Research Article **Open Access**

Allison Byars, Evan Camrud, Steven N. Harding, Sarah McCarty, Keith Sullivan, and Eric S. Weber*

Sampling and Interpolation of Cumulative Distribution Functions of Cantor Sets in $[0,1]$

DOI: DOI, Received ..; revised ..; accepted ..

Abstract: Cantor sets are constructed from iteratively removing sections of intervals. This process yields a cumulative distribution function (CDF), constructed from the invariant Borel probability measure associated with their iterated function systems. Under appropriate assumptions, we identify sampling schemes of such CDFs, meaning that the underlying Cantor set can be reconstructed from sufficiently many samples of its CDF. To this end, we prove that two Cantor sets have almost-nowhere intersection with respect to their corresponding invariant measures.

Keywords: Fractal, Cantor Set, Sampling, Interpolation, Normal Numbers

MSC: Primary: 94A20, 28A80; Secondary: 26A30, 11K16, 11K55

1 Introduction

A Cantor set is the result of an infinite process of removing sections of an interval—$[0,1]$ in this paper—in an iterative fashion. The set itself consists of the points remaining after the removal of intervals specified by two parameters: the scale factor N and digit set D. The positive integer N determines how many equal intervals each extant segment is divided into per iteration, while $D \subset \{0, ..., N-1\}$ enumerates which of the N intervals of the segments will be preserved in each iteration. Equivalently, a Cantor set is the subset of $[0,1]$ consisting of numbers whose base-N expansion uses only digits from D. Yet another description of Cantor sets is given by the invariant set for an iterated function system, which will be our view in this paper.

Each Cantor set yields a Cumulative Distribution Function (CDF), which we define formally in Definition 1.2. We denote the class of all such CDFs by $\mathscr{F}$. We consider the problems of sampling and interpolation of functions in $\mathscr{F}$. By sampling, we mean the reconstruction of an unknown function $F \in \mathscr{F}$ from its samples $\{F(x_i)\}_{i \in I}$ at known points $\{x_i\}_{i \in I}$ in its domain (for an introduction to sampling theory, see [1, 2]). By interpolation, we mean the construction of a function $F \in \mathscr{F}$ that satisfies the constraints $F(x_i) = y_i$ for *a priori* given data $\{(x_i, y_i)\}_{i \in I}$. Note that the premise of the sampling problem is that there is a unique

Allison Byars: Department of Mathematics, University of Wisconsin Madison, 480 Lincoln Drive, Madison, WI 53706; E-mail: abyars@wisc.edu

Evan Camrud: Department of Mathematics, Iowa State University, 411 Morrill Road, Ames, IA 50011; E-mail: ecamrud@iastate.edu

Steven N. Harding: Department of Mathematics, Milwaukee School of Engineering, 500 E. Kilbourn Ave., Milwaukee, WI 53202; E-mail: harding@msoe.edu

Sarah McCarty: Department of Mathematics, Iowa State University, 411 Morrill Road, Ames, IA 50011; E-mail: smmcc@iastate.edu

Keith Sullivan: Department of Mathematics, Concordia College, 901 8th St. S. Moorhead, MN 56562; E-mail: ksulliv1@cord.edu

***Corresponding Author: Eric S. Weber:** Department of Mathematics, Iowa State University, 411 Morrill Road, Ames, IA 50011; E-mail: esweber@iastate.edu

$F \in \mathscr{F}$ that satisfies the available data, whereas the interpolation problem may not have the uniqueness property. Depending on the context, I can be either finite or infinite.

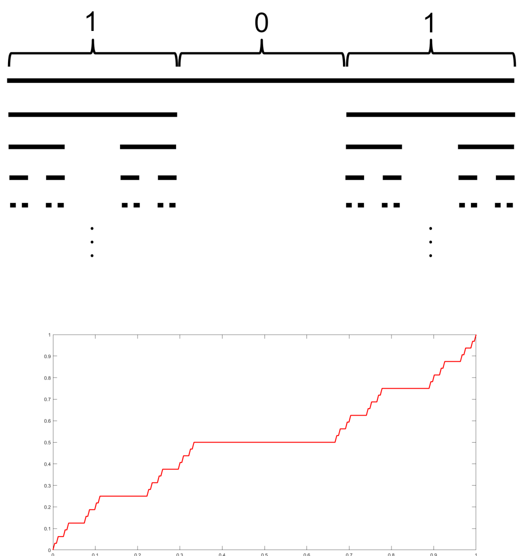

Figure 1. $C_{(1,0,1)}$ and $F_{(1,0,1)}$

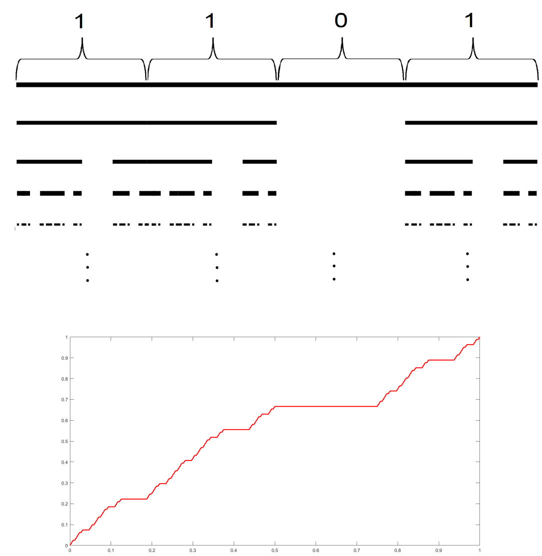

Figure 2. $C_{(1,1,0,1)}$ and $F_{(1,1,0,1)}$

To be more precise regarding sampling CDFs, we formulate the problem as follows: Fix $\mathscr{G} \subset \mathscr{F}$. For which sets of sampling points $\{x_i\}_{i \in I}$ does the following implication hold:

$$F, G \in \mathscr{G} \text{ and } F(x_i) = G(x_i) \ \forall i \in I \Rightarrow F = G \tag{1}$$

In the case where (1) holds, we call $\{x_i\}$ a *set of uniqueness* for $\mathscr{G}$.

Our main results in the paper concerning sampling include the following. In Theorem 2.5 we prove that if $\mathscr{G}$ consists of all CDFs for Cantor sets with unknown scale factor N, but the scale factor is known to be bounded by K, then there exists a set of uniqueness of size $O(K^3)$. We show that when the scale factor N is known, there exists a set of uniqueness of size $N-1$ that satisfies the implication in (1). We conjecture that there is a minimal set of uniqueness of size $\lfloor \frac{N}{2} \rfloor$, and prove that the minimal set of uniqueness cannot be smaller in Proposition 2.5. We also provide evidence of our conjecture by considering a conditional sampling procedure (meaning that the sampling points are data dependent) that can uniquely identify the CDF from $\lfloor \frac{N}{2} \rfloor$ samples in Theorem 2.2. Additionally, in Section 2.2, we include an interpolation procedure as an imperfect reconstruction of a CDF from samples, and provide an upper bound on the error that the reconstruction via interpolation could give.

1.1 Cantor Sets and Their Cumulative Distribution Functions

There are many ways to construct Cantor sets, and consequently many ways to denote a Cantor set. The Cantor sets we consider in this paper are those that corresponding to restricted digit sets. Thus, our set is defined by a choice of base (or scale) N and digits $D \subset \{0, \dots, N-1\}$ which are allowed. The Cantor set determined by such an N and D is denoted by $C_{N,D}$.

Another notation to describe N, D is to consider a vector $\overrightarrow{B} = (b_0, b_1, ..., b_{N-1}) \in (\mathbb{Z}_2)^N$, where $b_i = 1$ if $i \in D$ and $b_i = 0$ if $i \notin D$. We will also write $\overrightarrow{B}(i) := b_i$. Further, $\|\overrightarrow{B}\| := \sum_{i=0}^{N-1} b_i = |D|$. $\overrightarrow{B}$ is referred to as the binary digit vector, and we denote the Cantor set determined by $\overrightarrow{B}$ as $C_{\overrightarrow{B}}$. In this sense, both $C_{N,D}$ and $C_{\overrightarrow{B}}$ can be used to describe a Cantor set, and we naturally associate N, D with its corresponding $\overrightarrow{B}$. Note that in this work, all indexing will start with zero, such that b_0 the *first* entry of the vector $\overrightarrow{B}$.

In addition, special cases exist in which a Cantor set will be considered degenerate. In particular, $C_{\overrightarrow{B}}$ is not considered when the set is empty, a one-point set, or $[0,1]$. Under this definition, there does not exist a Cantor set with $N<3$ or $\|\overrightarrow{B}\|$ equal to 0, 1, or N. For an example of a legitimate Cantor set, $C_{(1,0,1)}$ is the well-known ternary Cantor set (Figure 1). We also provide an illustration of the iterative construction of the Cantor set corresponding to $\overrightarrow{B}=(1,1,0,1)$ (Figure 2).

Another description of the Cantor sets we consider is as the invariant set for an (affine) iterated function system (IFS).

Definition 1.1 (Iterated Function System)**.** *In general, an IFS is a collection of continuous contraction maps $\{\phi_d\}_{d\in D}$ on a complete metric space. Then, the invariant set C is the non-empty subset (if one exists) of the metric space satisfying*

$$C=\bigcup_{d\in D}\phi_d(C).$$

Our Cantor sets $C_{N,D}$ are invariant sets for an IFS in the following way. Let N be the scale factor, and let D be the digit set. For our purpose, we consider the particular IFS $\{\phi_d\}_{d\in D}$ on $\mathbb{R}$ where $\phi_d(x)=\frac{x+d}{N}$ for each $d\in D$. We allow ϕ_D to act on $[0,1]$, so the invariant set is a subset of $[0,1]$. Moreover, there exists an invariant measure on the invariant set $C_{N,D}$.

Theorem 1.1 (Hutchinson, [3])**.** *Let $\overrightarrow{B}$ be a binary digit vector. There exists a* unique *Borel probability measure $\mu_{\overrightarrow{B}}$ on $[0,1]$ such that $\mu_{\overrightarrow{B}}(C_{\overrightarrow{B}})=1$, $\mu_{\overrightarrow{B}}([0,1]\backslash C_{\overrightarrow{B}})=0$ and $\mu_{\overrightarrow{B}}=\frac{1}{\|\overrightarrow{B}\|}\sum_{d\in D}\mu_{\overrightarrow{B}}\circ\phi_d^{-1}$. That is, $\mu_{\overrightarrow{B}}$ is invariant under the iterated function system.*

Definition 1.2 (Cumulative Distribution Function)**.** *Let $\overrightarrow{B}$ be a binary digit vector. The Cantor set associated to $\overrightarrow{B}$ has a unique cumulative distribution function (CDF) $F:[0,1]\to[0,1]$ given by*

$$F(x)=\mu_{\overrightarrow{B}}([0,x])=\int_0^x d\mu_{\overrightarrow{B}}.$$

The CDF of $C_{\overrightarrow{B}}$ is denoted $F_{\overrightarrow{B}}$.

Note that the CDF of any of our Cantor sets is continuous. When convenient, we will extend $F_{\overrightarrow{B}}$ to all of $\mathbb{R}$ by $F_{\overrightarrow{B}}(x)=0$ if $x<0$ and 1 if $x>1$. It turns out that the invariant measure $\mu_{\overrightarrow{B}}$ is actually the pullback of Lebesgue measure under the CDF $F_{\overrightarrow{B}}$. For any Borel subset A of $C_{\overrightarrow{B}}$, $\mu_{\overrightarrow{B}}(A)=m(F_{\overrightarrow{B}}|_{C_{\overrightarrow{B}}}(A))$, where m is Lebesgue measure.

The Cantor ternary set is the invariant set for the iterated function system $\phi_0(x)=\frac{x}{3}$ and $\phi_2(x)=\frac{x+2}{3}$. The corresponding CDF is often referred to as the "Devil's staircase", and the invariant measure on the Cantor ternary set is the pullback of Lebesgue measure onto the Cantor set under the CDF.

The Cantor sets we consider in this paper are sometimes referred to as "thin" Cantor sets [4]. The Cantor sets we consider have Lebesgue measure 0; indeed, the Hausdorff dimension of $C_{N,D}$ is $\frac{\log|D|}{\log N}$.

Next, we describe an algorithm for approximating the CDF of a Cantor set. To be precise, we recursively define a sequence of piecewise linear functions $\{f_n\}$ which converges uniformly to the desired CDF. For this, we need the following definition.

Definition 1.3 (Cumulative Digit Function)**.** *Let $\overrightarrow{B}=(b_0,...,b_{N-1})$ be a binary digit vector. Define $g:\{0,...,N\}\to\{0,...,\|\overrightarrow{B}\|\}$ to be the* cumulative digit function *where $g(0)=0$ and $g(i):=\sum_{j=0}^{i-1}b_j\ \forall i\in\{1,...,N\}$.*

We can define a sequence of piecewise linear functions that approximate a CDF in the following manner. For the Cantor set $C_{\overrightarrow{B}}$ with cumulative digit function $g_{\overrightarrow{B}}$, we define $F_{\overrightarrow{B}}^{(1)}$ as the linear interpolation of the points

$$S_1 = \left\{ \left(\frac{i}{N}, \frac{g(i)}{\|\overrightarrow{B}\|} \right) \middle| i \in \{0, 1, ..., N\} \right\}.$$

Let

$$S_n = \left\{ \left(\sum_{i=1}^{n} \frac{a_i}{N^i}, \sum_{i=1}^{n} \frac{g(a_i)}{\|\overrightarrow{B}\|^i} \right) \middle| 0 \leq a_1 \leq N, 0 \leq a_i \leq N-1 \, \forall i \in \{2, ..., n\} \right\}$$

and define $F^{(n)}_{\overrightarrow{B}}$ to be the linear interpolation of S_n. It can be shown that

$$F_{\overrightarrow{B}}(x) = \lim_{n\to\infty} F^{(n)}_{\overrightarrow{B}}(x)$$

where the limit converges uniformly on $[0, 1]$.

1.2 Operations on Cantor Sets and IFS's

For convenience, we define several operations on Cantor sets and their associated CDF's and IFS's. We recall the Kronecker product of two vectors: Let $\overrightarrow{B} = (b_0, b_1, ..., b_{M-1})$ and $\overrightarrow{C} = (c_0, c_1, ..., c_{N-1})$. Then, the *Kronecker product* of $\overrightarrow{B}$ with $\overrightarrow{C}$, denoted $\overrightarrow{B} \otimes \overrightarrow{C}$, is defined as

$$(\overrightarrow{B} \otimes \overrightarrow{C})(i) = b_{\lfloor \frac{i}{N} \rfloor} c_{i(modN)},$$

or equivalently,

$$(\overrightarrow{B} \otimes \overrightarrow{C})(n + mN) = b_m c_n$$

where $n \in \{0, 1, ..., N-1\}$ and $m \in \{0, 1, ..., M-1\}$, such that the product vector has $M \cdot N$ entries [5]. Note that if $\overrightarrow{B}$ and $\overrightarrow{C}$ are binary digit vectors, then $\overrightarrow{B} \otimes \overrightarrow{C}$ is another binary digit vector.

Definition 1.4 (Kronecker Product of CDF's)**.** *We define the Kronecker product of two CDFs as follows: Let $F_{\overrightarrow{B}}$ and $F_{\overrightarrow{C}}$ be the CDFs corresponding to the binary digit vectors $\overrightarrow{B}$ and $\overrightarrow{C}$, respectively. The Kronecker product of $F_{\overrightarrow{B}}$ with $F_{\overrightarrow{C}}$, denoted $F_{\overrightarrow{B}} \otimes F_{\overrightarrow{C}}$, is the CDF whose binary digit vector is $\overrightarrow{B} \otimes \overrightarrow{C}$.*

We can define a Kronecker product on digit sets to retain the association of $\overrightarrow{B}, \overrightarrow{C}$ with N_1, D_1, N_2, D_2.

Definition 1.5 (Kronecker Product of digit sets)**.** *The Kronecker product of two digit sets D_1 and D_2, denoted $D_1 \otimes D_2$, is defined to be the Kronecker product of their associated binary digit vectors. That is, the scale factor of $D_1 \otimes D_2$ is $N_1 \cdot N_2$, and $i \in D_1 \otimes D_1$ if and only if $b_i = 1$ for $\overrightarrow{B}_1 \otimes \overrightarrow{B}_2 = (b_0, b_1, \ldots b_{N_1 \cdot N_2 - 1}$.*

Lemma 1.1. *$D_1 \otimes D_2 = \{c + bN_2 \mid c \in D_2, b \in D_1\}$ where N_2 is the scale factor corresponding to D_2. The scale factor associated to $D_1 \otimes D_2$ is $N_1 N_2$.*

Definition 1.6. $\overrightarrow{B}^{\otimes n} := \overbrace{\overrightarrow{B} \otimes \overrightarrow{B} \otimes ... \otimes \overrightarrow{B}}^{n \text{ times}}$. *For example, $\overrightarrow{B}^{\otimes 1} = \overrightarrow{B}$, and $\overrightarrow{B}^{\otimes 2} = \overrightarrow{B} \otimes \overrightarrow{B}$.*

Some assorted definitions and notations. We let $\phi_D(A) = \bigcup_{d\in D} \phi_d(A)$ and we will write

$$(\phi_D)^n(A) := \bigcup_{d_1,...,d_n \in D} \phi_{d_1} \circ ... \circ \phi_{d_n}(A).$$

Note, using this notation, $\bigcap_{n=1}^{\infty} (\phi_D)^n([0, 1]) = C_{N,D}$. We denote the exponential function $e^{2\pi ix}$ by $e(x)$.

Definition 1.7 (Multiplicative Dependence)**.** *Two integers r and s are* multiplicatively dependent*, denoted by $r \sim s$, if there exist integers m and n not both zero such that $r^m = s^n$. Else, if no such integers exist, then r and s are* multiplicatively independent*, denoted by $r \not\sim s$.*

1.3 Related Results

The results we obtain in this paper are the first of their kind, as far as we are aware. However, sampling of functions that are associated with fractals has been considered previously in various ways.

Sampling of functions with fractal spectrum was first investigated in [6, 7]. In those papers, the authors consider the class of functions F which are the Fourier transform of functions $f \in L^2(\mu)$. Here, the measure μ is a fractal measure that is *spectral*, meaning that the Hilbert space $L^2(\mu)$ possesses an orthonormal basis of exponential functions. Similar sampling theorems are obtained in [8] without the assumption that the measure is spectral. In higher dimensions, graph approximations of fractals (such as the Sierpinski gasket) are often considered; sampling of functions on such graphs has been considered in [9, 10].

Sampling of cumulative distribution functions appear in [11, 12] in the context of the Cumulative Distribution Transform (CDT). The CDT is nonlinear and can provide better separation for classification problems. Sampling of cumulative distribution functions occurs in the discretization of the CDT. Related results on interpolation of data using fractal functions and iterated function systems can be found in [13, 14]. Approximating the moments of the Cantor function is investigated in [15].

A much more general construction of Cumulative Distribution Functions, and approximations thereof, can be found in [16]. Sampling of probability distributions on Cantor-like sets is considered in [17, 18].

2 Main Results

2.1 Preliminary Theorems

The first Lemma of this section is a very useful invariance identity of the CDF.

Lemma 2.1 (Invariance Equation). *Let $F_{\overrightarrow{B}}$ be a CDF with scale factor N and binary digit vector $\overrightarrow{B}$, then*

$$F_{\overrightarrow{B}}(x) = \sum_{n=0}^{N-1} \frac{b_n}{\|\overrightarrow{B}\|} F_{\overrightarrow{B}}(Nx - n). \tag{2}$$

where we regard $F(x) = 0$ for all $x \leq 0$ and $F(x) = 1$ for all $x \geq 1$.

Proof. This follows nearly immediately from Theorem 1.1, however, we present the proof anyway. Observe,

$$F_{\overrightarrow{B}}(x) = \int_0^x d\mu_{\overrightarrow{B}} = \int_0^x d \sum_{d\in D} \frac{1}{\|D\|} \mu_{\overrightarrow{B}} \circ \phi_d^{-1} = \sum_{d\in D} \frac{1}{\|D\|} \int_0^x d\mu_{\overrightarrow{B}} \circ \phi_d^{-1}.$$

Hence under a change-of-variables

$$F_{\overrightarrow{B}}(x) = \sum_{d\in D} \frac{1}{\|D\|} \int_{\phi_d^{-1}(0)}^{\phi_d^{-1}(x)} d\mu_{\overrightarrow{B}} \circ \phi_d^{-1} \circ \phi_d = \sum_{d\in D} \frac{1}{\|D\|} \int_{-d}^{Nx-d} d\mu_{\overrightarrow{B}}$$

$$= \sum_{d\in D} \frac{1}{\|D\|} \left(\int_{-d}^{0} d\mu_{\overrightarrow{B}} + \int_0^{Nx-d} d\mu_{\overrightarrow{B}} \right) = \sum_{d\in D} \frac{1}{\|D\|} \int_0^{Nx-d} d\mu_{\overrightarrow{B}}.$$

Finally, since $\|\overrightarrow{B}\| = \|D\|$, $D \subset \{0, 1, ..., N-1\}$, and $b_n = 1$ for $n \in D$ and $b_n = 0$ for $n \notin D$,

$$F_{\overrightarrow{B}}(x) = \sum_{d\in D} \frac{1}{\|\overrightarrow{B}\|} \int_0^{Nx-d} d\mu_{\overrightarrow{B}} = \sum_{n=0}^{N-1} \frac{b_n}{\|\overrightarrow{B}\|} \int_0^{Nx-d} d\mu_{\overrightarrow{B}} = \sum_{n=0}^{N-1} \frac{b_n}{\|\overrightarrow{B}\|} F_{\overrightarrow{B}}(Nx - d).$$

□

Lemma 2.2. *Let $F_{\overrightarrow{B}}$ be a CDF with scale factor N and binary digit vector $\overrightarrow{B}$, then $F_{\overrightarrow{B}}\left(\frac{k}{N}\right) = \frac{g(k)}{\|\overrightarrow{B}\|}$ for $k \in \{0, ..., N\}$, where g is the cumulative digit function.*

Proof. Let $\overrightarrow{B} = (b_0, ..., b_{N-1})$ be the binary digit vector for $F_{\overrightarrow{B}}$. Then by the invariance equation 2,

$$F_{\overrightarrow{B}}\left(\frac{k}{N}\right) = \sum_{n=0}^{N-1} \frac{b_n}{\|\overrightarrow{B}\|} F_{\overrightarrow{B}}(k-n) = \sum_{n=0}^{k-1} \frac{b_n}{\|\overrightarrow{B}\|} = \frac{g(k)}{\|\overrightarrow{B}\|}.$$

□

Proposition 2.1. *A function $g : \{0, ..., N\} \to \{0, ..., d\}$ is a cumulative digit function for some valid CDF if and only if the following criteria are met.*

1. *$g(0) = 0$*
2. *$g(N) = d$, for some $d \in \{2, ..., N-1\}$*
3. *$0 \leq g(k+1) - g(k) \leq 1$ for all $k \in \{0, ..., N-1\}$.*

Moreover, if g satisfies conditions (1), (2), and (3), then the corresponding CDF $C_{\overrightarrow{B}}$ has binary representation $\overrightarrow{B} = (b_0, ..., b_{N-1})$ such that $b_k = 1$ if and only if $g(k+1) - g(k) = 1$ and $\|\overrightarrow{B}\| = d$.

Proof. ($\Rightarrow$) Let g be the cumulative digit function for $C_{\overrightarrow{B}}$. The first condition follows directly from the definition of g. Also, $g(N) = \sum_{j=0}^{N-1} b_j = d$ so the second condition holds. By definition of g, $g(i) = \sum_{j=0}^{i-1} b_j \leq \sum_{j=0}^{i} b_j = g(i+1)$, so $0 \leq g(i+1) - g(i)$

Finally, $g(i) + 1 = \sum_{j=0}^{i-1} b_j + 1 \geq \sum_{j=0}^{i} b_j = g(i+1)$ implies the third condition.

($\Leftarrow$) Construct a CDF with the binary representation $\overrightarrow{B} = (b_0, ..., b_{N-1})$ such that $b_k = 1$ if and only if $g(k+1) - g(k) = 1$. By the second and third conditions, at least two b_i will be 1, and this is a valid CDF. By the third condition and the range of g, either $g(k+1) - g(k) = 1$ and $b_k = 1$ or $g(k+1) - g(k) = 0$ and $b_k = 0$. By the first condition, $g(0) = 0$. For induction, suppose that for $0 \leq i \leq N-1$, $g(i) = \sum_{k=0}^{i-1} b_k$. Then, $g(i+1) - g(i) = 1$ if and only if $b_i = 1$. Therefore, $g(i+1) = g(i) + 1 = \sum_{k=0}^{i-1} b_k + 1$ if and only if b_i is 1. Then, $g(i+1) = \sum_{k=0}^{i} b_k$. By induction, it follows g is the cumulative digit function of $\overrightarrow{B}$ by definition.

□

2.1.1 Kronecker Product Results

We define $\phi_{D_1} \circ \phi_{D_2}(A) = \bigcup_{d \in D_1} \phi_d\left(\bigcup_{d' \in D_2} \phi_{d'}(A)\right)$.

Proposition 2.2. *Consider Cantor sets $C_{\overrightarrow{B}_1}$ and $C_{\overrightarrow{B}_2}$ such that the scale factor and binary digit vector for $\overrightarrow{B}_i$ are N_i, D_i. Then $\phi_{D_1} \circ \phi_{D_2} = \phi_{D_1 \otimes D_2}$*

Proof. First, $y \in \phi_{D_1} \circ \phi_{D_2}([0,1])$ if and only if there exists $x \in [0,1]$ such that $y = \phi_{D_1} \circ \phi_{D_2}(x)$. This occurs if and only if

$$y = \frac{\frac{x+\epsilon_2}{N_2} + \epsilon_1}{N_1} = \frac{x + \epsilon_2 + \epsilon_1 N_2}{N_1 N_2}$$

for some $\epsilon_1 \in D_1$, $\epsilon_2 \in D_2$.
This is the IFS for scale factor $N_1 N_2$ and binary digit vector $D_3 = \{\epsilon_2 + \epsilon_1 N_2 \mid \epsilon_2 \in D_2, \epsilon_1 \in D_1\} = D_1 \otimes D_2$ by definition of the Kronecker product. □

Corollary 2.1. $(\phi_D)^n = \phi_{D^{\otimes n}}$ *for all* $n \in \mathbb{Z}^+$.

Corollary 2.2. $F_{\overrightarrow{B}} = F_{\overrightarrow{B}^{\otimes k}}$.

Proof. Since $F_{\overrightarrow{B}}$ is uniquely determined by $C_{\overrightarrow{B}}$, and $C_{\overrightarrow{B}}$ is uniquely determined by the property that $\phi_D(C_{\overrightarrow{B}}) = C_{\overrightarrow{B}}$, we have that $C_{\overrightarrow{B}} = (\phi_D)^n(C_{\overrightarrow{B}}) = \phi_{D^{\otimes n}}(C_{\overrightarrow{B}})$. Hence $C_{\overrightarrow{B}}$ satisfies the invariance property of $\phi_{D^{\otimes n}}$. Since $D^{\otimes n}$ was defined to retain its association with $\overrightarrow{B}^{\otimes n}$ we have that $F_{\overrightarrow{B}} = F_{\overrightarrow{B}^{\otimes n}}$. □

Lemma 2.3. *Let $\overrightarrow{B} = (b_0, ..., b_{M-1})$ be a binary digit vector with cumulative digit function $g_{\overrightarrow{B}}$, $\overrightarrow{C} = (c_0, .., c_{N-1})$ be a binary digit vector with cumulative digit function $g_{\overrightarrow{C}}$, and $g_{\overrightarrow{B}\otimes\overrightarrow{C}}$ be the cumulative digit function for $\overrightarrow{B} \otimes \overrightarrow{C}$. Then, for $j \in \{0, ..., N\}$, $k \in \{0, ..., M\}$, $g_{\overrightarrow{B}\otimes\overrightarrow{C}}(kN + j) = \|\overrightarrow{C}\| g_{\overrightarrow{B}}(k) + b_k g_{\overrightarrow{C}}(j)$.*

Proof. The proof follows by induction on j.
When $j = k = 0$, $g_{\overrightarrow{B}\otimes\overrightarrow{C}}(0) = 0 = \|\overrightarrow{C}\| g_{\overrightarrow{B}}(0)$ by definition. When $k \geq 1$, then

$$\begin{aligned} g_{\overrightarrow{B}\otimes\overrightarrow{C}}(kN) &= \sum_{i=0}^{kN-1} (\overrightarrow{B} \otimes \overrightarrow{C})(i) = \sum_{m=0}^{k-1}\sum_{n=0}^{N-1} (\overrightarrow{B} \otimes \overrightarrow{C})(n + mN) \\ &= \sum_{m=0}^{k-1}\sum_{n=0}^{N-1} b_m c_n = \sum_{n=0}^{N-1} c_n \sum_{m=0}^{k-1} b_m \\ &= \|\overrightarrow{C}\| g_{\overrightarrow{B}}(k) \end{aligned}$$

as desired.
It follows the identity holds for all k when $j = 0$. This serves as the base case for induction on j. Now assume the identity for j. Then,

$$\begin{aligned} g_{\overrightarrow{B}\otimes\overrightarrow{C}}(kN + j + 1) &= g_{\overrightarrow{B}\otimes\overrightarrow{C}}(kN + j) + (\overrightarrow{B} \otimes \overrightarrow{C})(kN + j) \\ &= \|\overrightarrow{C}\| g_{\overrightarrow{B}}(k) + b_k g_C(j) + (\overrightarrow{B} \otimes \overrightarrow{C})(kN + j) \\ &= \|\overrightarrow{C}\| g_{\overrightarrow{B}}(k) + b_k g_{\overrightarrow{C}}(j) + b_k c_j. \end{aligned}$$

It follows, when $b_k = 1$,

$$g_{\overrightarrow{B}\otimes\overrightarrow{C}}(kN + j + 1) = \|\overrightarrow{C}\| g_{\overrightarrow{B}}(k) + g_{\overrightarrow{C}}(j) + c_j = \|\overrightarrow{C}\| g_{\overrightarrow{B}}(k) + b_k g_{\overrightarrow{C}}(j + 1).$$

Otherwise, when $b_k = 0$,

$$g_{\overrightarrow{B}\otimes\overrightarrow{C}}(kN + j + 1) = \|\overrightarrow{C}\| g_{\overrightarrow{B}}(k) = \|\overrightarrow{C}\| g_{\overrightarrow{B}}(k) + b_k g_{\overrightarrow{C}}(j + 1).$$

□

Proposition 2.3. *Let $F_{\overrightarrow{B}}$ be a CDF with scale factor N, binary digit vector $\overrightarrow{B} = (b_0, b_1, ..., b_{N-1})$, and cumulative digit function $g_{\overrightarrow{B}}$.*
For $x \in (0, 1)$, such that $x = \sum_{i=1}^{\infty} \frac{n_i}{N^i}$, $n_i \in \mathbb{Z}_N$, $F_{\overrightarrow{B}}(x) = \sum_{i=1}^{\infty} \left(\prod_{k=1}^{i-1} b_{n_k}\right) \frac{g(n_i)}{\|\overrightarrow{B}\|^i}$.

Proof. Fix the sequence $\{n_i\} \subset \mathbb{Z}_N$. We have, by Corollary 2.2 and Lemma 2.2, for all $j \in \mathbb{N}$

$$F_{\overrightarrow{B}}\left(\sum_{i=1}^{j} \frac{n_i}{N^i}\right) = F_{\overrightarrow{B}^{\otimes j}}\left(\sum_{i=1}^{j} \frac{n_i}{N^i}\right) = \frac{g_{\overrightarrow{B}^{\otimes j}}\left(\sum_{i=1}^{j} N^{j-i} n_i\right)}{\|\overrightarrow{B}\|^j}.$$

For an inductive base case, by Proposition 2.2,

$$F_{\overrightarrow{B}}\left(\frac{n_1}{N}\right) = \frac{g_{\overrightarrow{B}}(n_1)}{\|\overrightarrow{B}\|} = \sum_{i=1}^{1}\left(\prod_{k=1}^{i-1} b_{n_k}\right) \frac{g_{\overrightarrow{B}}(n_i)}{\|\overrightarrow{B}\|^i}.$$

For induction on j, suppose that

$$F_{\vec{B}}\left(\sum_{i=1}^{j}\frac{n_i}{N^i}\right)=\sum_{i=1}^{j}\left(\prod_{k=1}^{i-1}b_{n_k}\right)\frac{g_{\vec{B}}(n_i)}{\|\vec{B}\|^i}.$$

Then, with Lemma 2.3 and Lemma 2.2,

$$\begin{aligned}
F_{\vec{B}}\left(\sum_{i=1}^{j+1}\frac{n_i}{N^i}\right) &= F_{\vec{B}^{\otimes j+1}}\left(\sum_{i=1}^{j+1}\frac{n_i}{N^i}\right)=F_{\vec{B}^{\otimes j+1}}\left(\frac{\sum_{i=1}^{j+1}n_iN^{j+1-i}}{N^{j+1}}\right)\\
&=\frac{g_{\vec{B}^{\otimes j+1}}\left(\sum_{i=1}^{j+1}n_iN^{j+1-i}\right)}{\|\vec{B}\|^{j+1}}=\frac{g_{\vec{B}^{\otimes j}\otimes\vec{B}}\left(n_1N^j+\sum_{i=2}^{j+1}n_iN^{j+1-i}\right)}{\|\vec{B}\|^{j+1}}\\
&=\frac{\|\vec{B}\|^j g_{\vec{B}}(n_1)+b_{n_1}g_{\vec{B}^{\otimes j}}\left(\sum_{i=2}^{j+1}n_iN^{j+1-i}\right)}{\|\vec{B}\|^{j+1}}=\frac{g_{\vec{B}}(n_1)}{\|\vec{B}\|}+\frac{b_{n_1}}{\|\vec{B}\|}\frac{g_{\vec{B}^{\otimes j}}\left(\sum_{i=2}^{j+1}n_iN^{j+1-i}\right)}{\|\vec{B}\|^j}\\
&=\frac{g_{\vec{B}}(n_1)}{\|\vec{B}\|}+\frac{b_{n_1}}{\|\vec{B}\|}\sum_{i=2}^{j+1}\left(\prod_{k=2}^{i-1}b_{n_k}\right)\frac{g_{\vec{B}}(n_i)}{\|\vec{B}\|^{i-1}},\text{ by shifting indices in the inductive hypothesis}\\
&=\frac{g_{\vec{B}}(n_1)}{\|\vec{B}\|}+\sum_{i=2}^{j+1}\left(\prod_{k=1}^{i-1}b_{n_k}\right)\frac{g_{\vec{B}}(n_i)}{\|\vec{B}\|^{i}}=\sum_{i=1}^{j+1}\left(\prod_{k=1}^{i-1}b_{n_k}\right)\frac{g_{\vec{B}}(n_i)}{\|\vec{B}\|^{i}}.
\end{aligned}$$

Thus, by induction, for all j and $n_i\in\mathbb{Z}_N$

$$F_{\vec{B}}\left(\sum_{i=1}^{j}\frac{n_i}{N^i}\right)=\sum_{i=1}^{j}\left(\prod_{k=1}^{i-1}b_{n_k}\right)\frac{g_{\vec{B}}(n_i)}{\|\vec{B}\|^i}.$$

Next, note all $x\in(0,1)$ have the form $\sum_{i=1}^{\infty}\frac{n_i}{N^i}$ for some $n_i\in\mathbb{Z}_N$. Since $F_{\vec{B}}$ is a continuous function,

$$F_{\vec{B}}(x)=\lim_{j\to\infty}F_{\vec{B}}\left(\sum_{i=1}^{j}\frac{n_i}{N^i}\right)=\lim_{j\to\infty}\sum_{i=1}^{j}\left(\prod_{k=1}^{i-1}b_{n_k}\right)\frac{g_{\vec{B}}(n_i)}{\|\vec{B}\|^i}=\sum_{i=1}^{\infty}\left(\prod_{k=1}^{i-1}b_{n_k}\right)\frac{g_{\vec{B}}(n_i)}{\|\vec{B}\|^i}.$$

□

2.2 Interpolation

Proposition 2.4. *Let $\{(x_n,y_n)\}_{n=1}^k\subset(\mathbb{Q}\cap(0,1))\times(\mathbb{Q}\cap(0,1))$, i.e. rational pairs in the unit cube, with $x_m\neq x_n$ for $m\neq n$ and $y_m\geq y_n$ whenever $x_m\geq x_n$. Then there exists a CDF interpolating the data $\{(x_n,y_n)\}_{n=1}^k$; more specifically, there exists a binary digit vector $\vec{B}$ such that $F_{\vec{B}}(x_n)=y_n$ for all n.*

Proof. We may assume without loss of generality that $x_1<x_2<...<x_k$. Further, by considering equivalent fractions, we may assume for all n, that $x_n=\frac{a_n}{N}$ and $y_n=\frac{c_n}{C}$ where $a_{i+1}-a_i\geq c_{i+1}-c_i+1$ for $0\leq i\leq k$ with the following conventions: $a_0=c_0=0$, $a_{k+1}=N$, and $c_{k+1}=C$. We construct the binary digit vector $\vec{B}$ of length N as follows:

$$\vec{B}(a_i)=\vec{B}(a_i+1)=...=\vec{B}(a_i+c_{i+1}-c_i-1)=1$$
$$\vec{B}(a_i+c_{i+1}-c_i)=\vec{B}(a_i+c_{i+1}-c_i+1)=...=\vec{B}(a_{i+1}-1)=0.$$

Then, we observe the recurrence relation,

$$F_{\vec{B}}(x_1)=F_{\vec{B}}\left(\frac{a_1}{N}\right)=\frac{c_1}{C}=y_1$$
$$F_{\vec{B}}(x_{i+1})-F_{\vec{B}}(x_i)=F_{\vec{B}}\left(\frac{a_{i+1}}{N}\right)-F_{\vec{B}}\left(\frac{a_i}{N}\right)=\frac{g_{\vec{B}}(a_{i+1})-g_{\vec{B}}(a_i)}{C}=\frac{c_{i+1}-c_i}{C}=y_{i+1}-y_i$$

which concludes the proof.

□

Remark 2.1. *Let $\{(x_n, y_n)\}$ be a finite sampling set of rational pairs in the unit cube satisfying the hypotheses of Proposition 2.4. We note from the proof of the proposition that interpolation by a CDF is not unique.*

Corollary 2.3. *Let $\{(x_n, y_n)\}_{n=1}^k \subset (0,1) \times (\mathbb{Q} \cap (0,1))$ with $x_m \neq x_n$ for $m \neq n$ and $y_m \geq y_n$ whenever $x_m \geq x_n$. Then there exists a CDF that interpolates the data $\{(x_n, y_n)\}_{n=1}^k$; more specifically, there exists a binary digit vector $\overrightarrow{B}$ such that $F_{\overrightarrow{B}}(x_n) = y_n$ for all n.*

Proof. We may assume without loss of generality that $0 < x_1 < x_2 < ... < x_k < 1$. Now select a collection of rational pairs $\{(z_n, w_n)\}_{n=1}^{2k}$ such that $z_1 < x_1$, $x_n < z_{2n} < z_{2n+1} < x_{n+1}$ for $1 \leq n \leq k-1$, $x_k < z_{2k}$, and $w_{2n-1} = w_{2n} = y_n$ for all n. Then, by Proposition 2.4, there exists a binary digit vector $\overrightarrow{B}$ such that $F_{\overrightarrow{B}}(z_n) = w_n$ for all n and, in particular, $F_{\overrightarrow{B}}(x_n) = y_n$. □

Corollary 2.4. *Let $\{(x_n, y_n)\}_{n=1}^k$ be a set of samples of the CDF $F_{\overrightarrow{B}}$. Suppose $\mathcal{C}$ is the collection of binary digit vectors such that, for $\overrightarrow{C} \in \mathcal{C}$, $F_{\overrightarrow{C}}(x_n) = y_n$ for all n (as guaranteed by Corollary 2.3). Then*

$$\sup_{\overrightarrow{C} \in \mathcal{C}} \left(\sup_{x \in (0,1)} |F_{\overrightarrow{B}}(x) - F_{\overrightarrow{C}}(x)| \right) = \max_{n=1,...,k-1} (y_{n+1} - y_n).$$

Proof. Without loss of generality, let $(x_1, y_1), (x_2, y_2)$ be such that

$$(y_2 - y_1) = \max_{n=1,...,k-1} (y_{n+1} - y_n).$$

Then by adding the interpolation point $(x_2 - \frac{1}{n}, y_1)$ to be satisfied by $F_{\overrightarrow{B}_n}$, the sequence of CDFs $F_{\overrightarrow{B}_n}$ has the property that

$$\lim_{n \to \infty} F_{\overrightarrow{B}_n} \left(x_2 - \frac{1}{n} \right) = y_1.$$

Since CDFs are by definition increasing, this completes the proof. □

2.3 Sampling

We first show that if we know the scaling factor N, then $N-1$ well chosen sample points is enough to reconstruct $F_{\overrightarrow{B}}$.

Lemma 2.4. *For $m \in \{0, ..., N-1\}$, $F_{\overrightarrow{B}}\left(\frac{m+1}{N}\right) = F_{\overrightarrow{B}}\left(\frac{m}{N}\right)$ if and only if $b_m = 0$.*

Proof. Let $\overrightarrow{B} = (b_0, ..., b_{N-1})$ be the binary digit vector for $F_{\overrightarrow{B}}$. By Lemma 2.2, $F\left(\frac{m+1}{N}\right) - F\left(\frac{m}{N}\right) = \frac{b_m}{\|\overrightarrow{B}\|}$. Then, $F\left(\frac{m+1}{N}\right) = F\left(\frac{m}{N}\right)$ if and only if $b_m = 0$. □

Theorem 2.1. *Let $F_{\overrightarrow{B}}$ be a CDF with $\|\overrightarrow{B}\| = N$. Given $\{F_{\overrightarrow{B}}(\frac{k}{N})\}_{k=1}^{N-1}$, $\overrightarrow{B}$ can be uniquely determined.*

Proof. Since $F_{\overrightarrow{B}}(0) = 0$ and $F_{\overrightarrow{B}}(1) = 1$, this follows from Lemma 2.4. □

Corollary 2.5. *If $\mathscr{G}_N = \{F_{\overrightarrow{B}} : \|\overrightarrow{B}\| = N\}$, then $\{\left(\frac{k}{N}\right) : k = 1, ..., N-1\}$ is a set of uniqueness for $\mathscr{G}_N$.*

We will now consider the case when we do not know the scale factor.

2.3.1 Motivating a bound on scale factor

Remark 2.1 and Corollary 2.4 together establish that finite samples will never suffice without some sort of constraint. We contrast this with with Proposition 2.6 below as this shows a lower bound of $O(N)$ points is necessary, where N is the scale factor. The following proposition shows that to be able to uniquely determine a CDF with a finite number of points, there must be a bound on the scale factor.

Lemma 2.5. *Fix an integer $N \geq 4$, and suppose $\{x_n\}_{1 \leq n \leq k} \subset [0,1]$ where $0 \leq x_{n-1} \leq x_n \leq 1$ for all n and $k < \lfloor \frac{N}{2} \rfloor$. Then there exist two distinct CDFs $F_{\overrightarrow{B}}$ and $F_{\overrightarrow{C}}$, both with scale factor N, such that $F_{\overrightarrow{B}}(x_n) = F_{\overrightarrow{C}}(x_n)$ $\forall n \in \{1, ..., k\}$.*

Proof. First, we note that there exists an integer i such that $x_n \notin \left(\frac{i}{N}, \frac{i+2}{N}\right)$ for all $n \in \{1, 2, ..., k\}$ by the pigeon-hole principle.

Next, since $N \geq 4$, we also have that there exists an integer $j \in \{0, 1, ..., N-1\} \setminus \{i, i+1\}$ such that $x_n \notin \left(\frac{j}{N}, \frac{j+1}{N}\right)$ for all $n \in \{1, 2, ..., k\}$.

We construct two distinct binary digit vectors $\overrightarrow{B} = (b_0, b_1, ..., b_{N-1})$ and $\overrightarrow{C} = (c_0, c_1, ..., c_{N-1})$ as follows: Let $b_i = 0$, $b_{i+1} = 1$, $c_i = 1$, $c_{i+1} = 0$, $b_j = c_j = 1$, and $b_m = c_m = 0$ for all $m \notin \{i, i+1, j\}$. Note that both binary digit vectors are nondegenerate since two digits are kept and $\|\overrightarrow{B}\| = \|\overrightarrow{C}\| = 2$. We note that since $\|\overrightarrow{B}\| = \|\overrightarrow{C}\|$ and $\overrightarrow{B} \neq \overrightarrow{C}$, then $F_{\overrightarrow{B}} \neq F_{\overrightarrow{C}}$. We conclude the proof by showing that $F_{\overrightarrow{B}}(x_n) = F_{\overrightarrow{C}}(x_n)$ for all n.

Case 1: $i < j$
Let $x \leq \frac{i}{N}$. Then by Lemma 2.2

$$0 \leq F_{\overrightarrow{B}}(x) \leq F_{\overrightarrow{B}}\left(\frac{i}{N}\right) = \frac{g_{\overrightarrow{B}}(i)}{2} = 0.$$

Likewise, $F_{\overrightarrow{C}}(x) = 0$. Now let $\frac{i+2}{N} \leq x \leq \frac{j}{N}$. Then

$$\frac{1}{2} = \frac{g_{\overrightarrow{B}}(i+2)}{2} = F_{\overrightarrow{B}}\left(\frac{i+2}{N}\right) \leq F_{\overrightarrow{B}}(x) \leq F_{\overrightarrow{B}}\left(\frac{j}{N}\right) = \frac{g_{\overrightarrow{B}}(j)}{2} = \frac{1}{2}.$$

Likewise, $F_{\overrightarrow{C}}(x) = \frac{1}{2}$. Finally let $\frac{j+1}{N} \leq x \leq 1$. Then

$$1 = \frac{g_{\overrightarrow{B}}(j+1)}{2} = F_{\overrightarrow{B}}\left(\frac{j+1}{N}\right) \leq F_{\overrightarrow{B}}(x) \leq 1.$$

Likewise, $F_{\overrightarrow{C}}(x) = 1$. Thus, $F_{\overrightarrow{B}}(x_n) = F_{\overrightarrow{C}}(x_n)$ for all n.

Case 2: $j < i$
The argument is analogous to the one given for case 1, and we omit the details.

Figures 3 and 4 depict cases 1 and 2, respectively.

□

The next proposition observes the relationship between the CDFs of the binary digit vector $\overrightarrow{B}$ and its reverse $\overleftarrow{B}$, that is $\overleftarrow{B}(n) = \overrightarrow{B}(N-1-n)$ for all n where N is the length of $\overrightarrow{B}$.

Proposition 2.5. *Let $\overrightarrow{B}$ be a binary digit vector. Then,*

$$F_{\overleftarrow{B}}(x) = 1 - F_{\overrightarrow{B}}(1-x).$$

Proof. Since $F_{\overleftarrow{B}}(x) + F_{\overrightarrow{B}}(1-x)$ is continuous, it suffices to show the equality on a dense subset of the unit interval. Specifically, we show the identity on the set of N-adic numbers, that is

$$\left\{ \frac{1}{N^k} \sum_{\ell=0}^{k-1} n_\ell N^\ell \,\middle|\, k \in \mathbb{N}, n_\ell \in \{0, 1, ..., N-1\} \right\},$$

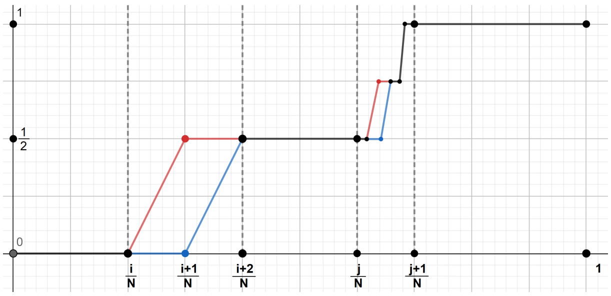

Figure 3. Case 1 — Sketch of piecewise linear approximations of $F_{\overrightarrow{B}}$ (blue) and $F_{\overrightarrow{C}}$ (red)

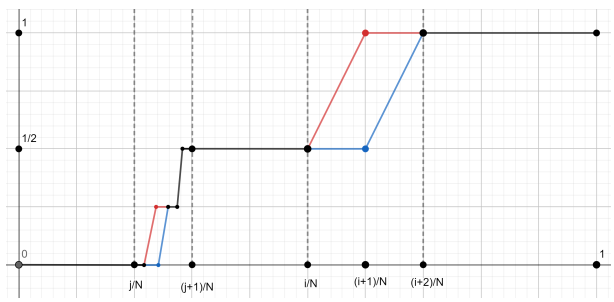

Figure 4. Case 2 — Sketch of piecewise linear approximations of $F_{\overrightarrow{B}}$ (blue) and $F_{\overrightarrow{C}}$ (red)

where N is the length of $\overrightarrow{B}$. We first observe that the simplest case, when $k = 1$, holds.

$$F_{\overleftarrow{B}}\left(\frac{n_0}{N}\right) + F_{\overrightarrow{B}}\left(1 - \frac{n_0}{N}\right) = F_{\overleftarrow{B}}\left(\frac{n_0}{N}\right) + F_{\overrightarrow{B}}\left(\frac{N-n_0}{N}\right) = \sum_{n=0}^{n_0-1} \frac{b_{N-1-n}}{\|\overrightarrow{B}\|} + \sum_{n=0}^{N-n_0-1} \frac{b_n}{\|\overrightarrow{B}\|} = 1.$$

We proceed by induction on the power of the N-adic number, assuming the identity is true for k. Then, by Lemma 2.1,

$$\begin{aligned}
&F_{\overleftarrow{B}}\left(\frac{1}{N^{k+1}}\sum_{\ell=0}^{k} n_\ell N^\ell\right) + F_{\overrightarrow{B}}\left(1 - \frac{1}{N^{k+1}}\sum_{\ell=0}^{k} n_\ell N^\ell\right) \\
&= F_{\overleftarrow{B}}\left(\frac{1}{N^{k+1}}\sum_{\ell=0}^{k} n_\ell N^\ell\right) + F_{\overrightarrow{B}}\left(\frac{1}{N^{k+1}} + \frac{1}{N^{k+1}}\sum_{\ell=0}^{k}(N-1-n_\ell)N^\ell\right) \\
&= \sum_{n=0}^{N-1} \frac{b_{N-1-n}}{\|\overrightarrow{B}\|} F_{\overleftarrow{B}}\left(n_k - n + \frac{1}{N^k}\sum_{\ell=0}^{k-1} n_\ell N^\ell\right) + \frac{b_n}{\|\overrightarrow{B}\|} F_{\overrightarrow{B}}\left(N - 1 - n_k - n + \frac{1}{N^k} + \frac{1}{N^k}\sum_{\ell=0}^{k-1}(N-1-n_\ell)N^\ell\right) \\
&= \frac{\|\overrightarrow{B}\| - b_{N-1-n_k}}{\|\overrightarrow{B}\|} + \frac{b_{N-1-n_k}}{\|\overrightarrow{B}\|}\left[F_{\overleftarrow{B}}\left(\frac{1}{N^k}\sum_{\ell=0}^{k-1} n_\ell N^\ell\right) + F_{\overrightarrow{B}}\left(\frac{1}{N^k} + \frac{1}{N^k}\sum_{\ell=0}^{k-1}(N-1-n_\ell)N^\ell\right)\right] \\
&= \frac{\|\overrightarrow{B}\| - b_{N-1-n_k}}{\|\overrightarrow{B}\|} + \frac{b_{N-1-n_k}}{\|\overrightarrow{B}\|}\left[F_{\overleftarrow{B}}\left(\frac{1}{N^k}\sum_{\ell=0}^{k-1} n_\ell N^\ell\right) + F_{\overrightarrow{B}}\left(1 - \frac{1}{N^k}\sum_{\ell=0}^{k-1} n_\ell N^\ell\right)\right] \\
&= \frac{\|\overrightarrow{B}\| - b_{N-1-n_k}}{\|\overrightarrow{B}\|} + \frac{b_{N-1-n_k}}{\|\overrightarrow{B}\|} = 1.
\end{aligned}$$

Thus, the identity holds on the N-adic numbers, and the proof is done.

□

We say that a sampling algorithm is *conditional* if previously attained samples inform the selection of the next sample. For the remainder of this section, we describe a conditional sampling algorithm that completely determines a binary digit vector $\overrightarrow{B}$ given its scale factor N. The algorithm as stated below requires at most $\lfloor \frac{N}{2} \rfloor$ samples to execute successfully which we note is the minimum number of samples that is required under non-conditional sampling to discern binary digit vectors of equal scale factor. We first state the result.

Theorem 2.2. *Fix an integer $N \geq 3$, and let $\overrightarrow{B} = (b_0, b_1, ..., b_{N-1})$ be a binary digit vector with $2 \leq \|\overrightarrow{B}\| \leq N-1$. Then there is a conditional sampling algorithm with at most $\lfloor \frac{N}{2} \rfloor$ points that completely determines $F_{\overrightarrow{B}}$.*

The conditional sampling algorithm that answers Theorem 2.2 is located in the appendix and split into two parts. Each part considers pairs of digits from $\overrightarrow{B}$ at a time, e.g. (b_0, b_1), (b_2, b_3), etc. The role of Algorithm 1 is to find the first nonzero digit of $\overrightarrow{B}$. As a consequence of the method, we can also find $\|\overrightarrow{B}\|$ from the sampling in Algorithm 1. Then the algorithm terminates if the first nonzero digit occurred in the last pair, i.e. (b_{N-2}, b_{N-1}) if N is even or (b_{N-3}, b_{N-2}) if N is odd, as $\overrightarrow{B}$ is then completely determined; otherwise, Algorithm 2 applies a similar procedure to $\overleftarrow{B}$. The sampling in Algorithm 2 is expressed in terms of $F_{\overleftarrow{B}}$ which translates to a sampling of $F_{\overrightarrow{B}}$ by Proposition 2.5. Then the maximum number of samples from both Algorithm 1 and Algorithm 2 is precisely the number of paired digits, that is there are at most $\lfloor \frac{N}{2} \rfloor$ samples. In the proof of the Theorem 2.2, we show that there exists a positive integer ℓ that is only dependent on N (the smallest positive ℓ such that $2^{\ell+1} > N-1$ is sufficient) such that Algorithm 1 and Algorithm 2 are well-defined and completely determine $\overrightarrow{B}$.

Proof. Let $m \in \left\{1, 2, ..., \lfloor \frac{N}{2} \rfloor\right\}$. For convenience, we denote

$$\psi_m(x) = \frac{g_{\overrightarrow{B}}(2m-1)}{\|\overrightarrow{B}\|} + \frac{b_{2m-1}}{\|\overrightarrow{B}\|} x,$$

and use the notation $\psi_m^\ell = \psi_m \circ ... \circ \psi_m$ to represent the composition of ℓ functions. We claim that

$$F_{\overrightarrow{B}}\left(\frac{2m}{N^{\ell+1}} + \sum_{n=1}^{\ell} \frac{2m-1}{N^n}\right) = \psi_m^{\ell+1}(1). \tag{3}$$

The case when $\ell = 0$ immediately follows from Lemma 2.2 since

$$F_{\overrightarrow{B}}\left(\frac{2m}{N}\right) = \frac{g_{\overrightarrow{B}}(2m)}{\|\overrightarrow{B}\|} = \psi_m(1).$$

To prove identity (3) in general, we proceed by induction, so assume that the identity holds for ℓ. Then, by Lemma 2.1, we find

$$\begin{aligned}
F_{\overrightarrow{B}}\left(\frac{2m}{N^{\ell+2}} + \sum_{n=1}^{\ell+1} \frac{2m-1}{N^n}\right) &= \sum_{k=0}^{N-1} \frac{b_k}{\|\overrightarrow{B}\|} F_{\overrightarrow{B}}\left(\frac{2m}{N^{\ell+1}} + \left[\sum_{n=1}^{\ell} \frac{2m-1}{N^n}\right] + 2m-1-k\right) \\
&= \left[\sum_{k=0}^{2m-2} \frac{b_k}{\|\overrightarrow{B}\|}\right] + \frac{b_{2m-1}}{\|\overrightarrow{B}\|} F_{\overrightarrow{B}}\left(\frac{2m}{N^{\ell+1}} + \sum_{n=1}^{\ell} \frac{2m-1}{N^n}\right) \\
&= \frac{g_{\overrightarrow{B}}(2m-1)}{\|\overrightarrow{B}\|} + \frac{b_{2m-1}}{\|\overrightarrow{B}\|} \psi_m^{\ell+1}(1) \\
&= \psi_m^{\ell+2}(1),
\end{aligned}$$

as desired.

There are four cases to consider:

Case 1: $b_{2m-2} = b_{2m-1} = 0$. Then

$$\psi_m^{\ell+1}(1) = \frac{g_{\overrightarrow{B}}(2m-2)}{\|\overrightarrow{B}\|}.$$

Case 2: $b_{2m-2} = 1$; $b_{2m-1} = 0$. Then

$$\psi_m^{\ell+1}(1) = \frac{g_{\overrightarrow{B}}(2m-2)+1}{\|\overrightarrow{B}\|}.$$

Case 3: $b_{2m-2} = 0$; $b_{2m-1} = 1$. Then

$$\psi_m^{\ell+1}(1) = \frac{g_{\overrightarrow{B}}(2m-2)+1}{\|\overrightarrow{B}\|^{\ell+1}} + \sum_{n=1}^{\ell} \frac{g_{\overrightarrow{B}}(2m-2)}{\|\overrightarrow{B}\|^n} = \frac{g_{\overrightarrow{B}}(2m-2)+1}{\|\overrightarrow{B}\|^{\ell+1}} + g_{\overrightarrow{B}}(2m-2)\frac{\|\overrightarrow{B}\|^{\ell}-1}{(\|\overrightarrow{B}\|-1)\|\overrightarrow{B}\|^{\ell}}.$$

Case 4: $b_{2m-2} = b_{2m-1} = 1$. Then

$$\psi_m^{\ell+1}(1) = \frac{g_{\overrightarrow{B}}(2m-2)+2}{\|\overrightarrow{B}\|^{\ell+1}} + \sum_{n=1}^{\ell} \frac{g_{\overrightarrow{B}}(2m-2)+1}{\|\overrightarrow{B}\|^n} = \frac{g_{\overrightarrow{B}}(2m-2)+2}{\|\overrightarrow{B}\|^{\ell+1}} + (g_{\overrightarrow{B}}(2m-2)+1)\frac{\|\overrightarrow{B}\|^{\ell}-1}{(\|\overrightarrow{B}\|-1)\|\overrightarrow{B}\|^{\ell}}.$$

In Algorithm 1, we have $g_{\overrightarrow{B}}(2m-2) = 0$. If $\psi_m^{\ell+1}(1) = 0$, then clearly $b_{2m-2} = b_{2m-1} = 0$; else

$$\psi_m^{\ell+1}(1) \in \left\{ \frac{1}{\|\overrightarrow{B}\|}, \frac{1}{\|\overrightarrow{B}\|^{\ell+1}}, \frac{1}{\|\overrightarrow{B}\|-1} + \frac{\|\overrightarrow{B}\|-2}{(\|\overrightarrow{B}\|-1)\|\overrightarrow{B}\|^{\ell+1}} \,\middle|\, 2 \le \|\overrightarrow{B}\| \le N-1 \right\}.$$

Using some basic algebra, we note that for $\ell \ge 1$,

$$\left\{ \frac{1}{\|\overrightarrow{B}\|} \,\middle|\, 2 \le \|\overrightarrow{B}\| \le N-1 \right\} \bigcap \left\{ \frac{1}{\|\overrightarrow{B}\|-1} + \frac{\|\overrightarrow{B}\|-2}{(\|\overrightarrow{B}\|-1)\|\overrightarrow{B}\|^{\ell+1}} \,\middle|\, 2 \le \|\overrightarrow{B}\| \le N-1 \right\} = \emptyset$$

since the numbers are properly interlaced

$$\sum_{n=1}^{\ell} \frac{1}{2^n} + \frac{2}{2^{\ell+1}} = 1 > \sum_{n=1}^{\ell} \frac{1}{3^n} + \frac{2}{3^{\ell+1}} > \frac{1}{2} > \sum_{n=1}^{\ell} \frac{1}{4^n} + \frac{2}{4^{\ell+1}} > \frac{1}{3} > \dots > \frac{1}{N-1}.$$

Thus, it suffices to find an integer L such that for $\ell \ge L$,

$$\left\{ \frac{1}{\|\overrightarrow{B}\|}, \frac{1}{\|\overrightarrow{B}\|-1} + \frac{\|\overrightarrow{B}\|-2}{(\|\overrightarrow{B}\|-1)\|\overrightarrow{B}\|^{\ell+1}} \,\middle|\, 2 \le \|\overrightarrow{B}\| \le N-1 \right\} \bigcap \left\{ \frac{1}{\|\overrightarrow{B}\|^{\ell+1}} \,\middle|\, 2 \le \|\overrightarrow{B}\| \le N-1 \right\} = \emptyset.$$

The simplest way to find such an L is to take the smallest positive integer L such that $2^{L+1} > N-1$.

It follows that we can then determine the parameters $(b_{2m-2}, b_{2m-1}, \|\overrightarrow{B}\|) \in \{0,1\} \times \{0,1\} \times \{2,3,\dots,N-1\}$.

In the validation of Algorithm 2, it is equivalent to consider the three situations:

Situation 1. $g_{\overrightarrow{B}}(2m-2) = 0$

Situation 2. $g_{\overrightarrow{B}}(2m-2) = \|\overrightarrow{B}\| - 1$

Situation 3. $0 < g_{\overrightarrow{B}}(2m-2) < \|\overrightarrow{B}\| - 1$

As for situation 1, we just showed that we may solve for b_{2m-2} and b_{2m-1}. It is clear that $b_{2m-2} = b_{2m-1} = 0$ in situation 2 since Algorithm 1 identified a nonzero digit. Under the assumption of situation 3, we have that all of the values of $\psi_m^{\ell+1}(1)$ in cases 1 through 4 are distinct. This follows from tedious algebra, so we only show that Case 2 and Case 3 are different and leave the remainder to the reader to verify. Since $g_{\overrightarrow{B}}(2m-2) < \|\overrightarrow{B}\| - 1$, we have $(g_{\overrightarrow{B}}(2m-2)+1)(\|\overrightarrow{B}\|^{\ell}-1) + \|\overrightarrow{B}\| < \|\overrightarrow{B}\|^{\ell+1}$. Rearranging and combining terms, we find

$$(g_{\overrightarrow{B}}(2m-2)+1)(\|\overrightarrow{B}\|-1) + g_{\overrightarrow{B}}(2m-2)(\|\overrightarrow{B}\|^{\ell}-1)\|\overrightarrow{B}\| < (g_{\overrightarrow{B}}(2m-2)+1)(\|\overrightarrow{B}\|-1)\|\overrightarrow{B}\|^{\ell}.$$

We conclude that Case 2 and Case 3 are distinct from dividing through by $(\|\overrightarrow{B}\|-1)\|\overrightarrow{B}\|^{\ell+1}$.

$\square$

Remark 2.2. *The sampling set*

$$\left\{ \frac{2m}{N^{\ell+1}} + \sum_{n=1}^{\ell} \frac{2m-1}{N^n} \,\middle|\, m \in \left\{1, 2, ..., \left\lfloor \frac{N}{2} \right\rfloor \right\} \right\}$$

completely determines $\overrightarrow{B}$ up to ambiguity of the last nonzero digit in $\overrightarrow{B}$. That is, suppose that for some $m \in \left\{1, 2, ..., \left\lfloor \frac{N}{2} \right\rfloor \right\}$, we have that $b_n = 0$ for all $n > 2m$. Then there is ambiguity in the binary digit vector elements (b_{2m-2}, b_{2m-1}) as they could be either $(1, 0)$ or $(0, 1)$ and the samples would agree.

2.3.2 Rationality and the CDF

Lemma 2.6. *Let $\overrightarrow{B}$ be a binary digit vector of length N. If $x \in \mathbb{Q} \cap [0, 1]$, then $F_{\overrightarrow{B}}(x) \in \mathbb{Q}$.*

Proof. We first note that $F_{\overrightarrow{B}}(0) = 0$ and $F_{\overrightarrow{B}}(1) = 1$.

Then let $x \in \mathbb{Q} \cap (0, 1)$, and consider its N-adic representation $x = \sum_{i=1}^{\infty} \frac{n_i}{N^i}$ where $n_i \in \{0, 1, ..., N-1\}$. Since x is rational, the sequence $\{n_i\}_{i=1}^{\infty}$ is eventually periodic. Recall from Proposition 2.3 that

$$F_{\overrightarrow{B}}(x) = \sum_{i=1}^{\infty} \left(\prod_{k=1}^{i-1} b_{n_k} \right) \frac{g_{\overrightarrow{B}}(n_i)}{\|\overrightarrow{B}\|^i}.$$

If there exists a positive integer ℓ such that $b_{n_\ell} = 0$, then

$$F_{\overrightarrow{B}}(x) = \sum_{i=1}^{\ell} \left(\prod_{k=1}^{i-1} b_{n_k} \right) \frac{g_{\overrightarrow{B}}(n_i)}{\|\overrightarrow{B}\|^i},$$

which is rational. Note that this is the case if $g_{\overrightarrow{B}}(n_i) = \|\overrightarrow{B}\|$ for some i as we may then take $\ell = i + 1$. Otherwise, assume that $b_{n_k} = 1$ for all k. Then $g_{\overrightarrow{B}}(n_i) \in \{0, 1, ..., \|\overrightarrow{B}\| - 1\}$ for all i, and we have the $\|\overrightarrow{B}\|$-adic representation,

$$F_{\overrightarrow{B}}(x) = \sum_{i=1}^{\infty} \frac{g_{\overrightarrow{B}}(n_i)}{\|\overrightarrow{B}\|^i}.$$

Since the sequence $\{g_{\overrightarrow{B}}(n_i)\}_{i=1}^{\infty}$ is eventually periodic, it follows that $F_{\overrightarrow{B}}(x)$ is rational.

□

Lemma 2.7. *Let $C_{\overrightarrow{B}}$ be a Cantor set and $F_{\overrightarrow{B}}$ the CDF. For $x \in \mathbb{Q}^c \cap [0, 1]$, $x \in C_{\overrightarrow{B}}$ if and only if $F_{\overrightarrow{B}}(x) \in \mathbb{Q}^c$.*

Proof. Let $\overrightarrow{B} = (b_0, ..., b_{N-1})$ be the binary representation of $F_{\overrightarrow{B}}$.
Suppose $x \in \mathbb{Q}^c \cap [0, 1]$. Since $x \in (0, 1)$ it follows $x = \sum_{i=1}^{\infty} \frac{n_i}{N^i}$, for some $\{n_i\}_{i=1}^{\infty}$. Further, since x is irrational, $\{n_i\}_{i=1}^{\infty}$ is never periodic. By Proposition 2.3, $F_{\overrightarrow{B}}(x) = \sum_{i=1}^{\infty} \left(\prod_{k=1}^{i-1} b_{n_k} \right) \frac{g_{\overrightarrow{B}}(n_i)}{\|\overrightarrow{B}\|^i}$.
Suppose $x \in C_{\overrightarrow{B}}$. Since $x \in C_{\overrightarrow{B}}$, it follows $n_k \in D$, $b_{n_k} = 1$, and $g(n_k) \in \{0, 1, ..., \|\overrightarrow{B}\| - 1\}$ for all k. Then, $F_{\overrightarrow{B}}(x) = \sum_{i=1}^{\infty} \frac{g_{\overrightarrow{B}}(n_i)}{\|\overrightarrow{B}\|^i}$.
Note, $g(j+1) > g(j)$ whenever $j \in D$ implies $g_{\overrightarrow{B}}|_D$ is injective. Then, since $\{n_i\}_{i=1}^{\infty}$ is never periodic and $n_i \in D$, it follows that $\{g_{\overrightarrow{B}}(n_i)\}_{i=1}^{\infty}$ is also never periodic. Then, $F_{\overrightarrow{B}}(x)$ is a never periodic decimal in base $\|\overrightarrow{B}\|$. Thus, $F_{\overrightarrow{B}}(x) \in \mathbb{Q}^c$.
Alternatively, suppose $x \notin C_{\overrightarrow{B}}$. Then, there exists a smallest K such that $n_K \notin D$ and $b_{n_K} = 0$. Then $\prod_{k=1}^{i-1} b_{n_k} = 0$ if and only if $i > K$ and $F_{\overrightarrow{B}}(x) = \sum_{i=1}^{K} \frac{g_{\overrightarrow{B}}(n_i)}{\|\overrightarrow{B}\|^i}$. Thus, $F_{\overrightarrow{B}}(x) \in \mathbb{Q}$. □

Corollary 2.6. *If $x \notin C_{\overrightarrow{B}}$, then $F_{\overrightarrow{B}}(x) \in \mathbb{Q}$.*

Proof. Let $x \notin C_{\overrightarrow{B}}$. If $x \in \mathbb{Q}$, by Lemma 2.6, $F_{\overrightarrow{B}}(x) \in \mathbb{Q}$. If $x \in \mathbb{Q}^c$, by Lemma 2.7, $F_{\overrightarrow{B}}(x) \in \mathbb{Q}$. □

2.3.3 Multiplicatively Dependent Scale Factors

Lemma 2.8. *Let $F_{\vec{B}_1}$ be a CDF with scale factor N^L and $F_{\vec{B}_2}$ be a CDF with scale factor N^M, for $L, M, N \in \mathbb{N}$. If $\vec{B}_1 \otimes \vec{B}_2 = \vec{B}_2 \otimes \vec{B}_1$, then $F_{\vec{B}_1} = F_{\vec{B}_2}$.*

Proof. We first note that the Kronecker product is associative. Let $\vec{B}_1 \otimes \vec{B}_2 = \vec{B}_2 \otimes \vec{B}_1$.
By Corollary 2.2, $F_{\vec{B}_1} = F_{\vec{B}_1^{\otimes L}}$ and $F_{\vec{B}_2} = F_{\vec{B}_2^{\otimes M}}$. Then, $\vec{B}_1^{\otimes L}$ and $\vec{B}_2^{\otimes M}$ have length N^{LM}.
We will show $\vec{B}_1^{\otimes L} \otimes \vec{B}_2^{\otimes M} = \vec{B}_2^{\otimes M} \otimes \vec{B}_1^{\otimes L}$, by first showing $\vec{B}_1^{\otimes L} \otimes \vec{B}_2 = \vec{B}_2 \otimes \vec{B}_1^{\otimes L}$ by inducting on L.
As the base case, when $L = 1$, $\vec{B}_1^{\otimes 1} \otimes \vec{B}_2 = \vec{B}_1 \otimes \vec{B}_2 = \vec{B}_2 \otimes \vec{B}_1 = \vec{B}_2 \otimes \vec{B}_1^{\otimes 1}$.
Now assume $\vec{B}_1^{\otimes L} \otimes \vec{B}_2 = \vec{B}_2 \otimes \vec{B}_1^{\otimes L}$. Then,

$$\begin{aligned}\vec{B}_1^{\otimes L+1} \otimes \vec{B}_2 &= \vec{B}_1 \otimes \vec{B}_1^{\otimes L} \otimes \vec{B}_2 = \vec{B}_1 \otimes \vec{B}_2 \otimes \vec{B}_1^{\otimes L} \\ &= \vec{B}_2 \otimes \vec{B}_1 \otimes \vec{B}_1^{\otimes L} = \vec{B}_2 \otimes \vec{B}_1^{\otimes L+1}.\end{aligned}$$

This proves $\vec{B}_1^{\otimes L} \otimes \vec{B}_2 = \vec{B}_2 \otimes \vec{B}_1^{\otimes L}$.
Now we will induct on M. For the base case, when $M = 1$, $\vec{B}_1^{\otimes L} \otimes \vec{B}_2 = \vec{B}_2 \otimes \vec{B}_1^{\otimes L}$.
Now assume $\vec{B}_1^{\otimes L} \otimes \vec{B}_2^{\otimes M} = \vec{B}_2^{\otimes M} \otimes \vec{B}_1^{\otimes L}$. Then,

$$\begin{aligned}\vec{B}_2^{\otimes M+1} \otimes \vec{B}_1^{\otimes L} &= \vec{B}_2 \otimes \vec{B}_2^{\otimes M} \otimes \vec{B}_1^{\otimes L} = \vec{B}_2 \otimes \vec{B}_1^{\otimes L} \otimes \vec{B}_2^{\otimes M} \\ &= \vec{B}_1^{\otimes L} \otimes \vec{B}_2 \otimes \vec{B}_2^{\otimes M} = \vec{B}_1^{\otimes L} \otimes \vec{B}_2^{\otimes M+1}.\end{aligned}$$

By induction, $\vec{B}_1^{\otimes L} \otimes \vec{B}_2^{\otimes M} = \vec{B}_2^{\otimes M} \otimes \vec{B}_1^{\otimes L}$. Since $\vec{B}_1^{\otimes L}$ and $\vec{B}_2^{\otimes M}$ have length N^{LM}, $\vec{B}_1^{\otimes L}$ and $\vec{B}_2^{\otimes M}$ can be represented as N^{LM} long row vectors. This gives an equivalent definition of the Kronecker product on matrices. Since $\vec{B}_1^{\otimes L} \otimes \vec{B}_2^{\otimes M} = \vec{B}_2^{\otimes M} \otimes \vec{B}_1^{\otimes L}$, either $\vec{B}_1^{\otimes L} = c\vec{B}_2^{\otimes M}$ or $\vec{B}_2^{\otimes M} = c\vec{B}_1^{\otimes L}$, for some $c \in \mathbb{Z}_2$ (see Theorem 24 of [19]). If $c = 0$, this implies $\vec{B}_1 = 0$ or $\vec{B}_2 = 0$, which is a contradiction. Therefore, $c = 1$, and

$$\vec{B}_1^{\otimes L} = \vec{B}_2^{\otimes M}.$$

Thus,

$$F_{\vec{B}_1} = F_{\vec{B}_1^{\otimes L}} = F_{\vec{B}_2^{\otimes M}} = F_{\vec{B}_2}.$$

□

Lemma 2.9. *Let $\vec{A}$ have scale factor N, and $\vec{B}$ and $\vec{C}$ both have scale factor M. If $\vec{A} \otimes \vec{B} = \vec{A} \otimes \vec{C}$, then $\vec{B} = \vec{C}$.*

Proof. Let $\vec{A} = (a_0, ..., a_{N-1})$, $\vec{B} = (b_0, ..., b_{M-1})$, and $\vec{C} = (c_0, ..., c_{M-1})$. From $\vec{A} \otimes \vec{B} = \vec{A} \otimes \vec{C}$, it follows $a_i b_j = a_i c_j$ $\forall i, j$ such that $0 \leq i \leq N-1$, $0 \leq j \leq M-1$. Since $\vec{A}$ is a valid binary representation, $\vec{A} \neq 0$ so $\exists I$ such that $a_I \neq 0$. Then, $a_I b_j = a_I c_j$ $\forall j \in \{0, ..., M-1\}$, and $b_j = c_j$ $\forall j \in \{0, ..., M-1\}$. Thus, $\vec{B} = \vec{C}$. □

Proposition 2.6. *Let $L, M, N \in \mathbb{N}$. Let $S = \{\frac{m}{N^{L+M}}\}_{m=1}^{N^{L+M}-1}$. Let $\vec{B}_L$ be a binary vector of a CDF with length N^L and $\vec{B}_M$ be a binary vector of a CDF with length N^M. Then, $F_{\vec{B}_L}(x) = F_{\vec{B}_M}(x)$ for all $x \in S$ if and only if $F_{\vec{B}_L} = F_{\vec{B}_M}$.*

Proof. Let $\vec{B}_L = (b_0, ..., b_{N^L-1})$ and $\vec{B}_M = (c_0, ..., c_{N^M-1})$. Let g_L be the cumulative digit function for $\vec{B}_L$ and g_M be the cumulative digit function for $\vec{B}_M$. If $F_{\vec{B}_L} = F_{\vec{B}_M}$, clearly $F_{\vec{B}_L}(x) = F_{\vec{B}_M}(x)$ when $x \in S$. Suppose that $F_{\vec{B}_L}(x) = F_{\vec{B}_M}(x)$ for all $x \in S$. Let $F_{\vec{B}_L \otimes \vec{B}_M}$ be the CDF for $\vec{B}_L \otimes \vec{B}_M$, and g_{LM} be the cumulative digit function. Therefore,

$$\|\vec{B}_L \otimes \vec{B}_M\| = \|\vec{B}_L\| \|\vec{B}_M\|.$$

Let $k \in \{0, ..., N^L - 1\}$. By Lemma 2.3, $g_{LM}(kN^M) = g_L(k)\|\vec{B}_M\|$.

It follows from Proposition 2.2 that

$$F_{\vec{B}_L \otimes \vec{B}_M}\left(\frac{k}{N^L}\right) = F_{\vec{B}_L \otimes \vec{B}_M}\left(\frac{kN^M}{N^{L+M}}\right) = \frac{g_{LM}(kN^M)}{\|\vec{B}_L\|\|\vec{B}_M\|}$$
$$= \frac{g_L(k)\|\vec{B}_M\|}{\|\vec{B}_L\|\|\vec{B}_M\|} = \frac{g_L(k)}{\|\vec{B}_L\|} = F_{\vec{B}_L}\left(\frac{k}{N^L}\right) = F_{\vec{B}_M}\left(\frac{k}{N^L}\right)$$

since $\frac{k}{N^L} \in S$. Then, if $F_{\vec{B}_L}\left(\frac{k}{N^L}\right) = F_{\vec{B}_L}\left(\frac{k+1}{N^L}\right)$, since all CDFs are increasing functions, $F_{\vec{B}_L \otimes \vec{B}_M}(x) = F_{\vec{B}_L}(x) = F_{\vec{B}_M}(x)$ for all $x \in \left[\frac{k}{N^L}, \frac{k+1}{N^L}\right]$. Next, suppose $F_{\vec{B}_L}\left(\frac{k}{N^L}\right) < F_{\vec{B}_L}\left(\frac{k+1}{N^L}\right)$. By Theorem 2.4, $b_k = 1$, so by Lemma 2.3, for $j < N^M$, $g_{LM}(kN^M + j) = g_L(k)\|\vec{B}_M\| + g_M(j)$. Also, $F_{\vec{B}_L}$ is self-similar on the interval $\left[\frac{k}{N^L}, \frac{k+1}{N^L}\right]$. Let $x \in S \cap \left(\frac{k}{N^L}, \frac{k+1}{N^L}\right)$. Then, $x = \frac{k}{N^L} + \frac{j}{N^{L+M}}$ for $j \in \{1, ..., N^M - 1\}$. It follows

$$F_{\vec{B}_L}(x) = F_{\vec{B}_L}\left(\frac{k}{N^L}\right) + \frac{1}{\|\vec{B}_L\|} F_{\vec{B}_L}\left(N^L \frac{j}{N^{L+M}}\right) = F_{\vec{B}_L}\left(\frac{k}{N^L}\right) + \frac{1}{\|\vec{B}_L\|} F_{\vec{B}_L}\left(\frac{j}{N^M}\right).$$

Since $\frac{j}{N^M} \in S$ and by Proposition 2.2,

$$F_{\vec{B}_L}(x) = F_{\vec{B}_L}\left(\frac{k}{N^L}\right) + \frac{1}{\|\vec{B}_L\|} F_{\vec{B}_M}\left(\frac{j}{N^M}\right) = \frac{g_L(k)}{\|\vec{B}_L\|} + \frac{1}{\|\vec{B}_L\|} \cdot \frac{g_M(j)}{\|\vec{B}_M\|} = \frac{g_L(k)\|\vec{B}_M\| + g_M(j)}{\|\vec{B}_L\|\|\vec{B}_M\|}.$$

Next, by Proposition 2.2,

$$F_{\vec{B}_L \otimes \vec{B}_M}(x) = F_{\vec{B}_L \otimes \vec{B}_M}\left(\frac{k}{N^L} + \frac{j}{N^{L+M}}\right) = \frac{g_{LM}(kN^M + j)}{\|\vec{B}_L\|\|\vec{B}_M\|} = \frac{g_L(k)\|\vec{B}_M\| + g_M(j)}{\|\vec{B}_L\|\|\vec{B}_M\|} = F_{\vec{B}_L}(x).$$

Therefore, $F_{\vec{B}_L \otimes \vec{B}_M}(x) = F_{\vec{B}_L}(x)$ for all $x \in S$.

Further, by switching L and M above, $F_{\vec{B}_M \otimes \vec{B}_L}(x) = F_{\vec{B}_M}(x)$ for all $x \in S$. However, $F_{\vec{B}_M}(x) = F_{\vec{B}_L}(x)$ for all $x \in S$. Therefore, $F_{\vec{B}_M \otimes \vec{B}_L}(x) = F_{\vec{B}_L \otimes \vec{B}_M}(x)$ for all $x \in S$, and both have scale factor N^{L+M}. By Corollary 2.1, $\vec{B}_M \otimes \vec{B}_L = \vec{B}_L \otimes \vec{B}_M$. It follows by Lemma 2.8, $F_{\vec{B}_M} = F_{\vec{B}_L}$. □

2.3.4 Almost nowhere intersection of Cantor Sets

We will use the fact that different Cantor sets have almost no intersection, i.e. the intersection has measure 0 under either of the invariant measures, to design sampling schemes. Intersections of Cantor sets have been extensively studied, e.g. [20, 21]. We prove here the property of the intersection of Cantor sets that we need.

Lemma 2.10. *Let $h, M, N \in \mathbb{N}$ such that $M \not\sim N$. Then, for all $L \in \mathbb{N}$, there exists a constant $\alpha(M, N) \in (0, 1)$ dependent on M and N such that*

$$\sum_{n=0}^{L-1} \prod_{k=1}^{\infty} |\cos(N^{-k} h M^n \pi)| \leq 2L^{1-\alpha(M,N)}.$$

Proof. From Lemma 5 of [22], translated in Lemma 1 of [23], there exists a constant $\beta(M, N) > 0$ dependent upon M and N such that

$$\sum_{n=0}^{L-1} \prod_{k=1}^{\infty} |\cos(N^{-k} h M^n \pi)| \leq 2L^{1-\beta(M,N)}.$$

Since $f(x) = 2L^{1-x}$ is a decreasing function, if it is true for $\beta(M, N) \geq 1$, then it must also be true for some $\alpha(M, N) < 1$. Then, letting

$$\alpha(M, N) = \begin{cases} \beta(M, N) & \beta(M, N) < 1 \\ \frac{1}{2} & \beta(M, N) \geq 1 \end{cases},$$

it follows

$$\sum_{n=0}^{L-1}\prod_{k=1}^{\infty}|\cos(N^{-k}hM^{n}\pi)| \leq 2L^{1-\alpha(M,N)}.$$

□

Lemma 2.11. *Let $N, t \in \mathbb{N}$. Let $D = \{\epsilon_0, ..., \epsilon_{d-1}\} \subset \mathbb{Z}_N$ (and $d \geq 2$). Then for any $j \in \mathbb{N}$*

$$\left|\frac{1}{d}\sum_{k=0}^{d-1} e^{2\pi i \frac{t}{N^j}\epsilon_k}\right| \leq \left|\cos\left(\pi t \frac{|\epsilon_a - \epsilon_b|}{N^j}\right)\right|$$

where $\epsilon_a, \epsilon_b \in D$ are such that

$$\left|e^{2\pi i \frac{t}{N^j}\epsilon_a} + e^{2\pi i \frac{t}{N^j}\epsilon_b}\right| = \max_{l,m\in\{0,\dots,d-1\}, l\neq m}\left|e^{2\pi i \frac{t}{N^j}\epsilon_l} + e^{2\pi i \frac{t}{N^j}\epsilon_m}\right|.$$

Proof.

$$\begin{aligned}
\left|\frac{1}{d}\sum_{k=0}^{d-1} e^{2\pi i t \frac{t}{N^j}\epsilon_k}\right| &= \left|\frac{1}{d}\cdot\frac{1}{2(d-1)}\sum_{k=0}^{d-1}\sum_{n\neq k}\left(e^{2\pi i \frac{t}{N^j}\epsilon_k} + e^{2\pi i \frac{t}{N^j}\epsilon_n}\right)\right| \\
&\leq \frac{1}{2d(d-1)}\left(\left|e^{2\pi i \frac{t}{N^j}\epsilon_0} + e^{2\pi i \frac{t}{N^j}\epsilon_1}\right| + \dots + \left|e^{2\pi i \frac{t}{N^j}\epsilon_{d-2}} + e^{2\pi i \frac{t}{N^j}\epsilon_{d-1}}\right|\right) \\
&\leq \frac{1}{2d(d-1)}(d(d-1))\left|e^{2\pi i \frac{t}{N^j}\epsilon_a} + e^{2\pi i \frac{t}{N^j}\epsilon_b}\right| \\
&= \frac{1}{2}\cdot\left|e^{2\pi i \frac{t}{N^j}\left(\frac{\epsilon_a+\epsilon_b}{2}\right)}\right|\left|e^{2\pi i \frac{t}{N^j}\left(\frac{\epsilon_a-\epsilon_b}{2}\right)} + e^{2\pi i \frac{t}{N^j}\left(\frac{-(\epsilon_a-\epsilon_b)}{2}\right)}\right| \\
&= 1\cdot\left|\cos\left(2\pi\frac{t}{N^j}\frac{\epsilon_a-\epsilon_b}{2}\right)\right| = \left|\cos\left(\pi t\frac{|\epsilon_a-\epsilon_b|}{N^j}\right)\right|
\end{aligned}$$

□

Lemma 2.12. *Let $C_{\overrightarrow{B}}$ be a Cantor set with scale factor N and binary digit vector $D = \{\epsilon_0, ..., \epsilon_{d-1}\}$. Let $M, L, h \in \mathbb{N}$ with $M > 1$ and $M \not\sim N$. Let $\alpha(d, M)$ be defined as in Lemma 2.10 and let $\delta = \frac{\alpha(d,M)}{3}$. Then the set of $x \in C_{\overrightarrow{B}}$ such that*

$$\left|\sum_{n=0}^{L-1} e(hM^n x)\right| \geq L^{1-\delta}$$

has $\mu_{\overrightarrow{B}}$-measure of at most $6L^{-\delta}$.

Proof. Adapted from Lemma 3 of [24]. Note, $e(\lambda x)$ is defined and continuous on the interval $0 \leq x \leq 1$. Let Z_t be the set of non-negative integers less than N^t containing only digits in D in their base N expansion. Therefore, by the invariance equation as applied to the push-forward measure $\mu_{\overrightarrow{B}}(A) = m(F_{\overrightarrow{B}}(A))$ where m is Lebesgue measure, we can calculate

$$\int_{x\in C_{\overrightarrow{B}}} e(\lambda x)d\mu_{\overrightarrow{B}} = \lim_{t\to\infty} d^{-t}\sum_{z\in Z_t} e(\lambda N^{-t}z) = \lim_{t\to\infty}\prod_{j<t}\left(\frac{1}{d}\sum_{i=0}^{d-1} e(N^{-j}\lambda\epsilon_i)\right).$$

The details of the above calculation are given in [24].

Let $\epsilon_a, \epsilon_b \in D$ be such that

$$\left|e^{2\pi i \frac{t}{N^j}\epsilon_a} + e^{2\pi i \frac{t}{N^j}\epsilon_b}\right| = \max_{l,m\in\{0,\dots,d-1\}, l\neq m}\left|e^{2\pi i \frac{t}{N^j}\epsilon_l} + e^{2\pi i \frac{t}{N^j}\epsilon_m}\right|.$$

Let $r = |\epsilon_a - \epsilon_b|$. Note, $r \in \mathbb{N}$, since $|D| \geq 2$ and contains only integers. Therefore, following from Lemma 2.11,

$$\left| \int_{x \in C_{\vec{B}}} e(\lambda x) d\mu_{\vec{B}} \right| \leq \prod_{j=0}^{\infty} |\cos(N^{-j} \lambda r \pi)|.$$

Further, $|z|^2 = z\overline{z}$ and $\overline{e(hM^n x)} = e(h(-M^n)x)$, so for any $L \in \mathbb{N}$

$$\begin{aligned} \int_{x \in C_{\vec{B}}} \left| \sum_{n=0}^{L-1} e(hM^n x) \right|^2 d\mu_{\vec{B}} &= \left| \sum_{m=0}^{L-1} \sum_{n=0}^{L-1} \int_{x \in C_{\vec{B}}} e(h(M^n - M^m)x) d\mu_{\vec{B}} \right| \\ &\leq \sum_{m=0}^{L-1} \sum_{n=0}^{L-1} \left| \int_{x \in C_{\vec{B}}} e(h(M^n - M^m)x) d\mu_{\vec{B}} \right| \\ &\leq \sum_{m=0}^{L-1} \sum_{n=0}^{L-1} \prod_{j=0}^{\infty} \left| \cos(N^{-j} h(M^n - M^m) r\pi) \right|. \end{aligned}$$

Consider $l = \min(m, n)$ and $k = \max(m, n) - l$, so that l and k determine m and n up to pairs. It follows,

$$\sum_{m=0}^{L-1} \sum_{n=0}^{L-1} \prod_{j=0}^{\infty} |\cos(N^{-j} h(M^n - M^m) r\pi)| \leq 2 \sum_{l=0}^{L-1} \sum_{k=0}^{L-1} \prod_{j=0}^{\infty} |\cos(N^{-j} h(M^l - 1) M^k r\pi)|.$$

When $l = 0$, all terms in the product are $\cos(0) = 1$ and the inner sum is no more than L. Otherwise, by Lemma 2.10, the inner sum is less than or equal to $2L^{1-\alpha(d,M)}$ where $\alpha(d, M) > 0$.
Therefore,

$$\int_{x \in C_{\vec{B}}} \left| \sum_{n=0}^{L-1} e(hM^n x) \right|^2 \leq 2(L + L(2L^{1-\alpha(d,M)})) = 2L + 4L^{2-\alpha(d,M)} < 6L^{2-\alpha(d,M)} = 6L^{2-3\delta}.$$

It follows that the $\mu_{\vec{B}}$-measure of $x \in C_{\vec{B}}$ such that $\left| \sum_{n=0}^{L-1} e(hM^n x) \right| \geq L^{1-\delta}$ is no more than $\frac{6L^{2-3\delta}}{L^{2(1-\delta)}} = 6L^{-\delta}$ by Chebychev's inequality. □

Our proof of the following theorem is adapted from [24], and a much stronger result has already been proven in [25]. See also [26] for a related result regarding the entropy of multiplication by an integer on $\mathbb{R}/\mathbb{Z}$.

Theorem 2.3. *Let $C_{\vec{B}}$ be a Cantor set. Then $\mu_{\vec{B}}$-almost all $x \in C_{\vec{B}}$ are normal to every base $M > 1$ such that $M \not\sim N$.*

Proof. Fix $M \in \mathbb{N}$. Let $\delta = \frac{\alpha(d,M)}{3}$, where α is defined as in Lemma 2.10. Then, $0 < \delta < 1$. Let $L_j = \lfloor e^{2\sqrt{j}} \rfloor$. Then, for $j > 1$, $L_j^{-\delta} \leq e^{-\delta\sqrt{j}}$, and $\int_0^\infty e^{-\delta\sqrt{j}} dj < \infty$, so $\sum_{j=0}^{\infty} L_j^{-\delta} < \infty$. It follows for every $\epsilon > 0$ there exists $J_\epsilon \in \mathbb{N}$ such that

$$\sum_{j=J_\epsilon}^{\infty} 6L_j^{-\delta} < \epsilon.$$

By Lemma 2.12, the sum of the $\mu_{\vec{B}}$-measures of the sets $\left\{ x : \left| \sum_{n=0}^{L_j-1} e(M^n x) \right| \geq L_j^{1-\delta} \right\}$ for some $j \geq J$ goes to 0 as $J \to \infty$. Therefore, for $\mu_{\vec{B}}$-almost all x there exists J_x such that

$$\left| \sum_{n=0}^{L_j-1} e(M^n x) \right| < L_j^{1-\delta} \text{ for all } j \geq J_x,$$

so $\left|\sum_{n=0}^{L_j-1} e(M^n x)\right| = o(L_j)$ as $L_j \to \infty$.

Further, for every L there exists j_L such that $L_{j_L} \leq L < L_{j_L+1}$. In addition,

$$\left|\sum_{n=0}^{L-1} e(M^n x) - \sum_{n=0}^{L_{j_L}-1} e(M^n x)\right| \leq L - L_{j_L}.$$

Note, $L - L_{j_L} = o(L)$ as $L \to \infty$, because L_j grows slower than a geometric series. Therefore,

$$\sum_{n=0}^{L-1} e(M^n x) = o(L)$$

as $L \to \infty$ for $\mu_{\vec{B}}$-almost all x.

Note, the set of x such that $\sum_{n=0}^{L-1} e(M^n x) \neq o(L)$ for a fixed $M \in \mathbb{N}$ has $\mu_{\vec{B}}$-measure 0, and the sets of possible h and M are countable. Then, the set of x such that $\sum_{n=0}^{L-1} e(M^n x) \neq o(L)$ for any $M \in \mathbb{N}$ has $\mu_{\vec{B}}$-measure 0 because it is the union of a countable number of sets with $\mu_{\vec{B}}$-measure 0. For $\mu_{\vec{B}}$-almost all x, $\sum_{n=0}^{L-1} e(hM^n x) = o(L)$ for all $M \in \mathbb{N}$ such that $M \not\sim N$. By Weyl's criterion in [10], then for $\mu_{\vec{B}}$-almost all x and any fixed $M \not\sim N$, the fractional part of the sequence $\{M^n x\}_{n=1}^{\infty}$ is uniformly distributed. Therefore, $\mu_{\vec{B}}$-almost all x are normal to all bases $M > 1$ such that $M \not\sim N$. □

Theorem 2.4. *Let M, N be scale factors of the Cantor sets $C_{\vec{B}}, C_{\vec{C}}$, respectively. If $M \not\sim N$, then $C_{\vec{B}} \cap C_{\vec{C}}$ is $\mu_{\vec{B}}$-almost empty and $\mu_{\vec{C}}$-almost empty.*

Proof. Let M, N be scale factors of the Cantor sets $C_{\vec{B}}, C_{\vec{C}}$, respectively, and $M \not\sim N$. By Theorem 2.3, $\mu_{\vec{B}}$-almost all of the elements in $C_{\vec{B}}$ are normal in base M. Since normal numbers contain all of the digits, it follows $\mu_{\vec{B}}$-almost all of the elements in $C_{\vec{B}}$ are not in $C_{\vec{C}}$. Similarly, $\mu_{\vec{C}}$-almost all of the elements in $C_{\vec{C}}$ are normal in base N and likewise are not elements of $C_{\vec{B}}$. Therefore, it follows their intersection is $\mu_{\vec{C}}$-almost empty and $\mu_{\vec{B}}$-almost empty. □

Note, normality is a stronger condition than necessary to show an element is not in any Cantor set with scale factor N. In fact, it only must have every digit appear at least once.

Corollary 2.7. *For every Cantor set $C_{\vec{B}}$, there exist irrational numbers in $C_{\vec{B}}$ normal to every base M such that $M \not\sim N$.*

Proof. There are uncountably many elements in $C_{\vec{B}}$, however, there are only countably many rationals. Further, by Theorem 2.3, $\mu_{\vec{B}}$-almost all of the elements in $C_{\vec{B}}$ are normal to multiplicatively independent bases. It follows that $\mu_{\vec{B}}$-almost all of the elements in $C_{\vec{B}}$ must be irrational and normal in multiplicatively independent bases. □

2.3.5 Using Samples

Lemma 2.13. *Let $C_{\vec{B}}$ be a Cantor set with scale factor N. Let $D_1, ..., D_k$ be all possible binary digit vectors of N with cardinality 2, and associate $\vec{B}_1, ..., \vec{B}_k$ with such binary digit vectors. Then for any set of irrationals $\{x_i \in C_{\vec{B}_i} : i = 1, ..., k\}$, D can be uniquely determined from $\{(x_i, F_{\vec{B}}(x_i))\}_{i=1}^{k}$. In particular, $D = \bigcup_{i \in A} D_i$ where $A = \{i \mid F_{\vec{B}}(x_i) \in \mathbb{Q}^c\}$*

Proof. Let $D_1, ..., D_k$ be all the digits sets of cardinality 2 for scale factor N. Let $\vec{B} = (b_0, ..., b_{N-1})$ be the binary representation of $C_{\vec{B}}$. Consider $x_i \in C_{\vec{B}_i}$. Since x_i is irrational and $|D_i| = 2$, the decimal expansion of x_i in base N must contain both digits in D_i. Then, $x_i \in C_{\vec{B}}$ if and only if $D_i \subseteq D$. Then, by Lemma 2.7, $F_{\vec{B}}(x_i)$ is irrational if and only if $D_i \subseteq D$. Let $A = \{i \mid F_{\vec{B}}(x_i) \in \mathbb{Q}^c\}$. Then, $\bigcup_{i \in A} D_i \subseteq D$.

Next, consider $\epsilon_1 \in D$. Since $\|\overrightarrow{B}\| \geq 2$, there exists an $\epsilon_2 \in D$, $\epsilon_1 \neq \epsilon_2$. Further, there exists a j such that $D_j = \{\epsilon_1, \epsilon_2\}$. Then, $D_j \subseteq D$, $x_j \in C_{\overrightarrow{B}}$ and $F_{\overrightarrow{B}}(x_j)$ will be irrational by Lemma 2.7. It follows $j \in A$, and therefore $D_j \subseteq \bigcup_{a \in A} D_a$. Thus, $\epsilon_1 \in \bigcup_{a \in A} D_a$. Since ϵ_1 is arbitrary, it follows $\bigcup_{a \in A} D_a = D$ where $A = \{i \mid F_{\overrightarrow{B}}(x_i) \in \mathbb{Q}^c\}$. □

Theorem 2.5. *Given K, there exists a constant $M := M(K)$ such that there exists $\{x_i\}_{i=1}^{M(K)} \subset (0,1)$ which is a set of uniqueness for $\mathscr{G}_K$. The constant $M(K) = O(K^3)$.*

Proof. Let $F_{\overrightarrow{B}}$ be a CDF with scale factor $N \leq K$. We proceed with $C_{N,D}$ (instead of $C_{\overrightarrow{B}}$ notation as it makes the proof clearer).

For every M, $3 \leq M \leq K$, there exist $\binom{M}{2} = \frac{M(M-1)}{2}$ unique Cantor sets C_{M,D_i} such that $|D_i| = 2$. Further, by Corollary 2.7, each of these Cantor sets contain an irrational element (in fact, almost all elements) which is normal to all bases multiplicatively independent of M. Then, for every M and $|D_i| = 2$, there exists $x \in C_{M,D_i} \cap \mathbb{Q}^c$ such that $\forall L \not\sim M, L \leq K$, x in base L has all possible digits in its representation.

Choose one such element for each C_{M,D_i}, and denote it x_{M,D_i}; let

$$S_1 = \{x_{M,D_i} \mid 3 \leq M \leq K, |D_i| = 2\}$$

Note, $|S_1| \leq \frac{K(K-1)(K-3)}{2}$.

For any $M \not\sim N$ and D_i with $|D_i| = 2$, x_{M,D_i} contains every possible digit in N. Then, since any $x \in C_{\overrightarrow{B}}$ cannot contain every digit in base N, $x_{M,D_i} \notin C_{\overrightarrow{B}}$. It follows from Corollary 2.6 that $F_{\overrightarrow{B}}(x_{M,D_i}) \in \mathbb{Q}$.

Suppose now that there exists a CDF with scale factor M passing through all of the points $\{(x_i, F_{\overrightarrow{B}}(x_i)) \mid x_i \in S_1\}$. Since this is true for all D_i corresponding with M, by Lemma 2.13 the binary digit vector of the CDF is be empty, a contradiction. Thus, no such CDF exists and M can be eliminated as a scale factor.

Thus, all possible scale factors remaining are multiplicatively dependent to N. Therefore, there is a fixed $J \in \mathbb{N}$ such that for each possible scale factor N', $N' = J^{L_{N'}}$ for some $L_{N'} \in \mathbb{N}$. Further, by Lemma 2.13, for each N' there exists at most one binary digit vector, $\overrightarrow{B'}$, such that $F_{\overrightarrow{B}}(x) = F_{\overrightarrow{B'}}(x)$ for all $x \in S_1$.

By Proposition 2.6, for all $L_1, L_2 \in \mathbb{N}$ and $\overrightarrow{B}_1, \overrightarrow{B}_2$ with scale factors J^{L_1}, J^{L_2}, respectively, either $F_{\overrightarrow{B}_1} = F_{\overrightarrow{B}_2}$ or only one agrees with $\left\{\left(\frac{m}{J^{L+M}}, F_{\overrightarrow{B}}\left(\frac{m}{J^{L+M}}\right)\right)\right\}_{m=1}^{J^{L+M}-1}$. Note that for any J, $2 \leq J \leq K$, there is $L_K \in \mathbb{N}$ such that $J^{L_K} \leq K < J^{L_K+1}$. The set of rational numbers expressible with denominator J^{2L_K} includes the set of rational numbers expressible with denominator J^L for $L \leq 2L_K$. Note, $J^{L_1}, J^{L_2} \leq K$ implies $L_1 + L_2 \leq 2L_K$.

Since $J^{L_K} \in \{2, 3, .., K\}$, sampling at

$$S_2 = \left\{\frac{m}{M}\right\}_{m=1, M \in \{2^2, 3^2, 4^2, \ldots, K^2\}}^{M-1}$$

is sufficient to differentiate all mutliplicatively dependent bases no more than K. Hence sampling at $\left\{\frac{m}{(J^{L_K})^2}\right\}_{m=1}^{(J^{L_K})^2-1}$ is sufficient for differentiating all bases multiplicatively dependent to J. It follows, of the remaining CDFs, only CDFs equivalent to $F_{\overrightarrow{B}}$ will pass through all the points $\{(x, F_{\overrightarrow{B}}(x)) \mid x \in S_2\}$, and all non-equivalent CDFs can be eliminated.

For any remaining CDFs F, $F = F_{\overrightarrow{B}}$. Thus, $S = S_1 \cup S_2$ is sufficient to reconstruct $F_{\overrightarrow{B}}$.

Finally, we note that since $|S_2| \leq 1^2 + 2^2 + 3^2 + \ldots + K^2 = \frac{K(K+1)(2K+1)}{6}$, $|S| \leq \frac{K(K-1)(K-3)}{2} + \frac{K(K+1)(2K+1)}{6} = \frac{5}{6}K^3 - \frac{3}{2}K^2 + \frac{5}{3}K = O(K^3)$.

□

Corollary 2.8. *There exists a set of uniqueness for $\mathscr{G}_K$ with sample complexity $O(K^3)$.*

Remark 2.3. *CDFs equivalent to $F_{\overrightarrow{B}}$ will not be eliminated by the algorithm described in Theorem 2.5, which only eliminates CDFs which do not pass through all the points. Then, the algorithm will produce all* *equivalent CDFs with scale factor less than K, which includes the CDF with the smallest possible scale factor, and the smallest possible scale factor can be determined.*

Remark 2.4. *Since CDFs are equivalent only if their underlying Cantor sets are equal, the algorithm also reconstructs the underlying Cantor set $C_{\vec{B}}$.*

3 Conclusion and Future Research

With a upper scale factor bound of K, and $O(K^3)$ points, a CDF of any Cantor set can be completely reconstructed. While an optimal order on the number of points has not been determined, we have shown that $\lfloor \frac{K}{2} \rfloor$ points is insufficient. Further, many of the points sampled in Theorem 2.5, those in S_1, are not specific, and must be chosen with *true randomness*.

If the scale factor N is known, then $N-1$ well chosen points is enough to determine the binary digit vector D. However, this is not the minimum number. A future research question would be to determine the minimum number of points necessary to determine the binary digit vector.

4 Acknowledgements

Allison Byars, Evan Camrud, Sarah McCarty, and Keith Sullivan were supported in part by the National Science Foundation under award #1457443.

Steven Harding and Eric Weber were supported in part by the National Science Foundation and the National Geospatial-Intelligence Agency under award #1830254.

Eric Weber was supported in part by the National Science Foundation under award #1934884.

Conflict of interest: Eric Weber is a member of the Editorial Board of Demonstratio Mathematica and was not involved in the review process of this article.

References

[1] J. Benedetto and P.J.S.G. Ferriera, editors. *Modern Sampling Theory*. Birkhauser, 2001.

[2] A. Aldroubi and K. Gröchenig. Nonuniform sampling and reconstruction in shift-invariant spaces. *SIAM Rev.*, 43(4):585–620 (electronic), 2001.

[3] John E. Hutchinson. Fractals and self-similarity. *Indiana Univ. Math. J.*, 30(5):713–747, 1981.

[4] Robert DiMartino and Wilfredo Urbina. On cantor-like sets and cantor-lebesgue singular functions, 2014.

[5] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013.

[6] Nina N. Huang and Robert S. Strichartz. Sampling theory for functions with fractal spectrum. *Experiment. Math.*, 10(4):619–638, 2001.

[7] Robert J. Ravier and Robert S. Strichartz. Sampling theory with average values on the Sierpinski gasket. *Constr. Approx.*, 44(2):159–194, 2016.

[8] John E. Herr and Eric S. Weber. Fourier series for singular measures. *Axioms*, 6(2:7):13 ppg., 2017. http://dx.doi.org/10.3390/axioms6020007.

[9] Richard Oberlin, Brian Street, and Robert S. Strichartz. Sampling on the Sierpinski gasket. *Experiment. Math.*, 12(4):403–418, 2003.

[10] Hermann Weyl. Über die Gleichverteilung von Zahlen mod. Eins. *Math. Ann.*, 77(3):313–352, 1916.

[11] Soheil Kolouri, Se Rim Park, and Gustavo K Rohde. The radon cumulative distribution transform and its application to image classification. *IEEE transactions on image processing*, 25(2):920–934, 2016. 10.1109/TIP.2015.2509419.

[12] Se Rim Park, Soheil Kolouri, Shinjini Kundu, and Gustavo K. Rohde. The cumulative distribution transform and linear pattern classification. *CoRR*, abs/1507.05936:38 pages, 2015.

[13] Michael F. Barnsley. Fractal functions and interpolation. *Constr. Approx.*, 2(4):303–329, 1986.

[14] Peter Massopust. *Interpolation and approximation with splines and fractals*. Oxford University Press, Oxford, 2010.
[15] Steven N. Harding and Alexander W. N. Riasanovsky. Moments of the weighted Cantor measures. *Demonstr. Math.*, 52(1):256–273, 2019.
[16] Abhay Parvate and A. D. Gangal. Calculus on fractal subsets of real line. I. Formulation. *Fractals*, 17(1):53–81, 2009.
[17] A. Khalili Golmankhaneh and A. Fernandez. Random variables and stable distributions on fractal cantor sets. *Fractal and Fractional*, 3(2)(31), 2019. https://doi.org/10.3390/fractalfract3020031.
[18] Alireza K. Golmankhaneh and Cemil Tunç. Stochastic differential equations on fractal sets. *Stochastics*, 92(8):1244–1260, 2020.
[19] B. Broxson. The Kronecker product. Master's thesis, University of North Florida, 2006. UNF Graduate Theses and Dissertations 25.
[20] Yuval Peres and Boris Solomyak. Self-similar measures and intersections of Cantor sets. *Trans. Amer. Math. Soc.*, 350(10):4065–4087, 1998.
[21] Carlos Gustavo T. de A. Moreira and Jean-Christophe Yoccoz. Stable intersections of regular Cantor sets with large Hausdorff dimensions. *Ann. of Math. (2)*, 154(1):45–96, 2001.
[22] Wolfgang M. Schmidt. Über die Normalität von Zahlen zu verschiedenen Basen. *Acta Arith.*, 7:299–309, 1961/1962.
[23] A. D. Pollington. The Hausdorff dimension of a set of normal numbers. II. *J. Austral. Math. Soc. Ser. A*, 44(2):259–264, 1988.
[24] J. W. S. Cassels. On a problem of Steinhaus about normal numbers. *Colloq. Math.*, 7:95–101, 1959.
[25] Michael Hochman and Pablo Shmerkin. Equidistribution from fractal measures. *Inventiones Mathematicae*, 202(1):427–479, 2015.
[26] Bernard Host. Nombres normaux, entropie, translations. *Israel J. Math.*, 91(1-3):419–428, 1995.

5 Appendix

Algorithm 1 Finding $\|\overrightarrow{B}\|$ and the first nonzero digit of $\overrightarrow{B}$

Since $g_{\overrightarrow{B}}(k) = \sum_{j=0}^{k-1} b_j$ we have that $g(2m-2) = \sum_{j=0}^{2m-3} b_j$ such that we know $g(2m-2)$ recursively dependent on sample values and their determination of each b_j. We further recall that ℓ is a large enough integer, and that the smallest positive ℓ such that $2^{\ell+1} > N-1$ is sufficient.

Initialize $m = 1$
while $g_{\overrightarrow{B}}(2m-2) = 0$ **do**
  Sample $F_{\overrightarrow{B}}$ at

$$x = \frac{2m}{N^{\ell+1}} + \sum_{n=1}^{\ell} \frac{2m-1}{N^n}.$$

  if $F_{\overrightarrow{B}}(x) = 0$ **then**
    $b_{2m-2} = 0$
    $b_{2m-1} = 0$
    $m = m + 1$
  else
    if $F_{\overrightarrow{B}}(x) = 1/\overrightarrow{B}$ for some $2 \leq \overrightarrow{B} \leq N-1$ **then**
      $\|\overrightarrow{B}\| = \overrightarrow{B}$
      $b_{2m-2} = 1$
      $b_{2m-1} = 0$
    else if $F_{\overrightarrow{B}}(x) = 1/\overrightarrow{B}^{\ell+1}$ for some $2 \leq \overrightarrow{B} \leq N-1$ **then**
      $\|\overrightarrow{B}\| = \overrightarrow{B}$
      $b_{2m-2} = 0$
      $b_{2m-1} = 1$
    else
      There is an integer $2 \leq \overrightarrow{B} \leq N-1$ such that

$$F_{\overrightarrow{B}}(x) = \frac{2}{\overrightarrow{B}^{\ell+1}} + \sum_{n=1}^{\ell} \frac{1}{\overrightarrow{B}^n}$$

      $\|\overrightarrow{B}\| = \overrightarrow{B}$
      $b_{2m-2} = 1$
      $b_{2m-1} = 1$
    end if
    Break
  end if
end while
if $N = 2m$ **then**
  Return
else if $N = 2m+1$ **then**
  if $g_{\overrightarrow{B}}(N-1) = \|\overrightarrow{B}\|$ **then**
    $b_{N-1} = 0$
  else
    $b_{N-1} = 1$
  end if
else
  Proceed to Algorithm 2
end if

Algorithm 2 Finding the remaining digits of $\overrightarrow{B}$

Initialize $M = \lfloor \frac{N}{2} \rfloor$
for $\mathfrak{m} = 1, 2, ..., M - m$ **do**
  if $g_{\overleftarrow{B}}(2m-2) = \|\overrightarrow{B}\| - 1$ **then**
    $b_{N-1-(2\mathfrak{m}-2)} = 0$
    $b_{N-1-(2\mathfrak{m}-1)} = 0$
  else
    Sample $F_{\overleftarrow{B}}$ at

$$x = \frac{2\mathfrak{m}}{N^{\ell+1}} + \sum_{n=1}^{\ell} \frac{2\mathfrak{m}-1}{N^n}$$

    if $F_{\overleftarrow{B}}(x) = \frac{g_{\overleftarrow{B}}(2\mathfrak{m}-2)}{\|\overrightarrow{B}\|}$ **then**
      $b_{N-1-(2\mathfrak{m}-2)} = 0$
      $b_{N-1-(2\mathfrak{m}-1)} = 0$
    else if $F_{\overleftarrow{B}}(x) = \frac{g_{\overleftarrow{B}}(2\mathfrak{m}-2)+1}{\|\overrightarrow{B}\|}$ **then**
      $b_{N-1-(2\mathfrak{m}-2)} = 1$
      $b_{N-1-(2\mathfrak{m}-1)} = 0$
    else if

$$F_{\overleftarrow{B}}(x) = \frac{g_{\overleftarrow{B}}(2\mathfrak{m}-2)}{\|\overrightarrow{B}\|^{\ell+1}} + \sum_{n=1}^{\ell} \frac{g_{\overleftarrow{B}}(2\mathfrak{m}-3)}{\|\overrightarrow{B}\|^n}$$

    then
      $b_{N-1-(2\mathfrak{m}-2)} = 0$
      $b_{N-1-(2\mathfrak{m}-1)} = 1$
    else
      $b_{N-1-(2\mathfrak{m}-2)} = 1$
      $b_{N-1-(2\mathfrak{m}-1)} = 1$
    end if
  end if
end for
if N is odd **then**
  if

$$\sum_{n=0}^{2m-1} b_n + \sum_{n=2m+1}^{N-1} b_n = \|\overrightarrow{B}\|$$

  then
    $b_{2m} = 0$
  else
    $b_{2m} = 1$
  end if
end if