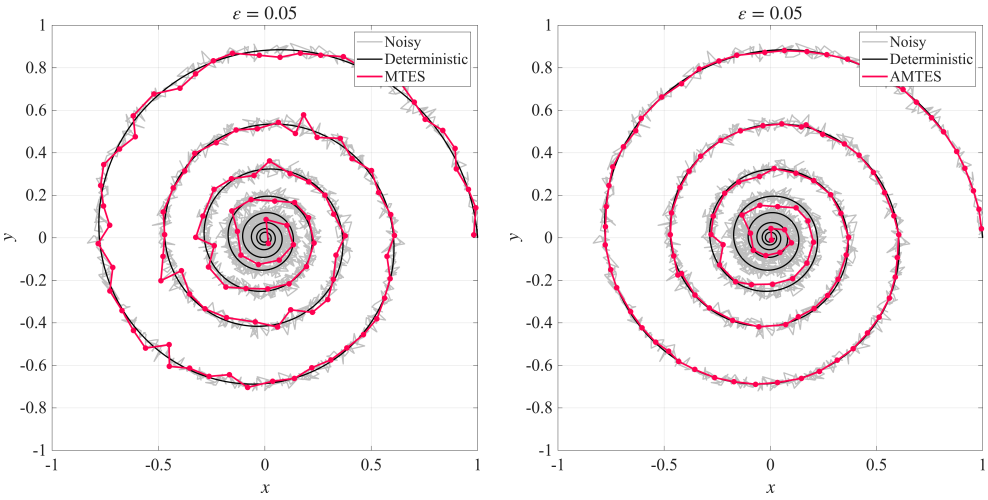

Fig. 2 MTES (left) and AMTES (right) representations of the noisy contracting spiral. The MTES representatives are selected directly from the observed orbit, whereas the AMTES representatives are obtained by averaging the corresponding tolerance classes, producing a smoother approximation of the underlying deterministic orbit.

The numerical results illustrate two complementary aspects of the metric realization. First, the MTES representatives remain on the observed orbit and provide a sparse yet informative description of the underlying dynamics. Second, the AMTES representatives produce a visibly smoother approximation by averaging neighbouring tolerance classes. Although AMTES effectively attenuates moderate observational noise through tolerance class averaging, its reconstruction accuracy is fundamentally limited by the chosen tolerance relation. When the noise level becomes sufficiently large so that the tolerance neighbourhoods overlap distinct branches of the deterministic orbit, the corresponding tolerance classes necessarily contain observations originating from different deterministic branches. Consequently, the averaged representatives need no longer approximate the underlying deterministic orbit, and accurate reconstruction cannot be expected.

The flexibility of the tolerance-space framework further allows the tolerance relation itself to be adapted to the underlying dynamics. As an illustration, we consider a radius-adaptive tolerance relation that gradually decreases the effective tolerance as the orbit approaches the attracting equilibrium. Let

$$R_i = \frac{\|x_i\|_2}{\max_k \|x_k\|_2} \in [0,1]$$

denote the normalized radial distance of the ith observation from the origin, and define

$$R_{ij} = \frac{R_i + R_j}{2}.$$

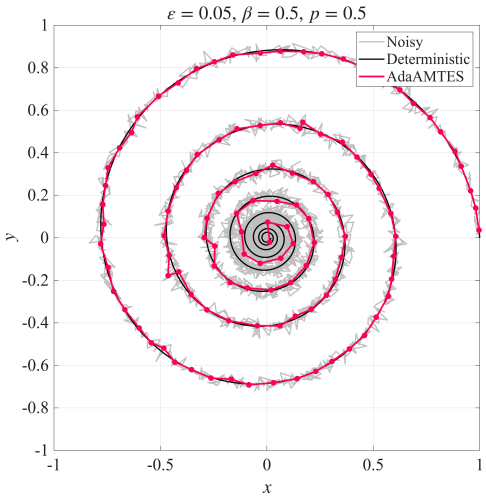

Fig. 3 AdaAMTES representation obtained using the radius-adaptive tolerance relation. Compared with the MTES construction, the shrinking tolerance neighbourhoods preserve additional representatives near the attracting equilibrium, providing a finer resolution of the asymptotic dynamics.

The adaptive tolerance relation is therefore defined by

$$x_i \sim_\varepsilon x_j \Leftrightarrow D_{ij} \leq \varepsilon \left[1 - \beta(1 - R_{ij})^p\right],$$

where $\beta \in (0, 1)$ controls the overall reduction of the tolerance radius and $p > 0$ determines how rapidly this reduction occurs as the orbit contracts toward the equilibrium. Since

$$1 - \beta \leq 1 - \beta(1 - R_{ij})^p \leq 1,$$

the effective tolerance varies continuously between $(1 - \beta)\varepsilon$ near the origin and ε far from the equilibrium. Consequently, observations in the outer part of the spiral are sampled similarly to the MTES construction, whereas the progressively shrinking tolerance neighbourhoods near the attracting equilibrium produce smaller tolerance classes and therefore preserve additional representatives in the region where the dynamics become increasingly concentrated. Applying the offline TES algorithm with this adaptive metric tolerance relation yields the adaptive averaged MTES (AdaAMTES) representation (Figure 3), which combines geometric smoothing with an increased resolution of the asymptotic dynamics.

5.2 A chaotic forced Duffing oscillator

As a second low-dimensional example, we consider the forced Duffing oscillator in a chaotic parameter regime. The purpose of this experiment is to compare two complementary notions of dynamical similarity on the same nonlinear trajectory. The first is the standard metric tolerance relation, which resolves geometric recurrence in the extended phase space. The second is an augmented energetic tolerance relation, which compares states through the stored, dissipated, and injected energy contributions appearing in the port-Hamiltonian energy balance. Introducing the forcing

phase as an additional state variable renders the non-autonomous Duffing equation autonomous in the extended phase space. In the port-Hamiltonian setting, the forced Duffing oscillator can therefore be written in the form

$$\dot{q} = (J - R)\nabla H(q) + b(q), \qquad q = (x, y, \theta)^\top,$$

where

$$J = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad R = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \delta & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad b(q) = \begin{pmatrix} 0 \\ F\cos\theta \\ \omega \end{pmatrix}.$$

The Hamiltonian of the system is

$$H(q) = \frac{1}{2}y^2 - \frac{1}{2}x^2 + \frac{1}{4}x^4,$$

which is independent of the phase variable θ. Hence,

$$\nabla H(q) = (-x + x^3, y, 0)^\top,$$

and the corresponding energy balance is

$$\dot{H}(q) = -\nabla H(q)^\top R \nabla H(q) + \nabla H(q)^\top b(q) = -\delta y^2 + Fy\cos\theta.$$

The quantities

$$Q(q) = \nabla H(q)^\top R \nabla H(q) = \delta y^2, \qquad I(q) = Fy\cos\theta,$$

represent the instantaneous dissipated and injected power, respectively. This energetic decomposition naturally motivates tolerance relations based on the local energetic behaviour of the trajectory. Throughout this example we employ the augmented energetic observable

$$E(q) = \begin{pmatrix} H(q) \\ Q(q)\Delta t \\ I(q)\Delta t \end{pmatrix} = \begin{pmatrix} \frac{1}{2}y^2 - \frac{1}{2}x^2 + \frac{1}{4}x^4 \\ \delta y^2 \Delta t \\ Fy\cos\theta\Delta t \end{pmatrix}.$$

Here $Q(q)\Delta t$ and $I(q)\Delta t$ represent the dissipated and injected energy over one sampling interval. The corresponding *augmented energetic tolerance relation* is

$$q_i \sim_{E_\varepsilon} q_j \Leftrightarrow \|E(q_i) - E(q_j)\|_2 \leq \varepsilon.$$

This relation compares the complete energetic signature of each state, consisting of its stored energy together with its instantaneous dissipated and injected energies. Consequently, it measures dynamical similarity in terms of local energetic behavior rather than geometric proximity.

In the computations below we use the parameter values $\delta = 1/4$, $\omega = 1.0$, and $F = 0.4$, for which the forced Duffing oscillator possesses a sustained chaotic attractor.

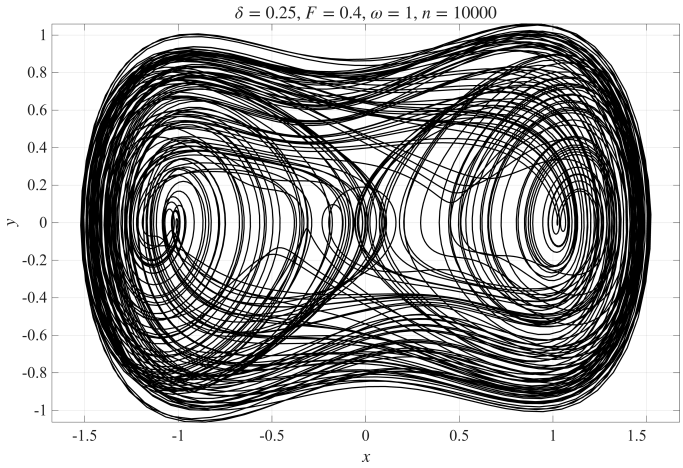

Fig. 4 Projection of the chaotic forced Duffing orbit onto the (x, y)-plane. The TES analysis is performed in the extended phase space (x, y, θ).

The system is integrated on the time interval $0 \leq t \leq 1000$ using Matlab's `ode45` solver with relative and absolute tolerances 10^{-10} and 10^{-12}, respectively, together with a uniform sampling interval $\Delta t = 0.1$. The initial condition is $q(0) = 0$, yielding 10^4 states in the extended phase space. For each state we evaluate the energetic observable $E(q)$ and construct two normalized pairwise distance matrices. One is induced by the Euclidean distance in the extended phase space, while the other by the Euclidean distance between the corresponding energetic observable. Normalizing both distance matrices by their maximal pairwise distance ensures that the tolerance parameter satisfies $0 \leq \varepsilon \leq 1$ for both tolerance relations, thereby allowing a direct comparison between the resulting TES filtrations.

Although the TES analysis is performed in the extended state space (x, y, θ), the resulting chaotic orbit projected onto the (x, y)-plane (Figure 4) clearly illustrates the characteristic folding structure of the Duffing attractor. The corresponding TES filtrations are computed over a uniform grid of 500 tolerance values on the interval $0 \leq \varepsilon \leq 1$, from which we obtain the representative cardinality function m_X for each tolerance relation (Figure 5, left). To quantify the scaling behavior, we perform a least-squares fit of the power-law model

$$m_X(\varepsilon) \propto \varepsilon^{-\Phi}$$

over the intervals

$$-4 \leq \log \varepsilon \leq -2, \qquad -5 \leq \log \varepsilon \leq -3$$

for the metric and energetic relations, respectively, where the associated log-log TES cardinality curves exhibit an approximately linear regime. The fitted TES indices (Figure 5, right) are subsequently used to estimate the expected tolerance scale associated with the power law approximation.

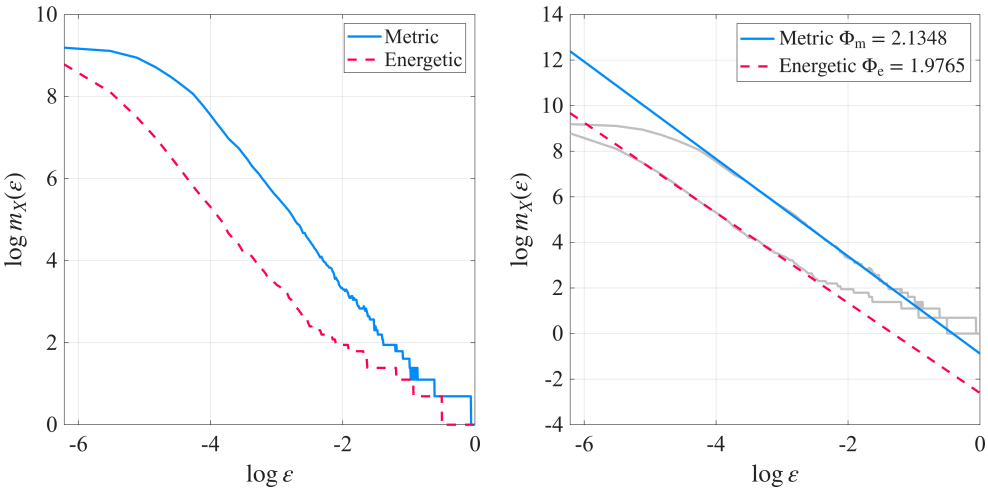

Fig. 5 TES cardinality curves $m_X(\varepsilon)$ for the metric and port-Hamiltonian tolerance relations (left), together with the corresponding least-squares power-law fits used to estimate the associated TES indices (right).

For the forced Duffing example, the metric and augmented energetic tolerance relations produce comparable, but distinct, scaling exponents on the fitted intervals,

$$\Phi_{\mathrm{m}} \approx 2.13, \qquad \Phi_{\mathrm{e}} \approx 1.98,$$

respectively. The metric tolerance relation resolves the complexity of the orbit in the extended phase space (x, y, θ), whereas the augmented energetic relation resolves complexity in the energetic observable space $(H, \Delta t\, Q, \Delta t\, I)$. The value $\Phi_{\mathrm{e}} \approx 1.98$ therefore should not be interpreted as a fractal dimension of the attractor, but rather as the TES scaling index associated with the energetic tolerance relation. Its proximity to two indicates that, when viewed through its energetic state variables, the chaotic Duffing dynamics possesses an approximately two-dimensional multiscale organization. The corresponding expected tolerance scales differ substantially,

$$\mathbb{E}(\varepsilon_{\mathrm{m}}) \approx 0.034, \qquad \mathbb{E}(\varepsilon_{\mathrm{e}}) \approx 0.013.$$

While the metric relation yields a relatively large characteristic scale, the augmented energetic relation indicates that energetically distinguishable states are separated at considerably finer scales. This reflects the increased sensitivity of the augmented energetic observable, where relatively small differences in stored, dissipated, or injected energy are sufficient to distinguish dynamically meaningful states. Consequently, the energetic filtration resolves the orbit more finely than the metric filtration while preserving essentially the same global scaling behavior.

More importantly than the numerical values themselves, both estimates emerge naturally from the scaling of the tolerance cardinality function, indicating that the tolerance index provides a meaningful quantitative descriptor of the intrinsic multiscale organization induced by the chosen tolerance relation. The close agreement

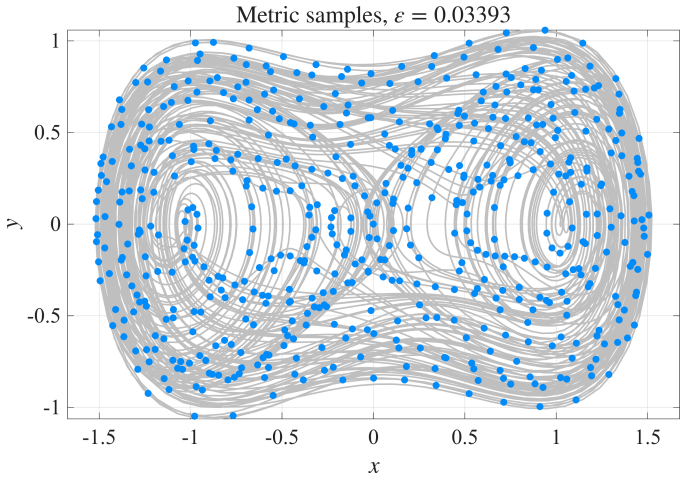

Fig. 6 Representative states selected by TES using the metric tolerance relation, projected onto the (x, y)-plane. The representatives are distributed approximately uniformly over the geometric attractor, reflecting geometric recurrence in the extended phase space.

between the metric and energetic scaling exponents further suggests that TES captures an intrinsic structural property of the underlying chaotic dynamics that is robust with respect to fundamentally different notions of dynamical similarity, even though the characteristic reduction scales differ significantly. A rigorous investigation of this apparent robustness is left for future work.

The similarity of the global scaling behavior, however, does not imply that the same representative states are selected. At their respective expected tolerance scales, the TES representatives obtained from the metric and augmented energetic tolerance relations are

$$m_{X,\mathrm{m}}(0.0339) = 558, \qquad m_{X,\mathrm{e}}(0.0138) = 350,$$

respectively. These values reflect the different characteristic scales induced by the two notions of indistinguishability; the augmented energetic relation operates at a substantially finer tolerance while requiring fewer representative states to describe the trajectory. This distinction becomes particularly apparent when the representatives are projected onto the physical (x, y)-plane. While the metric relation distributes representatives approximately uniformly over the geometric attractor (Figure 6), the augmented energetic relation concentrates representatives in regions exhibiting distinct energetic behavior (Figure 7). Consequently, similar global scaling behavior may coexist with distinct informative representations, depending on the notion of indistinguishability encoded by the chosen tolerance relation.

5.3 Discussion

The two numerical examples illustrate complementary aspects of the proposed entropy-guided sampling framework. The noisy contracting spiral demonstrates how the metric realization naturally produces sparse geometric representations, while its averaged variants simultaneously provide effective suppression of moderate observational noise. The chaotic Duffing oscillator shows that the same TES algorithm can be

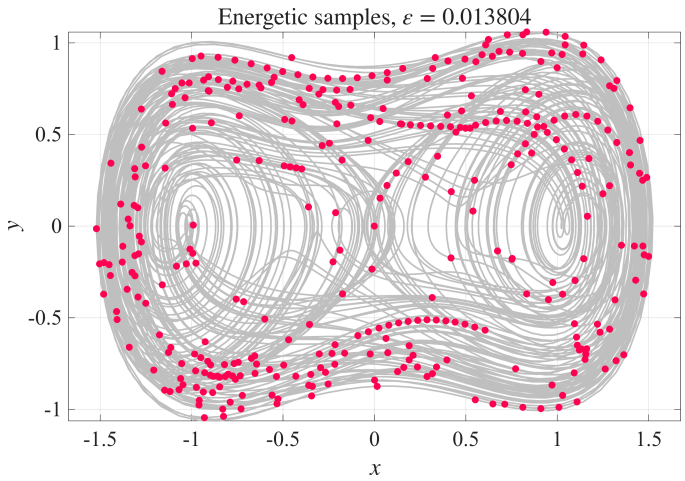

Fig. 7 Representative states selected by TES using the augmented energetic tolerance relation, projected onto the (x, y)-plane. Although the representatives are approximately uniformly distributed in the augmented energetic observable space, their projection onto the physical phase plane becomes non-uniform, illustrating the distinction between energetic and geometric notions of dynamical similarity.

combined with fundamentally different notions of dynamical similarity. Although the metric and augmented energetic tolerance relations exhibit comparable scaling exponents, they induce substantially different characteristic tolerance scales and produce different representative subsequences, reflecting the distinct geometric and energetic information encoded by the respective tolerance relations.

Taken together, these examples reinforce the central principle underlying the TES framework. The sampling algorithm itself remains unchanged, whereas the notion of information is determined entirely by the chosen tolerance relation. Consequently, different tolerance relations lead to different informative representations of the same dynamics without requiring any modification of the underlying mathematical or algorithmic framework. Rather than prescribing a universal notion of indistinguishability, the tolerance space formalism separates the definition of information from the process of eliminating informational redundancy, thereby providing a flexible and unified methodology for entropy-guided sampling across a broad range of nonlinear dynamical systems.

6 Conclusions

This paper has presented a unified mathematical framework for entropy-guided sampling in tolerance spaces. By treating tolerance relations as the primary objects encoding indistinguishability, the framework separates the notion of information from any particular metric, geometry, or similarity measure. Within this setting, tolerance entropy quantifies the informational content induced by the chosen tolerance relation, while tolerance entropy sampling constructs sparse representative sequences by eliminating informational redundancy relative to that notion of indistinguishability.

The resulting formalism extends naturally beyond metric spaces. Metric, energetic, linear, kernel-induced, graph-based, and other application-specific notions of similarity all arise as particular instances of the same general construction, allowing a common entropy-guided sampling principle to be adapted to diverse scientific problems. The introduction of tolerance filtrations further provides a multiscale description of information and establishes natural connections with nonlinear dynamics, recurrence analysis, manifold learning, and geometric data analysis.

The broader significance of the proposed framework lies in the separation between the definition of information and the algorithm used to eliminate redundancy. Once an application-specific notion of indistinguishability has been specified through a tolerance relation, the corresponding notions of entropy, representative sampling, and multiscale complexity follow within a common mathematical framework. Consequently, improvements in the modeling or learning of tolerance relations immediately induce new notions of information without requiring any modification of the underlying TES methodology.

We hope that the tolerance space viewpoint provides a useful foundation for developing application-specific notions of information and entropy-guided sampling across nonlinear dynamics, scientific computing, and data science.

Funding. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 531152215 - CRC 1701.

References

[1] F. Kasolis and M. Clemens, "Maximum Entropy Snapshot Sampling for Reduced Basis Generation," arXiv:2005.01280, 2020.

[2] M. W. F. M. Bannenberg, F. Kasolis, M. Günther, and M. Clemens, "Maximum Entropy Snapshot Sampling for Reduced Basis Modelling," *COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 41(2), 482–497, 2022.

[3] S. Lespinats, B. Colange, and D. Dutykh, *Nonlinear Dimensionality Reduction Techniques: A Data Structure Preservation Approach*, Springer Nature Switzerland AG, Cham, 2022.

[4] D. L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, 52(4), 1289–1306, 2006.

[5] E. J. Candès, J. Romberg, and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information," *IEEE Transactions on Information Theory*, 52(2), 489–509, 2006.

[6] D. P. Woodruff, "Sketching as a Tool for Numerical Linear Algebra," *Foundations and Trends in Theoretical Computer Science*, 10(1–2), 1–157, 2014.

[7] P. K. Agarwal, S. Har-Peled, and K. R. Varadarajan, "Geometric Approximation via Coresets," in *Combinatorial and Computational Geometry*, MSRI Publications, Vol. 52, Cambridge University Press, 2005.

[8] J. M. Phillips, "Coresets and Sketches," in *Handbook of Discrete and Computational Geometry*, 3rd ed., CRC Press, 2016.

[9] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, 27(3), 379–423, 1948; 27(4), 623–656, 1948.

[10] A. N. Kolmogorov, "A New Metric Invariant of Transitive Dynamical Systems and Automorphisms," *Doklady Akademii Nauk SSSR*, 119, 861–864, 1958.

[11] Y. G. Sinai, "On the Concept of Entropy for a Dynamic System," *Doklady Akademii Nauk SSSR*, 124, 768–771, 1959.

[12] R. L. Adler, A. G. Konheim, and M. H. McAndrew, "Topological Entropy," *Transactions of the American Mathematical Society*, 114, 309–319, 1965.

[13] P. Walters, *An Introduction to Ergodic Theory*, Springer, 1982.

[14] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press, 1995.

[15] F. Kasolis and M. Clemens, "Model Order Reducibility of Nonlinear Electro-Quasistatic Problems," *COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 38(5), 1453–1464, 2019.

[16] F. Kasolis and M. Clemens, "Information-Based Model Reduction for Nonlinear Electro-Quasistatic Problems," *Journal of Computational Physics*, 404, 109118, 2020.

[17] F. Kasolis and M. Clemens, "Entropy Snapshot Filtering for QR-Based Model Reduction of Transient Nonlinear Electro-Quasistatic Simulations," *IEEE Transactions on Magnetics*, 56(2), Art. 7400204, 2020.

[18] F. Kasolis and M. Clemens, "Energy-Variation Analysis and Orbit-Complexity Quantification," *Journal of Physics: Conference Series*, 2090(1), Art. 012086, 2021.

[19] F. Kasolis and M. Clemens, "Critical Recurrence-Scale Thresholds," TechRxiv, 2023.

[20] S. Stroka, F. Kasolis, N. Haußmann, and M. Clemens, "Efficient Low-Frequency Human Exposure Assessment With the Maximum Entropy Snapshot Sampling," *IEEE Transactions on Magnetics*, 60(12), 2024.

[21] S. Stroka, N. Haußmann, F. Kasolis, and M. Clemens, "Enhanced Low-Frequency Dosimetry Simulations with the Energy-Variation Maximum Entropy Snapshot Sampling," *IEEE Transactions on Magnetics*, 2025.

[22] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, 290, 2319–2323, 2000.

[23] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence Plots for the Analysis of Complex Systems," *Physics Reports*, **438**(5–6), 237–329 (2007).

[24] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, 290, 2323–2326, 2000.

[25] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural Computation*, 15(6), 1373–1396, 2003.

[26] R. R. Coifman and S. Lafon, "Diffusion Maps," *Applied and Computational Harmonic Analysis*, 21(1), 5–30, 2006.

[27] G. Carlsson, "Topology and Data," *Bulletin of the American Mathematical Society*, 46(2), 255–308, 2009.