View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Dagstuhl Research Online Publication Server

Mutual Dimension*

Adam Case and Jack H. Lutz

Department of Computer Science, Iowa State University
Ames, IA 50011 USA

Abstract

We define the lower and upper *mutual dimensions* $mdim(x : y)$ and $Mdim(x : y)$ between any two points x and y in Euclidean space. Intuitively these are the lower and upper densities of the algorithmic information shared by x and y. We show that these quantities satisfy the main desiderata for a satisfactory measure of mutual algorithmic information. Our main theorem, the *data processing inequality* for mutual dimension, says that, if $f : \mathbb{R}^m \to \mathbb{R}^n$ is computable and Lipschitz, then the inequalities $mdim(f(x) : y) \leq mdim(x : y)$ and $Mdim(f(x) : y) \leq Mdim(x : y)$ hold for all $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^t$. We use this inequality and related inequalities that we prove in like fashion to establish conditions under which various classes of computable functions on Euclidean space preserve or otherwise transform mutual dimensions between points.

1998 ACM Subject Classification F. Theory of Computation

Keywords and phrases computable analysis, data processing inequality, effective fractal dimensions, Kolmogorov complexity, mutual information

Digital Object Identifier 10.4230/LIPIcs.STACS.2013.116

1 Introduction

Recent interactions among computability theory, algorithmic information theory, and geometric measure theory have assigned a *dimension* $dim(x)$ and a *strong dimension* $Dim(x)$ to each individual point x in a Euclidean space $\mathbb{R}^n$. These dimensions, which are real numbers satisfying $0 \leq dim(x) \leq Dim(x) \leq n$, have been shown to be geometrically meaningful. For example, the *classical* Hausdorff dimension $dim_H(E)$ of any set $E \subseteq \mathbb{R}^n$ that is a union of Π^0_1 (computably closed) sets is now known [16, 10] to admit the pointwise characterization

$$dim_H(E) = \sup_{x \in E} dim(x).$$

More recent investigations of the dimensions of individual points in Euclidean space have shed light on connectivity [18, 22], self-similar fractals [17, 6], rectifiability of curves [9, 20, 8], and Brownian motion [11].

In their original formulations [16, 1], $dim(x)$ is $cdim(\{x\})$ and $Dim(x)$ is $cDim(\{x\})$, where *cdim* and *cDim* are constructive versions of classical Hausdorff and packing dimensions [7], respectively. Accordingly, $dim(x)$ and $Dim(x)$ are also called *constructive fractal dimensions*. It is often most convenient to think of these dimensions in terms of the Kolmogorov complexity characterization theorems

$$dim(x) = \liminf_{r \to \infty} \frac{K_r(x)}{r}, \; Dim(x) = \limsup_{r \to \infty} \frac{K_r(x)}{r}, \tag{1.1}$$

* This research was supported in part by National Science Foundation Grants 0652569, 1143830, and 1247051. Part of the second author's work was done during a sabbatical at Caltech and the Isaac Newton Institute for Mathematical Sciences at the University of Cambridge.

© Adam Case and Jack H. Lutz;
licensed under Creative Commons License BY-ND

where $K_r(x)$, the Kolmogorov complexity of x at precision r, is defined later in this introduction [19, 1, 17]. These characterizations support the intuition that $dim(x)$ and $Dim(x)$ are the lower and upper *densities of algorithmic information* in the point x.

In this paper we move the pointwise theory of dimension forward in two ways. We formulate and investigate the *mutual dimensions* — intuitively, the lower and upper densities of shared algorithmic information — between two points in Euclidean space, and we investigate the *conservation* of dimensions and mutual dimensions by computable functions on Euclidean space. We expect this to contribute to both computable analysis — the theory of scientific computing [3] — and algorithmic information theory.

The analyses of many computational scenarios call for quantitative measures of the degree to which two objects are correlated. In classical (Shannon) information theory, the most useful such measure is the *mutual information* $I(X : Y)$ between two probability spaces X and Y [5]. In the algorithmic information theory of finite strings, the (*algorithmic*) *mutual information* $I(x : y)$ between two *individual* strings $x, y \in \{0,1\}^*$ plays an analogous role [15]. Under modest assumptions, if x and y are drawn from probability spaces X and Y of strings respectively, then the expected value of $I(x : y)$ is very close to $I(X : Y)$ [15]. In this sense algorithmic mutual information is a refinement of Shannon mutual information.

Our formulation of mutual dimensions in Euclidean space is based on the algorithmic mutual information $I(x : y)$, but we do not use the seemingly obvious approach of using the binary expansions of the real coordinates of points in Euclidean space. It has been known since Turing's famous correction [23] that binary notation is not a suitable representation for the arguments and values of computable functions on the reals. (See also [12, 24].) This is why the characterization theorems (1.1) use $K_r(x)$, the *Kolmogorov complexity* of a point $x \in \mathbb{R}^n$ at precision r, which is the minimum Kolmogorov complexity $K(q)$ — defined in a standard way [15] using a standard binary string representation of q — for all rational points $q \in \mathbb{Q}^n \cap B_{2^{-r}}(x)$, where $B_{2^{-r}}(x)$ is the open ball of radius 2^{-r} about x. For the same reason we base our development here on the *mutual information* $I_r(x : y)$ between points $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$ at *precision* r. This is the minimum value of the algorithmic mutual information $I(p : q)$ for all rational points $p \in \mathbb{Q}^m \cap B_{2^{-r}}(x)$ and $q \in \mathbb{Q}^n \cap B_{2^{-r}}(y)$. Intuitively, while there are infinitely many pairs of rational points in these balls and many of these pairs will contain a great deal of "spurious" mutual information (e.g., *any* finite message can be encoded into both elements of such a pair), a pair of rational points p and q achieving the minimum $I(p : q) = I_r(x : y)$ will only share information that their proximities to x and y force them to share. Sections 2 and 3 below develop the ideas that we have sketched in this paragraph, along with some elements of the fine-scale geometry of algorithmic information in Euclidean space that are needed for our results. A modest generalization of Levin's coding theorem (Theorem 2.1) is essential for this work.

In analogy with the characterizations (1.1) we define our mutual dimensions as the lower and upper densities of algorithmic mutual information,

$$mdim(x : y) = \liminf_{r\to\infty} \frac{I_r(x : y)}{r}, \; Mdim(x : y) = \limsup_{r\to\infty} \frac{I_r(x : y)}{r}, \tag{1.2}$$

in section 4. We also prove in that section that these quantities satisfy all but one of the desiderata (e.g., see [2]) for any satisfactory notion of mutual information.

We save the most important desideratum — our main theorem — for section 5. This is the data processing inequality for mutual dimension (actually two inequalities, one for $mdim$ and one for $Mdim$). The data processing inequality of Shannon information theory [5] says that, for any two probability spaces X and Y and any function $f : X \to Y$,

$$I(f(X) : Y) \le I(X : Y), \tag{1.3}$$

i.e., the induced probability space $f(X)$ obtained by "processing the information in X through f" does not share any more information with Y than X shares with Y. The data processing inequality of algorithmic information theory [15] says that, for any computable partial function $f : \{0,1\}^* \to \{0,1\}^*$, there is a constant $c_f \in \mathbb{N}$ (essentially the number of bits in a program that computes f) such that, for all strings $x \in \mathit{dom}\, f$ and $y \in \{0,1\}^*$,

$$I(f(x) : y) \leq I(x : y) + c_f. \tag{1.4}$$

That is, modulo the constant c_f, $f(x)$ contains no more information about y than x contains about y.

The data processing inequality for points in Euclidean space is a theorem about functions $f : \mathbb{R}^m \to \mathbb{R}^n$ that are computable in the sense of computable analysis [3, 12, 24]. Briefly, an *oracle* for a point $x \in \mathbb{R}^m$ is a function $g_x : \mathbb{N} \to \mathbb{Q}^m$ such that $|g_x(r) - x| \leq 2^{-r}$ holds for all $r \in \mathbb{N}$. A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is *computable* if there is an oracle Turing machine M such that, for every $x \in \mathbb{R}^m$ and every oracle g_x for x, the function $r \to M^{g_x}(r)$ is an oracle for $f(x)$.

Given (1.2), (1.3), and (1.4), it is natural to conjecture that, for every computable function $f : \mathbb{R}^m \to \mathbb{R}^n$, the inequalities

$$\mathit{mdim}(f(x) : y) \leq \mathit{mdim}(x : y),\ \mathit{Mdim}(f(x) : y) \leq \mathit{Mdim}(x : y) \tag{1.5}$$

hold for all $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^t$. However, this is not the case. For a simple example, there exist computable functions $f : \mathbb{R} \to \mathbb{R}^2$ that are *space-filling*, e.g., satisfy $[0,1]^2 \subseteq \mathit{range}\, f$ [4]. For such a function f we can choose $x \in \mathbb{R}$ such that $\mathit{dim}(f(x)) = 2$. Letting $y = f(x)$, we then have

$$\mathit{mdim}(f(x) : y) = \mathit{dim}(f(x)) = 2 > 1 \geq \mathit{Dim}(x) \geq \mathit{Mdim}(x : y),$$

whence both inequalities in (1.5) fail.

The difficulty here is that the above function f is extremely sensitive to its input, and this enables it to compress a great deal of "sparse" high-precision information about its input x into "dense" lower-precision information about its output $f(x)$. Many theorems of mathematical analysis exclude such excessively sensitive functions by assuming a given function f to be *Lipschitz*, meaning that there is a real number $c > 0$ such that, for all x and x', $|f(x) - f(x')| \leq c|x - x'|$. This turns out to be exactly what is needed here. In section 5 we prove prove the *data processing inequality* for mutual dimension (Theorem 5.1), which says that the conditions (1.5) hold for every function $f : \mathbb{R}^m \to \mathbb{R}^n$ that is computable and Lipschitz. In fact, we derive the data processing inequality from the more general *modulus processing lemma* (Lemma 5.2). This lemma yields quantitative variants of the data processing inequality for other classes of functions. For example, we use the modulus processing lemma to prove that, if $f : \mathbb{R}^m \to \mathbb{R}^n$ is *Hölder with exponent* α (meaning that $0 < \alpha \leq 1$ and there is a real number $c > 0$ such that $|f(x) - f(x')| \leq c|x - x'|^\alpha$ for all $x, x' \in \mathbb{R}^m$), then the inequalities

$$\mathit{mdim}(f(x) : y) \leq \frac{1}{\alpha}\mathit{mdim}(x : y),\ \mathit{Mdim}(f(x) : y) \leq \frac{1}{\alpha}\mathit{Mdim}(x : y) \tag{1.6}$$

hold for all $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^t$.

In section 5 we also derive *reverse* data processing inequalities, e.g., giving conditions under which $\mathit{mdim}(x : y) \leq \mathit{mdim}(f(x) : y)$. We then use data processing inequalities and their reverses to explore conditions under which computable functions on Euclidean space preserve, approximately preserve, or otherwise transform mutual dimensions between points.

We expect mutual dimensions and the data processing inequalities to be useful for future research in computable analysis. We also expect the development of mutual dimensions in Euclidean spaces — highly structured spaces in which it is clear that *mdim* and *Mdim* are the right notions — to guide future explorations of mutual information in more challenging contexts, including computational complexity and the computational theory of chemical reaction networks.

2 Kolmogorov Complexity in Euclidean Space

We begin by developing some elements of the fine-scale geometry of algorithm information in Euclidean space. In this context it is convenient to regard the Kolmogorov complexity of a set of strings to be the number of bits required to specify *some* element of the set.

Definition (Shen and Vereshchagin [21]). The *Kolmogorov complexity* of a set $S \subseteq \{0,1\}^*$ is

$$K(S) = \min\{K(x) \mid x \in S\}.$$

Note that $S \subseteq T$ implies $K(S) \geq K(T)$. Intuitively, small sets may require "higher resolution" than large sets.

Similarly, we define the *algorithmic probability* of a set $S \subseteq \{0,1\}^*$ to be

$$\mathbf{m}(S) = \sum_{\substack{\pi \in \{0,1\}^* \\ U(\pi) \in S}} 2^{-|\pi|},$$

where U is the fixed universal Turing machine used in defining Kolmogorov complexity. For a single string $x \in \{0,1\}^*$, we write $\mathbf{m}(x) = \mathbf{m}(\{x\})$.

We need a generalization of Levin's coding theorem [13, 14] that is applicable to certain systems of disjoint sets.

Notation. Let $B \subseteq \mathbb{N} \times \mathbb{N} \times \{0,1\}^*$ and $r, s \in \mathbb{N}$.
1. The (r,s)-*block* of B is the set $B_{r,s} = \{x \in \{0,1\}^* \mid (r,s,x) \in B\}$.
2. The r^{th} *layer* of B is the sequence $B_r = (B_{r,t} \mid t \in \mathbb{N})$.

Definition. A *layered disjoint system* (LDS) is a set $B \subseteq \mathbb{N} \times \mathbb{N} \times \{0,1\}^*$ such that, for all $r, s, t \in \mathbb{N}$,

$$s \neq t \Rightarrow B_{r,s} \cap B_{r,t} = \emptyset.$$

Note that this definition only requires the sets within each layer of B to be disjoint.

▶ **Theorem 2.1** (LDS coding theorem). *For every computably enumerable layered disjoint system B there is a constant $c_B \in \mathbb{N}$ such that, for all $r, t \in \mathbb{N}$,*

$$K(B_{r,t}) \leq \log \frac{1}{\mathbf{m}(B_{r,t})} + K(r) + c_B.$$

If we take $B_{r,t} = \{s_t\}$, where s_t is the t^{th} element of the standard enumeration of $\{0,1\}^*$, then Theorem 2.1 tells us that $K(x) \leq \log \frac{1}{\mathbf{m}(x)} + O(1)$, which is Levin's coding theorem.

Our next objective is to use the LDS coding theorem to obtain useful bounds on the number of times that the value $K(S)$ is attained or approximated.

Definition. Let $S \subseteq \{0,1\}^*$ and $d \in \mathbb{N}$.
1. A *d-approximate K-minimizer* of S is a string $x \in S$ for which $K(x) \le K(S) + d$.
2. A *K-minimizer* of S is a 0-approximate K-minimizer of S.

We use the LDS coding theorem to prove the following.

▶ **Theorem 2.2.** *For every computably enumerable layered disjoint system B there is a constant $c_B \in \mathbb{N}$ such that, for all $r, t, d \in \mathbb{N}$, the block $B_{r,t}$ has at most $2^{d+K(r)+c_B}$ d-approximate K-minimizers.*

We now lift our terminology and notation to Euclidean space $\mathbb{R}^n$. In this context, a layered disjoint system is a set $B \subseteq \mathbb{N} \times \mathbb{N} \times \mathbb{R}^n$ such that, for all $r, s, t \in \mathbb{N}$,

$$s \neq t \Rightarrow B_{r,s} \cap B_{r,t} = \emptyset.$$

We lift our Kolmogorov complexity notation and terminology to $\mathbb{R}^n$ in two steps:
1. Lifting to $\mathbb{Q}^n$: Each rational point $q \in \mathbb{Q}^n$ is encoded as a string $x \in \{0,1\}^*$ in a natural way. We then write $K(q)$ for $K(x)$. In this manner, $K(S)$, $\mathbf{m}(S)$, K-minimizers, and d-approximate K-minimizers are all defined for sets $S \subseteq \mathbb{Q}^n$.
2. Lifting to $\mathbb{R}^n$. For $S \subseteq \mathbb{R}^n$, we define $K(S) = K(S \cap \mathbb{Q}^n)$ and $\mathbf{m}(S) = \mathbf{m}(S \cap \mathbb{Q}^n)$. Similarly, a K-minimizer for S is a K-minimizer for $S \cap \mathbb{Q}^n$, etc.

For each $r \in \mathbb{N}$ and each $m = (m_1, \ldots, m_n) \in \mathbb{Z}^n$, let

$$Q_m^{(r)} = [m_1 \cdot 2^{-r},\ (m+1) \cdot 2^{-r}) \times \cdots \times [m_n \cdot 2^{-r},\ (m_n + 1) \cdot 2^{-r})$$

be the *r-dyadic cube* at m. Note that each $Q_m^{(r)}$ is "half-open, half-closed" in such a way that, for each $r \in \mathbb{N}$, the family

$$\mathcal{Q}^{(r)} = \{Q_m^{(r)} \mid m \in \mathbb{Z}^n\}$$

is a partition of $\mathbb{R}^n$. It follows that (modulo trivial encoding) the collection

$$\mathcal{Q} = \{Q_m^{(r)} \mid r \in \mathbb{N} \text{ and } m \in \mathbb{Z}^n\}$$

of all *dyadic cubes* is a layered disjoint system whose rth layer is $\mathcal{Q}^{(r)}$. Moreover, the set

$$\{(r, m, q) \in \mathbb{N} \times \mathbb{Z}^n \times \mathbb{Q}^n \mid q \in Q_m^{(r)}\}$$

is decidable, so Theorem 2.2 has the following useful consequence.

▶ **Corollary 2.3.** *There is a constant $c \in \mathbb{N}$ such that, for all $r, d \in \mathbb{N}$, no r-dyadic cube has more than $2^{d+K(r)+c}$ d-approximate K-minimizers. In particular, no r-dyadic cube has more than $2^{K(r)+c}$ K-minimizers.*

The Kolmogorov complexity of an arbitrary point in Euclidean space depends on both the point and a precision parameter.

Definition. Let $x \in \mathbb{R}^n$ and $r \in \mathbb{N}$. The *Kolmogorov complexity* of x at precision r is

$$K_r(x) = K(B_{2^{-r}}(x)).$$

That is, $K_r(x)$ is the number of bits required to specify *some* rational point in the open ball $B_{2^{-r}}(x)$. Note that, for each $q \in \mathbb{Q}^n$, $K_r(q) \nearrow K(q)$ as $r \to \infty$.

A careful analysis of the relationship between cubes and balls enables us to derive the following from Corollary 2.3.

▶ **Theorem 2.4.** *There is a constant $c \in \mathbb{N}$ such that, for all $r, d \in \mathbb{N}$, no open ball of radius 2^{-r} has more than $2^{d+2K(r)+c}$ d-approximate K-minimizers. In particular, no open ball of radius 2^{-r} has more than $2^{2K(r)+c}$ K-minimizers.*

3 Mutual Information in Euclidean Space

This section develops the mutual dimensions of points in Euclidean space at a given precision. As in section 2 we assume that rational points $q \in \mathbb{Q}^n$ are encoded as binary strings in some natural way. Mutual information between rational points is then defined from conditional Kolmogorov complexity in the standard way [15] as follows.

Definition. Let $p \in \mathbb{Q}^m$, $r \in \mathbb{Q}^n$, $s \in \mathbb{Q}^t$.

1. The *mutual information* between p and q is

$$I(p : q) = K(q) - K(q \,|\, p).$$

2. The *mutual information* between p and q *given* s is

$$I(p : q|s) = K(q \,|\, s) - K(q \,|\, p, s).$$

The following properties of mutual information are well known [15].

▶ **Theorem 3.1.** *Let $p \in \mathbb{Q}^m$ and $q \in \mathbb{Q}^n$.*

1. $I(p, K(p) : q) = K(p) + K(q) - K(p, q) + O(1)$.
2. $I(p, K(p) : q) = I(q, K(q) : p) + O(1)$.
3. $I(p : q) \leq \min\{K(p), K(q)\} + O(1)$.

(Each of the properties 1 and 2 above is sometimes called *symmetry of mutual information.*)

Mutual information between points in Euclidean space at a given precision is now defined as follows.

Definition. The *mutual information* of $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^t$ at *precision* $r \in \mathbb{N}$ is

$$I_r(x : y) = \min\{I(q_x : q_y) \,|\, q_x \in B_{2^{-r}}(x) \cap \mathbb{Q}^n \text{ and } q_y \in B_{2^{-r}}(y) \cap \mathbb{Q}^t\}.$$

As noted in the introduction, the role of the minimum in the above definition is to eliminate "spurious" information that points $q_x \in B_{2^{-r}} \cap \mathbb{Q}^n$ and $q_y \in B_{2^{-r}}(y) \cap \mathbb{Q}^t$ might share for reasons not forced by their proximities to x and y, respectively.

Notation. We also use the quantity

$$J_r(x : y) = \min\{I(q_x : q_y) \,|\, p_x \text{ is a K-minimizer of } B_{2^{-r}}(x) \text{ and } p_y \text{ is a K–minimizer of } B_{2^{-r}}(y)\}.$$

Although $J_r(x : y)$, having two "layers of minimization", is somewhat more involved than $I_r(x : y)$, one can imagine using it as the definition of mutual information. Using Theorem 2.4 (and hence the LDS coding theorem), we prove the useful fact that $J_r(x : y)$ does not differ greatly from $I_r(x : y)$.

▶ **Theorem 3.2.** *For all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^t$,*

$$I_r(x : y) = J_r(x : y) + o(r)$$

as $r \to \infty$.

Using this we establish the following useful properties of $I_r(x : y)$.

▶ **Theorem 3.3.** *For all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^t$, the following hold as $r \to \infty$.*

1. $I_r(x : y) = K_r(x) + K_r(y) - K_r(x, y) + o(r)$.
2. $I_r(x : y) \leq \min\{K_r(x), K_r(y)\} + o(r)$.
3. *If x and y are independently random, then $I_r(x : y) = o(r)$.*
4. $I_r(x : y) = I_r(y : x) + o(r)$.

4 Mutual Dimension in Euclidean Space

We now define mutual dimensions between points in Euclidean space(s).

Definition. The *lower* and *upper mutual dimensions* between $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^t$ are

$$mdim(x : y) = \liminf_{r\to\infty} \frac{I_r(x : y)}{r}$$

and

$$Mdim(x : y) = \limsup_{r\to\infty} \frac{I_r(x : y)}{r},$$

respectively.

With the exception of the data processing inequality, which we prove in section 5, the following theorem says that the mutual dimensions $mdim$ and $Mdim$ have the basic properties that any mutual information measure should have. (See, for example, [2].)

▶ **Theorem 4.1.** *For all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^t$, the following hold.*
1. $mdim(x : y) \geq dim(x) + dim(y) - Dim(x, y)$.
2. $Mdim(x : y) \leq Dim(x) + Dim(y) - dim(x, y)$.
3. $mdim(x : y) \leq \min\{dim(x), dim(y)\}$, $Mdim(x : y) \leq \min\{Dim(x), Dim(y)\}$.
4. $0 \leq mdim(x : y) \leq Mdim(x : y) \leq \min\{n, t\}$.
5. *If x and y are independently random, then $Mdim(x : y) = 0$.*
6. $mdim(x : y) = mdim(y : x)$, $Mdim(x : y) = Mdim(y : x)$.

5 Data Processing Inequalities and Applications

Our objectives in this section are to prove data processing inequalities for lower and upper mutual dimensions in Euclidean space and to use these inequalities to investigate how certain functions on Euclidean space preserve or predictably transform mutual dimensions.

The following result is our main theorem. The meaning and necessity of the Lipschitz hypothesis are explained in the introduction.

▶ **Theorem 5.1** (data processing inequality). *If $f : \mathbb{R}^n \to \mathbb{R}^t$ is computable and Lipschitz, then, for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^t$,*

$$mdim(f(x) : y) \leq mdim(x : y)$$

and

$$Mdim(f(x) : y) \leq Mdim(x : y).$$

We in fact prove a stronger result.

Definition. A *modulus (of uniform continuity)* for a function $f : \mathbb{R}^n \to \mathbb{R}^k$ is a nondecreasing function $m : \mathbb{N} \to \mathbb{N}$ such that, for all $x, y \in \mathbb{R}^n$ and $r \in \mathbb{N}$,

$$|x - y| \leq 2^{-m(r)} \Rightarrow |f(x) - f(y)| \leq 2^{-r}.$$

▶ **Lemma 5.2** (modulus processing lemma). *If $f : \mathbb{R}^n \to \mathbb{R}^k$ is a computable and uniformly continuous function, and m is a computable modulus for f, then for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^t$,*

$$mdim(f(x) : y) \leq mdim(x : y)\left(\limsup_{r\to\infty} \frac{m(r+1)}{r}\right)$$

and

$$Mdim(f(x) : y) \leq Mdim(x : y)\left(\limsup_{r\to\infty} \frac{m(r+1)}{r}\right).$$

Theorem 5.1 follows immediately from Lemma 5.2 and the following.

▶ **Observation 5.3.** *If a function $f : \mathbb{R}^n \to \mathbb{R}^k$ is Lipschitz, then there exists $s \in \mathbb{N}$ such that $m(r) = r + s$ is a modulus for f.*

In similar fashion, we can derive the following fact from the modulus processing lemma. (Recall the definition of Hölder functions given in the introduction.)

▶ **Corollary 5.4.** *If $f : \mathbb{R}^n \to \mathbb{R}^k$ is computable and Hölder with exponent α, then, for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^t$,*

$$mdim(f(x) : y) \leq \frac{1}{\alpha} mdim(x : y)$$

and

$$Mdim(f(x) : y) \leq \frac{1}{\alpha} Mdim(x : y).$$

We next develop reverse versions of the above inequalities.

Notation. Let $n \in \mathbb{Z}^+$.

1. $[n] = \{1, \cdots, n\}$.
2. For $S \subseteq [n]$, $x \in \mathbb{R}^{|S|}$, $y \in \mathbb{R}^{n-|S|}$, the string

   $$x *_S y \in \mathbb{R}^n$$

   is obtained by placing the components of x into the positions in S (in order) and the components of y into the positions in $[n] \setminus S$ (in order).

Definition. Let $f : \mathbb{R}^n \to \mathbb{R}^k$.

1. f is *co-Lipschitz* if there is a real number $c > 0$ such that for all $x, y \in \mathbb{R}^n$,

   $$|f(x) - f(y)| \geq c|x - y|.$$

2. f is *bi-Lipschitz* if f is both Lipschitz and co-Lipschitz.
3. For $S \subseteq [n]$, f is *S-co-Lipschitz* if there is a real number $c > 0$ such that, for all $u, v \in \mathbb{R}^{|S|}$ and $y \in \mathbb{R}^{n-|S|}$,

   $$|f(u *_S y) - f(v *_S y)| \geq c|u - v|.$$

4. For $i \in [n]$, f is *co-Lipschitz in its i^{th} argument* if f is $\{i\}$-co-Lipschitz.

Note that f is $[n]$-co-Lipschitz if and only if f is co-Lipschitz.

Example. The function $f : \mathbb{R}^n \to \mathbb{R}$ defined by

$$f(x_1, \cdots, x_n) = x_1 + \cdots + x_n$$

is S-co-Lipschitz if and only if $|S| \leq 1$. In particular, if $n \geq 2$, then f is co-Lipschitz in every argument, but f is not co-Lipschitz.

We next relate co-Lipschitz conditions to moduli.

Definition. Let $f : \mathbb{R}^n \to \mathbb{R}^k$.

1. An *inverse modulus* for f is a nondecreasing function $m : \mathbb{N} \to \mathbb{N}$ such that, for all $x, y \in \mathbb{R}^n$ and $r \in \mathbb{N}$,
$$|f(x) - f(y)| \leq 2^{-m(r)} \Rightarrow |x - y| \leq 2^{-r}.$$
2. Let $S \subseteq [n]$. An *S-inverse modulus* for f is a nondecreasing function $m : \mathbb{N} \to \mathbb{N}$ such that, for all $u, v \in \mathbb{R}^{|S|}$, all $y \in \mathbb{R}^{n-|S|}$, and all $r \in \mathbb{N}$,
$$|f(u *_S y) - f(v *_S y)| \leq 2^{-m(r)} \Rightarrow |u - v| \leq 2^{-r}.$$
3. Let $i \in [n]$. An *inverse modulus* for f *in its i^{th} argument* is an $\{i\}$-inverse modulus for f.

▶ **Observation 5.5.** *Let $f : \mathbb{R}^n \to \mathbb{R}^k$ and $S \subseteq [n]$.*

1. *If f is S-co-Lipschitz, then there is a positive constant $t \in \mathbb{N}$ such that $m'(r) = r + t$ is an S-inverse modulus of f.*
2. *If f is co-Lipschitz, then there is a positive constant $t \in \mathbb{N}$ such that $m'(r) = r + t$ is an inverse modulus of f.*

▶ **Lemma 5.6** (reverse modulus processing lemma). *If $f : \mathbb{R}^n \to \mathbb{R}^k$ is a uniformly continuous function and m' is a computable S-inverse modulus for f, then, for all $S \subseteq [n]$, $x \in \mathbb{R}^{|S|}$, $y \in \mathbb{R}^t$, and $z \in \mathbb{R}^{n-|S|}$,*
$$mdim(x : y) \leq mdim((f(x *_S z), z) : y)\left(\limsup_{r\to\infty} \frac{m'(r+1)}{r}\right)$$
and
$$Mdim(x : y) \leq Mdim((f(x *_S z), z) : y)\left(\limsup_{r\to\infty} \frac{m'(r+1)}{r}\right).$$

By Observation 5.5 and Lemma 5.6, we have the following.

▶ **Theorem 5.7** (reverse data processing inequality). *If $S \subseteq [n]$ and $f : \mathbb{R}^n \to \mathbb{R}^k$ is computable and S-co-Lipschitz, then, for all $x \in \mathbb{R}^{|S|}$, $y \in \mathbb{R}^t$, and $z \in \mathbb{R}^{n-|S|}$,*
$$mdim(x : y) \leq mdim((f(x *_S z), z) : y)$$
and
$$Mdim(x : y) \leq Mdim((f(x *_S z), z) : y).$$

Definition. Let $f : \mathbb{R}^n \to \mathbb{R}^k$ and $0 < \alpha \leq 1$.

1. f is *co-Hölder with exponent α* if there is a real number $c > 0$ such that for all $x, y \in \mathbb{R}^n$,
$$|x - y| \leq c|f(x) - f(y)|^\alpha.$$
2. For $S \subseteq [n]$, f is *S-co-Hölder with exponent α* if there is a real number $c > 0$ such that, for all $u, v \in \mathbb{R}^{|S|}$ and $y \in \mathbb{R}^{n-|S|}$,
$$|u - v| \leq c|f(u *_S y) - f(v *_S y)|^\alpha.$$

▶ **Corollary 5.8.** *If $S \subseteq [n]$ and $f : \mathbb{R}^n \to \mathbb{R}^k$ is computable and S-co-Hölder with exponent α, then, for all $x \in \mathbb{R}^{|S|}$, $y \in \mathbb{R}^t$, and $z \in \mathbb{R}^{n-|S|}$,*
$$mdim(x : y) \leq \frac{1}{\alpha} mdim((f(x *_S z), z) : y)$$
and
$$Mdim(x : y) \leq \frac{1}{\alpha} Mdim((f(x *_S z), z) : y).$$

The rest of this section is devoted to applications of the data processing inequalities and their reverses.

▶ **Theorem 5.9** (mutual dimension conservation inequality). *If $f : \mathbb{R}^n \to \mathbb{R}^k$ and $g : \mathbb{R}^t \to \mathbb{R}^l$ are computable and Lipschitz, then, for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^t$,*

$$mdim(f(x) : g(y)) \leq mdim(x : y)$$

and

$$Mdim(f(x) : g(y)) \leq Mdim(x : y).$$

▶ **Theorem 5.10** (reverse mutual dimension conservation inequality). *Let $S_1 \subseteq [n]$ and $S_2 \subseteq [t]$. If $f : \mathbb{R}^n \to \mathbb{R}^k$ is computable and S_1-co-Lipschitz, and $g : \mathbb{R}^t \to \mathbb{R}^l$ is computable and S_2-co-Lipschitz, then, for all $x \in \mathbb{R}^{|S_1|}$, $y \in \mathbb{R}^{|S_2|}$, $w \in \mathbb{R}^{n-|S_1|}$, and $z \in \mathbb{R}^{t-|S_2|}$,*

$$mdim(x : y) \leq mdim((f(x *_S w), w) : (g(y *_S z), z))$$

and

$$Mdim(x : y) \leq Mdim((f(x *_S w), w) : (g(y *_S z), z))$$

▶ **Corollary 5.11** (preservation of mutual dimension). *If $f : \mathbb{R}^n \to \mathbb{R}^k$ and $g : \mathbb{R}^t \to \mathbb{R}^l$ are computable and bi-Lipschitz, then, for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^t$,*

$$mdim(f(x) : g(y)) = mdim(x : y)$$

and

$$Mdim(f(x) : g(y)) = Mdim(x : y).$$

▶ **Corollary 5.12.** *If $f : \mathbb{R}^n \to \mathbb{R}^k$ and $g : \mathbb{R}^t \to \mathbb{R}^l$ are computable and Hölder with exponents α and β, respectively, then, for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^t$,*

$$mdim(f(x) : g(y)) \leq \frac{1}{\alpha\beta} mdim(x : y)$$

and

$$Mdim(f(x) : g(y)) \leq \frac{1}{\alpha\beta} Mdim(x : y).$$

▶ **Corollary 5.13.** *Let $S_1 \subseteq [n]$ and $S_2 \subseteq [t]$. If $f : \mathbb{R}^n \to \mathbb{R}^k$ is computable and S_1-co-Hölder with exponent α, and $g : \mathbb{R}^t \to \mathbb{R}^l$ is computable and S_2-co-Hölder with exponent β, then, for all $x \in \mathbb{R}^{|S_1|}$, $y \in \mathbb{R}^{|S_2|}$, $w \in \mathbb{R}^{n-|S_1|}$, and $z \in \mathbb{R}^{t-|S_2|}$,*

$$mdim(x : y) \leq \frac{1}{\alpha\beta} mdim((f(x *_S w), w) : (g(y *_S z), z))$$

and

$$Mdim(x : y) \leq \frac{1}{\alpha\beta} Mdim((f(x *_S w), w) : (g(y *_S z), z)).$$

Acknowledgments. We thank Elvira Mayordomo, A. Pavan, Giora Slutzki, and Jim Lathrop for useful discussions.

References

1 Krishna B. Athreya, John M. Hitchcock, Jack H. Lutz, and Elvira Mayordomo. Effective strong dimension in algorithmic information and computational complexity. *SIAM Journal of Computing*, 37(3):671–705, 2007.

2 C.B. Bell. Mutual information and maximal correlation as measures of dependence. *Annals of Mathematical Statistics*, 33(2):587–595, 1962.

3 Mark Braverman and Stephen Cook. Computing over the reals: foundations for scientific computing. *Notices of the American Mathematical Society*, 53(3):318–329, 2006.

4 P. J. Couch, B. D. Daniel, and Timothy H. McNicholl. Computing space-filling curves. *Theory of Computing Systems*, 50(2):370–386, 2012.

5 Thomas R. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., second edition, 2006.

6 Randall Dougherty, Jack H. Lutz, Daniel R. Mauldin, and Jason Teutsch. Translating the Cantor set by a random real. *Transactions of the American Mathematical Society*, to appear.

7 Kenneth Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. Wiley, second edition, 2003.

8 Xiaoyang Gu, Jack Lutz, and Elvira Mayordomo. Curves that must be retraced. *Information and Computation*, 209(6):992–1006, 2011.

9 Xiaoyang Gu, Jack H. Lutz, and Elvira Mayordomo. Points on computable curves. In *Foundations of Computer Science*, pages 469–474. IEEE Computer Society, 2006.

10 John M. Hitchcock. Correspondence principles for effective dimensions. *Theory of Computing Systems*, 38(5):559–571, 2005.

11 Bjørn Kjos-Hanssen and Anil Nerode. Effective dimension of points visited by Brownian motion. *Theoretical Computer Science*, 410(4-5):347–354, 2009.

12 Ker-I Ko. *Complexity Theory of Real Functions*. Birkhäuser, first edition, 1991.

13 Leonid A. Levin. On the notion of a random sequence. *Soviet Mathematics Doklady*, 14(5):1413–1416, 1973.

14 Leonid A. Levin. Laws of information conservation (nongrowth) and aspects of the foundation of probability theory. *Problemy Peredachi Informatsii*, 10(3):30–35, 1974.

15 Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, third edition, 2008.

16 Jack H. Lutz. The dimension of individual strings and sequences. *Information and Computation*, 187(1):49–79, 2003.

17 Jack H. Lutz and Elvira Mayordomo. Dimensions of points in self-similar fractals. *SIAM Journal on Computing*, 38(3):1080–1112, 2008.

18 Jack H. Lutz and Klaus Weihrauch. Connectivity properties of dimension level sets. *Mathematical Logic Quarterly*, 54(5):483–491, 2008.

19 Elvira Mayordomo. A Kolmogorov complexity characterization of constructive Hausdorff dimension. *Information Processing Letters*, 84(1):1–3, 2002.

20 Robert Rettinger and Xizhong Zheng. Points on computable curves of computable lengths. In *MFCS*, pages 736–743. Springer, 2009.

21 Alexander Shen and Nikolai K. Vereshchagin. Logical operations and Kolmogorov complexity. *Theoretical Computer Science*, 271(1-2):125–129, 2002.

22 Daniel Turetsky. Connectedness properties of dimension level sets. *Theoretical Computer Science*, 412(29):3598–3603, 2011.

23 Alan M. Turing. On computable numbers, with an application to the Entscheidungsproblem. A correction. *Proceedings of the London Mathematical Society*, 43(2):544–546, 1937.

24 Klaus Weihrauch. *Computable Analysis: An Introduction*. Springer, first edition, 2000.