OPEN

Information theory and dimensionality of space

Subhash Kak

We present an information-theoretic approach to the optimal representation of the intrinsic dimensionality of data and show it is a noninteger. Since optimality is accepted as a physical principle, this provides a theoretical explanation for why noninteger dimensions are useful in many branches of physics, where they have been introduced based on experimental considerations. Noninteger dimensions correlate with lesser density as in the Hausdorff dimension and this can have measurable effects. We use the lower density of noninteger dimension to resolve the problem of two different values of the Hubble constant obtained using different methods.

Noninteger dimensions are used in various branches of physics and engineering to explain the emergence of scale invariant phenomena^{1–3}. In quantum field theories, they are introduced through dimensional regularization^{4–7}; since Gaussian integrals can be written as a d -dimensional product of a single dimensional Gaussian integral, one can generalize them to noninteger dimensions. The continuous squeezing of a three-dimensional space into a two-dimensional space also generates intermediate scenarios with noninteger dimensions^{8–10}. They are introduced elsewhere to address the origins of self-similar structures in cosmology^{11–14}, where the exact dimensions are chosen to fit the experimental data.

In this paper, we approach dimensionality from an information theoretic perspective related to optimal representation of data. We show that the optimal dimension associated with the representation of information is $e = 2.71828\dots$, a noninteger number^{15,16}, and we derive it also from the perspective of the experimenter. Knowledge, as justified true belief, is normally seen as resulting from lucky guesses, but optimality provides a sounder basis to epistemology and, therefore, this can be the standard against which the inherent dimensionality of data may be compared.

In principle, the optimal dimension may be tested on data from a variety of areas. Here we consider the problem of the large-scale structure of the universe. Sky surveys and mappings of the various wavelength bands of electromagnetic radiation have yielded information on the content and character of this structure that appears to follow a hierarchical model with organization up to the scale of superclusters and filaments. A fractal dimension was computed based on the data up to a certain scale¹⁷. But such computations of fractal dimensions are yet not fully developed because of the incompleteness of the available data.

We apply the optimal dimensionality measure to a much simpler problem that side-steps the question of the actual dimension to be associated with the large-scale structure of the universe. We propose that physical data of expansion of the large-scale universe may be related to this optimal number, and so we should expect a discrepancy in the computations that implicitly assign the dimension of three to the data. In particular, we show that the discrepancy in the data of the Hubble constant obtained using two different methods^{18–20} can be resolved considering local- and large-scale perspectives on dimensionality of data.

Information and dimensions

The assumption that space is three-dimensional Euclidean is the idealistic position of mathematical philosophy according to which mathematical concepts have an independent reality. But as we know it is necessary to examine all underlying assumptions in formal systems to see if they square with our intuition, which, in itself, must be tested against experimental observations. Newton in his *Principia* defined motion of material objects with respect to absolute time and space, which were left undefined. Leibniz, on the other hand, proposed a different view in which events are more fundamental than time instants. According to Leibniz, the primary role is to be ascribed to matter and its properties and it is in relation to these properties that one can speak of time and space and not in any absolute sense. Time is simply succession of instantaneous configurations of matter and not something that flows independent of the bodies in the universe.

By considering the event as primary, we can measure its information as related to its frequency: the less likely the event, the information is greater. The logarithm measure associated with information has become widely accepted because it is additive. In other words, *additivity* of information is the underlying unstated assumption

Oklahoma State University, Stillwater, USA. email: subhash.kak@okstate.edu

Figure 1. Two points separated by distance r .

in the information-theoretic approach to structure and since information is related to the experimenter, this constitutes a subject-centered approach to reality. The entropy, or the average information associated with a source, is maximized when the events have the same probability.

Although it is troublesome to visualize noninteger number of dimensions, formal analysis of such systems has been done and several studies have investigated their topological properties²¹, and also questions of differentiability^{22–24}, and integration^{25,26}. Noninteger dimensions provide explanation for recursion and scale-invariance in complex systems—biological, physical, or engineered—and explain fractal behavior, examples of which include the Mandelbrot set, patterns in Romanesco broccoli, snowflakes, the Nautilus shell, complex computer networks, brain structure, as well as the filament structure and distribution of matter in cosmology.

The fractal dimension is generally the same as the box-counting dimension $\dim_{\text{box}}(S)$ of a set S in a Euclidean space $\mathbf{R}^n$. Suppose that $N(\epsilon)$ is the number of boxes of side length ϵ required to cover the set. Then the box-counting dimension is defined as:

$$\dim_{\text{box}}(S) = \lim_{\epsilon \rightarrow \infty} \frac{\ln N(\epsilon)}{\ln(\frac{1}{\epsilon})} \quad (1)$$

The computed dimension is a measure of the density of the set, and to that extent it appears to embody the intuition of dimension that is based on our sense of the nature of physical space.

The algorithms to generate mathematical examples of noninteger sets incorporate self-similarity. Thus, in the well-known one-dimensional Cantor set, in which at every step we leave out the middle-third^{1,2}, the total length of iteration n is $(2/3)^n$. We need to calculate the box-counting dimension using boxes (sticks) of length $\epsilon_n = 3^{-n}$ to cover the set of size $N_{\text{box}}(\epsilon_n) = 2^n$.

Since the iteration imply leaving out one part for every three, we get

$$d = \lim_{\epsilon \rightarrow 0} \left(\frac{\ln N_{\text{box}}(\epsilon)}{\ln(1/\epsilon)} \right) = \lim_{n \rightarrow \infty} \frac{\ln 2^n}{\ln 3^n} = \frac{\ln 2}{\ln 3} \approx 0.6309 \quad (2)$$

But note that the idea of noninteger dimensions should not be viewed only through the lens of algorithmically generated fractals and in spite of many studies on the topological properties of such spaces^{21–23}, its deeper implications for understanding physical phenomena are yet to be fully investigated. In particular, one may relate the process of recursive generation of the algorithmically generated fractal to fundamental properties of space.

To look at another perspective on noninteger space, assume an infinite space of dimension d . Let there be two objects separated by distance r . Building on the intuition that the density of objects in a noninteger space goes down with each iteration as in the Cantor set, the two objects of Fig. 1 would tend to come nearer, which constitutes a potential between them.

One interpretation of this d -dimensional space is that the objects at the two points A and B will have a potential equal to $U_{AB} = \frac{k(1-d)}{r}$, where k is a constant that depends on the nature of the two objects. When $d = 1$, that is the space is integer dimensional, there is zero potential and so there is no dynamics. The force associated with this potential will be $F_{AB} = \frac{dU}{dr} = \frac{-k(1-d)}{r^2}$, which constitutes an inverse-square law.

Now let us consider the question of information associated with the space. If space were d -dimensional, one could label the dimensions as 1, 2, 3, ..., d . The probability of the use of each of the d dimensions may, from considerations related to maximization of entropy, be taken to be the same and equal to $1/d$. Therefore, the information associated with each dimension is $\ln d$.

Clearly, this information increases as d increases. But this increase must be balanced against the cost of the use of the larger dimension set. Information efficiency per dimension is:

$$E(d) = \frac{\ln d}{d} \quad (3)$$

Its maximum value is obtained by taking the derivative of $E(d)$ and equating that to zero, which yields $d_{\text{optimum}} = e = 2.71828...$. In other words:

Theorem *The optimal number of dimensions associated with representation is e .*

The optimal base value for representation of numbers is $e = 2.718$, and the integer closest to it, namely 3, is more efficient than binary^{27,28}. The superiority of the efficiency of e -dimensions over two or three is evident in Fig. 2. The mathematical argument is that the e -dimensions are the most efficient representation of information and since Nature chooses optimality, as in physical laws such as the *Principle of Least Action*, the actual physical data should have this dimensionality.

From the perspective of visualization, the dimensionality is quite close to the fractal dimension associated with the Menger sponge^{29–31} which has a dimension of roughly 2.727. But this structure should not be viewed as anything more than a discrete analog of the physical noninteger dimensional system which is clear from the fact that different sets can have the same fractal dimension.

Figure 2. Efficiency of dimensions for $d=2$, e , 3, and 4.

To determine the fractal dimension of a random self-similar system, one needs to do a computation over the actual experimental data. The Cantor set is one-dimensional data, but one can see how a similar approach leads to three-dimensional structures.

Let us now consider this problem from the perspective of the experimenter. Let the system objects be labeled 1, 2, 3, ..., n , and the objects in the measuring instrument be labeled $n+1$, ..., $n+m$. The information associated with the system will be maximized when the objects are associated with equal probability; this means that the contents of the objects, which may be represented by nodes in a network, have equal significance. When the measurement nodes interact with the system, the information obtained by its nodes will be maximized when they are characterized quite the same way as the system nodes. This means that a *posteriori* probability of the nodes upon measurement goes down from $1/n$ to $1/(n+m)$.

The a priori information in each object is the logarithm of the inverse of its probability, or the information is $-\ln \frac{1}{n} = \ln n$.

The information in node k is $I(k) = \ln n$, and the information in the node k together with the measurement apparatus is $I(k+M) = \ln(n+m)$. The information obtained by the observer upon measurement is $I(k+M) - I(k) = \ln(n+m) - \ln n = \ln \frac{n+m}{n}$.

We can also examine it in terms of bins of potential information, and these bins are, ordinarily, the nodes on which the data resides. This data is transformed into information by virtue of the experimenter's interaction with the physical system. We call the ratio of the system nodes and the measurement nodes as variety $V(k) = \left(\frac{n+m}{n}\right)$, which reflects the result of the exploration of one system node in terms of the dimensionality of data.

It is reasonable to assume that the novelty from each node measurement is independent and therefore the individual values should be multiplied:

$$V_{data} = \left(\frac{n+m}{n}\right)^n$$

The true variety of the data is obtained when $n \rightarrow \infty$:

$$V = \lim_{n \rightarrow \infty} \left(1 + \frac{m}{n}\right)^n = e^m \quad (4)$$

Clearly, interrogating the system with greater number of measurement nodes will provide increased information. An intuitively satisfactory way to define dimensionality is to compute the infimum of the *variety*, V , that can be associated with all the object (or node)-states within the system.

Definition The dimension, D , of the data is the minimum possible value of V .

The dimension of all linear data will be one, because such data can be placed in a single bin, and that of data associated with a plane will be two. This idea of dimension derived from information considerations appears to be consistent with its intuitive meaning.

Considering $n=1$, $m=1$ the space has a dimension of 2, for it is associated with the pair of states associated with the system and the observer. With $n=2$, $D=2.25$, that indicate correlations between the two objects and the one observer. Beyond this the value builds up to 2.718...

$$D = \min_m \{e^m\} = e \quad (5)$$

This value of D may be interpreted in different ways: (i) as representing the statistical average of the fundamental relationships amongst the data, or (ii) the fundamental characteristics of the data space.

If we interpret D to represent the dimensions of the data space, then data obtained from astronomical problems should be associated with this dimension. In order to test this hypothesis, we consider the question of the measurement of the Hubble constant (H_0) because it deals with the physical distribution of objects in the universe.

Two different estimates of the Hubble constant

There are two conflicting values of the Hubble constant, based on whether one analyzes the cosmic microwave background (CMB) (the “early” universe estimate) or observes motions of stars and galaxies (the “late” universe estimate). According to the Planck Collaboration¹⁸ exploring the early universe, H_0 is about $67 \text{ km s}^{-1} \text{ Mpc}^{-1}$, whereas a late universe estimate, Supernova H_0 for the Equation of State (SHoES)¹⁹, has it at about $74 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

A succession of distance indicators, which constitutes the distance ladder, is used for determining distances to other stars and galaxies. Because the more distant steps of the cosmic distance ladder depend upon the nearer ones, they include the effects of both systematic and statistical errors in the nearer steps, and due to these propagating errors, the precision is necessarily poorer for the more distant objects. Fortunately, the significance of location errors is lessened in computation of rates of change as in the case of the Hubble constant.

Planck’s estimate of H_0 relies on measuring features in the CMB using Lambda-CDM, the standard model of cosmology. On the other hand, SHoES estimates the universe’s expansion rate by measuring distances to other galaxies from the supernovas’ distance and redshift, where brightness is related to luminosity by calibration against standard candles, which for SHoES are the Cepheid variables.

The independent Megamaser Cosmology Project²⁰ considers galaxies with disks of water-bearing molecular gas orbiting supermassive black holes at galaxy centers. Considering distance measurements to four galaxies, and combined with previous distance measurements of two other galaxies, the calculations produce a value for H_0 of $73.9 \text{ km s}^{-1} \text{ Mpc}^{-1}$ that is nearly identical to the SHoES estimate. The discrepancy between the early and late universe estimates of the Hubble constant has not yielded to any analysis and it has been termed a crisis in physics³².

We now present a resolution to the problem of the diverging estimates of the Hubble constant H_0 , based on early- or late- universe models. The discrepancy is between the values of $67 \text{ km s}^{-1} \text{ Mpc}^{-1}$ from the early universe and $74 \text{ km s}^{-1} \text{ Mpc}^{-1}$ from the late universe. This resolution comes from our information-theoretic perspective that implies that the optimal dimension to be associated with physical space is $e = 2.71828$. If physical reality is e -dimensional and we insist on viewing it as being 3-dimensional then there is a discrepancy equal to $\frac{e}{3} = 0.9060$. This number is very close to the divergence of $\frac{67}{74} = 0.9054$ from the experimental data.

The efficiency of the three-dimensional space is off from the e -dimensional optimal space by about 0.003, or about 0.6 percent¹⁶. One may speculate that since our cognitions are based on counting, we associate the nearest integer space of 3 dimensions to space which leads to our normal sense about its nature.

Conclusions

The paper presented arguments on the relationship between geometry of numbers and corresponding dimensionality. Using information theory, we showed that data from physical space has an optimal dimensionality of $e \approx 2.71828$; this was also considered from the perspective of the experimenter and we obtained the same figure when the size of the data set became boundless. This provides the mathematical basis for many instances of noninteger dimensionality that are encountered in physical and engineered systems.

As illustration of this theory, we applied this idea to the question of expansion rate of the universe. Specifically, we considered the problem of the discrepancy of the Hubble constant based on early and late universe models and showed that this discrepancy disappears if one recognizes that the early universe model provides the true e -dimensional estimate whereas the late universe model imposes a 3-dimensional gloss on our measurements, which arises from the mathematical models that are fitted into the actual measurements.

Received: 18 September 2020; Accepted: 10 November 2020

Published online: 26 November 2020

References

1. Mandelbrot, B. B. *The Fractal Geometry of Nature* (W. H. Freeman, New York, 1983).
2. Falconer, K. J. *Fractal Geometry: Mathematical Foundations and Applications* (Wiley, New York, 2003).
3. Zaslavsky, G. M. *Hamiltonian Chaos and Fractional Dynamics* (Oxford University Press, Oxford, 2005).
4. 't Hooft, G. & Veltman, M. Regularization and renormalization of gauge fields. *Nucl. Phys. B* **44**, 189–213 (1972).
5. Wilson, K. G. & Kogut, J. The renormalization group and the q expansion. *Phys. Rep.* **12**, 75–199 (1974).
6. Bollini, C. G. & Giambiagi, J. J. Dimensional renormalization: the number of dimensions as a regularizing parameter. *Nuovo Cimento B* **12**, 20–26 (1972).
7. 't Hooft, G. & Veltman, M. Scalar one-loop integrals. *Nucl. Phys. B* **153**, 365–401 (1979).
8. Armstrong, J. R. *et al.* Analytic solutions of topologically disjoint systems. *J. Phys. A* **48**, 085301 (2015).
9. Garrido, E., Jensen, A. S. & Alvarez-Rodriguez, R. Few-body quantum method in a d -dimensional space. *Phys. Lett. A* **383**, 2021 (2019).
10. Garrido, E. & Jensen, A. S. Confinement of two-body systems and calculations in d dimensions. *Phys. Rev. Res.* **1**, 023009 (2019).
11. Einasto, J., Htsi, G., Kuutma, T. & Einasto, M. Correlation function: biasing and fractal properties of the cosmic web. *Astron. Astrophys.* **640**, A47 (2020).
12. Tarasov, V. E. Anisotropic fractal media by vector calculus in non-integer dimensional space. *J. Math. Phys.* **55**, 083510 (2014).
13. Gurevich, A. V. & Zybin, K. P. Large-scale structure of the Universe: Analytic theory. *Phys. Usp.* **38**, 687–722 (1995).
14. Valdarnini, R., Borgani, S. & Provenzale, A. Multifractal properties of cosmological N -body simulations. *Astrophys. J.* **394**, 422–441 (1992).
15. Kak, S. The base- e representation of numbers and the power law. *Circuits Syst. Signal Process.* <https://doi.org/10.1007/s00034-020-01480-0> (2020).

16. Kak, S. Information, representation, and structure. *International Conference on Recent Trends in Mathematics and Its Applications to Graphs, Networks and Petri Nets*, New Delhi, India (2020); Preprint at https://www.techrxiv.org/articles/preprint/Information_Representation_and_Structure/12722549 (2020).
17. Joyce, M., Sylos Labini, F., Gabrielli, A., Montuori, M. & Pietronero, L. Basic properties of galaxy clustering in the light of recent results from the Sloan Digital Sky Survey. *Astron. Astrophys.* **443**, 11–16 (2005).
18. Aghanim, N. *et al.* Planck 2018 results. VI. Cosmological parameters. *Astro. Astrophys.* **641**, A6 (2020); preprint at <https://arxiv.org/abs/1807.06209> (2018).
19. Riess, A. G. *et al.* Large magellanic cloud cepheid standards provide a 1% foundation for the determination of the hubble constant and stronger evidence for physics beyond Lambda CDM. *Astrophys. J.* **876**, 1 (2019).
20. Pesce, D. W. *et al.* The Megamaser Cosmology Project. XIII. Combined hubble constant constraints. *Astrophys. J.* **891**(1), L1 (2020).
21. Stillinger, F. H. Axiomatic basis for spaces with noninteger dimensions. *J. Math. Phys.* **18**, 1224–1234 (1977).
22. Tarasov, V. E. Vector calculus in non-integer dimensional space and its applications to fractal media. *Commun. Nonlinear Sci. Numer. Simul.* **20**, 360–374 (2015).
23. Kilbas, A. A., Srivastava, H. M. & Trujillo, J. J. *Theory and Application of Fractional Differential Equations* (Elsevier, Amsterdam, 2006).
24. Podlubny, I. *Fractional Differential Equations* (Academic Press, San Diego, 1999).
25. Carpinteri, A. & Mainardi, F. (eds) *Fractals and Fractional Calculus in Continuum Mechanics* (Springer, Wien, 1997).
26. Sabatier, J. *et al.* (eds) *Advances in Fractional Calculus* (Springer, Dordrecht, 2007).
27. Hayes, B. Third base. *Am. Sci.* **89**, 490–494 (2001).
28. Kak, S. *On Ternary Coding and Three-Valued Logic*. <https://arxiv.org/abs/1807.06419> (2018).
29. Beck, C. & Schögl, F. *Thermodynamics of Chaotic Systems: An Introduction* (Cambridge University Press, Cambridge, 1995).
30. Bunde, A. & Havlin, S. *Fractals in Science* (Springer Science & Business Media, Berlin, 2013).
31. Menger, K. *Reminiscences of the Vienna Circle and the Mathematical Colloquium* (Springer Science & Business Media, Berlin, 2013).
32. Panek, R. A cosmic crisis. *Sci. Am.* **322**(3), 30–37 (2020).

Author contributions

The work is author's own research.

Competing interests

The author declares no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020