Published in

Nonlinear Structures in Physical Systems—Pattern Formation, Chaos and Waves,

L. Lam and H. C. Morris, Springer-Verlag, New York (1990) 119-130.

INFORMATION AND ITS METRIC

J. P. Crutchfield[*]

Information is taken as the primary physical entity from which probabilities can be derived. Information produced by a source is defined as the class of sources that are recoding-equivalent. Shannon entropy is one of a family of formal Renyi information measures on the space of unique sources. Each of these measures quantifies the volume of the source's recoding-equivalence class. A space of information sources is constructed from elements each of which is the class of recoding-equivalent, but otherwise unique sources. The norm in this space is the source entropy. A measure of distance between information sources is derived from an algebra of measurements. With this, the space of information sources is shown to be a metric space whose logic is described by a metric lattice. Applications of the information metric to quantum informational uncertainty and to information densities in multicomponent dynamical systems are outlined.

1. INTRODUCTION

This brief note introduces a measure of distance between information sources. It demonstrates that the space of information sources has much topological structure which heretofore has not been utilized directly in applications of information theory or in the study of complex systems. The space of information sources is, in fact, a metric space to which a geometric picture can be ascribed. Furthermore, if one considers the elemental events from the information source to be experimental measurements, then the logic of inference is described by a metric lattice. Although this note is restricted to a presentation of formal details and to their geometric interpretation, a few motivational remarks are in order.

There are four aspects of the use of information theory in the physics of complex systems that suggest a need for a better understanding of its formal structure and of the concept of information itself. Briefly, these are the observational theory of chaotic systems,[1] the *subjectivity* of information, the modeling of complex dynamical systems, and multi-component information sources.

[*] J. P. Crutchfield: Physics Department, University of California, Berkeley, California 94720 USA; Internet: chaos@gojira.berkeley.edu.

An explicit physical theory of observing chaotic behavior is necessitated by (i) a chaotic process's extreme (exponential) sensitivity to extrinsic fluctuations, (ii) the impossibility of an infinitely precise determination of a system's state, and (iii) the impossibility of employing infinitely large computational resources for the prediction of states at arbitrary future times. Although the latter two limitations have no rigorous lower bounds in a classical, deterministic universe, they are at the very least practical, macroscopic facts. The proper framework for the description of such behavior is information theory.[2]

As Renyi has noted,[3] information is no more subjective than probability. The position taken here is a simple corollary of Renyi's: that information theory provides a quantitative and consistent framework with which to describe physical processes that admit only partial knowledge. Information, in this framework, has two functions. First, it quantifies the utility of observations in the prediction of behavior, and so, conversely, is a quantifier of behavioral complexity. Second, on a higher level of abstraction, it measures the utility of observations in the construction of models.[4]

In chaotic data analysis, and in other fields as well, a number of information theoretic quantities have been referred to as distances. Most notable among these quantities have been the conditional mutual information,[5] Jeffrey's divergence,[6] and Kullback's information gain.[7] As noted by authors in communication theory and mathematical statistics, however, none of these satisfy all of the metric properties required of a distance function. The following discussion alleviates this with the introduction of an informational quantity that is a distance function. Aside from clarifying problems in chaotic data analysis, this information metric presumably will be of use to information theory proper, and provide a starting point for extending information theory to many sources producing, transmitting, and receiving information simultaneously. A more complete discussion of these topics and a more detailed discussion of the following results will be presented in a sequel.

2. INFORMATION AND CODING

The first step is to discuss the sources of information. These might be deterministic or non-deterministic. Examples of the former are pseudo-random number generators and chaotic dynamical systems; of the latter, Markovian or general stochastic processes and noisy dynamical systems.[1]

The case of sources producing distinct, discrete events will be considered here. An information source shall be denoted as S. Each measurement of the state of the information source yields one of a finite number of symbols. The symbols a_i, or simply **measurements**, come from a finite **alphabet** $\mathbf{A}$ of N elements or events: $\mathbf{A} = \{a_0, \ldots, a_{N-1}\}$. This is tantamount to a model of the measuring instrument used to observe the source. If the source has more states than the measuring instrument can uniquely identify, (say) if it were continuous, then the model of the measuring instrument needs to specify how the information source states are grouped into measurement symbols by the observation process. In any case, this association of source states with instrument measurements is called a **partition** P of the information source state space.[8, 9]

The space of unique sources $I = \{S_i\}$ is the collection of all information sources. The dimension of this space is unspecified.

Beyond the sources themselves, one must consider what can and should be done with information. Acquisition of, processing, and inferring from information are the functions of an observer. An **observer** is essentially defined by its available **observation resources**. These consist of the measurement resolution and rate, the storage capacity, and the computational power. Delineating these resources constitutes the barest outline of a model of an observer. Although the following discussion does not require further development of such a model, it is useful to keep in mind the general goal. The central problem of observational theory can be stated as follows. Given an *arbitrarily* large number of measurements and given *finite* observational resources, how much algorithmic structure in the information source can be deduced? There are a number of related questions such as, the existence of optimum partitions or instruments, bounds on obtainable information, and how does size of available statistics affect these quantities?[9] But these are subsidiary questions that will have to wait until the main questions have been addressed.

For the present purposes the dependence on a partition and its geometry can be ignored. The information sources should then be considered as the resultant measurements produced by some instrument. In other words, the source is the output of the instrument, which has filtered the observed behavior through the partition.

Conventional developments of information theory measure the complexity of a source in terms of the **entropy**. Entropy is a function on the space of sources I that yields real numbers, $H(S) : I \to \Re$, with the source characterized by its probability measure. With respect to the entropy, there is a degeneracy of information sources with the same entropy. Given two sources S_1 and S_2, $H(S_1) = H(S_2)$ does not mean that the sources produce the same information, merely that the quantity of information produced by the sources are the same, and possibly that their information production **rates** are the same. The nature of this degeneracy can be explored via the existence of transformations, or encodings, of the sources into each other.

Recall Shannon's coding theorem for a noiseless channel. With a uni-directional channel of capacity $C \geq H(S_1)$, an "input" source S_1 can be encoded into an output "source" S_2, such that $H(S_1) = H(S_2)$ and the input source can be "reconstructed" from the output. The latter is to say that the channel looses no information, or that there are no "errors" in transmission. S_2, considered as an information source, then, is a faithful reproduction of S_1, with the proper channel encoding. The channel is considered here as mechanism that transforms one source into another.

Somewhat anticipating a later development, when applied to the observation process, note that encodings capture the conditional nature of measurements. Any time a measurement is made, the particular measurement's occurrence is conditioned on the physical process being in some state.

To clarify this situation, define an **encoding** of a source S_1 into another S_2 as a many-to-one transformation of the first source's alphabet into the second's. Encodings describe asymmetric intersource relationships and will be denoted $S_1 \rightarrow S_2$. So defined, encoding $\rightarrow$ is an order relation on the space I that partially orders information sources.

The notion of source equivalence can now be established using a bi-directional channel. Define a **recoding** R as a one-to-one coding such that $S_1 \rightarrow S_2$ and $S_2 \rightarrow S_1$. Recoding makes the two sources isomorphic $S_1 \sim S_2$; it also follows that $H(S_1) = H(S_2)$. In terms of a bi-directional channel with channel capacities for each direction of transformation $C_{1\rightarrow 2}$ and $C_{2\rightarrow 1}$, Shannon's theorem requires that a recoding must have $C_{1\rightarrow 2} > H(S_1)$ and $C_{2\rightarrow 1} > H(S_2)$. These bound the complexity of the recoding considered as a transformation.

Two additional restrictions must be placed on recodings. They must not allow temporal translation, since a_i and a_{i+k} may be decorrelated for some k and should not be considered equivalent by a time shift coding. Recodings must be **instantaneous** encodings, although we may allow for a finite time for their construction. Thus, recodings are much more restricted than the encodings consistent with Shannon's first theorem.[5]

With this framework a number of formal properties are straightforwardly established. First, a recoding R is a binary relation on two sources, $R \subset S_1 \times S_2$, or equivalently

$$R = \{(a_i, b_j) : \ a_i \in S_1 \text{ and } b_j \in S_2\} \tag{1}$$

where $\{a_i\}$ and $\{b_j\}$ are sets of possible measurements.

Theorem: A recoding R is an equivalence relation on I.

Proof:

1. R is reflexive: $\{(a,a) : a \in R\} \subset R$, or equivalently $R \subset \mathbf{A} \times \mathbf{A}$. This follows from noting that the identity encoding is a recoding.
2. R is symmetric: $\{(b,a) : a, b \in R\} = R$. This follows from reversibility of recoding.
3. R is transitive: $\{(a,c) : \exists b \in \mathbf{A}$ and $(a,b) \in R$ and $(b,c) \in R\} \subset R$. This follows from composition of one-to-one codings.

If $(S_1, S_2) \in R$, then $S_1 \sim S_2$.

This establishes an equivalence relation $\sim$ on information sources: $S_i \sim S_j$. Two sources are equivalent if there exists a one-to-one coding that transforms measurement sequences from one to the other. In other words, measurement sequences from one source can be inferred from the other with probability one and with finite computational resources. The latter means converges with error probability $1 - \delta$ with ϵ resolution. Resolution here refers to the average closeness of the two representations. A one-to-one coding is called a **re**coding to emphasize that all observations exist in some code before they can be perceived, processed, stored, and so on.

If the sources are discrete then recoding is a permutation matrix; if they are continuous, recoding is a homeomorphism (1-1 invertible). Note that with this definition of equivalence no notion of the computational difficulty of computing the transformation is included. Thus, quantitative aspects of computational complexity will be missing in the geometric picture presented below. Extending the definition of equivalent sources to account for this appears to be an interesting direction for future work.

Let $[S]$ denote the subset of sources in I that are recoding-equivalent to source S. Then $[S_1] = \{S_2 : S_1 \sim S_2\}$ is the equivalence class of S_1.

When one speaks of information one refers to that entity shared by all sources that are equivalent up to recoding. This observation suggests a definition of information. **Information** of the source S is the equivalence class of all recodings of the symbol sequences from S. This is noteworthy since information generally is left undefined in information theory. Information theory only considers the amount of information or rates of production, loss, and transmission, as measured by various entropies.

The recoding equivalence reflects a further degeneracy due to a transformation of the symbols of the first source S_1 into those of the second S_2. One consequence of the existence of such an encoding is that any possible “meaning” of the first source’s symbols and symbol sequences may not have any relevance in the context of S_2’s symbols. We leave out at this stage the consideration of anthropomorphic projections of “context” or “meaning”, but return to these questions at the end, except to mention a simple example. Consider the following situation in the context of mathematical language recognition. There exists a trajectory of the chaotic Rössler attractor and a coarse measuring instrument with only 26 measurement symbols such that the symbol sequence encodes the system’s equations of motion. The latter has meaning to the mathematical observer, the former is meaningless to the same observer although it encodes the same information.

A standard result in binary relations gives the

Theorem: Recoding R **partitions** the space I of unique information sources into mutually disjoint subsets.

Proof: Every element of each subset is recoding-equivalent to every other.

This suggests the definition of the more abstract **information space** $\mathbf{I} = I/R$ as the set of equivalence classes of I under recodings R. That is, $\mathbf{I} = \{[S_i] : S_i \in I\}$. Elements of $\mathbf{I}$ shall be denoted X, Y, and so on. This will distinguish these classes of sources from the unique sources S_i of I. The elements of $\mathbf{I}$ shall be the objects of interest in the following. The necessary logical distinction between the class X and its constituent sources will be blurred in the following. A reference to X as an information source should be construed as connoting the common properties of its members. As a source, X is the generic source. We can speak, in a similar vein, of the events or measurements of source X.

3. TOWARD INFORMATION GEOMETRY

The preceding development of information theory neither relies on nor makes reference to a probability measure. This is no coincidence, however, since a formal equivalence has been established.[10] Under appropriate restrictions, probability can be derived from information. Probability here is taken to be a secondary concept. Indeed, the quantity of information, the **entropy**, is a measure on the space of information sources, in the same sense that probability is a measure on event space. The entropy of source $X \in \mathbf{I}$ quantifies the size or volume of the equivalence class. The following establishes entropy as a measure and sets up a partially "geometric" picture.

The starting point of the development is the following four definitions.

1. The origin $\emptyset$ is the measurement set that is predictable: $H(\emptyset) = 0$.
2. The **norm** $\|X\|$ of a source X is its entropy $H(X)$.
3. The addition of two sources is the union of measurements: $X + Y = \{$all events in either X or $Y\}$.
4. The product of two sources is the intersection of their measurements: $X \cdot Y = \{$all those events common to X and $Y\}$.

These operations yield an algebra of measurements. The first step is to establish that the entropy is a measure.

1. $O \leq H(X) \leq \infty$ for every source X and $\exists\, X_0$ such that $H(X_0) < \infty$.
2. $H(X+Y) = H(X) + H(Y)$, whenever $X \cdot Y = \emptyset$; that is, the sources are independent.

It also follows from these that

1. If $X \rightarrow Y$, then $H(Y) \leq H(X)$;
2. $H(\emptyset) = 0$; and
3. $H(X+Y) \leq H(X) + H(Y)$ for any sources X and Y.

To determine the distance $d(X,Y)$ between two sources X and Y requires a measure of their difference $X \triangle Y$, where $\triangle$ is the symmetric difference in the set theoretic sense. $X \triangle Y$ is itself an information source and so formally we define $d(X,Y) = \|X \triangle Y\|$. This yields a generalized picture of the norm of a source as being the "distance" from the origin of predictable, zero-entropy sources, since

$$\|X\| = \|X \triangle \emptyset\| = d(X, \emptyset) \ . \tag{2}$$

What is the interpretation of the information source $X \triangle Y$? Roughly speaking, common events are missing. In set theoretic terms there are two constituent, independent sources: (i) $X - Y$ are those events in X and not in Y and (ii) $Y - X$ corresponds to those in Y and not in X. The entropy $H(X-Y)$ of the source measurements in $X - Y$ defines a conditional entropy $H(X|Y)$ for measurements x_i of X given y_j of Y, such that x_i is not determined from y_j with probability one.

Define a source Z that is the union of these two sources,

$$Z \equiv X \triangle Y = Z_1 + Z_2 = (X - Y) + (Y - X) \ . \qquad (3)$$

From the algebra of measurements it follows that

$$Z = (X + Y) \cdot \overline{X \cdot Y} = X \cdot \overline{X \cdot Y} + Y \cdot \overline{X \cdot Y} \ . \qquad (4)$$

In informational terms, we have

$$H(X - Y) = H\big(X \cdot \overline{X \cdot Y}\big) \text{ and } H(Y - X) = H\big(Y \cdot \overline{X \cdot Y}\big) \ . \qquad (5)$$

A measure of the size or entropy of Z will be a measure of the non-commonality or distance between X and Y,

$$H(Z) = H(Z_1 + Z_2) = H(Z_1) + H(Z_2|Z_1) = H(Z_1) + H(Z_2) \ . \qquad (6)$$

The last step follows from the independence of Z_1 and Z_2,

$$Z_1 \cdot Z_2 = \big(X \cdot \overline{X \cdot Y}\big) \cdot \big(Y \cdot \overline{X \cdot Y}\big) = (X \cdot Y) \cdot \overline{X \cdot Y} = \emptyset \ . \qquad (7)$$

Continuing, we find

$$H(Z) = H\big(X \cdot \overline{X \cdot Y}\big) + H\big(Y \cdot \overline{X \cdot Y}\big) = H(X|Y) + H(Y|X) = d(X, Y) \ . \qquad (8)$$

With this we have established, starting with the entropy as a norm of an information source, that the algebra of measurements allows us to define the conditional sources $X - Y$ and $Y - X$. From this, it readily follows that $\|X - Y\| = d(X - Y, \emptyset)$.

The associated pseudo-geometric picture is shown in Fig. 1.

4. THE INFORMATION METRIC

The information quantity $d(X, Y) \equiv H(X|Y) + H(Y|X)$ can be interpreted as the total independent information. We now establish its metric properties. Although the proofs are straightforward, the basic steps are given in order to elucidate the central role of recoding.

Theorem d is a metric and $(\mathbf{I}, d)$ is a metric space.

Proof: **Symmetry**: $d(X, Y) = d(Y, X)$. This follows directly from the symmetry of the definition.

Equivalence: $d(X, Y) = 0$ if and only if $X \sim Y$.

"Only if": Assume $d(X, Y) = 0$. As the conditional entropies themselves are positive or zero and their sum is zero, they individually vanish. Consider one of the zero conditional entropies $H(X|Y) = 0$. The measurements of X knowing those of Y provide no new information and so may be inferred with probability one from Y. Thus, there is a recoding, that may be many-to-one, of measurements of Y into X measurements. Similarly, since $H(Y|X)$ vanishes, there is a recoding of X measurements into those from source Y. Taking

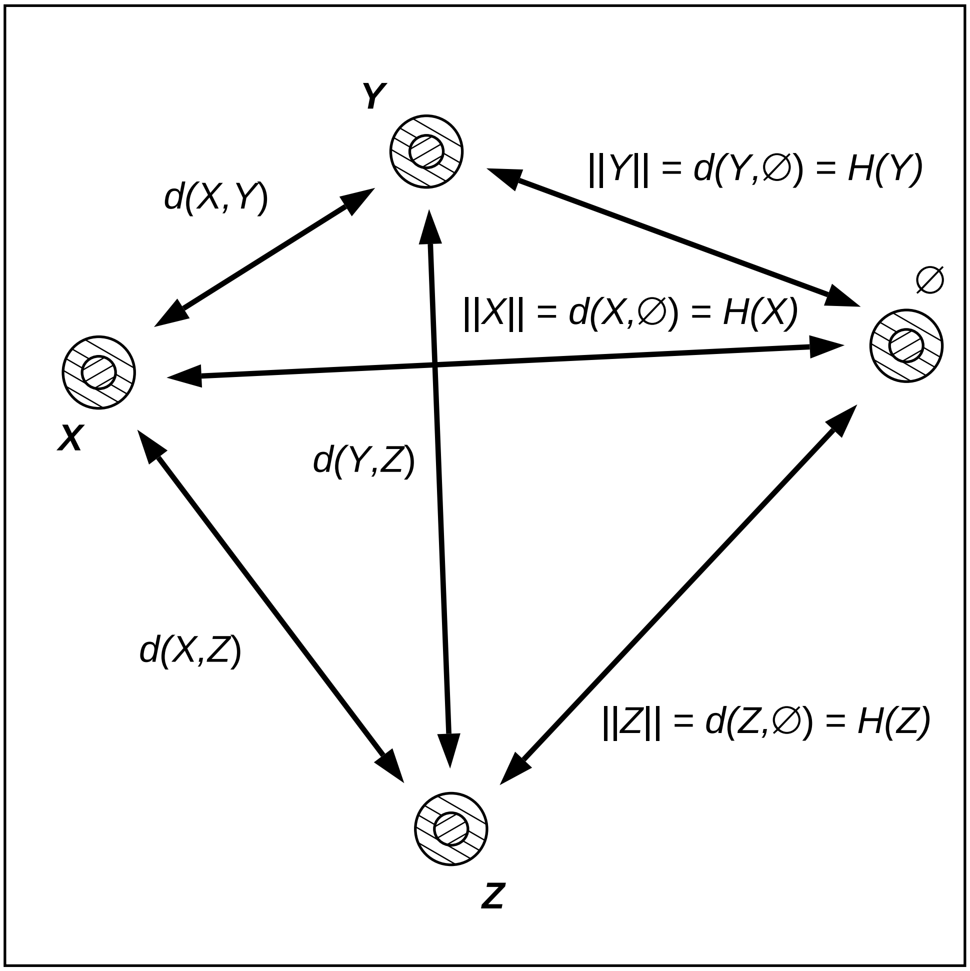

Figure 1 The geometric interpretation of the space of information sources. Shown are the distances and information "vectors" for three sources.

these together, there is a one-to-one coding between measurements from X and from Y and so they are equivalent sources: $X \sim Y$.

For the "If" portion: Assume $X \sim Y$, then there is a one-to-one recoding between measurements from X and from Y. Measurements of source X can be deduced with probability one from those of Y and *visa versa*. It follows from this that the conditional entropies vanish, as does the distance between them.

Triangle inequality: $d(X,Z) \leq d(X,Y) + d(Y,Z)$.

We consider expansions of the three variable joint entropy.

$$\begin{aligned} & H(X,Z) \leq H(X,Y,Z) \\ \text{or } & H(X,Z) \leq H(X|YZ) + H(Y,Z) \ . \end{aligned} \tag{9}$$

Noting that additional measurements cannot increase the entropy, i.e. $H(X|Y) \geq H(X|YZ)$, we have

$$\begin{aligned} & H(X,Z) \leq H(X|Y) + H(Y,Z) \\ \text{or } & H(X,Z) \leq H(X,Y) - H(Y) + H(Y,Z) \ . \end{aligned} \tag{10}$$

Subtracting the average independent entropy $\frac{1}{2}(H(X)+H(Y))$ yields

$$\begin{aligned} & H(X,Z)-\frac{1}{2}H(X)-\frac{1}{2}H(Z) \\ & \qquad \leq H(X,Y)-\frac{1}{2}H(Y)-\frac{1}{2}H(X)+H(Y,Z)-\frac{1}{2}H(Z)-\frac{1}{2}H(Y) \\ & \text{or } H(X|Z)+H(Z|X) \leq H(X|Y)+H(Y|X)+H(Y|Z)+H(Z|Y) \\ & \qquad \text{or } d(X,Z) \leq d(X,Y)+d(Y,Z)\ . \end{aligned} \tag{11}$$

Thus, $d(\cdot,\cdot)$ is a metric. With this, it follows that the pair $(\mathbf{I},d)$, where **I** is the space of recoding-equivalent information sources, is a metric space. This completes the proof.

The theorem indicates that the space of information sources has quite a bit of topological structure. For example, the notion of ϵ-balls of "close" information sources, the continuity of functions on information sources, and the limits and convergence of sequences of information sources, can be developed. These and numerical computations of information distances will follow in a sequel.

We can define a normalized metric as follows

$$\hat{d}(X,Y) = \frac{H(X|Y)+H(Y|X)}{H(X,Y)}\ . \tag{12}$$

Note that in the case of independent sources $\hat{d} = 1$.

5. INFORMATION FLUCTUATIONS

The information metric can be readily extended to account for fluctuations in information by using Renyi's generalization[1, 3] of Shannon information as the starting point. For brevity's sake we shall simply quote the results here in terms of probabilities rather than developing them axiomatically. First, recall the definition of the α-order Renyi information of a probability distribution $P(x)$,

$$H_\alpha(X) = (1-\alpha)^{-1}\log\sum_{x\in X} P^\alpha(x) \tag{13}$$

and the α-order conditional information,

$$H_\alpha(Y|X) = \frac{\alpha}{1-\alpha}\log\sum_{y\in Y}\left[\sum_{x\in X} P^\alpha(x)P^\alpha(y|x)\right]^{\alpha^{-1}}\ . \tag{14}$$

For $\alpha > 0$, we have the following Renyi metric

$$d_\alpha(X,Y) = H_\alpha(X|Y)+H_\alpha(Y|X)\ . \tag{15}$$

As is standard, we note that for $\alpha \to 1$ d_α reduces to the normalized Shannon metric $d(\cdot,\cdot)$. The spectrum of information fluctuations for a source is investigated via the α dependence of the Renyi metric H_α. Thus d_α provides a metric for (say) comparing the fluctuations between two different sources.

6. APPLICATIONS

The preceding has established a "geometric" picture underlying information theory and measurements of unpredictable processes. In closing, we briefly mention three applications and then conclude with a few remarks on the philosophical context.

1. As already noted, the recoding as a binary relation induces a partial ordering on the space of information sources. From this and the existence of the information metric it follows that the metric space $(\mathbf{I}, d)$ is a metric lattice.[11] With this we have established a framework with which to discuss the relationship between the inferential logic underlying observations of information sources and that of quantum logic as developed as developed by Birkhoff and von Neumann.[12] In the context of observing chaotic dynamical systems, this similarity bolsters others between chaos and quantum mechanics. For example, in both initial information from a prepared state can decay necessitating further measurements to determine future states. When these additional measurements are made the observer's ignorance "collapses" revealing the system's actual state. Said more simply, just as the process of observation is central in quantum mechanics, a model of the measurement process is required for chaotic physical systems.[1]
2. The second application of the metric is the derivation of an informational form of quantum mechanical uncertainty. Consider the Wigner wave function on the state space and its associated joint probability density $P(q,p)$ as a function of position q and momentum p. An example, would be to consider the latter to be a two-dimensional Gaussian. The Gaussian gives the maximum information consistent with a given mean and standard deviation of position and momentum. The conditional distributions required for the metric are readily formed,

$$P(p|q) = P^{-1}(q)P(q,p) \text{ and } P(q|p) = P^{-1}(p)P(q,p) \; . \tag{16}$$

The informational distance between measurements of the position and of the momentum, two different sources, is then

$$d(q,p) = H(p|q) + H(q|p) \; . \tag{17}$$

To give some significant to the result of this we change coordinates from a unit-sized physical system to one with Planck's constant h as the scale. With this we find the informational uncertainty principle,

$$d(p,q) \geq \log(h) \; . \tag{18}$$

Note that this applies to any time-shift invariant measurement system, like quantum mechanics. This suggests the possibility of measuring the effective quantization in any data set by measuring for all given observables their mutual distance. Observable pairs for which the metric does not vanish would be effectively conjugate. The interpretation of this result for quantum mechanics is that conjugate variables in quantum mechanics cannot be any closer that approximately 37 bits. They must have at least 37 bits of noncommon information. And so both must be measured in order to characterize a quantum system's state.

3. As a final arena of application, we introduce a general notion of information densities. This will be of use in multicomponent systems where one wishes to measure their information production and transport properties. To take an example, consider spatially-extended dynamical systems. Generally, we define the information density at a space-time point $\vec{p} = (\vec{x}, t)$ in terms of the metric as

$$\hat{I}(\vec{p}) = \lim_{\delta\vec{p}\to 0} \frac{d\big(S_{\vec{p}}, S_{\vec{p}+\delta\vec{p}}\big)}{\delta\vec{p}} \tag{19}$$

where $\delta\vec{p}$ is a space-time separation. If the source $S_{\vec{p}}$ is the asymptotic distribution on the reconstructed data at point $\vec{p}$, then this yields the dimension density. If the source $S_{\vec{p}}$ is the asymptotic distribution on the measurement sequences obtained at point $\vec{p}$, then this yields the entropy density.

Further exposition of these applications will appear elsewhere.

7. CONCLUDING REMARKS

One question that arises in this development is why not simply use mutual information instead of the information metric. Aside from the pseudo-geometric picture we have presented, we note that the former measures only a kind of informational correlation. The information metric, however, quantifies the degree of recoding equivalence. And so, it provides some insight into the nature of information itself. Mutual information is a derivative concept that simply reflects the properties Shannon entropy and no more.

The foregoing mathematical development instantiates a particular philosophical viewpoint, that of phenomenology. All that an observer has to work with in developing an understanding of the world are finite measurements and the attendant information. This intrinsic finiteness derives first and foremost from the limited computation resources available to an observer in a finite space-time region. The information space, as developed here, is the substrate for all perception, quantification, and modeling building. This is then structured with the pseudo-geometry as we have just shown. Only under suitable restrictions is one justified in using observations to form probabilities via (say) frequencies of events.

Information theory was founded on a quantitative measure of the **amount** of information. The foregoing has given a formal definition of information itself in terms of the equivalence class structure of sources. But what of the "meaning" of this information? A motivation of this work, unstated until this point, was the conviction that an understanding of the topological structure of the metric lattice of inferential logic is necessary for developing a quantitative measure of meaning and of context. Thus, we offer no immediate answer to the question, only the hope that progress can be made. We shall return to this question in the future.

The author thanks Persi Diaconis for his extended loan of reference [7]; divergent variations on which lead to the current work. A post-doctoral fellowship from International Business Machines Corporation, supported this work during 1985 - 1986 academic year. Further development was funded in part by ONR contract N00014-86-K-0154.

Postscript: This work was first distributed in October 1987. It appears here in its original form with minor corrections.

REFERENCES

1. J. P. Crutchfield, *Noisy Chaos*. PhD thesis, University of California, Santa Cruz, 1983. Published by University Microfilms Intl, Ann Arbor, Michigan.
2. R. Shaw, “Strange attractors, chaotic behavior, and information flow,” *Z. Naturforsh.*, vol. 36a, p. 80, 1981.
3. A. Renyi, “Some fundamental questions of information theory,” in *Selected Papers of Alfred Renyi, Vol. 2*, p. 526, Budapest: Akademii Kiado, 1976.
4. J. P. Crutchfield and B. S. McNamara, “Equations of motion from a data series,” *Complex Systems*, vol. 1, pp. 417 – 452, 1987.
5. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Champaign-Urbana: University of Illinois Press, 1962.
6. H. Jeffreys, *Theory of Probability*. Oxford: Oxford Clarendon Press, second ed., 1948.
7. S. Kullback, *Information Theory and Statistics*. New York: Dover, 1968.
8. V. M. Alekseyev and M. V. Jacobson, “Symbolic dynamics,” *Phys. Rep.*, vol. 25, p. 287, 1981.
9. J. P. Crutchfield and N. H. Packard, “Symbolic dynamics of one-dimensional maps: Entropies, finite precision, and noise,” *Intl. J. Theo. Phys.*, vol. 21, p. 433, 1982.
10. R. S. Ingarden and K. Urbanik, “Information without probability,” *Colloq. Math.*, vol. IX, p. 131, 1962.
11. G. Birkhoff, *Lattice Theory*. Providence: American Mathematical Society, third ed., 1967.
12. G. Birkhoff and J. von Neumann, “On the logic of quantum mechanics,” *Ann. Math.*, vol. 37, p. 823, 1936.