Optimal Codebook and its Applications

By

VINH NGUYEN

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

MATHEMATICS

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Qinglan Xia, Chair

Bruno Nachtergaele

Adam Jacob

Committee in Charge

2024

To my parents

Contents

Abstract

The process of encoding and decoding an original dataset is fundamental to a variety of applications, including lossy compression in information theory and autoencoders in machine learning. In this thesis, we propose a framework to study this process through the concepts of codepage and codebook. Given a data set N and a space of codes F, an ϵ-codepage A is a subset of F that can encode and then decode N with an acceptable error of ϵ. A codebook, in turn, is a function that assigns each acceptable error ϵ a corresponding codepage in F. By extending the concept of the external covering number for a totally bounded set to an arbitrary set, we establish the existence criteria for both codepages and codebooks in more general settings.

We also explore the problem of optimality, specifically the task of identifying codepages or codebooks that minimize certain cost functions. We define the cost of a codebook as the integral of a cost function applied to the individual codepages. Two types of cost functions are proposed for the codepages. In the first type, the cost of each codepage is determined by the cost of the codes it contains, which is represented by a Borel measure on the space of codes. For the case where N is totally bounded, we provide criteria for the existence of an optimal codebook. Assuming further that N belongs to a Heine-Borel space, we derive a formula for the minimum cost. In the second type, the cost of each codepage takes into account the combined costs of encoding, decoding, and the error in the encoding-decoding process. We define a topology on the set of codepages based on the Hausdorff distance and use it to establish criteria for the existence of an optimal codebook under this cost function. Throughout the thesis, we illustrate these concepts by applying them to the MNIST dataset.

Acknowledgments

I would like to begin by expressing my heartfelt gratitude to my advisor, Qinglan Xia, for his guidance and support throughout my time at the University of California, Davis. Without his patience, generosity, and expertise, I would not have been able to complete this journey.

I am also deeply grateful to Bruno Nachtergaele and Adam Jacob for reviewing my thesis and providing invaluable feedback, as committee members. My thanks go to everyone at the UC Davis Mathematics Department for their resources and the assistance they provided during my studies.

To my family, words cannot fully convey my gratitude for your unwavering love and support throughout the years. Thank you, Mom! Thank you, Dad! And thank you, Brother! I would not be where I am today without you.

CHAPTER 1

Introduction

1.1. Motivation from information theory and machine learning

1.1.1. Information Theory. Information Theory is a field that is chiefly concerned with the mathematics of measuring, storing, and communicating information [7]. Examples of this include transferring sound over telephone lines, cell reproduction processes based on DNA information, or storing data on disks. At the core of both of these is an encoding-decoding process: we want to send a collection of data through a channel by first using an 'encoder' to map the data points into a space of 'codes', which will then be transferred to a receiver, where a 'decoder' reconstructs the original data from the codes, i.e., mapping the codes to the original data space. For storing information this means compressing data to a smaller size, then decompressing it when used. For communicating information this means turning original data into a form that is better suited for communicating over a noisy channel, i.e. one that has the potential to corrupt data traveling through it.

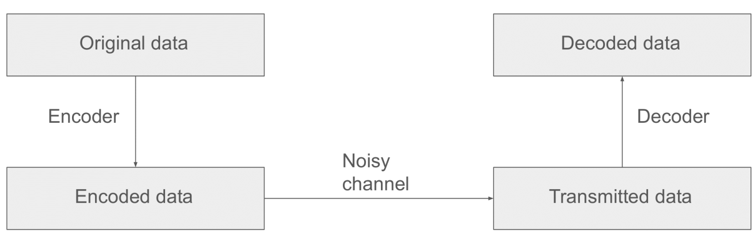

FIGURE 1.1. At the core of Information Theory is the process of encoding-transmitting-decoding information.

Lossy compression. One form of data compression is called lossy compression. Suppose we want to compress a data set (X, P), where P is a probability measure representing the frequency of data points in X. In lossy compression, the set E of codes has a smaller size than that of the source data X, and thus not all data points can be uniquely encoded, i.e. there is no 1-1 mapping from X to E when $|X| > |E|$. In this case, an alternative is to encode a subset S of X with $|S| \leq |E|$, while leaving the data points outside of S possibly without a code. We want the probability that a data point without an associated code in E to be small. Formally, let $\delta \in [0, 1]$ represents the risk factor of the compression process, i.e. the largest acceptable probability that an original data point cannot be uniquely encoded. Notice that this is equivalent to $d_P(X, S) \leq \delta$, where d_p is the symmetric distance defined in Definition 1.2.7. The compression process must then guarantee that $P(S) \geq 1 - \delta$, and the set of codes E must have at least as many elements as S to ensure that S is uniquely encoded. Consequently, there are many ways to encode X with a risk factor δ. For each of those ways, the corresponding set of codes E can be thought of as a 'δ-codepage', a set of codes that guarantee X can be encoded with risk factor δ. As an example, suppose $X^* = \{a, b, c, d, e\}$, $P = \{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\}$. For $\delta = 0.125$, hence $P(S) \geq 0.875$, $S = X^*, \{a, b, c\}, \{a, b, c, d\}, \{a, b, c, e\}, \{a, b, d, e\}$ are all possible subsets of X^* that can be uniquely encoded. If we choose a binary code, the corresponding 0.125-codepage for X^* for those S is shown in Table 1.1.

TABLE 1.1. Examples of 0.125-codepages for X.

S	0.125-codepage with binary codes
$\{a, b, c\}$	$\{0, 1\}$
$\{a, b, c, d\}$	$\{00, 01, 10, 11\}$
$\{a, b, c, e\}$	$\{00, 01, 10, 11\}$
$\{a, b, d, e\}$	$\{00, 01, 10, 11\}$
$\{a, b, c, d, e\}$	$\{000, 001, 010, 100, 011\}$

Let S_δ be one with the smallest size among such sets S, i.e.

$$S_\delta \in \text{argmin}\{|S| : S \text{ measurable subsets of } X, P(S) \geq 1 - \delta\}.$$

Notice that as δ decreases, the size of $|S_\delta|$ must necessarily increase. Specifically, for a finite set X, the set S_δ can be constructed by progressively adding elements from X in order of their decreasing probability. This process continues until the probability of S, $P(S)$, satisfies the condition $P(S) \leq 1 - \delta$. In the case of the set X^* constructed earlier, Table 1.2 illustrates the values of S_δ for various choices of δ, providing a clearer view of how S_δ evolves with different values of δ.

Table 1.2. Relation between δ and S_δ.

δ	S_δ	$P(S_\delta)$	Some other possible S's
1	$\varnothing$	0	$\{b,c\},\{a\},\{a,b,c\},\{b,c,d,e\}$
0.5	$\{a\}$	$\frac{1}{2}$	$\{a,c\},\{b,c,d,e\},\{a,b,e\}$
0.25	$\{a,b\}$	$\frac{1}{2}+\frac{1}{4}$	$\{a,b,e\},\{a,b,c\},\{c,d,e\}$
0.125	$\{a,b,c\}$	$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}$	$\{a,b,c,d\}$
0.0625	$\{a,b,c,d\}$	$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}$	$\{a,b,c,d,e\}$
0	$\{a,b,c,d,e\}$	1	None

For any set of data points X, the *raw bit of content* [7, Equation (4.15)] of X, $H_0(X) := \log_2|X|$ represents a lower bound for the number of yes-no questions that are guaranteed to identify an element of X, and is used as a measurement of information content of X. For any δ, $H_\delta(X) := \log_2|S_\delta|$ is called the *essential bit content* [7, Equation (4.19)] of X.

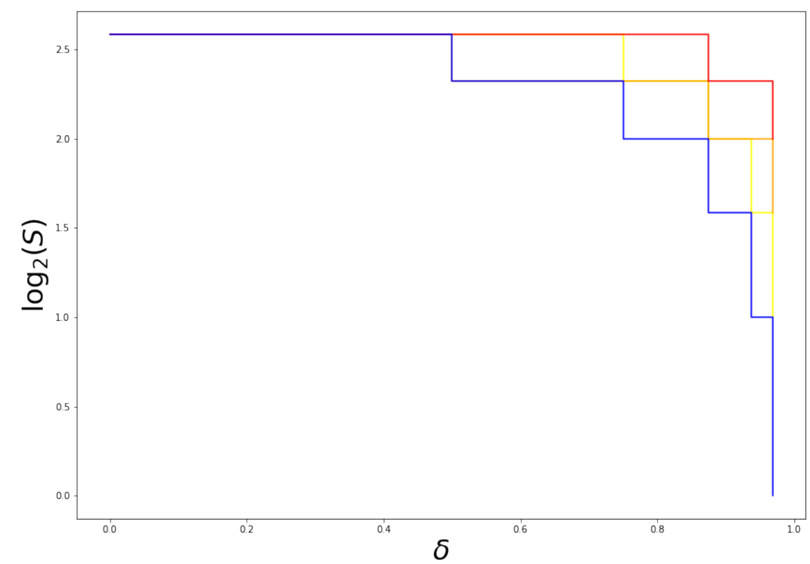

Figure 1.2. Blue curve illustrates relation between δ and $H_\delta(X)$. Yellow, orange, and red curves illustrate $H_0(S)$ for another set of possible S.

In general, since each codepage can only encode a subset of X up to a certain risk factor δ, to encode a larger subset of X we need a "finer" codepage. Hence for any X it is natural to have a 'codebook' which is a collection of codepages. We think of a 'codebook' as a formula that assigns with risk factor an associated codepages. Since there are many possible codepages for each risk factor δ, there are more than one possible codebooks.

Noisy Channel Coding. One simple example of noisy channel coding is the repetition code. Assume that we want to transmit a source message that consists entirely of numbers 0 and 1 over a noisy channel. To represent the noise level of the channel, we let $p \in [0, 1]$ be the probability that the channel will flip the value of transmitted data either from 0 to 1, or 1 to 0. We first encode the source message by a repetition code, where each digit is repeated N times. To decode the transmitted message, a decoder would then look at the received message N digits at the time and take a majority vote. Table 1.3 illustrates an example of repetition code for noisy channel coding with the message '011001', with $N = 3$.

TABLE 1.3. An example of repetition code for noisy channel coding with the message '011001', with $N = 3$.

Original message	0	1	1	0	0	1
Encoded message	000	111	111	000	000	111
Transmitted message	010	011	110	001	010	100
Decoded message	0	1	1	0	0	0

Assuming N is odd, by direct computation the error probability $p_E(N, p)$ can be computed as

$$p_E(N, p) = \sum_{n=(N+1)/2}^{N} \binom{N}{n} p^n (1-p)^{N-n}.$$

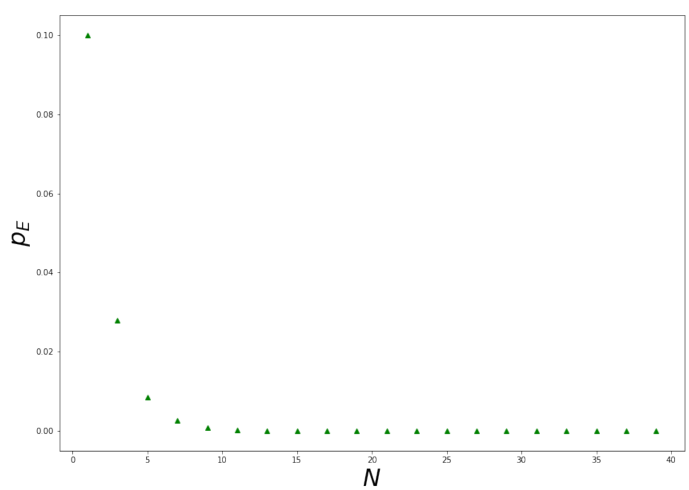

FIGURE 1.3. In this graph the x-axis is the number of repetitions N, and y-axis is the error probability p_E, for $p = 0.1$.

Figure 1.3 shows that for any fixed p, as N increases, p_E tends to decrease. Notice that this means for any given acceptable error $p_E^* > 0$, there are more than one repetition code (more than one N) that guarantees the coding process satisfies the error term p_E^*, i.e. $p_E(N,p) \leq p_E^*$. Again, the values N such that $p_E(N,p) \leq p_E^*$ represent the 'p_E^*-codepages', i.e. the associated set of codes must have length N times the length of the elements in the original source. The integer $N^* = \min\{N : p_E(N,p) \leq p_E^*\}$ represents an 'optimal codepage', one that minimizes the necessary length of the codes in the codepage. Any function $p_E^* \mapsto N$ such that $p_E(N,p) \leq p_E^*$ represents a 'codebook', a formula for choosing a codepage for different error terms.

Autoencoder. [4] Similar ideas can also be found in Machine learning, in particular Autoencoder. An autoencoder is an unsupervised neural network. Consider an original set of data that belongs to a finite-dimensional metric space (X, d) and is equipped with a probability measure μ. An autoencoder consists of a finite dimensional latent space Y and 2 parametrized functions: $f_\theta : X \to Y$ (encoder), and $g_\theta : Y \to X$(decoder).

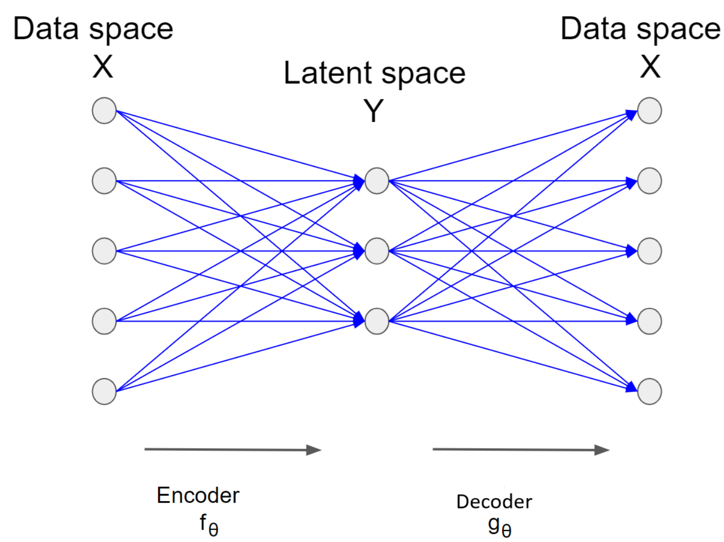

FIGURE 1.4. Autoencoder.

In a simple autoencoder, the machine learns by adjusting the parameter θ in order to minimize the expected difference between the original data points x and the reconstructed data $g_\theta \circ f_\theta(x)$; $\mathbb{E}L\big(x, g_\theta \circ f_\theta(x)\big)$, where L is some loss function that measures the difference between the original and reconstructed data. In other words, we want to solve the problem

$$\underset{\theta}{\operatorname{argmin}}\, \mathbb{E}L\big(x, g_\theta \circ f_\theta(x)\big). \tag{1.1}$$

The idea of autoencoder is that by restricting the latent space Y in some way, the machine is forced to learn the deeper structure of the original data source rather than simply copying the original data one-to-one. The most straightforward way of doing this is making $\dim(Y) < \dim(X)$, which is called *undercomplete autoencoder.*

Similar to the examples of lossy compression and noisy channel coding, we observe an encoding-decoding process with an error term that we want to minimize. One challenge in using an autoencoder is determining an appropriate dimension for the latent space Y. If the dimension of Y is too small, the error term might be excessively large, leading to a poor reconstruction of the original data. Conversely, if the dimension of Y is too large, the autoencoder may struggle to learn the underlying structure of the data. For example, setting $\dim(Y) = \dim(X)$ would most likely result in $g_\theta \circ f_\theta$ behaving just like the identity function. Thus, an interesting question arises: how do we determine a suitable $\dim(Y)$ that is neither too small nor too large? To this end, we may apply a similar procedure discussed in lossy compression and noisy channel coding. We first set a limit for the error, specifying $\epsilon > 0$, and require that

$$\underset{\theta}{\operatorname{argmin}}\, \mathbb{E}L\big(x, g_\theta \circ f_\theta(x)\big) \le \epsilon.$$

In other words, there exists a pair $f_\theta : X \to Y$ and $g_\theta : Y \to X$ such that $\mathbb{E}L\big(x, g_\theta \circ f_\theta(x)\big) \le \epsilon$. Any Y where such a pair f_θ and g_θ exist is referred to as an 'ϵ-codepage', in a context similar to those previously discussed. For any $\epsilon > 0$ given, there are many such 'ϵ-codepages'. Among them, an optimal one minimizing a suitable cost function may be chosen as the latent space Y with the minimum dimension. Also, since smaller ϵ typically gives "finer" codepages, we can consider a collection of 'codepages', where each $\epsilon > 0$ has an associated 'ϵ-codepage', forming a 'codebook'. Like using a dictionary, one may find out the ϵ-codepage in the codebook that best meets their needs based on the desired error tolerance ϵ.

Ultimately, at the heart of these processes lies an encoding-decoding mechanism. Given any error term ϵ, there is usually more than one way to encode the original source, i.e. more than one possible 'codepage'. If we define a 'codebook' as a formula to assign for each error term an associated codepage, then there are possibly many such codebooks. As we also see in the previous example, given their existence, we are also interested in finding an optimal codepage or codebook, i.e. one that minimizes some cost functions. These are the questions that motivate the work in this thesis. In other words, given an original data source and a space of codes, we sought to answer the following questions:

(1) Given an error term $\epsilon > 0$, does there exist a 'codepage', a set of codes that allow for the encoding-decoding process of the original data source such that the error of the process is smaller than ϵ? If such codepage exists, what are its properties?
(2) Does there exist a 'codebook', a formula that assigns for each error term its associated codepage? If such codebook exists, what are its properties?
(3) Given some cost function, does there exist a codepage/codebook that minimizes this cost function? If such optimal codepage/codebook exists, what are their properties?

Notice that the error term is calculated differently in the examples above. For lossy compression and noisy channel coding, the error term is defined by the symmetric distance d_P. Meanwhile, the error term in autoencoder is $\mathbb{E}L\big(x, g_\theta \circ f_\theta(x)\big)$, with L being some loss function. If L is the metric d on X, then the error term $\mathbb{E}L\big(x, g_\theta \circ f_\theta(x)\big)$ is the L_1 norm. Within the context of this thesis, the error terms will be calculated with the L_∞ distance.

1.2. Preliminaries

1.2.1. Cardinality. In chapter 2 we will examine the question of the existence of codepage and codebook in the most general setting, one that imposes no additional conditions beyond the basic definitions of the relevant concepts. Specifically, we will show that the criteria for the existence of codepage are determined by counting the number of close balls with a fixed radius that covers the data set. Since we do not assume the data set is totally bounded in this general case, the number of such balls can be infinite. Similarly, the existence of a codebook is related to counting a dense set, which, again, can also be infinite. Therefore, the concept of cardinality is applicable. Recall that in set theory, cardinality is a concept that is used to measure the number of elements in a set. For any two sets A and B, we use the expression $A \leq_c B$ to say that A has cardinal less than or equal to B, i.e. there exists an injective map $A \mapsto$ B. The expression $A =_c B$ indicates A and B has the same cardinal, i.e. there exists a bijective map $A \mapsto B$.

Theorem 1.2.1. *[8, Theorem 2.26.(Schroeder-Bernstein)]. For any two sets A, B, if $A \leq_c B$ and $B \leq_c A$, then $A =_c B$.*

Definition 1.2.2. *A cardinal assignment is a definite operation on sets $A \mapsto |A|$ which satisfies*
(1) $A =_c |A|$,
(2) if $A =_c B$, then $|A| = |B|$,
(3) for each set $\mathcal{E}$ of sets, $\{|X| : X \in \mathcal{E}\}$ is a set.
The cardinal numbers, or cardinals, are the values of this assignment, i.e. κ is a cardinal number if $\kappa = |A|$, for some set A.

Thus we can write $|A| \leq |B|$ to indicate $A \leq_c B$, and $|A| = |B|$ for $A =_c B$. Recall also that the Principle of Purity states that all mathematical objects are sets, including cardinal numbers.

Definition 1.2.3. *Given a set X, a well ordering $\leq$ on X is a binary relation that satisfies the following: For any $x, y, z \in X$:*
(1) (Reflexivity) $x \leq x$,
(2) (Transitivity) $x \leq y$ and $y \leq z \Rightarrow x \leq z$,
(3) (Antisymmetry) $x \leq y$ and $y \leq x \Rightarrow x = y$,
(4) (Any 2 elements are comparable) Either $x \leq y$ or $y \leq x$.
(5) (Every nonempty subset has a least element) Given a nonempty $A \subset X$, there exists an $a^* \in A$ such that $a^* \leq a$, for all $a \in A$.
A set that admits a well-ordering is called a well-orderable set.

Theorem 1.2.4. *[8, Theorem 9.16.] If κ and λ are non-zero, well-orderable cardinals and at least one of them is infinite, then*

$$\kappa + \lambda = \kappa \cdot \lambda = \max\{\kappa, \lambda\}.$$

An important concept in Set Theory is the Axiom of Choice. A set S is a choice set for a family of sets $\mathcal{E}$ if $S \subset \bigcup \mathcal{E}$, and for every $X \in \mathcal{E}$, $S \bigcap X$ is a singleton. The Axiom of Choice states that every family of non-empty and pairwise disjoint sets admits a choice set. Accepting the Axiom of Choice, we have the following important corollary.

Corollary 1.2.1. *[8, Corollary 9.17.] Suppose we accept the Axiom of Choice. For every indexed family of sets* $(i \mapsto \kappa_i)_{i\in I}$ *and every infinite* κ, *if* $|I| \leq \kappa$ *and for each* $i \in I$, $\kappa_i \leq \kappa$, *then* $\sum_{i\in I} \kappa_i \leq \kappa$.

The two following lemmas establish the concepts of infimum and supremum of sets of cardinalities.

Lemma 1.2.1. *[8, Lemma 9.18.] There is a definite operation* $\inf_c$, *such that for each non-empty family* $\mathcal{E}$ *of well orderable sets, the value* $\kappa = \inf_c(\mathcal{E})$ *has the following properties:*

(1) κ *is a well-orderable cardinal number.*

(2) For some $A \in \mathcal{E}, \kappa = |A|$.

(3) For every $B \in \mathcal{E}, \kappa \leq |B|$.

(4) If κ' *is any cardinal that satisfies (1)-(3), then* $\kappa' =_c \kappa$.

Lemma 1.2.2. *[8, Lemma 9.20.] There is a definite operation* $\sup_c$, *such that for each non-empty family* ϵ *of well orderable sets, the value* $\kappa = \sup_c(\mathcal{E})$ *has the following properties:*

(1) κ *is a well-orderable cardinal number.*

(2) For every $A \in \mathcal{E}, |A| \leq \kappa$.

(3) If B *is well-orderable and for all* $A \in \mathcal{E}, |A| \leq |B|$, *then* $\kappa \leq B$.

(4) If κ' *is any cardinal that satisfies (1)-(3), then* $\kappa' =_c \kappa$.

1.2.2. Covering number.

Definition 1.2.5. *[11, Definition 3] (External covering number). Let* (X, d) *be a totally bounded metric space. Given* $\epsilon > 0$, *the external covering number* $\sigma_\epsilon(X)$ *is the smallest number of points* $x_1, x_2, x_3, ..., x_n$ *such that*

$$X \subset \bigcup_{i=1}^{n} B(x_i, \epsilon).$$

1.2.3. Compact-open topology.

Definition 1.2.6. *(Compact-open topology). Let* X, Y *be two topological spaces, and* $C(X, Y)$ *denotes the set of all continuous maps from* X *to* Y. *The compact-open topology on* $C(X, Y)$ *is generated by subsets of the following form*

$$B(K, U) := \{f \in C(X, Y) : f(K) \subset U\},$$

where K *is a compact subset of* X, *and* U *is a compact subset of* Y.

If Y is a metric space, the compact-open topology is equivalent to the topology of compact convergence.

PROPOSITION 1.2.1. *[6, Theorem 7.11] If Y is a metric space, f_n converges to f in the compact-open topology if and only if f_n converges to f uniformly on any compact subset of X.*

Suppose X is also a metric space, we say that a sequence of functions f_n converges continuously if for all convergent sequence x_n in X, the sequence $f_n(x_n)$ also converges. In the case of complex functions, converging in the topology of compact convergence and converging continuously is one and the same [9, Section 3.1.5]. Expanding on that idea, the following property will be useful.

PROPOSITION 1.2.2. *Let X, Y be two metric spaces, and $C(X, Y)$ denotes the set of all continuous maps from X to Y. Suppose $f_n \to f$ in the compact-open topology on $C(X, Y)$, and $x_n \to x$ in X, then $f_n(x_n) \to f(x)$ in Y.*

PROOF. The set $S = \{x, x_1, x_2, x_3, ...\}$ is sequentially compact, hence compact in the metric space X. As $f_n \to f$ in the compact-open topology, from Proposition 1.2.1, $f_n \to f$ uniformly on S. Combined with $x_n \to x$, $f_n(x_n)$ converges to $f(x)$. □

1.2.4. Distances between measures and sets.

1.2.4.1. *Symmetric distance on measure space.* Recall that for any two sets A, B,

$$A \triangle B = \big(A/B\big) \bigcup \big(B/A\big).$$

DEFINITION 1.2.7. *On a finite measure space $(X, \mathcal{F}, \mu)$, we define the distance between A, B in $\mathcal{F}$ as*

$$d_\mu\big(A, B\big) := \mu\big(A \triangle B\big).$$

This is called the Fréchet–Nikodym metric. It is a pseudometric on $\mathcal{F}$ [2, Section 1.12].

PROPOSITION 1.2.3. *d_μ is a pseudometric on $\mathcal{F}$.*

PROOF. For any two μ-measurable subsets A and B of X, $d_\mu(A, B) = \mu(A \triangle B) = \mu(B \triangle A) = d_\mu(B, A)$. Suppose C is another μ-measurable subset of X, as $A \triangle B = (A \triangle C) \triangle (C \triangle B) \subset (A \triangle C) \cup$

$(C\triangle B)$,

$$\begin{aligned} d_\mu(A,B) &= \mu(A\triangle B) = \mu\big((A\triangle C)\triangle(C\triangle B)\big) \\ &\leq \mu\big((A\triangle C)\cup(C\triangle B)\big) \\ &\leq \mu\big(A\triangle C\big) + \mu\big(C\triangle B\big) = d_\mu(A,C) + d_\mu(C,B). \end{aligned}$$

This shows d_μ possesses symmetry and triangle inequality properties, thus it is a pseudometric. □

1.2.4.2. *Wasserstein Distance.* The error in the encoding-decoding process is defined as a distance. One such distance concept is called Wasserstein distance, which is a distance between measures on the same metric space ([10, Section 5.1]). For $\Omega \subset \mathbb{R}^n$, let $\mathcal{P}(\Omega)$ be the collection of all Borel probability measure on Ω. For $p \geq 1$, we define

$$\mathcal{P}_p(\Omega) := \{\mu \in \mathcal{P}(\Omega) : \int_\Omega |x|^p d\mu < +\infty\}.$$

We also define $\prod(\mu,\nu)$ to be the set of all Borel probability measures γ on $\Omega \times \Omega$ such that

$$\gamma(A\times\Omega) = \mu(A),\ \text{for all Borel set } A \text{ in } \Omega;\ \gamma(\Omega\times B) = \nu(B),\ \text{for all Borel set } B \text{ in } \Omega.$$

Wasserstein distance is then defined as follows.

Definition 1.2.8. *For $\mu, \nu \in \mathcal{P}_p(\Omega)$, we define the Wasserstein distance W_p between μ,ν as*

$$W_p(\mu,\nu) := \left(\inf_{\gamma\in\prod(\mu,\nu)}\left\{\int |x-y|^p d\gamma\right\}\right)^{\frac{1}{p}}.$$

The Wasserstein Distance is a metric.

Proposition 1.2.4. *[12, Theorem 7.3] For any $p \geq 1$, the Wasserstein distance W_p defines a metric on $\mathcal{P}_p(\Omega)$.*

1.2.4.3. *Hausdorff Distance.* Hausdorff distance is a concept of distance between subsets of a metric space. Recall that in a metric space (X,d), the distance between a point x and a subset A is defined as

$$d(x,A) := \inf_{a\in A} d(x,a).$$

The Hausdorff distance is defined as follows.

DEFINITION 1.2.9. *Let (X,d) be a metric space. For any pair of non-empty subsets A, B of X, the Hausdorff distance between A and B are defined as*

$$d_{\mathcal{H}}(A,B) := \max\{\sup_{a\in A} d(a,B), \sup_{b\in B} d(b,A)\}.$$

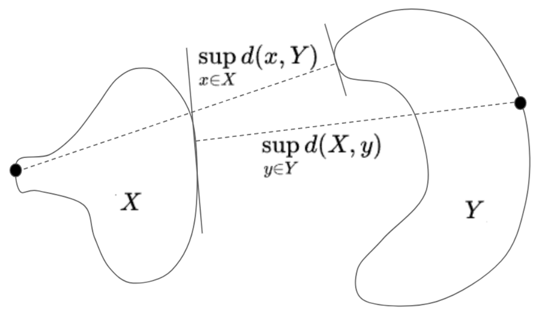

FIGURE 1.5. Hausdorff distance.

On the set of bounded subsets of X, the Hausdorff distance is well-defined and finite. Unfortunately, it does not define a metric. In particular, $d_{\mathcal{H}}(A,B) = 0$ does not necessary guarantee that $A = B$, only that $\overline{A} = \overline{B}$. Hence, Hausdorff distance defines a metric on the space of non-empty, closed, and bounded subsets of X. Still, on the space of non-empty, bounded subsets of a metric space, the Hausdorff distance defines a pseudometric, i.e. it satisfies the symmetry and triangle inequality. Such pseudometric is then sufficient to induce a topology on the space of bounded subsets of a metric space. The following proposition summarizes these properties.

PROPOSITION 1.2.5. *[5] Let (X,d) be a non-empty metric space. On the space of non-empty bounded subsets of X, the Hausdorff distance is well defined, finite, and satisfies the following properties:*

(1) For any A, B in S, $d_{\mathcal{H}}(A,B) = 0$ if and only if $\overline{A} = \overline{B}$;

(2) For any A, B in S, $d_{\mathcal{H}}(A,B) = d_{\mathcal{H}}(B,A)$;

(3) For any A, B, C in S, $d_{\mathcal{H}}(A,C) \le d_{\mathcal{H}}(A,B) + d_{\mathcal{H}}(B,C)$.

Consequently, the Hausdorff distance $d_{\mathcal{H}}$ defines a pseudometric on the space of non-empty, bounded subsets of X. Moreover, it defines a metric on the space of nonempty, close and bounded subsets of X.

One property of the Hausdorff distance is that for two sequentially compact sets, their Hausdorff distance equals the distance between some two points in each set.

Proposition 1.2.6. *Let A, B be two sequentially compact sets of a metric space (X, d). Then there exists $a^* \in A$ and $b^* \in B$ such that*

$$d_{\mathcal{H}}(A, B) = d(a^*, b^*).$$

Proof. Recall that

$$d_{\mathcal{H}}(A, B) := \max\{\sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A)\}.$$

Without loss of generality, suppose $\max\{\sup_{a\in A} d(a, B), \sup_{b\in B} d(b, A)\} = \sup_{a\in A} d(a, B)$. Let a_n be the minimizing sequence of $d(a, B)$, i.e.

$$\lim_{n\to\infty} d(a_n, B) = \sup_{a\in A} d(a, B).$$

From sequentially compactness of A, a_n converges subsequently to some a^* in A. From continuity of d,

$$d(a^*, B) = \lim_{n\to\infty} d(a_n, B) = \sup_{a\in A} d(a, B).$$

Recall now that $d(a^*, B) = \inf_{b\in B} d(a^*, b)$. Let b_n be a minimizing sequence for $d(a^*, b)$, i.e.

$$\lim_{n\to\infty} d(a^*, b_n) = \inf_{b\in B} d(a^*, b).$$

From sequentially compactness of B, b_n converges subsequently to some b^* in B. From continuity of d,

$$d(a^*, b^*) = \lim_{n\to\infty} d(a^*, b_n) = \inf_{b\in B} d(a^*, b).$$

Thus,

$$\begin{aligned} d(a^*, b^*) &= \inf_{b\in B} d(a^*, b) = d(a^*, B) = \sup_{a\in A} d(a, B) \\ &= \max\{\sup_{a\in A} d(a, B), \sup_{b\in B} d(b, A)\} \\ &= d_{\mathcal{H}}(A, B). \end{aligned}$$

□

1.2.5. Metric derivative and length of curve in a metric space.

DEFINITION 1.2.10. *Let (X,d) be a metric space. Given a curve $\gamma:[a,b]\to X$, we denote its range by $\Gamma:=\gamma([a,b])$, and we define*

$$\mathrm{Var}_{a'}^{b'} := \sup\left\{\sum_{i=1}^{n-1} d\big(\gamma(t_i),\gamma(t_{i+1})\big) : a'\le t_1<...<t_n\le b'\right\},$$

for every pair a',b' with $a\le a'\le b'\le b$. $\mathrm{Var}_{a'}^{b'}(\gamma)$ represents the length of the curve γ from $\gamma(a')$ to $\gamma(b')$. We denote $\mathrm{Var}_a^b(\gamma)$ as $\mathrm{Var}(\gamma)$. If $\mathrm{Var}(\gamma)<\infty$, we say that γ is rectifiable.

DEFINITION 1.2.11. *Given a curve $\gamma:[a,b]\to X$, we define the metric derivative of γ at the point $t\in(a,b)$ as*

$$|\gamma'|(t) := \lim_{h\to 0}\frac{d\big(\gamma(t+h),\gamma(t)\big)}{|h|}.$$

THEOREM 1.2.12. *[1, Theorem 4.1.6.] Suppose $\gamma:[a,b]\to X$ is a Lipschitz curve. The metric derivative exists at $\mathcal{L}^1-$ almost every point in $[a,b]$. Moreover, it holds that*

$$\mathrm{Var}(\gamma)=\int_a^b |\gamma'|(t)dt.$$

CHAPTER 2

Codepage and Codebook: General Set-up

This chapter formally introduces the concepts of codepage, codebook, and discusses their existence in the most general setup. Section 2.1 describes the formal definition of codepage, codebook. Section 2.2 illustrates the use of codebooks for compressing the MNIST dataset. Finally, section 2.3 discusses the criteria for the existence of codepage and codebook.

2.1. Definition of codepage and codebook

In this section, we formally define codepage and codebook.

Definition 2.1.1. *Suppose N is a subset of a metric space (M, d), and F is a non-empty set. Given $\epsilon \geq 0$, a set $A \subseteq F$ is called an ϵ-codepage of N if there exist two functions:*

$$f : N \to A \text{ and } g : A \to M$$

such that $d\Big(g \circ f(n), n\Big) \leq \epsilon$ for all $n \in N$.
If such a codepage exists, we say that F affords an ϵ-codepage for N.

Here, M represents the data space. N is the set of data we want to encode. F is the space of codewords, and A is the set of codewords chosen from F for the coding of N. The function f is called an encoder, and g is called a decoder. Namely, we first encode every element in N by a codeword in A via the encoder f, and then decode it back to data in M via the decoder g, with the requirement that the error, represented by the metric d, has to be within an acceptable range ϵ. We also want to assign a proper term to the pair of functions (f, g) in Definition 2.1.1.

Definition 2.1.2. *Let $f : N \to A$ and $g : A \to M$. We say (f, g) is an ϵ-linkable pair of A if*

$$d\Big(g \circ f(n), n\Big) \leq \epsilon, \forall n \in N.$$

Remark 2.1.1. *By definition, A being an ϵ-codepage means that A has at least one ϵ-linkable pair. Equivalently, a pair (f, g) is ϵ-linkable if and only if*

$$f^{-1}(a) \subset \overline{B\Big(g(a), \epsilon\Big)}, \forall a \in A.$$

Every 0-codepage of N is called a perfect reconstruction of N. Note that if A is an ϵ-codepage of N, then for any $\tilde{\epsilon} \geq \epsilon$, A is automatically an $\tilde{\epsilon}$-codepage of N. In cases where perfect reconstruction does not exist, the next best thing to hope for would be to find an $\epsilon-$ codepage, for small $\epsilon > 0$. This inspires the following definition.

Definition 2.1.3. *A function $\mathcal{A} : (0, \infty) \to 2^F$ is called a codebook of N if for every $\epsilon > 0$, $\mathcal{A}(\epsilon)$ is an ϵ-codepage of N. If such a codebook exists, we say that F affords a codebook for N.*

2.2. Constructing codepage and codebook for the MNIST dataset

In this section, we show a simple example of building a codebook for encoding the MNIST dataset. The MNIST dataset is a collection of 60,000 images of handwritten numbers from 0 to 9. Each image is in the form of a 28×28 pixel grayscale bounding box, where the value of each pixel ranges from 0 (indicating black) to 255 (indicating white). Figure 2.1 shows some images taken from MNIST.

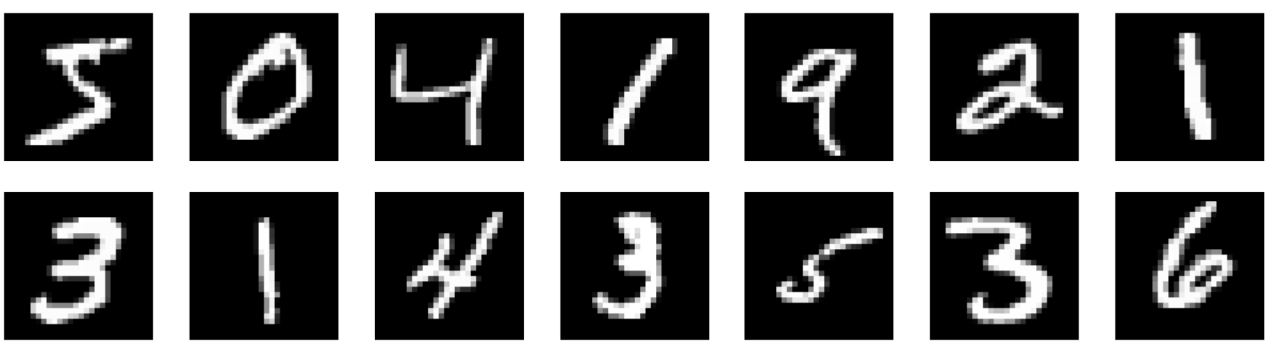

Figure 2.1. Some images of handwritten numbers from MNIST.

Let $M = [0, 1, ..., 255]^{28\times 28}$, then we consider N :=MNIST, which is a subset of M. We define a metric d_{Ave} on N as

$$d_{Ave}(n_1, n_2) := \frac{\sum_{i=1}^{28\times 28} |n_1^i - n_2^i|}{28 \times 28}.$$

Let $A_H = [0, 1, ..., 255]^H$ for $H \in \{1, 2, ..., 28 \times 28\}$. For $m = 1, 2, ..., 28$, we consider a mapping $f_m : N \to A_H$, where $H = H(m) = [\frac{28}{m}] \times [\frac{28}{m}]$. This map partitions each MNIST image into $m \times m$

blocks and then takes the average value of each block. More precisely,

$$f_m(n)_{ij} := \frac{\sum_{h=1}^{m}\sum_{p=1}^{m} n_{(mi+p)(mj+h)}}{m^2},$$

for each $n = [n_{kl}] \in N$, $1 \leq i \leq [\frac{28}{m}]$, $1 \leq j \leq [\frac{28}{m}]$. We also consider the mapping $g_m : A_H \to M$ that expands each $H \times H$ image in A_H back to a 28×28 image in M, by expanding each pixel in the $H \times H$ image into an $m \times m$ block with the same value. More precisely,

$$g_m(a)_{ij} := \begin{cases} a_{[\frac{i-1}{m}]+1,[\frac{j-1}{m}]+1}, & \text{if } [\frac{i-1}{m}]+1 \leq [\frac{28}{m}], [\frac{j-1}{m}]+1 \leq [\frac{28}{m}], \\ 0, & \text{otherwise.} \end{cases}$$

Figure 2.2 displays two images of handwritten numbers 5 and 4 from MNIST, along with examples of the reconstructed images applying $g_m \circ f_m$. Notice that as the dimension H increases the quality of the reconstructed images improves.

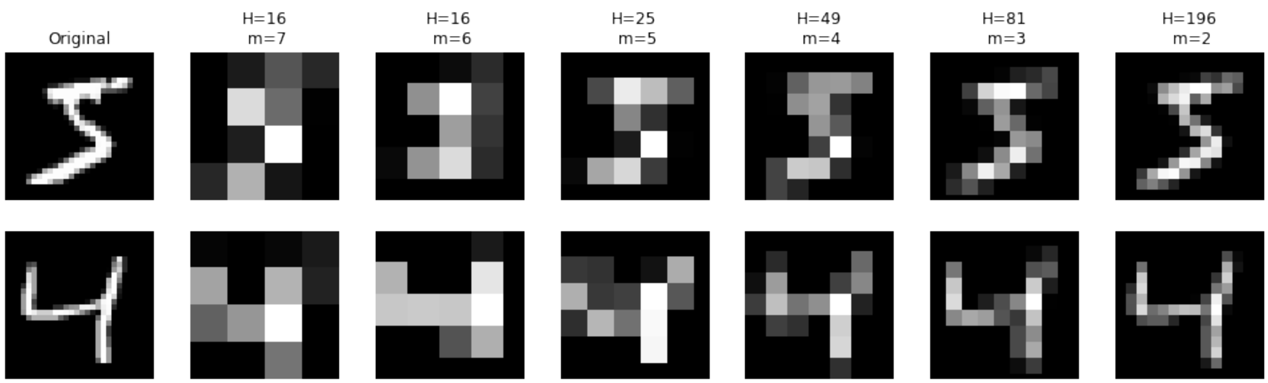

FIGURE 2.2. Pictures of handwritten numbers 5 and 4 from MNIST dataset, along with examples of the reconstructed images under the mapping $g_m \circ f_m$. Here $H = H(m)$ represents A_H, and m denotes the pair (f_m, g_m) used.

For each value of m, Table A.1 shows the maximum values of $d_{Ave}(n, g_m \circ f_m(n))$ and the Wasserstein 1 distance (Definition 1.2.8) among all elements n inside the MNIST. Based on this, Table A.2 and A.3 display some possible ϵ-codepages corresponding to each range of ϵ. Table 2.1 and 2.2 illustrate two examples of codebooks, based on d_{Ave} and the Wasserstein 1 distance, respectively.
In figure 2.3 we illustrate using the codebook in table 2.1 and 2.2 on an image from MNIST.

TABLE 2.1. An example of a codebook $\mathcal{A}(\epsilon) := A_{H_\epsilon}$. In the first column, the value of ϵ's shown are calculated based on d_{Ave} and are rounded up to 2 decimals. In the second column, the corresponding H_ϵ represents the ϵ-codepage $A_H = [0, 1, ..., 255]^H$. In the third column, m's represents the pairs (f_m, g_m), as defined above.

ϵ	H_ϵ	Possible m's for H_ϵ
$[98.21, \infty)$	1	$15 \leq m \leq 28$
$[86.92, 98.21)$	4	$10 \leq m \leq 14$
$[82.35, 86.01)$	9	8, 9
$[58.78, 82.35)$	16	6, 7
$[53.47, 58.78)$	25	5
$[49.68, 53.47)$	49	4
$[37.61, 49.68)$	81	3
$[24.95, 37.61)$	196	2
$[0, 24.95)$	784	1

TABLE 2.2. An example of a codebook $\mathcal{A}(\epsilon) := A_{H_\epsilon}$. In the first column, the value of ϵ's shown are calculated based on the Wasserstein distance W_1 and are rounded up to 2 decimals. In the second column, the corresponding H_ϵ represents the ϵ-codepage $A_H = [0, 1, ..., 255]^H$. In the third column, m's represents the pairs (f_m, g_m), as defined above.

ϵ	H_ϵ	Possible m's for H_ϵ
$[101.72, \infty)$	1	$15 \leq m \leq 28$
$[97.12, 101.72)$	4	$10 \leq m \leq 14$
$[90.62, 97.12)$	9	8, 9
$[58.78, 90.62)$	16	6, 7
$[53.47, 58.78)$	25	5
$[49.68, 53.47)$	49	4
$[37.61, 49.68)$	81	3
$[24.95, 37.61)$	196	2
$[0, 24.95)$	784	1

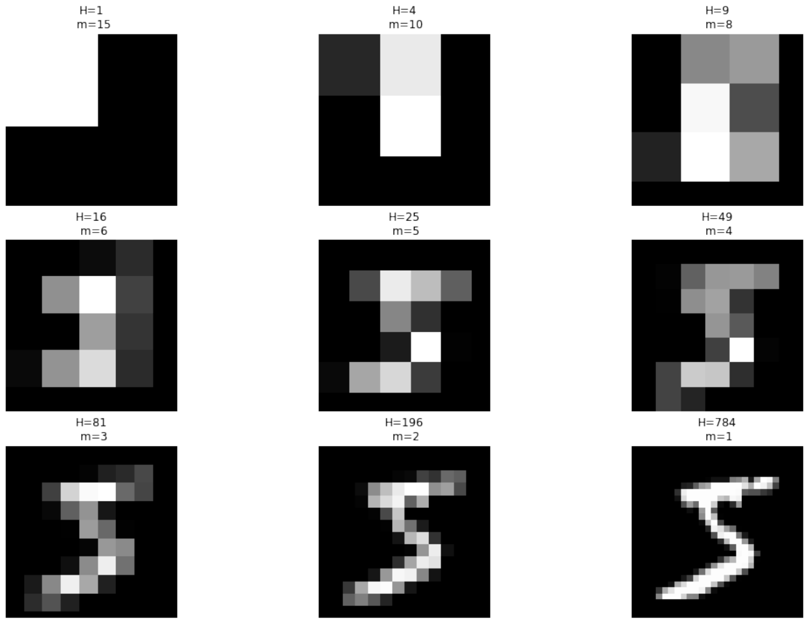

FIGURE 2.3. Illustration of applying the codebook described in table 2.1 and 2.2 to a handwritten image of the number 5 from MNIST.

2.3. Existence of codepages and codebook

Given $N \subseteq M$ and F as above, we want to explore the existence of ϵ-codepage and codebooks of N. We begin with one simple example of when a codepage or codebook exists. Recall that for a set X, $|X|$ represents the cardinality of X.

PROPOSITION 2.3.1. *If $|F| \geq |N|$, then N has a 0-codepage $A \subseteq F$, and the function $\mathcal{A}_0 : (0, \infty) \to 2^F$, $\mathcal{A}_0(\epsilon) = A$ for all $\epsilon \in (0, \infty)$, is a codebook for N.*

PROOF. Since $|F| \geq |N|$, by the definition of cardinality, there exists an injective function $\varphi : N \to F$. Let $A = \varphi(N)$. Consider $f = \varphi$ and $g = (\varphi|_A)^{-1}$. For every $n \in N$, we have $g \circ f(n) = (\varphi^{-1}) \circ \varphi(n) = n$ and $d\Big(g \circ f(n), n\Big) = d(n, n) = 0$. As a result, A is a 0-codepage of N. Since A is also an ϵ-codepage of N for all $\epsilon > 0$, $\mathcal{A}_0$, as described in the proposition, is a (trivial) codebook of N. □

The previous proposition suggests that the criteria for the existence of either codepage or codebook involve the comparison between the cardinality of the space of codewords F and some subset of the data space. First, we need some definitions. Recall the concept of external covering number for a totally bounded set (Definition 1.2.5). To expand this concept to any set, we introduce the following definitions.

Definition 2.3.1. *We define the following terms:*

- *For any set N, we formally denote $\sigma_\epsilon(N)$ by*

$$\sigma_\epsilon(N) := \inf{}_c\{|K| : K \subseteq M \text{ such that } N \subseteq \bigcup_{k \in K} \overline{B(k,\epsilon)}\}. \tag{2.1}$$

- *Suppose A and B are subsets of a metric space M. We say that B is dense with respect to A if $A \subset \overline{B}$, the closure of B. With that, for any subset N of a metric space (M, d), we denote the notion $\sigma(N)$ as follows:*

$$\begin{aligned} \sigma(N) &:= \inf{}_c\{|N'| : N' \subset M,\, N' \text{ is dense with respect to } N\} \\ &= \inf{}_c\{|N'| : N' \subset M,\, N \subset \overline{N'} \subset M\}. \end{aligned} \tag{2.2}$$

Remark 2.3.1. *For any set N fixed, $\sigma_\epsilon(N)$ can be thought of as a non-decreasing function of ϵ, in the sense that for any $0 \le \epsilon_1 \le \epsilon_2$, $\sigma_{\epsilon_1}(N) \ge \sigma_{\epsilon_2}(N)$.*

Accepting the Axiom of Choice, we obtain the following criterion for the existence of codepage.

Theorem 2.3.2 (Existence Criterion for Codepage). *Accepting the Axiom of Choice. Given $\epsilon > 0$, a subset A of F is an ϵ-codepage of N if and only if $|A| \ge \sigma_\epsilon(N)$. Therefore, an ϵ-codepage of N exists in F if and only if*

$$|F| \ge \sigma_\epsilon(N).$$

Proof. '⇒': Suppose A is an ϵ-codepage of N. By the definition of ϵ-codepage, there exist two functions $f : N \to A$ and $g : A \to M$ such that $d\Big(g \circ f(n), n\Big) \le \epsilon$ for all $n \in N$. As a result, $\{\overline{B\big(g[f(n)], \epsilon\big)} : n \in N\}$ is a covering of N. From the definition of $\sigma_\epsilon(N)$ and Lemma 1.2.1, this implies

$$\sigma_\epsilon(N) \le |\{g[f(n)] : n \in N\}| \le |\{f(n) : n \in N\}| \le |A| \le |F|.$$

'⇐': Suppose $|F| \ge \sigma_\epsilon(N)$. Then there exists a subset $\tilde{N} \subseteq M$ with $|\tilde{N}| \le |F|$ and $\{\overline{B(x,\epsilon)} : x \in \tilde{N}\}$ covers N. Thus, there exists an injective map $\varphi : \tilde{N} \to F$. Define $A = \varphi(\tilde{N})$, and $g = \varphi^{-1}$. f is defined as follows: For each n in N, n belong to a ball $\overline{B(x,\epsilon)}$, for some $x \in \tilde{N}$, and we use the axiom of choice to define $f(n) = \varphi(x)$. Then, for each n in N, $g \circ f(n) = \varphi^{-1} \circ \varphi(x) = x$, and so $d\Big(g \circ f(n), n\Big) = d(n, x) \leq \epsilon$. This shows A is indeed an ϵ-codepage of N. □

Next we prove an existence criterion for codebook.

THEOREM 2.3.3 (First Existence Criterion for Codebook). *Accepting the Axiom of Choice. F affords a codebook for N if and only if $|F| \geq \sup_c \{\sigma_\epsilon(N) : \epsilon > 0\}$.*

PROOF. F affords a codebook for N if and only if, for all $\epsilon > 0$, F affords an ϵ-codepage. This is equivalent to the condition that for all $\epsilon > 0$, $|F| \geq \sigma_\epsilon(N)$. This, in turn, is equivalent to $|F| \geq \sup_c \{\sigma_\epsilon(N) : \epsilon > 0\}$. □

One example of a case where a codebook exists is when the dataset N is totally bounded. In this case, any infinite set F can afford a codebook.

COROLLARY 2.3.1. *Suppose N is totally bounded, and F is an infinite set, then F affords a codebook of N.*

PROOF. Since N is totally bounded, the function $\sigma_\epsilon(N)$ is an integer-valued decreasing function of ϵ. So,

$$|F| \geq |\mathbb{N}| \geq \sup_{\epsilon > 0}{}_c \sigma_\epsilon(N).$$

Thus, by Theorem 2.3.3, there exists a codebook of N in F. □

Accepting the Axiom of Choice again, we obtain the equality between $\sigma(N)$and $\sup_c \{\sigma_\epsilon(N) : \epsilon > 0\}$ for any set N.

PROPOSITION 2.3.2. *Accepting the Axiom of Choice, for any given N it holds that*

$$\sigma(N) = \sup_{\epsilon > 0}{}_c \sigma_\epsilon(N).$$

PROOF. We first note that for any $N' \subseteq M$ dense with respect to N and any $\epsilon > 0$, it holds that $\{\overline{B(x, \epsilon)} : x \in N'\}$ covers N. Thus, $|N'| \geq \sigma_\epsilon(N)$ for every N'. By definition of $\sigma(N)$ this implies that $\sigma(N) \geq \sigma_\epsilon(N)$ for every $\epsilon > 0$. From Lemma 1.2.2, $\sigma(N) \geq \sup_c \{\sigma_\epsilon(N) : \epsilon > 0\}$.
We now claim that $\sigma(N) \leq \sup_c \{\sigma_\epsilon(N) : \epsilon > 0\}$ and prove it in two cases:

Case 1. $\sup_c \{\sigma_\epsilon(N) : \epsilon > 0\} = m$ is finite: For any $\epsilon > 0$, $\sigma_\epsilon(N) \leq m$. Therefore, there exists a set $K_\epsilon \subseteq M$ with $|K_\epsilon| \leq m$ such that $N \subseteq \bigcup_{k \in K_\epsilon} \overline{B(k, \epsilon)}$. We claim that N has at most

m elements. Indeed, towards a contradiction, suppose N contains at least $m+1$ distinct points $\{n_1, ..., n_{m+1}\}$. By the pigeonhole principle, at least two elements of N are contained in the same ball $\overline{B(k,\epsilon)}$ for some $k \in K_\epsilon$. This is impossible when $\epsilon < \frac{\min_{i,j} d(n_i,n_j)}{2}$. As a result, since N is also dense in itself, we have

$$\sup_{\epsilon>0}{}_c \sigma_\epsilon(N) \geq |N| \geq \sigma(N).$$

Case 2. $\sup_c\{\sigma_\epsilon(N) : \epsilon > 0\}$ is infinite: For any positive integer p, since $\sup_c\{\sigma_\epsilon(N) : \epsilon > 0\} \geq \sigma_{\frac{1}{p}}(N)$, by the definition of $\sigma_{\frac{1}{p}}$ and Lemma 1.2.1 there exists a set $K_p \subset M$ such that $\{\overline{B(k, \frac{1}{p})} : k \in K_p\}$ covers N, and $\sup_c\{\sigma_\epsilon(N) : \epsilon > 0\} \geq |K_p|$. By Corollary 1.2.1,

$$\sup_{\epsilon>0}{}_c \sigma_\epsilon(N) \geq \sum_{p\in\mathbb{Z}^+} |K_p| \geq |\bigcup_{p\in\mathbb{Z}^+} K_p|.$$

As $\bigcup_{p\in\mathbb{Z}^+} K_p$ is dense with respect to N, $|\bigcup_{p\in\mathbb{Z}^+} K_p| \geq \sigma(N)$, and we conclude that $\sup_c\{\sigma_\epsilon(N) : \epsilon > 0\} \geq \sigma(N)$.

□

Combining Theorem 2.3.3 and Proposition 2.3.2 gives us a second criterion for the existence of codebook.

Theorem 2.3.4 (Second Existence Criterion for Codebook). *Accepting the Axiom of Choice, for any given N, a set F affords a codebook for N if and only if*

$$|F| \geq \sigma(N).$$

Remark 2.3.2. *We illustrate the behavior of $\sigma_\epsilon(N)$ and $\sigma(N)$ for finite and infinite cases in several simple examples. Suppose $M = \mathbb{R}$, equipped with the usual Euclidean metric.*

- *Suppose $N = [0,1]$. In this case, a close ball with radius ϵ is a close interval with length 2ϵ. As $\lceil\frac{1}{2\epsilon}\rceil = \inf\{k \in \mathbb{N} : k \cdot 2\epsilon \geq 1\}$, $\sigma_\epsilon(N) = \lceil\frac{1}{2\epsilon}\rceil$. From Proposition 2.3.2,*

$$\sigma(N) = \sup_{\epsilon>0}{}_c \sigma_\epsilon(N) = \sup_{\epsilon>0}{}_c \lceil\frac{1}{2\epsilon}\rceil = \aleph_0,$$

where $\aleph_0$ is the countably infinite cardinality. Indeed, as the set $\mathbb{Q}\bigcap[0,1]$ is countable and dense in $[0,1]$, and no finite sets are dense in $[0,1]$, it follows from the definition of σ (Definition 2.2) that $\sigma(N) = \aleph_0$.

• *Suppose* $N = M$. *Since* N *is unbounded, for any* $\epsilon > 0$, *we require at least a countably infinitely many close intervals with length* 2ϵ *to cover* N. *Thus, for any* $\epsilon > 0$, $\sigma_\epsilon(N) = \aleph_0$. *From Proposition 2.3.2,*

$$\sigma(N) = \sup_{\epsilon>0}{}_c \sigma_\epsilon(N) = \sup_{\epsilon>0}{}_c \aleph_0 = \aleph_0.$$

Indeed, as the set $\mathbb{Q}$, *which is countable, is dense in* $N = M = \mathbb{R}$, *and no finite sets are dense in* $\mathbb{R}$, *again from the definition of* σ, $\sigma(N) = \aleph_0$.

• *Suppose* $M = N = L^2[-\pi, \pi]$, *equipped with the* L^2 *metric. As* $\left\{\sum_{k=-K}^{K} c_k e^{ikx} : c_k \in \mathbb{Q}, K \in \mathbb{N}\right\}$ *is dense in* $L^2[-\pi, \pi]$, *and no finite set is dense in* $L^2[-\pi, \pi]$,

$$\sigma(N) = \left|\left\{ \sum_{k=-K}^{K} c_k e^{ikx} : c_k \in \mathbb{Q}, K \in \mathbb{N}\right\}\right| = \aleph_0.$$

From Proposition 2.3.2, $\sup_c\{\sigma_\epsilon(N) : \epsilon > 0\} = \sigma(N) = \aleph_0$. *Thus, for any* $\epsilon > 0$, $\sigma_\epsilon(N) \leq \aleph_0$. *On the other hand, since* N *is unbounded, for any* $\epsilon > 0$, *no finite set of close balls of radius* ϵ *can cover* $L^2[-\pi, \pi]$. *Consequently,* $\sigma_\epsilon(N) = \aleph_0$, *for any* $\epsilon > 0$.

Indeed, suppose $\epsilon > 0$ *is given. Again as* $\left\{\sum_{k=-K}^{K} c_k e^{ikx} : c_k \in \mathbb{Q}, K \in \mathbb{N}\right\}$ *is dense in* $L^2[-\pi, \pi]$, *the set of close balls of the form*

$$\left\{ \overline{B\left(\sum_{k=-K}^{K} c_k e^{ikx}, \epsilon\right)} : c_k \in \mathbb{Q}, K \in \mathbb{N}\right\}$$

does cover $L^2[-\pi, \pi]$. *As the cardinality of* $\left\{\sum_{k=-K}^{K} c_k e^{ikx} : c_k \in \mathbb{Q}, K \in \mathbb{N}\right\}$ *is* $\aleph_0$, $\sigma_\epsilon(N) = \aleph_0$.

CHAPTER 3

Optimal Codepage and Codebook: First-Type Cost Function

In Chapter 2, we provide the formal definitions of codepage and codebook, and discuss their properties in a general context. In Chapters 3 and 4, we explore the existence of optimal codepages and codebooks—i.e., those that minimize specific cost functions. Chapter 3 considers a simple cost function where the cost on each codepage only takes into account the cost of the codes used. In section 3.1, we define the cost function $C(\mathcal{A})$ on codebook and establish the general existence criteria for this cost function (Theorem 3.1.2). In section 3.2 we simplify these existence criteria in the case data set N is totally bounded. Assuming also that data space M is a Heine-Borel metric space, in section 3.3 we derive a formula for the minimum $C(\mathcal{A})$. Section 3.4 then briefly discusses the case N not totally bounded.

Throughout Chapter 3, we continue to use the notations introduced in Definition 2.1.1.

3.1. Cost function and existence of optimal codebook

We begin this section with the definition of a cost function.

Definition 3.1.1. *Given a Borel measure μ on the metric space F and a Borel measure ρ on $(0, \infty)$, a map $\mathcal{A} : (0, \infty) \to 2^F$ is called a (μ, ρ)-codebook if each $\mathcal{A}(\epsilon)$ is a μ-measurable ϵ-codepage of N in F, and $\mu(\mathcal{A}(\epsilon))$ is a ρ-measurable function of ϵ.*

Let Ω be the collection of all (μ, ρ)-codebooks. For any $\mathcal{A} \in \Omega$, define its cost by

$$C(\mathcal{A}) := \int_0^\infty \mu\Big(\mathcal{A}(\epsilon)\Big) d\rho(\epsilon).$$

We want to consider the minimization problem

$$\min_{\mathcal{A} \in \Omega} C(\mathcal{A}). \tag{3.1}$$

The optimal codebooks are therefore those that solve Problem (3.1).

Lemma 3.1.1. *Suppose the minimization problem*

$$\min\{\mu(A) : A \textit{ is a } \mu\textit{-measurable } \epsilon\textit{-codepage of } N \textit{ in } F\}$$

has a solution $A^*(\epsilon)$ *for both* $\epsilon = \epsilon_1, \epsilon_2$ *with* $0 < \epsilon_1 \leq \epsilon_2$*, then* $\mu\Big(A^*(\epsilon_1)\Big) \geq \mu\Big(A^*(\epsilon_2)\Big)$.

Proof. By Theorem 2.3.2, a μ-measurable subset A of F is an ϵ-codepage if and only if $|A| \geq \sigma_\epsilon(N)$. By Remark 2.3.1, when $0 < \epsilon_1 \leq \epsilon_2$, we have $\sigma_{\epsilon_1}(N) \geq \sigma_{\epsilon_2}(N)$. Consequently,

$$\begin{aligned} \mu\Big(A^*(\epsilon_1)\Big) &= \min\{\mu(A) : A \text{ is a } \mu\text{-measurable } \epsilon_1\text{-codepage of } N \text{ in } F\} \\ &= \min\{\mu(A) : A \text{ is a } \mu\text{-measurable and } |A| \geq \sigma_{\epsilon_1}(N)\} \\ &\geq \min\{\mu(A) : A \text{ is a } \mu\text{-measurable and } |A| \geq \sigma_{\epsilon_2}(N)\} \\ &= \mu\Big(A^*(\epsilon_2)\Big). \end{aligned}$$

□

The following is the existence criteria for Problem (3.1) in a general setting.

Theorem 3.1.2. *Suppose for* ρ*-a.e.* $\epsilon > 0$*, the minimization problem*

$$\min\{\mu(A) : A \textit{ is a } \mu\textit{-measurable } \epsilon\textit{-codepage of } N \textit{ in } F\} \tag{3.2}$$

has a solution. Let $\mathcal{A}^*$ *be a* (μ, ρ)*-codebook.*

(a) *If* $\mathcal{A}^*(\epsilon)$ *is a solution to (3.2) for* ρ*-a.e.* $\epsilon > 0$*, then* $\mathcal{A}^*$ *is a solution to the minimization Problem (3.1).*

(b) *If* $\mathcal{A}^*$ *is a solution to the minimization Problem (3.1) with* $C(\mathcal{A}^*)$ *finite, then* $\mathcal{A}^*(\epsilon)$ *is a solution to (3.2) for* ρ*-a.e.* $\epsilon > 0$.

Proof. (a) Suppose $\mathcal{A}^*(\epsilon)$ is a solution to (3.2) for ρ-a.e. $\epsilon > 0$. That is, for ρ-a.e. $\epsilon > 0$,

$$\mathcal{A}^*(\epsilon) \in \textbf{argmin}\{\mu(A) : A \text{ is a } \mu\text{-measurable } \epsilon\text{-codepage of } N \text{ in } F\}.$$

By Lemma 3.1.1, $\mathcal{A}^*(\epsilon)$ is a non-increasing function of ϵ on $(0, \infty)$ $\rho-$a.e. and thus it is a (μ, ρ)-codepage. By the μ-minimality of each $\mathcal{A}^*(\epsilon)$, it follows that $\mathcal{A}^*$ is a C-optimal codebook.

(b) On the other hand, suppose $\mathcal{A}^*$ is solution to (3.1). By assumption, one may construct a (μ,ρ)-codepage $\mathcal{A}^{**}$ such that $\mathcal{A}^{**}(\epsilon)$ is a solution to (3.2) for ρ-a.e. $\epsilon > 0$. Thus, $0 \leq \mu\Big(\mathcal{A}^{**}(\epsilon)\Big) \leq \mu\Big(\mathcal{A}^{*}(\epsilon)\Big)$. Note that when $C(\mathcal{A}^*) < +\infty$,

$$C(\mathcal{A}^{**}) = \int_0^\infty \mu\Big(\mathcal{A}^{**}(\epsilon)\Big) d\rho(\epsilon) \leq \int_0^\infty \mu\Big(\mathcal{A}^{*}(\epsilon)\Big) d\rho(\epsilon) = C(\mathcal{A}^*) \leq C(\mathcal{A}^{**}).$$

This implies that for ρ-a.e. $\epsilon > 0$,

$$\mu\Big(\mathcal{A}^{*}(\epsilon)\Big) = \mu\Big(\mathcal{A}^{**}(\epsilon)\Big).$$

That is, $\mathcal{A}^*(\epsilon)$ is a solution to (3.2) for ρ-a.e. $\epsilon > 0$. □

REMARK 3.1.1. *Theorem 3.1.2 illustrates that in the case (3.2) has solution for $\rho - a.e.$, solving (3.1) reduces to solving (3.2). From Theorem 2.3.2, solving (3.2) is equivalent to solving*

$$\min\{\mu(A) : A \subseteq F \text{ is } \mu\text{-measurable with } |A| \geq \sigma_\epsilon(N)\}. \tag{3.3}$$

This leads us to consider the following Problem

$$\min\{\mu(A) : A \subseteq F \text{ is } \mu\text{-measurable with } |A| \geq K\}, \tag{3.4}$$

for some cardinality K.

3.2. The case of totally bounded data set

We first consider the case when N is totally bounded, i.e. $\sigma_\epsilon(N)$ is finite for any $\epsilon > 0$. Note that since μ is Borel on the metric space F, all singletons of F are μ-measurable. As $\sigma_\epsilon(N)$ is finite, based on remark 3.1.1 we are want to solve (3.4) when K is finite.

PROPOSITION 3.2.1. *If K is finite, Problem (3.4) has a solution if and only if there exists a subset $A^* = \{a_1, ..., a_K\} \subseteq F$ such that*

$$\max\{\mu(\{a_i\}) : i = 1, 2, \cdots, K\} \leq \inf\{\mu(\{a'\}) : a' \in F \setminus A^*\}, \tag{3.5}$$

in which case A^* is a solution to (3.4).

PROOF. Firstly, suppose such an A^* exists, i.e. $A^* = \{a_1, ..., a_K\}$ with $\mu(\{a_1\}) \leq \mu(\{a_2\}) \leq ... \leq \mu(\{a_K\}) \leq \mu(\{a'\})$ for all $a' \in F \setminus A^*$. Since μ is a Borel measure, A^* is μ-measurable.

For any $A \subset F$ with $|A| \geq K$, there exists a subset $A' \subseteq A$ with $|A'| = K$. Suppose $A' = \{a'_1, ..., a'_p, a'_{p+1}, ..., a'_K\}$ with $\{a'_1, ..., a'_p\} = A^* \cap A'$ and $\{a'_{p+1}, ..., a'_K\} \subset F \setminus A^*$. Then,

$$\mu(A^*) = \sum_{i=1}^{p} \mu(\{a_i\}) + \sum_{i=p+1}^{K} \mu(\{a_i\}) \leq \sum_{i=1}^{p} \mu(\{a'_i\}) + \sum_{i=p+1}^{K} \mu(\{a'_i\}) = \mu(A') \leq \mu(A).$$

This shows that A^* is a solution for (3.4).

On the other hand, towards a contradiction, we assume that $A^* = \{a_1, ..., a_K\} \subset F$ is a solution to (3.4) but does not satisfy (3.5). That is, there exists a point $a_i \in A^*$ and a point $a' \in F \setminus A^*$ such that $\mu(\{a'\}) < \mu(\{a_i\})$. Set $A = (A^* \setminus \{a_i\}) \cup \{a'\}$. Then,

$$\mu(A) = \mu(A^*) - \mu(\{a_i\}) + \mu(\{a'\}) < \mu(A^*),$$

a contradiction to the optimality of A^*. This shows that any solution to (3.4) must satisfy (3.5). □

Corollary 3.2.1. *Suppose N is totally bounded. If (3.2) has no solution at $\epsilon = \epsilon^* > 0$, then it also has no solution for any $0 < \epsilon < \epsilon^*$.*

Proof. Towards a contradiction, assume that for some positive $\epsilon < \epsilon_*$, (3.2) has a solution. From Proposition 3.2.1, this implies the existence of a set $A_\epsilon = \{a_1, ..., a_{\sigma_\epsilon}\} \subset F$ with $\mu(\{a_1\}) \leq \mu(\{a_2\}) \leq ... \leq \mu(\{a_{\sigma_\epsilon}\}) \leq \mu(\{a'\})$ for all $a' \in F \setminus A_\epsilon$. As $\epsilon^* > \epsilon$, $\sigma_{\epsilon^*} \leq \sigma_\epsilon$. Hence $\mu(\{a_1\}) \leq \mu(\{a_2\}) \leq ... \leq \mu(\{a_{\sigma_{\epsilon^*}}\}) \leq \mu(\{a'\})$ for all $a' \in F \setminus A_{\epsilon^*}$. This then implies that the set $\{a_1, ..., a_{\sigma_{\epsilon^*}}\}$ solves (3.2) for $\epsilon = \epsilon^*$, a contradiction. □

The following lemma will help simplify Theorem (3.1.2) under some conditions, which will be helpful in solving Problem (3.1).

Lemma 3.2.1. *Suppose N is totally bounded and (3.1) has a solution with finite C-cost, then for any $\epsilon > 0$ with $\rho\big((0, \epsilon)\big) > 0$, (3.2) has a solution.*

Proof. Let $\mathcal{A}^*$ be a minimizer for (3.1) with $C(\mathcal{A}^*) < \infty$. Towards a contradiction, suppose there exists an $\epsilon^* > 0$ for which $\rho\big((0, \epsilon^*)\big) > 0$ but (3.2) has no solution. From Corollary (3.2.1), there is also no solution for (3.2) for any $\epsilon \in (0, \epsilon^*)$.

Let $0 < \epsilon_0 < \epsilon^*$ be small enough so that $\rho([\epsilon_0, \epsilon^*)) > 0$. Let $\{\epsilon_0, ..., \epsilon_H = \epsilon^*\}$ be a partition of the interval $[\epsilon_0, \epsilon^*]$ so that $\sigma_\epsilon(N)$ is constant on each subinterval $(\epsilon_k, \epsilon_{k+1})$ for $k = 0, 1, \cdots, H-1$.

Since $\rho([\epsilon_0, \epsilon^*)) > 0$, without loss of generality, we may assume that $\rho\big([\epsilon_0, \epsilon_1)\big) > 0$. Note that since $\sigma_\epsilon(N)$ is a constant on (ϵ_0, ϵ_1), the value

$$\tau := \inf\{\mu(A) : A \subseteq F \text{ is } \mu\text{-measurable with } |A| \geq \sigma_\epsilon(N)\}$$

is independent of the choice of $\epsilon \in (\epsilon_0, \epsilon_1)$. For each $\epsilon \in (\epsilon_0, \epsilon_1)$, since (3.2) has no solution at ϵ, we have $\mu(\mathcal{A}^*(\epsilon)) > \tau$. Let $\{A_k\}_{k=1}^\infty$ be a sequence of μ-measurable subsets in F with $|A_k| \geq \sigma_\epsilon(N)$ for $\epsilon \in (\epsilon_0, \epsilon_1)$ and $\{\mu(A_k)\}$ is a decreasing sequence with limit τ. Also, since (3.2) has no solution at ϵ_0, there exists a μ-measurable subset A^0 with $|A^0| \geq \sigma_{\epsilon_0}(N)$ and $\mu(A^0) < \mu(\mathcal{A}^*(\epsilon_0))$. For each k, consider the (μ, ρ)-codebook $\mathcal{A}_k$ given by

$$\mathcal{A}_k(\epsilon) := \begin{cases} A^0, & \epsilon = \epsilon_0, \\ A_k, & \epsilon \in (\epsilon_0, \epsilon_1), \\ \mathcal{A}^*(\epsilon), & \text{otherwise.} \end{cases}$$

Note that,

$$\begin{aligned} C(\mathcal{A}^*) - C(\mathcal{A}_k) &= \int_{[\epsilon_0, \epsilon_1)} \Big(\mu(\mathcal{A}^*(\epsilon)) - \mu(\mathcal{A}_k(\epsilon))\Big) d\rho \\ &= \int_{(\epsilon_0, \epsilon_1)} \Big(\mu(\mathcal{A}^*(\epsilon)) - \mu(A_k)\Big) d\rho + \Big(\mu(\mathcal{A}^*(\epsilon_0)) - \mu(\mathcal{A}^0)\Big)\rho(\{\epsilon_0\}). \end{aligned}$$

By the Monotone Convergence Theorem,

$$\lim_{k\to\infty} \Big(C(\mathcal{A}^*) - C(\mathcal{A}_k)\Big) = \int_{(\epsilon_0, \epsilon_1)} \Big(\mu(\mathcal{A}^*(\epsilon)) - \tau\Big) d\rho + \Big(\mu(\mathcal{A}^*(\epsilon_0)) - \mu(\mathcal{A}^0)\Big)\rho(\{\epsilon_0\}) > 0,$$

because $\rho([\epsilon_0, \epsilon_1)) > 0, \mu(\mathcal{A}^*(\epsilon_0)) > \mu(\mathcal{A}^0)$ and $\mu(\mathcal{A}^*(\epsilon)) > \tau$ on (ϵ_0, ϵ_1). As a result, when k is large enough, $C(\mathcal{A}^*) > C(\mathcal{A}_k)$, a contradiction to the minimality of $\mathcal{A}^*$. □

Using Lemma 3.2.1, in the case where ρ satisfies $\rho\big((0, \epsilon)\big) > 0$ for any $\epsilon > 0$, Theorem 3.1.2 can be simplified as follow.

Theorem 3.2.1. *Suppose N is totally bounded and $\rho\big((0, \epsilon)\big) > 0$ for any $\epsilon > 0$. Let $\mathcal{A}^*$ be a (μ, ρ)-codebook.*

- *(a) If $\mathcal{A}^*(\epsilon)$ is a solution to (3.2) for ρ-a.e. $\epsilon > 0$, then $\mathcal{A}^*$ is a solution to the minimization problem (3.1).*

(b) *If $\mathcal{A}^*$ is a solution to the minimization problem (3.1) with $C(\mathcal{A}^*)$ finite, then $\mathcal{A}^*(\epsilon)$ is a solution to (3.2) for ρ-a.e. $\epsilon > 0$.*

The next theorem then solves Problem (3.1) for the case N totally bounded.

THEOREM 3.2.2. *Suppose N is totally bounded and $\rho\big((0,\epsilon)\big) > 0$ for any $\epsilon > 0$. Suppose also there exists a (μ,ρ)-codebook $\mathcal{A}$ such that $C(\mathcal{A})$ is finite. Then the minimization problem (3.1) has a solution if and only if there exists a μ-measurable subset A^* of F with $|A^*| = \sigma(N)$ such that*

$$\sup\{\mu(\{a\}) : a \in A^*\} \leq \inf\{\mu(\{a'\}) : a' \in F \setminus A^*\}. \tag{3.6}$$

PROOF. From Theorem 3.2.1, $C(\mathcal{A})$ admits a minimizer in Ω if and only if for $\rho-$a.e. $\epsilon > 0$, Problem (3.3) has a solution. As N is totally bounded, $\sigma_\epsilon(N) < \infty$ for any $\epsilon > 0$. As $\epsilon \downarrow 0$, $\sigma_\epsilon(N) \uparrow \sup_c\{\sigma_\epsilon(N) : \epsilon > 0\} = \sigma(N)$, and we consider two cases:

- *The case $\sigma(N) < \aleph_0$:* From the proof of Case 1 in Proposition 2.3.2, it follows that

$$\sup_{\epsilon>0}{}_c\sigma_\epsilon(N) = |N| = \sigma(N) < \infty.$$

  Since each $\sigma_\epsilon(N)$ is integer-valued, there exists an $\epsilon' > 0$ such that $\sigma_\epsilon(N) = \sigma(N) = |N|$ for all $\epsilon \in (0,\epsilon')$. Now, suppose Problem (3.1) has a solution. By Theorem 3.2.1 and Remark 3.1.1, Problem (3.3) has a solution for all $\epsilon > 0$. Since $\rho\big((0,\epsilon')\big) > 0$, for at least one $\epsilon^* \in (0,\epsilon')$, (3.3) has a solution. By Proposition 3.2.1, it follows that there exists an $A^* \subset F$ with $|A^*| = \sigma_{\epsilon^*}(N) = \sigma(N)$ that satisfies (3.6). Conversely, suppose there exists an $A^* = \{a_1, ..., a_{\sigma(N)}\} \subseteq F$ satisfying (3.6). Without loss of generality, we may assume that

$$\mu(\{a_1\}) \leq \mu(\{a_2\}) \leq ... \leq \mu(\{a_{\sigma(N)}\}) \leq \mu(\{a'\}),$$

  for any $a' \in F \setminus A^*$. For any $\epsilon > 0$, let $A^*(\epsilon) := \{a_1, ..., a_{\sigma_\epsilon(N)}\}$, which is a Borel and hence μ-measurable subset of F. Since $\sigma_\epsilon(N) \leq \sigma(N)$, it follows that $A^*(\epsilon)$ satisfies (3.5), and hence is a minimizer for $\min_{|A|\geq\sigma_\epsilon(N)} \mu(A)$ by Proposition 3.2.1. Theorem 3.2.1 then tells us that $C(\mathcal{A})$ indeed has a minimizer $A^*(\epsilon)$ for $\epsilon > 0$.
- *The case $\sigma(N) = \aleph_0$:* Assume that (3.1) has a solution. From Proposition 3.2.1, for (3.3) to admit a solution, we need the existence of a set $A_\epsilon = \{a_1, a_2, ..., a_{\sigma_\epsilon(N)}\}$ such that $\mu(\{a_1\}) \leq \mu(\{a_2\}) \leq ... \leq \mu(\{a_{\sigma_\epsilon(N)}\}) \leq \mu(\{a'\}), \forall a' \in F - A_\epsilon$. As $\epsilon \downarrow 0, \sigma_\epsilon(N) \to \sigma(N) =$

$|\mathbb{N}|$, and we obtain the set $A^* = \bigcup_{\epsilon>0} A_\epsilon$, with $|A^*| = |\mathbb{N}|$, and A^* satisfies (3.6).

Conversely, suppose there exists an $A^* = \{a_1, a_2, a_3, ...\}$ such that $\mu(\{a_i\}) \le \mu(\{a_{i+1}\})$, $\forall i \in \mathbb{N}$, and $\mu(\{a_i\}) \le \mu(\{a'\})$, for any $i \in \mathbb{N}$ and any $a' \in F \setminus A^*$, then for any $\epsilon > 0$, $\{a_1, ..., a_{\sigma_\epsilon(N)}\}$ is a minimizer for (3.3). Theorem 3.2.1 now tells us that $C(\mathcal{A})$ indeed has a minimizer.

□

Remark 3.2.1. *As shown in the following example, the existence of a μ-measurable subset A^* of F, with $|A^*| = \sigma(N)$ and satisfying condition (3.6), depends on the choice of the measure μ on F. Let $M = N = \{n_1, n_2\}$, and define a metric d_M on M such that $d_M(n_1, n_2) = 1$. Let*

$$F = \{\frac{1}{k} : k = 1, 2, 3, ...\} \cup \{0\}.$$

Note that $\sigma_\epsilon(N) = 1$ for any $\epsilon \ge 1$ and $\sigma_\epsilon(N) = 2$ for any $0 < \epsilon < 1$, and thus $\sigma(N) = 2$.

(a) Let μ be a measure on F such that $\mu(a) = a^2$ for any $a \in F$. For any set $A \subset F$ with $|A| = 2$, it holds that

$$\sup\{\mu(\{a\}) : a \in A\} > 0 = \inf\{\mu(\{a'\}) : a' \in F \setminus A\}.$$

Thus, no subset A of F satisfies both $|A| = \sigma(N) = 2$ and condition (3.6).

(b) Let μ be the counting measure on F. Then, for any set $A \subset F$ with $|A| = 2$, it holds that

$$\sup\{\mu(\{a\}) : a \in A\} = 1 = \inf\{\mu(\{a'\}) : a' \in F \setminus A\}.$$

Thus, every subset A of F with $|A| = \sigma(N) = 2$ also satisfies condition (3.6).

3.3. A formula for the minimum cost in Heine-Borel data space

The following theorem establishes a formula for the minimum value of $C(\mathcal{A})$ in the case M is Heine-Borel metric space.

Theorem 3.3.1. *Suppose M is a Heine-Borel metric space, ρ is a measure on $[0, \infty)$ with $\rho\big((0, \epsilon)\big) > 0$ for all $\epsilon > 0$ and $\rho(\{0\}) = 0$. Also, suppose N is totally bounded and let $\{\epsilon_i\}_{i=1}^P$ be the corresponding jump-discontinuity points, listed in decreasing order, of the right-continuous function $\sigma_\epsilon(N)$ on $\epsilon \in (0, \infty)$ with $P \in \mathbb{N} \cup \{\infty\}$. If (3.1) has a solution $\mathcal{A}^*$ with $C(\mathcal{A}^*) < \infty$, then there exists a* *subset* $A^* = \{a_i\}_{i=1}^{\sigma(N)}$ *of* F *with* $\{\mu(\{a_i\})\}_{i=1}^{\sigma(N)}$ *increasing, and*

$$\mu(\{a_i\}) \leq \inf\{\mu(a') : a' \in F \setminus A^*\}, \forall i = 1, 2, \cdots, \sigma(N);$$

such that the minimum value of $C(\mathcal{A})$ *is*

$$\min_{\mathcal{A}\in\Omega} C(\mathcal{A}) = \sum_{i=1}^{P+1} \left(\rho\big([\epsilon_i, \epsilon_{i-1})\big) \sum_{j=1}^{K_i} \mu(a_j) \right),$$

where $\epsilon_0 = \infty$*,* $K_i := \sigma_{\epsilon_i}(N)$ *for each* $1 \leq i \leq P$ *and* $K_{P+1} = \sigma(N), \epsilon_{P+1} = 0$ *when* P *is finite.*

First, we need the following proposition, which also gives us more insight into $\sigma_\epsilon(N)$ as a function of ϵ.

Proposition 3.3.1. *Suppose* M *is a Heine-Borel metric space and* N *is totally bounded. The function* $\epsilon \to \sigma_\epsilon(N)$ *is right continuous.*

Proof. For any fixed $\epsilon > 0$, since N is totally bounded, $\sigma_\epsilon(N) < +\infty$. For ease of notation, we denote $K := \sigma_\epsilon(N)$. By the definition of $\sigma_\epsilon(N)$, the collection of

$$A_\epsilon := \left\{(x_1, ..., x_K) \in M^K : N \subseteq \bigcup_{i=1}^{K} \overline{B(x_i, \epsilon)}\right\}$$

is non-empty. We claim that A_ϵ is a bounded and closed subset in the Heine-Borel space $M^k := M \times M \times \cdots \times M$.

- To prove A_ϵ is bounded, we first fix an $(x_1^*, ..., x_K^*) \in A_\epsilon$. Let $(x_1, ..., x_K)$ be another point in A_ϵ. If there is an x_i such that no points of N is covered by $\overline{B(x_i, \epsilon)}$, the set $\{\overline{B(x_1, \epsilon)}, ..., \overline{B(x_K, \epsilon)}\} \setminus \{\overline{B(x_i, \epsilon)}\}$ covers N, and thus $\sigma_\epsilon(N) \leq K - 1 < K$, a contradiction. Therefore, for all $i \in \{1, ..., K\}$, there exists an $n \in N$ such that $d(x_i, n) \leq \epsilon$. As $\{\overline{B(x_1^*, \epsilon)}, ..., \overline{B(x_K^*, \epsilon)}\}$ covers N, for some x_j^*, $d(x_j^*, n) \leq \epsilon$, and thus $d(x_i, x_j^*) \leq d(x_i, n) + d(n, x_j^*) \leq 2\epsilon$. As this is true for all x_i, and $(x_1^*, ..., x_k^*)$ is fixed, A_ϵ is indeed bounded.
- To prove A_ϵ is closed, suppose $\{(x_1^i, ..., x_K^i)\}$ is a sequence in A_ϵ that converges to some $(x_1^*, ..., x_K^*)$. Given any point $n \in N$, and any $i \in \mathbb{N}$, as $\{\overline{B(x_1^i, \epsilon)}, ..., \overline{B(x_K^i, \epsilon)}\}$ covers N,

$\min\{d(x_1^i,n),...,d(x_K^i,n)\} \le \epsilon$ for all $i \in \mathbb{N}$. Therefore,

$$\begin{aligned}\min\{d(x_1^*,n),\dots,d(x_K^*,n)\} &= \min\{\lim_{i\to\infty} d(x_1^i,n),\dots,\lim_{i\to\infty} d(x_K^i,n)\}\\ &= \lim_{i\to\infty}\min\{d(x_1^i,n),\dots,d(x_K^i,n)\} \le \epsilon.\end{aligned}$$

As this is true for all n in N, N is covered by $\{\overline{B(x_1^*,\epsilon)},...,\overline{B(x_K^*,\epsilon)}\}$, and thus $(x_1^*,...,x_K^*) \in A_\epsilon$. This shows that A_ϵ is closed.

As A_ϵ is closed and bounded in the Heine-Borel metric space M^K, it is compact.

We now fix an $H \in \mathbb{N}$ such that for some $\epsilon > 0$, $\sigma_\epsilon(N) = H$, and let $\epsilon^* = \inf\{\epsilon' > 0 : \sigma_{\epsilon'}(N) = H\}$. As the function $\epsilon \to \sigma_\epsilon(N)$ is a decreasing step function with positive integer values, to prove that it is right continuous, we need only show that $\sigma_{\epsilon^*}(N) = H$.

Consider a sequence ϵ_j in $\mathbb{R}$ such that $\sigma_{\epsilon_j}(N) = H$ for all $j \in \mathbb{N}$, and $\epsilon_j \downarrow \epsilon^*$. The sets A_{ϵ_j} then form a decreasing nested sequence(i.e. $A_{\epsilon_1} \supset A_{\epsilon_2} \supset A_{\epsilon_3} \supset ...$). Since each A_{ϵ_n} is compact and nonempty, the intersection $\bigcap_{j\in\mathbb{N}} A_{\epsilon_j}$ is nonempty. Let $(x_1',...,x_H')$ be a point in $\bigcap_{j\in\mathbb{N}} A_{\epsilon_j}$. Then, for each $j \in \mathbb{N}$, $(x_1',...,x_H')$ lies in A_{ϵ_j} , and thus for any $n \in N$, we have

$$\min\{d(x_1',n),...,d(x_H',n)\} \le \epsilon_j.$$

Taking the limit as $j \to \infty$ on both sides, we obtain

$$\min\{d(x_1',n),...,d(x_H',n)\} \le \epsilon^*.$$

Since this holds for any $n \in N$, $\{\overline{B(x_1',\epsilon^*)},...,\overline{B(x_H',\epsilon^*)}\}$ covers N. Therefore, $\sigma_{\epsilon^*}(N) \le H$. But $\sigma_\epsilon(N)$ is a decreasing function of ϵ, so $\sigma_{\epsilon^*}(N) \ge H$. Therefore, $\sigma_{\epsilon^*}(N) = H$. This shows that the function $\epsilon \to \sigma_\epsilon(N)$ is indeed right continuous. □

Proof of Theorem 3.3.1. We first consider the case where $\sigma(N)$ is finite. Let $\mathcal{A}^*$ be the minimizer of (3.1) with finite $C(\mathcal{A}^*)$. By Theorem 3.2.2, there exists a subset $A^* = \{a_1, a_2, \cdots, a_{\sigma(N)}\}$ of F such that

$$\mu(\{a_1\}) \le \mu(\{a_2\}) \le \mu(\{a_3\}) \le ... \le \mu(a_{\sigma(N)}) \le \inf\{\mu(a') : a' \in F \setminus A^*\}.$$

Since $\epsilon_P > 0$ is the smallest jump discontinuity point of the function $\sigma_\epsilon(N)$ (as a function of ϵ), we have $\sigma_\epsilon(N) = \sigma(N)$ for any $\epsilon \in (0, \epsilon_P)$. Note that, by Proposition 3.2.1, A^* is also a solution to

(3.3) for each $\epsilon \in (0, \epsilon_P)$. Thus,

$$\mu(\mathcal{A}^*(\epsilon)) = \mu(A^*) = \sum_{i=1}^{\sigma(N)} \mu(\{a_i\}).$$

Similarly, for each $i \in \{1, 2, \cdots, P\}$, since the function $\epsilon \to \sigma_\epsilon(N)$ is constant on $(\epsilon_i, \epsilon_{i-1})$ and right continuous at ϵ_i, it follows that $\sigma_\epsilon(N) = K_i$ for each $\epsilon \in [\epsilon_i, \epsilon_{i-1})$. By Proposition 3.2.1, $\{a_1, a_2, \cdots, a_{K_i}\}$ is also a solution to (3.3). Hence,

$$\mu(\mathcal{A}^*(\epsilon)) = \sum_{j=1}^{K_i} \mu(\{a_i\})$$

for each $\epsilon \in [\epsilon_i, \epsilon_{i-1})$. Therefore,

$$\begin{aligned}
\min_{\mathcal{A}\in\Omega} C(\mathcal{A}) &= C(\mathcal{A}^*) = \int_0^\infty \mu\Big(\mathcal{A}^*(\epsilon)\Big) d\rho(\epsilon) \\
&= \sum_{i=1}^{P} \int_{[\epsilon_i, \epsilon_{i-1})} \mu\Big(\mathcal{A}^*(\epsilon)\Big) d\rho(\epsilon) + \int_{[0,\epsilon_P)} \mu\Big(\mathcal{A}^*(\epsilon)\Big) d\rho(\epsilon) \\
&= \sum_{i=1}^{P} \int_{[\epsilon_i, \epsilon_{i-1})} \sum_{j=1}^{K_i} \mu(\{a_i\}) d\rho(\epsilon) + \int_{[0,\epsilon_P)} \sum_{j=1}^{\sigma(N)} \mu(\{a_j\}) d\rho(\epsilon) \\
&= \sum_{i=1}^{P} \Big(\rho[\epsilon_i, \epsilon_{i-1}) \sum_{j=1}^{K_i} \mu(a_j)\Big) + \rho\big([0, \epsilon_P)\big) \sum_{j=1}^{\sigma(N)} \mu(a_j) \\
&= \sum_{i=1}^{P+1} \Big(\rho[\epsilon_i, \epsilon_{i-1}) \sum_{j=1}^{K_i} \mu(a_j)\Big), \text{ since } \rho(\{0\}) = 0.
\end{aligned}$$

A similar argument confirms the formula for the case $\sigma(N) = \infty$. □

3.4. The case of not totally bounded data set

For the case N is not totally bounded, there exists an $\epsilon > 0$ for which $\sigma_\epsilon(N)$ is not finite. Thus we are thus interested in solving (3.4),

$$\min\{\mu(A) : A \subseteq F \text{ is } \mu\text{-measurable with } |A| \geq K\},$$

when K is not finite. Still assuming that μ is a Borel measure on F, and that the measure ρ satisfies $\rho\big((0, \epsilon)\big) > 0$, for all $\epsilon > 0$, we will show in this section that if a solution of (3.1) exists, the value min C-cost is necessarily zero.

We begin with the following proposition, which shows that if the solution to (3.4) exists and its measure is finite, then it must be 0. Recall that for a measure μ, a μ-atom is defined as a positive μ-measurable set that contains no subset of smaller positive μ-measure.

Proposition 3.4.1. *Suppose K is infinite. If A^* is the solution to (3.4) with $\mu(A^*)$ finite, then $\mu(A^*) = 0$.*

Proof. Towards a contradiction, suppose $0 < \mu(A^*) < \infty$. Notice that since μ is a Borel measure on the metric space F, all μ-atoms of F are singletons. If $\mu\lfloor_{A^*}$ is concentrated on such a singleton $a \in A^*$, then $\mu(A^* \setminus \{a\}) = 0 < \mu(A^*)$. But as $|A^*|$ is infinite, $|A^* \setminus \{a\}| = |A^*| \geq K$, thus $A^* \setminus \{a\}$ is also a solution to (3.4), contradicting the minimality of $\mu(A^*)$. Therefore, there exist two disjoint μ-measurable subsets A_1 and A_2 of A^* such that $A_1 \cup A_2 = A^*$, with both $\mu(A_1)$ and $\mu(A_2)$ strictly smaller than $\mu(A^*)$. As one of the sets A_1 or A_2 has the same cardinality as A^*, that set is a solution of (3.4), which again contradicts the minimality of $\mu(A^*)$. □

Remark 3.4.1. *With Proposition 3.4.1 in mind, for the case of infinite K, solving*

$$\max\{|A| : A \ \mu\text{-measurable such that } \mu(A) = 0\} \tag{3.7}$$

can help us solve (3.4), since if $\max|A| \geq K$ then (3.4) and (3.7) have the same solution. Otherwise, (3.4) has no solution with finite measure.

As a consequence of Proposition 3.4.1, if $\mathcal{A}^*$ solves (3.1) with finite $C(\mathcal{A}^*)$, then $C(\mathcal{A}^*) = 0$.

Proposition 3.4.2. *Suppose N is not totally bounded. Suppose ρ is a measure on $(0, \infty)$ such that for all $\epsilon > 0, \rho\big((0, \epsilon)\big) > 0$. Suppose also that for ρ-a.e. $\epsilon > 0$ (3.2) has a solution. If $\mathcal{A}^*$ solves (3.1) with $C(\mathcal{A}^*) < \infty$, then $C(\mathcal{A}^*) = 0$.*

Proof. Let $\mathcal{A}^*$ be a minimizer of $C(\mathcal{A})$ with $C(\mathcal{A}^*) < \infty$. By Theorem 3.1.2, for ρ-a.e. $\epsilon > 0$, $\mathcal{A}^*(\epsilon)$ is a solution of (3.3). Since

$$\int_0^\infty \mu\big(\mathcal{A}^*(\epsilon)\big) d\rho(\epsilon) = C(\mathcal{A}^*) < \infty,$$

$\mu\big(\mathcal{A}^*(\epsilon)\big)$ is a non-increasing function of ϵ (Lemma 3.1.1), and $\rho\big((0, \epsilon)\big) > 0$ for all $\epsilon > 0$, it follows that $\mu\big(\mathcal{A}^*(\epsilon)\big) < \infty$ for all $\epsilon > 0$. As N is not totally bounded, there exists an $\epsilon^* > 0$ such that

$\sigma_{\epsilon^*}(N)$ is infinite. For all $0 < \epsilon \leq \epsilon^*$, since $\sigma_\epsilon(N)$ is a non-increasing function of ϵ, it follows that $\sigma_\epsilon(N) \geq \sigma_{\epsilon^*}(N)$, and hence $\sigma_\epsilon(N)$ is also infinite. By Proposition 3.4.1, $\mu\big(\mathcal{A}^*(\epsilon)\big) = 0$ for all $0 < \epsilon \leq \epsilon^*$. Again, as $\mu\big(\mathcal{A}^*(\epsilon)\big)$ is non-increasing in ϵ, we have $\mu\big(\mathcal{A}^*(\epsilon)\big) = 0$ for all $\epsilon > 0$. As a result,

$$\min_{\mathcal{A}\in\Omega} C(\mathcal{A}) = C(\mathcal{A}^*) = \int_0^\infty \mu\big(\mathcal{A}^*(\epsilon)\big) d\rho(\epsilon) = \int_0^\infty 0 d\rho(\epsilon) = 0.$$

□

CHAPTER 4

Optimal Codepage and Codebook: Second-Type Cost Function

In Chapter 3, to compute the cost function we assign $\mu\big(\mathcal{A}(\epsilon)\big)$ as the cost for each codepage $\mathcal{A}(\epsilon)$, where μ is a pre-determined measure on F. The underlying assumption is that the cost of each codepage depends solely on the cost of the codewords themselves. In general, instead of representing the cost of A by $\mu(A)$ using a measure μ, we can represent it using a non-negative function I, denoted as $I(A)$. The function I that we focus on in Chapter 4 takes into account the reconstruction error, as well as the cost of the encoding and decoding process. Section 4.1 gives a proper definition for this cost function and proves the existence of an optimal codepage. Section 4.2 introduces a topology on the space of codepages based on $d_{\mathcal{H},L_q}$, the Hausdorff distance between the sets of pairs of encoder-decoder associated with those codepages. Section 4.3 describes a cost function on codebooks and proves the existence of an optimal codebook. In section 4.4 we illustrate an example of finding the optimal codebook. Finally, section 4.5 discusses the estimation of $d_{\mathcal{H},L_q}$ under various conditions. Throughout Chapter 4 we continue to use the notations introduced in Definition 2.1.1.

4.1. Cost function and existence theorem for optimal codepage

Suppose $A \subset F$ is an ϵ-codepage. Consider the function $C : N \times F \times M \to [0, \infty)$ defined by

$$C(x, y, z) = C_1(x, y) + C_2(y, z) + C_3(x, z),$$

where $C_1 : N \times F \to [0, \infty)$, $C_2 : F \times M \to [0, \infty)$ and $C_3 : N \times M \to [0, \infty)$ are lower semi-continuous. Suppose A is a subset of F. Note that for each $n \in N$, for each $f : N \to A$ and $g : A \to M$,

$$C\Big(n, f(n), g \circ f(n)\Big) = C_1\Big(n, f(n)\Big) + C_2\Big(f(n), g \circ f(n)\Big) + C_3\Big(n, g \circ f(n)\Big)$$

represents the sum of the encoding cost, the decoding cost, and the error cost, respectively. Suppose ν is a measure on N, let $\mathcal{E}(N, A)$ be the space of all ν-measurable encoding functions from N to

A. Let $\mathcal{D}(A, M)$ be the space of continuous decoding functions from A to M.

For Chapter 4, we relax the definition of ϵ-linkable pairs (Definition 2.1.2) as follows.

DEFINITION 4.1.1. *Let $f : N \to A$ and $g : A \to M$. We say (f, g) is an a.e. ϵ-linkable if*

$$d\Big(g \circ f(n), n\Big) \leq \epsilon, \ \textit{for } \nu - a.e. n \in N.$$

Suppose $\mathcal{P}$ is a subset of $\mathcal{E}(N, A) \times \mathcal{D}(A, M)$. Let $\mathcal{L}(\epsilon, A; \mathcal{P})$ denote the subset of $\mathcal{P}$ which consists of all a.e. $\epsilon-$linkable pairs (f, g) in $\mathcal{P}$. Let $S : \mathcal{E}(N, A) \to [0, \infty)$ represents the storage cost of $f \in \mathcal{E}(N, A)$. We propose to consider the following *Optimal Coding-Pair Problem*:

$$\text{Minimize } J(f, g) := \int_N C\Big(n, f(n), g \circ f(n)\Big) d\nu(n) + S(f), \tag{4.1}$$

among all $(f, g) \in \mathcal{L}(\epsilon, A; \mathcal{P})$. The cost of the codepage A is then defined as

$$I(A) := \inf_{(f,g) \in \mathcal{L}(\epsilon, A; \mathcal{P})} J(f, g). \tag{4.2}$$

In order to establish the existence of a solution to Problem (4.1), we first need an appropriate topology on $\mathcal{P}$. We endow $\mathcal{E}(N, A)$ with the pointwise convergence topopolgy, and $\mathcal{D}(A, M)$ with the compact-open topology (Definition 1.2.6). The topology on $\mathcal{E}(N, A) \times \mathcal{D}(A, M)$ is the product topology. Theorem 4.1.2 provides the conditions that guarantee the existence of a J-minimizer for (4.1).

THEOREM 4.1.2. *Let A be an $\epsilon-$codepage for some $\epsilon > 0$. Suppose $\mathcal{P}$ is sequentially compact, and C and S are lower semi-continuous. If the collection $\mathcal{L}(\epsilon, A; \mathcal{P})$ is nonempty, then Problem (4.1) has a solution.*

PROOF. Let (f_k, g_k) be a $J-$minimizing sequence in $\mathcal{L}(\epsilon, A; \mathcal{P}) \subseteq \mathcal{P}$. By the sequential compactness of $\mathcal{P}$, there exists a subsequence of (f_k, g_k), still denoted by (f_k, g_k), that converges to some $(f^*, g^*) \in \mathcal{P}$. That is, $f_k(n) \to f^*(n)$ pointwise in N and $g_k \to g^*$ in the compact-open topology of $\mathcal{D}(A, M)$. Thus, from Proposition 1.2.2, it follows that $g_k(f_k(n)) \to g^*(f^*(n))$ for all $n \in N$. Since each (f_k, g_k) is a.e. ϵ-linkable,

$$d\Big(g_k \circ f_k(n), n\Big) \leq \epsilon, \ \text{for a.e } n \in N.$$

Let $k \to \infty$, it gives

$$d\Big(g^* \circ f^*(n), n\Big) \leq \epsilon, \text{ for a.e. } n \in N.$$

Therefore $(f^*, g^*) \in \mathcal{L}(\epsilon, A; \mathcal{P})$. By Fatou's Lemma [3, Theorem 1.3.1] and the lower semi-continuity of both C and S, we have

$$\begin{aligned}
& J(f^*, g^*) \\
= & \int_N C\Big(n, f^*(n), g^* \circ f^*(n)\Big) d\nu(n) + S(f^*) \\
\leq & \int_N \liminf_{k\to\infty} C\Big(n, f_k(n), g_k \circ f_k(n)\Big) d\nu(n) + \liminf_{k\to\infty} S(f_k) \\
\leq & \liminf_{k\to\infty} \int_N C\Big(n, f_k(n), g_k \circ f_k(n)\Big) d\nu(n) + \liminf_{k\to\infty} S(f_k) \\
\leq & \liminf_{k\to\infty} \left[\int_N C\Big(n, f_k(n), g_k \circ f_k(n)\Big) d\nu(n) + S(f_k)\right] \\
= & \liminf_{k\to\infty} J(f_k, g_k) = \inf_{(f,g)\in\mathcal{L}(\epsilon,A;\mathcal{P})} J(f, g).
\end{aligned}$$

Thus we achieve the desired result

$$J(f^*, g^*) = \min_{(f,g)\in\mathcal{L}(\epsilon,A;\mathcal{P})} J(f, g).$$

□

REMARK 4.1.1. *Under the conditions of Theorem 4.1.2, the cost function I from (4.2) becomes*

$$I(A) := \min_{(f,g)\in\mathcal{L}(\epsilon,A,\mathcal{P}_A)} J(f, g). \tag{4.3}$$

For the special case where $\mathcal{P} = \mathcal{E}(N, A) \times \mathcal{D}(A, M)$, we may denote $\mathcal{L}(\epsilon, A; \mathcal{P}) = \mathcal{L}(\epsilon, A; \mathcal{E}, \mathcal{D})$.

COROLLARY 4.1.1. *Let A be an $\epsilon-$codepage for some $\epsilon > 0$. Suppose both $\mathcal{E}(N, A)$ and $\mathcal{D}(A, M)$ are sequentially compact. If the collection $\mathcal{L}(\epsilon, A; \mathcal{D}, \mathcal{E})$ is nonempty, then Problem (4.1) has a solution for $\mathcal{P} = \mathcal{E}(N, A) \times \mathcal{D}(A, M)$.*

EXAMPLE 4.1.3. *In this example, we illustrate Theorem 4.1.2 for a simple Autoencoder. Suppose $X \subset \mathbb{R}^n$ is a finite set, where X represents the data set. We equip X with the counting measure. Let $Y = \mathbb{R}^m$, with $m < n$, be the latent space. The encoders and decoders of the Autoencoder are* *defined based on the Sigmoid function $S : \mathbb{R} \to \mathbb{R}$, which is defined by*

$$S(z) := \frac{1}{1+e^{-z}}, \text{ for any } z \in \mathbb{R}.$$

Let $\mathcal{W} \subset \mathbb{R}^{m\times n}$ and $\mathcal{V} \subset \mathbb{R}^{n\times m}$ represent the sets of weight matrices. Let $B \subset \mathbb{R}^m$ and $A \subset \mathbb{R}^n$ represent the sets of biases. The weight matrices and biases represent the parameters of the Autoencoder. For any $W \in \mathcal{W}$ and $\beta \in B$, we define the encoding function $E_{W,\beta} : X \to \mathbb{R}^m$ by

$$E_{W,\beta}(x) := \Big(S(-w_1 \cdot x - b_1), \ldots, S(-w_n \cdot x - \beta_m)\Big), \text{ for any } x \in X,$$

where $w_i (i = 1, \ldots, m)$ represents the i-th row of the matrix W, and $\beta_j (j = 1, ..., m)$ represents the j-th entry of the vector β. Let E be the set of all such $E_{W,\beta}$.

Similarly, for any $V \in \mathcal{V}$ and $\alpha \in A$, we define the decoding $D_{V,\alpha} : \mathbb{R}^m \to \mathbb{R}^n$ by

$$D_{V,\alpha}(y) := \Big(S(-v_1 \cdot y - \alpha_1), \ldots, S(-v_n \cdot y - \alpha_n)\Big), \text{ for any } y \in \mathbb{R}^m,$$

where $v_i (i = 1, \ldots, n)$ represents the i-th row of the matrix V, and $\beta_j (j = 1, ..., n)$ represents the j-th entry of the vector α. Let D be the set of all such $D_{V,\alpha}$.

Let $\epsilon > 0$ be given. Suppose $m, \mathcal{W}, \mathcal{V}, B, A$ are such that $\mathcal{L}(\epsilon, A; D, E)$ is non empty. Recall that the learning process for the Autoencoder aims at solving Problem (1.1). In this case, we consider solving the following problem:

$$\min_{(E_{W,\beta}, D_{V,\alpha}) \in E\times D} \sum_{x\in X} |x - D_{V,\alpha} \circ E_{W,\beta}(x)| \tag{4.4}$$

Suppose the sets $\mathcal{W}, \mathcal{V}, B, A$ are compact. For any $x \in X$, since $\mathcal{W}$ and B are compact, the sets $\{-w_i \cdot x - b_i : w_i \in \mathcal{W}, b_i \in B\}$, for $i = 1, \ldots, m$, are sequentially compact. As S is continuous $\{E_{W,\beta}(x) : E_{W,\beta} \in E\}$ is then also sequentially compact. As X is finite, E is sequentially compact in the pointwise convergence topology.

Similarly, the set $\{D_{V,\alpha}(x) : D_{V,\alpha} \in D\}$ is sequentially compact for any $x \in X$. As X is finite, pointwise convergence on X implies uniform convergence. Consequently, D is sequentially compact in the compact open topology. From Corollary 4.1.1, Problem (4.4) has a solution.

4.2. A topology on the space of codepages

This section describes a topology on the set of ϵ-codepages based on the Hausdorff distance (Definition 1.2.9). We then show that the cost function I (4.2) is lower-semicontinuous under this topology.

For any $\epsilon > 0$, let Γ_ϵ be a collection of ϵ-codepages of N in F and

$$\Gamma := \bigcup_{\epsilon>0} \Gamma_\epsilon.$$

For any ϵ-codepage A, we associate with it a unique $\mathcal{P}_A \subset \mathcal{E}(N, A) \times \mathcal{D}(A, M)$, and we denote

$$\mathcal{L}_\Gamma := \bigcup_{A \in \Gamma} \mathcal{L}(\epsilon, A, \mathcal{P}_A).$$

Definition 4.2.1. *Let $(X, d_X), (Y, d_Y)$ be two metric spaces, ν is a measure on X, and $q > 1$. For any two ν-measurable functions $\phi, \psi : X \to Y$, their L_q distance is defined as*

$$L_q(\phi, \psi) := \Big(\int_X d_Y \Big(\phi(x), \psi(x)\Big)^q d\nu(x) \Big)^{1/q}.$$

Remark 4.2.1. *Two ν-measurable functions ψ and ϕ from X to Y are equivalent if $\psi(x) = \phi(x)$ for ν-a.e $x \in X$. We establish the following properties for L_q.*

- *Given $\phi, \psi, \theta : X \to Y$ ν-measurable, from Minkowski's and the triangle inequality*

$$\begin{aligned} L_q(\phi, \psi) &= \Big(\int_X d\Big(\phi(n), \psi(n)\Big)^q d\nu(n) \Big)^{1/q} \\ &\le \Big(\int_X \Big(d\Big(\phi(n), \theta(n)\Big) + d\Big(\theta(n), \psi(n)\Big) \Big)^q d\nu(n) \Big)^{1/q} \\ &\le \Big(\int_X d\Big(\phi(n), \theta(n)\Big)^q d\nu(n) \Big)^{1/q} + \Big(\int_X d\Big(\theta(n), \psi(n)\Big)^q d\nu(n) \Big)^{1/q} \\ &= L_q(\phi, \theta) + L_q(\theta, \psi). \end{aligned}$$

  This establishes the triangle inequality for L_q, thus showing that L_q defines a metric on all equivalent ν-measurable functions from X to Y.

- *For any two sequences of functions (ϕ_k) and (ψ_k) with $L_q(\phi_k, \psi_k) \to 0$, it follows that $d_Y(\phi_k(x), \psi_k(x)) \to 0$ subsequently ν-a.e $x \in X$.*

Using L_q, we now introduce a distance on $\mathcal{L}_\Gamma$. For any two pairs of functions (f,g) and $(f',g') \in \mathcal{L}_\Gamma$, by viewing $f, f' : N \to F$ and $g \circ f, g' \circ f' : N \to M$ as ν-measurable maps between metric spaces, one can define the distance between them as

$$d_q\Big((f,g),(f',g')\Big) := L_q(f,f') + L_q(g \circ f, g' \circ f').$$

It is straightforward to verify that d_q is a pseudometric on $\mathcal{L}_\Gamma$, with $(f,g) = (f',g')$ if $f = f'$ and $g \circ f = g' \circ f'$ ν-a.e. If we define the equivalence relation $\sim$ on $\mathcal{L}_\Gamma$ by $(f.g) = (f'g')$ when $f = f'$ and $g \circ f = g' \circ f'$ ν-a.e., d_q defines a metric on $\mathcal{L}_\Gamma / \sim$. We now use this metric and the Hausdorff distance (Definition 1.2.9) to introduce a distance between codepages. For any codepage A, the corresponding collection $\mathcal{L}(\epsilon_A, A, \mathcal{P}_A)$ is a subset of the metric space $(\mathcal{L}_\Gamma, d_q)$. Using it, we introduce a distance between codepages as follows.

Definition 4.2.2. *For any two codepages A and B in Γ and $q > 1$, we define the distance $d_{H,q}$ between A and B by*

$$d_{H,q}(A,B) := d_{\mathcal{H},L_q}\Big(\mathcal{L}(\epsilon_A, A, \mathcal{P}_A)/\sim, \mathcal{L}(\epsilon_B, B, \mathcal{P}_B) \sim\Big).$$

Remark 4.2.2. *Note that for $d_{H,q}$, $d_{H,q}(A,B) = 0$ does not necessarily mean $A = B$, but only $\overline{\mathcal{L}(\epsilon_A, A, \mathcal{P}_A)/\sim} = \overline{\mathcal{L}(\epsilon_B, B, \mathcal{P}_B)/\sim}$ (Proposition 1.2.5). Consequently, $d_{H,q}$ does not define a metric, only a pseudometric. However, this pseudometric can still define a topology on Γ, whose base consists of the open balls whose centers are elements in Γ.*

The remainder of the section will be dedicated to showing that the cost function I is lower semi-continuous. First, we establish the following proposition, which describes the lower semi-continuity of $J : \mathcal{L}_\Gamma \to \mathbb{R}$ given by

$$J(f,g) = \int_N C\Big(n, f(n), g \circ f(n)\Big) d\nu(n) + S(f). \tag{4.5}$$

Proposition 4.2.1. *Suppose $C : N \times F \times M \to \mathbb{R}$ and $S : \mathcal{L}_\Gamma \to \mathbb{R}$ are both lower-semicontinuous. For any sequence $\{(f_k, g_k)\}$ and (f_0, g_0) in $\mathcal{L}_\Gamma$, if $f_k \to f_0$ and $g_k \circ f_k \to g_0 \circ f_0$ pointwise ν-a.e., then*

$$J(f_0, g_0) \le \liminf_{k \to \infty} J(f_k, g_k).$$

Proof. Suppose $(f_k, g_k) \to (f, g) \in \mathcal{L}_\Gamma$. From lower-semicontinuity of C and S,

- $\liminf_{k\to\infty} C\Big(n, f_k(n), g_k \circ f_k(n)\Big) \geq C\Big(n, f(n), g\circ f(n)\Big)$,
- $\liminf_{k\to\infty} S(f_k) \geq S(f)$.

From Fatou's Lemma,

$$\begin{aligned}\liminf_{k\to\infty} \int C\Big(n, f_k(n), g_k\circ f_k(n)\Big) d\nu(n) &\geq \int \liminf_{k\to\infty} C\Big(n, f_k(n), g_k \circ f_k(n)\Big) d\nu(n)\Big) d\nu(n) \\ &\geq \int C\Big(n, f(n), g\circ f(n)\Big) d\nu(n).\end{aligned}$$

Hence,

$$\begin{aligned}\liminf_{k\to\infty} J(f_k, g_k) &= \liminf_{k\to\infty} \Big(\int C\Big(n, f_k(n), g_k\circ f_k(n)\Big) d\nu(n) + S(f_k)\Big) \\ &\geq \liminf_{k\to\infty} \int C\Big(n, f_k(n), g_k\circ f_k(n)\Big) d\nu(n) + \liminf_{k\to\infty} S(f_k). \\ &\geq \int C\Big(n, f(n), g\circ f(n)\Big) d\nu(n) + S(f) = J(f,g),\end{aligned}$$

which is our desired result. □

Lemma 4.2.1 describes a useful property when A_k converges to A_0 in $d_{H,q}$.

LEMMA 4.2.1. *Let $A_0, A_k \in \Gamma$ be codepages and $(f_k, g_k) \in \mathcal{L}(\epsilon_{A_k}, A_k, \mathcal{P}_{A_k})$ for each $k = 1, 2, \cdots$. If $A_k \xrightarrow{d_{H,q}} A_0$ and $\mathcal{P}_{A_0}$ is sequentially compact, then there exists an element $(f_0, g_0) \in \mathcal{L}(\epsilon_{A_0}, A_0, \mathcal{P}_{A_0})$ such that*

$$f_k(n) \to f_0(n), g_k \circ f_k(x) \to g_0 \circ f_0(x)$$

subsequently pointwise ν-a.e. $n \in N$.

PROOF. Since $\lim_{k\to 0} d_{H,q}(A_k, A_0) = 0$, by the definition of the distance $d_{H,q}$, there is a sequence $(f_k^0, g_k^0) \in \mathcal{L}(\epsilon_0, A_0, \mathcal{P}_0)$ such that $\lim_{k\to 0} d_p\Big((f_k, g_k), (f_k^0, g_k^0)\Big) = 0$. Hence,

$$L_q(f_k, f_k^0) \to 0, \text{ and } L_q(g_k \circ f_k, g_k^0 \circ f_k^0) \to 0.$$

From Remarks 4.2.1, this implies

$$d\big(f_k(n), f_k^0(n)\big) \to 0, \text{ and } d\big(g_k\circ f_k(n), g_k^0 \circ f_k^0(n)\big) \to 0, \text{ subsequently } \nu - a.e.$$

On the other hand, as $\mathcal{L}(\epsilon_0, A_0, \mathcal{P}_0)$ is a closed subset of the sequentially compact set $\mathcal{P}_0$, it is itself sequentially compact. Hence (f_k^0, g_k^0) converges subsequently to some (f_0, g_0) in $\mathcal{L}(\epsilon_0, A_0, \mathcal{P}_0)$, which implies that

$$f_k^0(n) \to f_0(n), \text{ and } g_k^0 \circ f_k^0(n) \to g_0 \circ f_0(n),$$

for $n \in N$.

As a result, subsequently,

$$f_k \to f_0, \text{ and } g_k \circ f_k \to g_0 \circ f_0,$$

subsequently pointwise ν-a.e. $n \in N$. □

The following proposition establishes the semicontinuousness of I.

PROPOSITION 4.2.2. *Assuming $\mathcal{P}_{A_k}$ is sequentially compact for all k. If $A_k \xrightarrow{d_{H,q}} A_0$, then*

$$I(A_0) \leq \liminf_{k\to\infty} I(A_k).$$

PROOF. By picking a subsequence if necessary, without loss of generality, we may assume that

$$\lim_{k\to\infty} I(A_k) = \liminf_{k\to\infty} I(A_k). \tag{4.6}$$

For each k, since $\mathcal{P}_{A_k}$ is sequentially compact, by Theorem 4.1.2, there exists a pair $(f_k, g_k) \in \mathcal{L}(\epsilon_k, A_k, \mathcal{P}_k)$ such that $I(A_k) = J(f_k, g_k)$. From Lemma 4.2.1, there exists a pair $(f_0, g_0) \in \mathcal{L}(\epsilon_0, A_0, \mathcal{P}_0)$ such that $f_k(n) \to f_0(n)$ and $g_k \circ f_k(x) \to g_0 \circ f_0(x)$ subsequently pointwise ν-a.e. $n \in N$. By the minimality of $I(A_0)$, Proposition 4.2.1 and (4.6),

$$I(A_0) \leq J(f_0, g_0) \leq \liminf_{k\to\infty} J(f_k, g_k) = \liminf_{k\to\infty} I(A_k).$$

□

4.3. Existence of optimal codebook

In the previous section, we have defined a topology on the space of codepage. In this section, we define a cost function on codebook and prove the existence of an optimal codebook. First, we need some notations. Recall that the Hausdorff pseudometric on $\mathcal{L}_\Gamma$ generates a topology on Γ. We let $\mathcal{B}_{H,q}$ denote the σ-algebra on Γ generated by this topology.

Now let

$$\Omega_0 := \{\mathcal{A} : (0, \infty) \to \Gamma; \mathcal{A}(\epsilon) \in \Gamma_\epsilon, \forall \epsilon > 0, \mathcal{A} \text{ is } (\rho, \mathcal{B}_{H,q})\text{-measurable}\}.$$

Suppose $I : \Gamma \to [0, \infty]$ is a cost function, and ρ is a measure on $(0, \infty)$, we define the cost on the codebook $\mathcal{A}$ as

$$\mathcal{S}(\mathcal{A}) := \int_0^\infty I\Big(\mathcal{A}(\epsilon)\Big) d\rho(\epsilon). \tag{4.7}$$

whenever $I \circ \mathcal{A}$ is ρ-measurable.

Let $\Omega \subset \Omega_0$ be non-empty, we consider the problem:

$$\text{Minimize } \mathcal{S}(\mathcal{A}) \text{ among all } \mathcal{A} \in \Omega. \tag{4.8}$$

In the following theorem $\mathcal{P}_A$ is sequentially compact for any $A \in \Gamma$, and I is defined as in (4.3).

THEOREM 4.3.1. *Suppose Ω is either compact or sequentially compact in the ρ-a.e. pointwise topology, then Problem (4.8) has a solution.*

PROOF. From Proposition 4.2.2, I is lower-semicontinuous on Γ, thus it is Borel measurable. For any $\mathcal{A}$ in Ω, $\mathcal{A}$ is $(\rho, \mathcal{B}_{H,q})$-measurable, thus $I \circ \mathcal{A}$ is ρ-measurable. $\mathcal{S}$ is therefore well-defined. When Ω is either compact or sequentially compact, to establish the existence of an $\mathcal{S}$-minimizer for the Problem (4.8), it is sufficient to show that $\mathcal{S} : \Omega \to [0, \infty]$ is lower-semicontinuous. Indeed, suppose $(\mathcal{A}_n)$ converges pointwisely ρ-a.e. to $\mathcal{A}$ in Ω, then for ρ-a.e. $\epsilon > 0$, $\mathcal{A}_n(\epsilon) \xrightarrow{d_{H,q}} \mathcal{A}(\epsilon)$ in Γ_ϵ. From Proposition 4.2.2,

$$I \circ \mathcal{A}(\epsilon) \leq \liminf_{n \to \infty} I \circ \mathcal{A}_n(\epsilon).$$

By Fatou's Lemma,

$$\begin{aligned} \mathcal{S}(\mathcal{A}) &= \int_0^\infty I \circ \mathcal{A}(\epsilon) d\rho(\epsilon) \leq \int_0^\infty \liminf_{n \to \infty} I \circ \mathcal{A}_n(\epsilon) d\rho(\epsilon) \\ &\leq \liminf_{n \to \infty} \int_0^\infty I \circ \mathcal{A}_n(\epsilon) d\rho(\epsilon) = \liminf_{n \to \infty} \mathcal{S}(\mathcal{A}_n). \end{aligned}$$

Therefore, $\mathcal{S}$ is lower semicontinuous, and hence Problem (4.8) has a solution. □

REMARK 4.3.1. *Note that in general, sequentially compactness is not equivalent to compactness.*

COROLLARY 4.3.1. *Suppose for ρ-a.e. $\epsilon > 0$, Γ_ϵ is compact. Suppose the measure ρ is such that any $\mathcal{A} : (0, \infty) \to \Gamma : \mathcal{A}(\epsilon) \in \Gamma_\epsilon, \forall \epsilon > 0$, is $(\rho, \mathcal{B}_{H,q})$-measurable. Then (4.8) has a solution.*

Proof. As

$$\Omega = \{\mathcal{A} : \text{supp}(\rho) \to \Gamma, \mathcal{A}(\epsilon) \in \Gamma_\epsilon, \forall \epsilon\} = \prod_{\epsilon \in \text{supp}(\rho)} \Gamma_\epsilon,$$

Ω is compact(Tychonoff's Theorem). By Theorem 4.3.1, Problem (4.8) has a solution. □

4.4. An illustrative example

Let $M = \mathbb{R}, N = [0,1]$, $F = \mathbb{Q}$ and ν be a non-atomic measure on N. For each $k \in \mathbb{N}$, define

$$A_k = \{0, \frac{1}{k}, \frac{2}{k}, \cdots, 1, 1 + \frac{1}{k}\}.$$

Lemma 4.4.1. *For any $\epsilon > 0$, if $k \geq \frac{1}{2\epsilon}$, then A_k is an ϵ-codepage.*

Proof. We define $f : N \to A_k$ and $g_k : A_k \to M$ as follow:

$$f_k(n) := \frac{[kn + \frac{1}{2}]}{k}, \forall n \in N \text{ and } g_k(a) := a, \forall a \in A_k.$$

Then, $f_k \in \mathcal{E}(N, A_k)$ and $g_k \in \mathcal{D}(A_k, M)$ with

$$|g_k(f_k(n)) - n| = \left|\frac{[kn + \frac{1}{2}]}{k} - n\right| = \left|\frac{[kn + \frac{1}{2}] - kn}{k}\right| \leq \frac{1/2}{k} = \frac{1}{2k}.$$

Thus, when $k \geq \frac{1}{2\epsilon}$, (f_k, g_k) is an ϵ-linkable pair and hence A_k is an ϵ-codepage. □

We now consider pairs of functions in the following form: for each $\alpha \in [0,1]$ and $\beta \in [-1,1]$, consider the functions

$$f_k^\alpha : [0,1] \to A_k \text{ with } f_k^\alpha(n) = \frac{[kn + \alpha]}{k},$$

and

$$g_k^\beta : A_k \to \mathbb{R} \text{ with } g_k^\beta(a) = a + \frac{\beta}{k}.$$

Remark 4.4.1. *We want to illustrate how the pairs (f_k^α, g_k^β) can be used for encoding images using the MNIST and CIFAR-10 datasets. Recall The MNIST dataset consists of images of handwritten numbers from 0 to 9. Each image is in the form of a 28×28 pixel grayscale bounding box, with pixel values ranging from 0 (black) to 255 (white). The CIFAR 10 dataset on the other hand is a collection of color images, each measuring 32×32 pixels. Each pixel contains three channels (red, green, blue), with values ranging from 0 to 255.*

For the MNIST dataset, we consider $N = \{0, 1, 2, 3, ..., 255\}$. Let $F_k^\alpha : N^{28\times 28} \to \mathbb{Q}^{28\times 28}$, where

$F_k^\alpha = (f_k^\alpha)^{28\times 28}$, and $G_k^\beta = (g_k^\beta)^{28\times 28}$. For CIFAR 10, again we consider $N = \{0, 1, 2, 3, ..., 255\}$. Let $F_k^\alpha : N^{28\times 28\times 3} \to \mathbb{Q}^{28\times 28\times 3}$, where $F_k^\alpha = (f_k^\alpha)^{28\times 28\times 3}$, and $G_k^\beta = (g_k^\beta)^{28\times 28\times 3}$. Figure 4.1 illustrates an example of applying $G_k^\beta \circ F_k^\alpha$ to an MNIST and a CIFAR image.

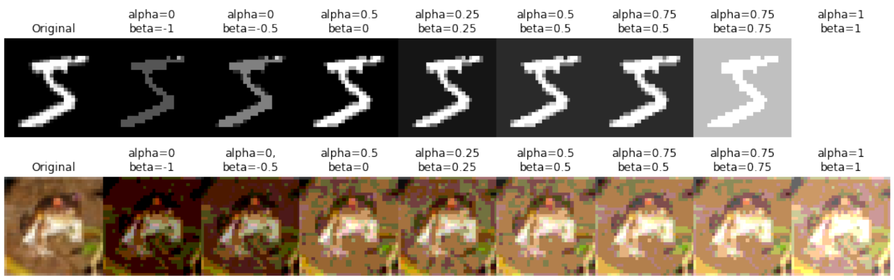

FIGURE 4.1. Example of applying $G_k^\beta \circ F_k^\alpha$ to an MNIST and CIFAR image.

Consider the collection

$$\mathcal{P}_k := \{(f_k^\alpha, g_k^\beta) : \alpha \in [0, 1], \beta \in [-1, 1]\}.$$

LEMMA 4.4.2. *$\mathcal{P}_k$ is a sequentially compact subset of $\mathcal{E}(N, A_k) \times \mathcal{D}(A_k, M)$.*

PROOF. Suppose $(f_k^{\alpha_i}, g_k^{\beta_i})$ is a sequence in $\mathcal{P}_k$. As $[0, 1] \times [-1, 1]$ is compact, (α_i, β_i) converges subsequently to some (α, β) in $[0, 1] \times [-1, 1]$. We still denote the converging subsequence as (α_i, β_i). For each $n \in N = [0, 1]$ with $kn + \alpha \notin \mathbb{Z}$, it follows that $[kn + \alpha_i] \to [kn + \alpha]$, and hence $f_k^{\alpha_i}(n) \to f_k^\alpha(n)$ as $i \to \infty$. This shows that $f_k^{\alpha_i}$ converges to f_k^α pointwise $\nu-$a.e. since ν has no atoms. Also, since $\frac{\beta_i}{k} \to \frac{\beta}{k}$, the continuous functions $g_k^{\beta_i}(a)$ converges to $g_k^\beta(a)$ uniformly on A_k. Thus $(f_k^{\alpha_i}, g_k^{\beta_i})$ converges to (f_k^α, g_k^β), and $\mathcal{P}_k$ is indeed sequentially compact. □

The following lemma establishes the criteria under which a pair (f_k^α, g_k^β) is ϵ-linkable.

LEMMA 4.4.3. *For any $\epsilon > 0$, a pair (f_k^α, g_k^β) in $\mathcal{P}_k$ is ϵ-linkable if and only if*

$$k\epsilon \geq \max\{|\alpha + \beta|, |\alpha + \beta - 1|\}.$$

PROOF. By definition, (f_k^α, g_k^β) is ϵ-linkable if and only if

$$\sup_{n\in[0,1]} \left|g_k^\beta \circ f_k^\alpha(n) - n\right| \leq \epsilon.$$

For any $n \in [0,1]$,

$$\left|g_k^\beta \circ f_k^\alpha(n) - n\right| = \left|\frac{[kn+\alpha]}{k} + \frac{\beta}{k} - n\right| = \frac{1}{k}\left|[kn+\alpha] - kn + \beta\right|.$$

Also, since $kn + \alpha - 1 < [kn+\alpha] \leq kn + \alpha$, it follows that

$$\left|[kn+\alpha] - kn + \beta\right| \leq \max\left\{|\alpha+\beta|, |\alpha+\beta-1|\right\}.$$

Indeed,

$$\sup_{n\in[0,1]} \left|[kn+\alpha] - kn + \beta\right| = \max\left\{|\alpha+\beta|, |\alpha+\beta-1|\right\}.$$

Therefore, (f_k^α, g_k^β) is ϵ-linkable if and only if $\max\left\{|\alpha+\beta|, |\alpha+\beta-1|\right\} \leq k\epsilon$. □

The following lemma establishes the criteria for $\mathcal{L}(\epsilon, A_k; \mathcal{P}_k)$ to be nonempty.

Lemma 4.4.4. *For any $\epsilon > 0$, the collection $\mathcal{L}(\epsilon, A_k; \mathcal{P}_k)$ is nonempty if and only if $k \geq \lceil \frac{1}{2\epsilon} \rceil$.*

Proof. If $k \geq \frac{1}{2\epsilon}$, by the proof of Lemma 4.4.1, it follows that $(f_k^{1/2}, g_k^0) = (f_k, g_k)$ is an ϵ-linkable pair in $\mathcal{P}_k$, and hence $\mathcal{L}(\epsilon, A_k; \mathcal{P}_k)$ is nonempty. On the other hand, suppose $\mathcal{L}(\epsilon, A_k; \mathcal{P}_k)$ is nonempty. Let (f_k^α, g_k^β) be any ϵ-linkable pair in $\mathcal{P}_k$. By Lemma 4.4.3,

$$k\epsilon \geq \max\left\{|\alpha+\beta|, |\alpha+\beta-1|\right\} = \left|(\alpha+\beta) - \frac{1}{2}\right| + \frac{1}{2} \geq \frac{1}{2}.$$

Thus, $k \geq \lceil \frac{1}{2\epsilon} \rceil$. □

We now consider the cost function on $\mathcal{P}_k$ given by

$$J(f_k^\alpha, g_k^\beta) := \int_N |g_k^\beta \circ f_k^\alpha(n) - n| d\nu(n) + \Lambda k^p.$$

J represents the sum of the total error cost and the storage cost, where $\Lambda, p > 0$ are fixed constants. By the above lemmas and Theorem 4.1.2, whenever $k \geq \frac{1}{2\epsilon}$, the corresponding Optimal Coding-Pair Problem

$$\min\{J(f_k^\alpha, g_k^\beta) : (f_k^\alpha, g_k^\beta) \in \mathcal{L}(\epsilon, A_k; \mathcal{P}_k)\} \tag{4.9}$$

has a solution. Suppose $d\nu(n) = \theta(n)dn$, this solution can be explicitly computed as follows: for each k,

$$
\begin{aligned}
&\int_N |g_k^\beta \circ f_k^\alpha(n) - n| d\nu(n) \\
= \;& \int_0^1 \left| \frac{[kn+\alpha]}{k} + \frac{\beta}{k} - n \right| \theta(n) dn = \frac{1}{k^2} \int_0^k \big|[y+\alpha] + \beta - y\big| \theta\Big(\frac{y}{k}\Big) dy \\
= \;& \frac{1}{k^2} \sum_{i=0}^{k-1} \int_i^{i+1} \big|[y+\alpha] + \beta - y\big| \theta\Big(\frac{y}{k}\Big) dy = \frac{1}{k^2} \sum_{i=0}^{k-1} \int_0^1 |[y+\alpha] + \beta - y| \theta\Big(\frac{y+i}{k}\Big) dy \\
= \;& \frac{1}{k^2} \sum_{i=0}^{k-1} \left(\int_0^{1-\alpha} |[y+\alpha] + \beta - y| \theta\Big(\frac{y+i}{k}\Big) dy + \int_{1-\alpha}^1 |[y+\alpha] + \beta - y| \theta\Big(\frac{y+i}{k}\Big) dy \right) \\
= \;& \frac{1}{k^2} \sum_{i=0}^{k-1} \left(\int_0^{1-\alpha} |\beta - y| \theta\Big(\frac{y+i}{k}\Big) dy + \int_{1-\alpha}^1 |1 + \beta - y| \theta\Big(\frac{y+i}{k}\Big) dy \right) \\
= \;& \frac{1}{k^2} \sum_{i=0}^{k-1} \left(\int_0^{1-\alpha} |\beta - y| \theta\Big(\frac{y+i}{k}\Big) dy + \int_{-\alpha}^0 |\beta - y| \theta\Big(\frac{y+i+1}{k}\Big) dy \right) \\
= \;& \frac{1}{k^2} \left(\int_0^{1-\alpha} |\beta - y| \sum_{i=0}^{k-1} \theta\Big(\frac{y+i}{k}\Big) dy + \int_{-\alpha}^0 |\beta - y| \Big[\theta\Big(\frac{y+k}{k}\Big) - \theta\Big(\frac{y}{k}\Big) + \sum_{i=0}^{k-1} \theta\Big(\frac{y+i}{k}\Big) \Big] dy \right) \\
= \;& \frac{1}{k^2} \left(\int_{-\alpha}^{1-\alpha} \Big|\beta - y\Big| \sum_{i=0}^{k-1} \theta\Big(\frac{y+i}{k}\Big) dy + \int_{-\alpha}^0 \Big|\beta - y\Big| \left(\theta\Big(\frac{y}{k} + 1\Big) - \theta\Big(\frac{y}{k}\Big) \right) dy \right) \\
= \;& \frac{1}{k^2} \left(\int_0^1 \Big|y - (\alpha+\beta)\Big| \sum_{i=0}^{k-1} \theta\Big(\frac{y+i-\alpha}{k}\Big) dy + \int_{-\alpha}^0 \Big|\beta - y\Big| \left(\theta\Big(\frac{y}{k} + 1\Big) - \theta\Big(\frac{y}{k}\Big) \right) dy \right).
\end{aligned}
$$

In particular, when ν is the Lebesgue measure restricted on N, $\theta(n) = 1$ for all $n \in N$. Thus,

$$\int_N |g_k^\beta \circ f_k^\alpha(n) - n| dn = \frac{1}{k} \int_0^1 \Big|y - (\alpha+\beta)\Big| dy,$$

whose minimum value is $\frac{1}{4k}$, achieved when $\alpha + \beta = \frac{1}{2}$. As a result, in (4.9), the minimum value of

$$J(f_k^\alpha, g_k^\beta) = \frac{1}{k} \int_0^1 \Big|y - (\alpha+\beta)\Big| dy + \Lambda k^p$$

is $\frac{1}{4k} + \Lambda k^p$, achieved when $\alpha + \beta = \frac{1}{2}$. Hence, in this example, the cost of the codepage A_k, in the form of 4.3, is

$$I(A_k) = \frac{1}{4k} + \Lambda k^p. \tag{4.10}$$

Next, we describe the cost on codebook in the form of (4.7). Suppose a measure ρ on $(0,\infty)$ is given. For any function $h:(0,\infty)\to\mathbb{N}$ with $h(\epsilon)\geq\frac{1}{2\epsilon}$, it associates with a codebook function $\mathcal{A}^h:(0,\infty)\to 2^F$, where each $\mathcal{A}^h(\epsilon)=A_{h(\epsilon)}$. Let Ω denotes the set of all such ρ-measurable $\mathcal{A}^h$. The cost on each individual codepage $\mathcal{A}^h(\epsilon)$ is defined as

$$I(A)=\inf_{(f_k^\alpha,g_k^\beta)\in\mathcal{P}_k} J(f_k^\alpha,g_k^\beta).$$

From (4.10),

$$I(\mathcal{A}^h(\epsilon))=\frac{1}{4h(\epsilon)}+\Lambda h(\epsilon)^p.$$

The cost of $\mathcal{A}^h$ is then defined as

$$P(\mathcal{A}^h):=\int_0^\infty I(\mathcal{A}^h(\epsilon))d\rho(\epsilon)=\int_0^\infty\Big(\frac{1}{4h(\epsilon)}+\Lambda h(\epsilon)^p\Big)d\rho(\epsilon). \tag{4.11}$$

And we consider the problem

$$\min_{\mathcal{A}^h\in\Omega} P(\mathcal{A}^h). \tag{4.12}$$

Proposition 4.4.1. *Problem (4.12) has a solution.*

Proof. Note that for any $\epsilon>0$, the set $\Gamma_\epsilon=\{A_k : k\in\mathbb{N}, k\geq\lceil\frac{1}{2\epsilon}\rceil\}$ is not compact, so Corollary 4.3.1 cannot be directly applied. However, it can still be applied indirectly as follows: Define $A_\infty:=[0,1]$. If we define $f_\infty:[0,1]\to A_\infty$ and $g_\infty:A_\infty\to[0,1]$ by

$$f_\infty(n):=n,\forall n\in[0,1], \text{ and } g_\infty(a):=a,\forall a\in A_\infty,$$

(f_∞,g_∞) is an ϵ-linkable pair. This shows that A_∞ is indeed an ϵ-codepage. For any $\epsilon>0$, let $\Gamma'_\epsilon=\Gamma_\epsilon\cup\{A_\infty\}$, $\Gamma'=\bigcup_{\epsilon>0}\Gamma'_\epsilon$, and $\Omega'=\prod_{\epsilon\in[0,\infty)}\Gamma'_\epsilon$. From direct calculation,

$$|f_k^\alpha(n)-f_h^{\alpha'}(n)|=\Big|\frac{[kn+\alpha]}{k}-\frac{[hn+\alpha']}{h}\Big|\leq 6\Big|\frac{1}{k}-\frac{1}{h}\Big|,$$

$$|g_k^\beta\circ f_k^\alpha(n)-g_h^{\beta'}\circ f_h^{\alpha'}(n)|=\Big|\frac{[kn+\alpha]}{k}+\frac{\beta}{k}-\frac{[hn+\alpha']}{h}-\frac{\beta'}{h}\Big|\leq 6\Big|\frac{1}{k}-\frac{1}{h}\Big|,$$

$$|f_k^\alpha(n)-f_\infty(n)|=\Big|\frac{[kn+\alpha]}{k}-n\Big|\leq\frac{2}{k},$$

$$|g_k^\beta\circ f_k^\alpha(n)-g_\infty\circ f_\infty(n)|=\Big|\frac{[kn+\alpha]}{k}+\frac{\beta}{k}-n\Big|\leq\frac{2}{k}.$$

For $n \in [0,1], \alpha \in [0,1]$, and $\beta \in [-1,1]$. Therefore,

$$||f_k^\alpha - f_h^{\alpha'}||_1 \leq 6\Big|\frac{1}{k} - \frac{1}{h}\Big|\rho\Big([0,1]\Big),$$

$$||g_k^\beta \circ f_k^\alpha - g_h^{\beta'} \circ f_h^{\alpha'}||_1 \leq 6\Big|\frac{1}{k} - \frac{1}{h}\Big|\rho\Big([0,1]\Big),$$

$$||f_k^\alpha - f_\infty||_1 \leq \frac{2}{k}\rho\Big([0,1]\Big),$$

$$||g_k^\beta \circ f_k^\alpha - g_\infty \circ f_\infty||_1 \leq \frac{2}{k}\rho\Big([0,1]\Big).$$

Thus,

$$d(A_k, A_{k'}) \leq \begin{cases} 12|\frac{1}{k} - \frac{1}{k'}|\rho\Big([0,1]\Big), k \text{ and } k' \in \mathbb{N} \\ \frac{4}{k}\rho\Big([0,1]\Big), k \in \mathbb{N} \text{ and } k' = \infty. \end{cases}$$

Hence, if we endow Γ with the topology induced by the psedometric $d_{H,1}$, each set $\Gamma'_\epsilon = \{A_k; k \in \mathbb{Z} \cup \{\infty\}, k \geq \frac{1}{2\epsilon}\}$ is compact. Define $I(A_\infty) = \infty$. As I is lower-semicontinuous, and Γ'_ϵ is compact for all $\epsilon > 0$, by Corollary 4.3.1, Problem (4.8) has a solution on Ω', henceforth denoted as $\mathcal{A}_*$. As $\int I \circ \mathcal{A}_*(\epsilon) d\rho(\epsilon) < \infty$, the set of $\epsilon's$ where $I\Big(\mathcal{A}_*(\epsilon)\Big) = \infty$, thus $\mathcal{A}_*(\epsilon) = A_\infty$, must be ρ-null. By re-assigning the values of $\mathcal{A}_*$ on this ρ-null set appropriately, we obtain a codebook $\mathcal{A}'_* \in \Omega$. As $\mathcal{A}'_*$ differs from $\mathcal{A}_*$ on this ρ-null set, $\int I \circ \mathcal{A}'_*(\epsilon) d\rho(\epsilon) = \int I \circ \mathcal{A}_*(\epsilon) d\rho(\epsilon)$. $\mathcal{A}'_*$ is thus a solution for (4.8) on Ω. □

In this case, we can compute one such solution directly, by noticing that any such codebook must minimize the integrand of (4.11) for ρ-a.e. By direct calculation this is obtained when

$$h(\epsilon) = \max\Big\{\text{argmin}\{I(\mathcal{A}^h(\epsilon)) : h(\epsilon) = \lfloor (4\Lambda p)^{-\frac{1}{p+1}} \rfloor, \lceil (4\Lambda p)^{-\frac{1}{p+1}} \rceil\}, \lceil \frac{1}{2\epsilon} \rceil\Big\}.$$

4.5. Estimation of the Hausdorff distance

In section 4.2 we defined the distance between two codepages A, B using $d_{\mathcal{H},L_q}$, the Hausdorff distance between two sets of encoding-decoding pairs on them. In this section, we will explore how to compute or estimate this distance in various specific scenarios. Specifically, suppose we want to encode a set X using two codepages A, B subsets of the metric space (Y, d_Y).

4.5.1. The case X is a singleton. Let $X = \{x\}$ be a singleton, endowed with a probability measure. Let A, B are two subsets of a metric space (Y, d_Y). Suppose P_A is the set of all pairs (ϕ_A, ψ_A) where $\phi_A : X \to A$ and $\psi_A : A \to X$, and P_B is the set of all pairs (ϕ_B, ψ_B) where

$\phi_B : X \to B$ and $\psi_B : B \to X$. We claim that, for any $q \geq 1$,

$$d_{\mathcal{H},L_q}(P_A, P_B) = d_{\mathcal{H},d}(A, B).$$

PROOF. By direct calculation,

$$d_q\Big((\phi_A,\psi_A),(\phi_B,\psi_B)\Big) = d_Y\Big(\phi_A(x),\phi_B(x)\Big) + d_X\Big(\psi_A\circ\phi_A(x),\psi_B\circ\phi_B(x)\Big).$$

Thus, for any pair $(\phi_A, \psi_A) \in P_A$,

$$\begin{aligned} d_q\Big((\phi_A,\psi_A),P_B\Big) &= \inf_{(\phi_B,\psi_B)\in P_B} d_q\Big((\phi_A,\psi_A),(\phi_B,\psi_B)\Big) \\ &= \inf_{(\phi_B,\psi_B)\in P_B} d_Y\Big(\phi_A(x),\phi_B(x)\Big) = d_Y\Big(\phi_A(x),B\Big). \end{aligned}$$

As a consequence,

$$\begin{aligned} d_{\mathcal{H},L_q}(P_A,P_B) &= \max\Big\{\sup_{(\phi_A,\psi_A)\in P_A} d_q\Big((\phi_A,\psi_A),P_B\Big), \sup_{(\phi_B,\psi_B)\in P_B} d_q\Big((\phi_B,\psi_B),P_A\Big)\Big\} \\ &= \max\Big\{\sup_{(\phi_A,\psi_A)\in P_A} d_Y(\phi_A(x),B), \sup_{(\phi_B,\psi_B)\in P_B} d_Y(\phi_B(x),A)\Big\} \\ &= \max\Big\{\sup_{a\in A} d_Y(a,B), \sup_{b\in B} d_Y(b,A)\Big\} = d_{\mathcal{H},d}(A,B). \end{aligned}$$

□

4.5.2. The case X is finite. Let $X = \{x_1, ..., x_n\}$ subset of a metric space Z, endowed with a measure $\mu = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$. Let A, B be two subsets of a metric space (Y, d_Y). Suppose $\mathcal{L}_A$ is the set of all ϵ-linkable pairs (ϕ_A, ψ_A) where $\phi_A : X \to A$ and $\psi_A : A \to Z$, and $\mathcal{L}_B$ is the set of all ϵ-linkable pairs (ϕ_B, ψ_B) where $\phi_B : X \to B$ and $\psi_B : B \to Z$. We claim that for each $q \geq 1$,

$$d_{\mathcal{H},d}(A,B) \leq d_{\mathcal{H},L_q}(\mathcal{L}_A,\mathcal{L}_B) \leq n^{1-1/q}\Big(d_{\mathcal{H},d}(A,B) + 2\epsilon\Big). \tag{4.13}$$

PROOF. Indeed, by direct calculation,

$$d_q\Big((\phi_a,\psi_A),(\phi_B,\psi_B)\Big) = \left(\frac{1}{n}\sum_{i=1}^n d_Y\Big(\phi_A(x_i),\phi_B(x_i)\Big)^q\right)^{1/q} + \left(\frac{1}{n}\sum_{i=1}^n d_X\Big(\psi_A\circ\phi_A(x_i),\psi_B\circ\phi_B(x_i)\Big)^q\right)^{1/q}.$$

Thus, from Holder's inequality,

$$d_q\Big((\phi_A,\psi_A),\mathcal{L}_B\Big)$$

$$
\begin{aligned}
&= \inf_{(\phi_B,\psi_B)\in\mathcal{L}_B}\left[\left(\frac{1}{n}\sum_{i=1}^n d_Y\Big(\phi_A(x_i),\phi_B(x_i)\Big)^q\right)^{1/q} + \left(\frac{1}{n}\sum_{i=1}^n d_X\Big(\psi_A\circ\phi_A(x_i),\psi_B\circ\phi_B(x_i)\Big)^q\right)^{1/q}\right] \\
&\geq \inf_{(\phi_B,\psi_B)\in\mathcal{L}_B}\left(\frac{1}{n}\sum_{i=1}^n d_Y\Big(\phi_A(x_i),\phi_B(x_i)\Big)^q\right)^{1/q} \\
&\geq \frac{1}{n}\inf_{(\phi_B,\psi_B)\in\mathcal{L}_B}\left(\sum_{i=1}^n d_Y\Big(\phi_A(x_i),\phi_B(x_i)\Big)\right) = \frac{1}{n}\sum_{i=1}^n d_Y\Big(\phi_A(x_i),B\Big),
\end{aligned}
$$

where the last equality is obtained by picking $\phi_B(x_i) = \text{argmin}_{b\in B}\, d_Y\Big(\phi_A(x_i),b\Big)$. By a similar argument, we have

$$
d_q\Big((\phi_B,\psi_B),\mathcal{L}_A\Big) \geq \frac{1}{n}\sum_{i=1}^n d_Y\Big(\phi_B(x_i),A\Big).
$$

Consequently,

$$
\begin{aligned}
d_{\mathcal{H},L_q}(\mathcal{L}_A,\mathcal{L}_B) &= \max\Big\{\sup_{(\phi_A,\psi_A)\in\mathcal{L}_A} d_q\Big((\phi_A,\psi_A),P_B\Big), \sup_{(\phi_B,\psi_B)\in\mathcal{L}_B} d_q\Big((\phi_B,\psi_B),P_A\Big)\Big\} \\
&\geq \max\Big\{\frac{1}{n}\sup_{(\phi_A,\psi_A)\in\mathcal{L}_A}\sum_{i=1}^n d_Y\Big(\phi_A(x_i),B\Big), \frac{1}{n}\sup_{(\phi_B,\psi_B)\in\mathcal{L}_B}\sum_{i=1}^n d_Y\Big(\phi_B(x_i),A\Big)\Big\} \\
&= \frac{1}{n}\max\Big\{n\max_{a\in A} d_Y(a,B), n\max_{b\in B} d_Y(b,A)\Big\} = d_{\mathcal{H},d}(A,B),
\end{aligned}
$$

where the second to last equality is obtained by picking

$$
\phi_A(x_i) = \text{argmax}_{a\in A}\, d_Y(a,B) \text{ and } \phi_B(x_i) = \text{argmax}_{b\in B}\, d_Y(b,A),
$$

for any x_i in X.

On the other hand,

$$
\begin{aligned}
&d_q\Big((\phi_A,\psi_A,\mathcal{L}_B\Big) \\
&= \inf_{(\phi_B,\psi_B)\in\mathcal{L}_B}\left[\left(\frac{1}{n}\sum_{i=1}^n d_Y\Big(\phi_A(x_i),\phi_B(x_i)\Big)^q\right)^{1/q} + \left(\frac{1}{n}\sum_{i=1}^n d_X\Big(\psi_A\circ\phi_A(x_i),\psi_B\circ\phi_B(x_i)\Big)^q\right)^{1/q}\right] \\
&\leq n^{-1/q}\inf_{(\phi_B,\psi_B)\in\mathcal{L}_B}\left[\left(\sum_{i=1}^n d_Y\Big(\phi_A(x_i),\phi_B(x_i)\Big) + \sum_{i=1}^n d_X\Big(\psi_A\circ\phi_A(x_i),\psi_B\circ\phi_B(x_i)\Big)\right]\right. \\
&\leq n^{-1/q}\Big(\sum_{i=1}^n d_Y\Big(\phi_A(x_i),B\Big) + 2n\epsilon\Big),
\end{aligned}
$$

where the last inequality is obtained by noticing the existence of a pair (ϕ_B, ψ_B) such that

$$\phi_B(x_i) = \text{argmin}_{b\in B}\Big(\phi_A(x_i), b\Big),$$

and that $d_X\Big(\psi_A \circ \phi_A(x_i), \psi_B \circ \phi_B(x_i)\Big) \leq 2\epsilon$, for any x_i in X. By a similar argument,

$$d_q\Big((\phi_B, \psi_B, \mathcal{L}_A\Big) \leq n^{-1/q}\Big(\sum_{i=1}^{n} d_Y\Big(\phi_B(x_i), A\Big) + 2n\epsilon\Big).$$

Consequently,

$$\begin{aligned}
& d_{\mathcal{H},L_q}(\mathcal{L}_A, \mathcal{L}_B) \\
= \; & \max\Big\{\sup_{(\phi_A,\psi_A)\in\mathcal{L}_A} d_q\Big((\phi_A, \psi_A), \mathcal{L}_B\Big), \sup_{(\phi_B,\psi_B)\in\mathcal{L}_B} d_q\Big((\phi_B, \psi_B), \mathcal{L}_A\Big)\Big\} \\
\leq \; & \max\Big\{n^{-1/q}\Big(\sup_{(\phi_A,\psi_A)\in\mathcal{L}_A} \sum_{i=1}^{n} d_Y\Big(\phi_A(x_i), B\Big) + 2n\epsilon\Big), n^{-1/q}\Big(\sup_{(\phi_B,\psi_B)\in\mathcal{L}_B} \sum_{i=1}^{n} d_Y\Big(\phi_B(x_i), A\Big) + 2n\epsilon\Big)\Big\} \\
= \; & n^{-1/q} \max\Big\{n \max_{a\in A} d_Y(a, B) + 2n\epsilon, n \max_{b\in B} d_Y(b, A) + 2n\epsilon\Big\}m = n^{1-1/q}\Big(d_{\mathcal{H},d}(A, B) + 2\epsilon\Big),
\end{aligned}$$

where the second to last equality is obtained by picking $\phi_A(x_i) = \text{argmax}_{a\in A}\, d_Y(a, B)$, and $\phi_B(x_i) = \text{argmax}_{b\in B}\, d_Y(b, A)$. □

REMARK 4.5.1. *Suppose we want to estimate the Hausdorff distance between two objects A and B that are far away from our position, X. Suppose we can send signals from our position at X to A and B, represented by the sets of functions ϕ_A's and ϕ_B's, respectively. We then record the reflection of these signals at X, represented by the functions ψ_A and ψ_B, respectively. Let $\mathcal{L}_A$ be the set of ϵ-linkable pairs (ϕ_A, ψ_A), and $\mathcal{L}_B$ be the set of ϵ-linkable pairs (ϕ_B, ψ_B). If we can compute the Hausdorff distance $d_{\mathcal{H},L_q}(\mathcal{L}_A, \mathcal{L}_B)$, we can estimate the distance $d_{\mathcal{H}}(A, B)$ by rearranging equation (4.13) as*

$$n^{\frac{1-q}{q}} d_{\mathcal{H},q}(\mathcal{L}_A, \mathcal{L}_B) - 2\epsilon \leq d_{\mathcal{H},d}(A, B) \leq d_{\mathcal{H},q}(\mathcal{L}_A, \mathcal{L}_B).$$

4.5.3. The case $\mathcal{L}_A$ and $\mathcal{L}_B$ are sequentially compact. From Proposition 1.2.6 the Hausdorff distance between two sequentially compact sets equals the distance between some two points in each set. Under the right conditions, as described in Proposition 4.5.1, the sets of pairs $\mathcal{L}_A$ and $\mathcal{L}_B$ are sequentially compact by the topology generated by d_q.

Proposition 4.5.1. *Suppose ν is finite, A is bounded, and $\mathcal{L}_A$ is sequentially compact in the product topology of the pointwise convergence topology on A^X and the compact-open topology on $C(A,Z)$. Then $\mathcal{L}_A$ is also sequentially compact with respect to the topology generated by d_q.*

Proof. For any sequence (ϕ_A, ψ_A) in $\mathcal{L}_A$, the sequential compactness of $\mathcal{L}_A$ in the compact open topology implies (ϕ_A, ψ_A) converges subsequently to some (ϕ_{A_*}, ψ_{A_*}) in $\mathcal{L}_A$. Still denoting this subsequence (ϕ_A, ψ_A), this means $\phi_{A_k} \to \phi_{A_*}$, and $\psi_{A_k} \circ \phi_{A_k} \to \psi_{A_*} \circ \phi_{A_*}$ pointwisely. As A is bounded, there exists an $M \in \mathbb{R}^+$ such that $d_Y\Big(\phi_{A_k}(x), \phi_{A_*}(x)\Big) \leq M$ for all $x \in X$, $k \in \mathbb{N}$. As $\mathcal{L}_A$ is ϵ-linkable, $d_X\Big(\psi_{A_k} \circ \phi_{A_k}(x), \psi_{A_*} \circ \phi_{A_*}(x)\Big) \leq 2\epsilon$. As ν is finite, we can then apply the dominated converge Theorem [3, Theorem 1.3.3] to get

$$\begin{aligned} & d_q\Big((\phi_{A_k}, \psi_{A_k}), (\phi_{A_*}, \psi_{A_*})\Big) \\ = \ & \left(\int_X d_Y\Big(\phi_{A_k}(x), \phi_{A_*}(x)\Big)^q d\nu(x)\right)^{1/q} + \left(\int_X d_X\Big(\psi_{A_k} \circ \phi_{A_k}(x), \psi_{A_*} \circ \phi_{A_*}(x)\Big)^q d\nu(x)\right)^{1/q} \to 0. \end{aligned}$$

Hence $\mathcal{L}_A$ is sequentially compact in the topology generated by d_q. ☐

Proposition 4.5.1, together with Proposition 1.2.6, allow us to make the following conclusion.

Corollary 4.5.1. *Suppose ν is finite, A and B are bounded; $\mathcal{L}_A$ is sequentially compact in the product topology of the pointwise convergence topology on A^X and the compact-open topology on $C(A,Z)$, and $\mathcal{L}_B$ is sequentially compact in the product topology of the pointwise convergence topology on B^X and the compact-open topology on $C(B,Z)$. Then there exists a pair $(\phi_A, \psi_A) \in \mathcal{L}_A$ and a pair $(\phi_B, \psi_B) \in \mathcal{L}_B$ such that $d_{\mathcal{H},L_q}(\mathcal{L}_A, \mathcal{L}_B) = d_q\Big((\phi_A, \psi_A), (\phi_B, \psi_B)\Big)$.*

CHAPTER 5

Future Works

In this chapter, we explore several possible future directions to expand the work discussed in this thesis.

5.1. First-type cost function: Additional questions about existence

Recall in chapter 3 on the set Ω of (μ, ρ)-codebook (Definition 3.1.1), we define the cost function

$$C(\mathcal{A}) := \int_0^\infty \mu\Big(\mathcal{A}(\epsilon)\Big) d\rho(\epsilon).$$

And we want to consider the minimization problem (3.1),

$$\min_{\mathcal{A}\in\Omega} C(\mathcal{A}).$$

In section (3.2) we solved Problem (3.1) for the case μ, ρ are Borel measures, and N is totally bounded. In section (3.4) we briefly discussed the case N not totally bounded (still assuming that the measure μ on F is Borel). This leads to the following problem:

$$\min\{\mu(A) : A \subset F \text{ is } \mu\text{-measurable with } |A| \geq K\}. \tag{5.1}$$

when K is infinite. In Proposition (3.4.1) we showed that under this circumstance any solution to Problem (5.1) must have zero measure. We have not however solved this problem. Thus we may ask the following question.

Question 5.1.1. *Under what conditions does Problem (5.1) have a solution?*

Proposition 3.4.2 also shows that in the case N not totally bounded, and $\rho\Big((0, \epsilon)\Big) > 0$ for all $\epsilon > 0$, any $\mathcal{A}^*$ that solves (3.1) then must satisfy $C(\mathcal{A}^*) = 0$. Yet again we have not solved this problem. Thus, we can ask:

Question 5.1.2. *For the case N is not totally bounded, under what conditions does Problem (3.1) have a solution?*

5.2. Second-type cost function: Finding the optimal codebook

In Chapter 4 we discussed the cost function

$$J(f,g) := \int_N C\Big(n, f)n), g\circ f(n)\Big) d\nu(n) + S(f)$$

on a set $\mathcal{L}(\epsilon, A; \mathcal{P})$ of ϵ-linkable pairs. We then consider the the problem of minimizing $J(f,g)$ among all pairs (f,g) in $\mathcal{L}(\epsilon, A; \mathcal{P})$ (Problem 4.1). Theorem 4.1.2 proves this problem has a solution in the case $\mathcal{P}$ is sequentially compact. However, we have not yet understood what the solution looks like, or what the minimum value is. Thus we may ask the following questions.

Question 5.2.1. *Can the solutions to Problem (4.1) be computed or estimated?*

Question 5.2.2. *Suppose $\mathcal{P}$ is sequentially compact. Can we compute or estimate the value of* $\min\{J(f,g) : (f,g) \in \mathcal{L}(\epsilon, A; \mathcal{P})\}$*?*

We then define the cost function on a codepage A as

$$I(A) := \inf_{(f,g)\in\mathcal{L}(\epsilon,A;\mathcal{P})} J(f,g),$$

and the cost function on a codebook $\mathcal{A}$ as

$$\mathcal{S}(\mathcal{A}) := \int_0^\infty I\Big(\mathcal{A}(\epsilon)\Big) d\rho(\epsilon).$$

In section (4.3) we prove the existence of a solution to Problem (4.8), $\min\{\mathcal{S}(\mathcal{A}) : \mathcal{A} \in \Omega\}$, for a compact or sequentially compact Ω. Yet again the question of how to compute the actual optimal codebook or the minimal value of $\mathcal{S}$ is left unanswered. Therefore, we may ask the following questions.

Question 5.2.3. *Can we compute, or estimate, the codebook $\mathcal{A}$ that solves Problem (4.8)?*

Question 5.2.4. *Suppose Ω is compact or sequentially compact, can we compute or estimate the value of* $\min\{\mathcal{S}(\mathcal{A}) : \mathcal{A} \in \Omega\}$?

5.3. Cost function of other types

We continue using the notations introduced in Definition 2.1.1. Let $(\mathcal{C}(2^F), d)$ be a metric subspace of 2^F. A codebook $\mathcal{A} : (0,\infty) \to \mathcal{C}(2^F)$ is called Lipschitz continuous if there exists a $K_{\mathcal{A}} > 0$ such that $d\Big(\mathcal{A}(\epsilon_1), \mathcal{A}(\epsilon_2)\Big) \leq K_{\mathcal{A}}|\epsilon_1 - \epsilon_2|$, for any ϵ_1 and ϵ_2 in $(0, \infty)$. Let Ω_{Lip} be the set of Lipschitz continuous codebooks $\mathcal{A} : (0, \infty) \to \mathcal{C}(2^F)$.

QUESTION 5.3.1. *Under what conditions is a codebook $\mathcal{A} : (0, \infty) \to \mathcal{C}(2^F)$ a Lipschitz continuous?*

Given $0 < \alpha < \beta < \infty$, for any partition $P = \{\epsilon_0, \epsilon_1, ..., \epsilon_k\}$ of $[\alpha, \beta]$ and $\mathcal{A} \in \Omega_{Lip}$, we define

$$L_\alpha^\beta(P, \mathcal{A}) := \sum_{i=1}^{k} d\Big(\mathcal{A}(\epsilon_i), \mathcal{A}(\epsilon_{i-1})\Big).$$

And define the length of codebook $\mathcal{A}$ as

$$L_\alpha^\beta(\mathcal{A}) := \sup_{\text{All partitions P of [0,1]}} L_\alpha^\beta(P, \mathcal{A}).$$

Since $\mathcal{A}$ is Lipschitz continuous, $L_\alpha^\beta(P, \mathcal{A}) \leq \sum_{i=1}^{k} K_{\mathcal{A}}|\epsilon_i - \epsilon_{i-1}| = K_{\mathcal{A}}$, therefore $L_\alpha^\beta(\mathcal{A})$ is finite. We consider the minimization problem:

$$\min_{\mathcal{A} \in \Omega} L_\alpha^\beta(\mathcal{A}) \tag{5.2}$$

for some $\Omega \subset \Omega_{Lip}$.

QUESTION 5.3.2. *Under what conditions does Problem (5.2) have a solution?*

At any point $\epsilon > 0$, the Metric Derivative of $\mathcal{A}$ at ϵ, henceforth denoted as $|\dot{\mathcal{A}}|(\epsilon)$, is defined as

$$|\dot{\mathcal{A}}|(\epsilon) := \lim_{h \to 0} \frac{d\Big(\mathcal{A}(\epsilon + h), \mathcal{A}(\epsilon)\Big)}{|h|}.$$

If a metric derivative exists at some $\epsilon > 0$ we say that $\mathcal{A}$ is differentiable at ϵ.

For any Lipschitz $\mathcal{A} : (0, \infty) \to C(2^F)$, $\mathcal{A}$ is differentiable $\mathcal{L}$-a.e. . From Theorem 1.2.12, for any interval $[\alpha, \beta]$,

$$L_\alpha^\beta(\mathcal{A}) = \int_\alpha^\beta |\dot{\mathcal{A}}|(\epsilon) d\epsilon.$$

And thus if we consider the set $\Omega_{Lip} \subset \Omega$ consisting of Lipschitz curves $\mathcal{A}$, Problem (5.2) can be rephrased as

$$\min_{\mathcal{A} \in \Omega_{Lip}} \int_\alpha^\beta |\dot{\mathcal{A}}|(\epsilon) d\epsilon. \tag{5.3}$$

QUESTION 5.3.3. *Under what conditions does Problem (5.3) have a solution?*

APPENDIX A

Tables and figures

TABLE A.1. Maximum value of d_{Ave} and W_1 for each value of m. The values $d_{Ave}(n, g \circ f(n))$ and $W_1(n, g \circ f(n))$ are computed across all 60,000 elements of MNIST and are rounded up to 2 decimals. Here $H = H(m) = [\frac{28}{m}] \times [\frac{28}{m}]$.

m	H	$\max d_{Ave}(n, g \circ f(n))$	$\max W_1(n, g \circ f(n))$
1	784	0	0
2	196	24.95	16.37
3	81	37.61	26.69
4	49	49.68	40.77
5	25	53.47	43.34
6	16	58.78	46.87
7	16	82.34	67.05
8	9	74.86	57.27
9	9	86.92	73.75
10	4	85.65	54.71
11	4	87.57	67.29
12	4	86.01	71.85
13	4	98.20	89.37
14	4	113.56	108.00
15	1	104.47	71.68
16	1	103.27	65.40
17	1	101.72	58.43
18	1	99.63	59.20
19	1	97.12	62.42
20	1	94.68	66.82
21	1	92.68	72.31
22	1	90.62	78.36
23	1	89.69	84.36
24	1	91.56	90.01
25	1	95.23	95.23
26	1	101.87	101.87
27	1	108.71	108.71
28	1	114.86	114.86

TABLE A.2. Some possible ϵ-codepages corresponding to each range of ϵ. The value of ϵ's shown are calculated based on d_{Ave}, and are rounded up to 2 decimals. m represents the pair (f_m, g_m), and $H = H(m) = [\frac{28}{m}] \times [\frac{28}{m}]$ represents corresponding $A_H = [0, 1, ..., 255]^H$.

ϵ	Possible m's	Corresponding H
$[114.86, \infty)$	1-28	1, 4, 9, 16, 25, 49, 81, 196, 784
$[113.56, 114.86)$	1-27	1, 4, 9, 16, 25, 49, 81, 196, 784
$[108.71, 113.56)$	1-13,15-27	1, 4, 9, 16, 25, 49, 81, 196, 784
$[104.47, 108.71)$	1-13, 15-26	1, 4, 9, 16, 25, 49, 81, 196, 784
$[103.27, 104.47)$	1-13,16-26	1, 4, 9, 16, 25, 49, 81, 196, 784
$[101.87, 103.27)$	1-13,17-26	1, 4, 9, 16, 25, 49, 81, 196, 784
$[101.72, 101.87)$	1-13, 17-25	1, 4, 9, 16, 25, 49, 81, 196, 784
$[99.63, 101.72)$	1-13,18-25	1, 4, 9, 16, 25, 49, 81, 196, 784
$[98.21, 99.63)$	1-13, 19-25	1, 4, 9, 16, 25, 49, 81, 196, 784
$[97.12, 98.21)$	1-12, 20-25	1, 4, 9, 16, 25, 49, 81, 196, 784
$[95.23, 97.12)$	1-12, 20-25	1, 4, 9, 16, 25, 49, 81, 196, 784
$[94.68, 95.23)$	1-12,20-24	1, 4, 9, 16, 25, 49, 81, 196, 784
$[92.68, 94.68)$	1-12, 21-24	1, 4, 9, 16, 25, 49, 81, 196, 784
$[91.56, 92.68)$	1-12, 22-24	1, 4, 9, 16, 25, 49, 81, 196, 784
$[90.62, 91.56)$	1-12, 23	1, 4, 9, 16, 25, 49, 81, 196, 784
$[89.69, 90.62)$	1-12, 23	1, 4, 9, 16, 25, 49, 81, 196, 784
$[87.57, 89.69)$	1-12	4, 9, 16, 25, 49, 81, 196, 784
$[86.92, 87.57)$	1-10, 12	4, 9, 16, 25, 49, 81, 196, 784
$[86.01, 86.92)$	1-10, 12	4, 9, 16, 25, 49, 81, 196, 784
$[85.65, 86.01)$	1-8, 10	4, 9, 16, 25, 49, 81, 196, 784
$[82.35, 85.65)$	1-8	9, 16, 25, 49, 81, 196, 784
$[74.86, 82.35)$	1-6, 8	9, 16, 25, 49, 81, 196, 784
$[58.78, 74.86)$	1-6	16, 25, 49, 81, 196, 784
$[53.47, 58.78)$	1-5	25, 49, 81, 196, 784
$[49.68, 53.47)$	1-4	49, 81, 196, 784
$[37.61, 49.68)$	1, 2, 3	81, 196, 784
$[24.95, 37.61)$	1,2	196, 784
$[0, 24.95)$	1	784

TABLE A.3. Some possible ϵ-codepages corresponding to each range of ϵ. The value of ϵ's shown are calculated based on W_1, and are rounded up to 2 decimals. m represents the pair (f_m, g_m), and $H = H(m) = [\frac{28}{m}] \times [\frac{28}{m}]$ represents corresponding $A_H = [0, 1, ..., 255]^H$.

ϵ	Possible m's	Corresponding H
$[114.86, \infty)$	1-28	1, 4, 9, 16, 25, 49, 81, 196, 784
$[108.71, 114.86)$	1-27	1, 4, 9, 16, 25, 49, 81, 196, 784
$[108.00, 108.71)$	1-26	1, 4, 9, 16, 25, 49, 81, 196, 784
$[101.87, 108.00)$	1-13, 15-26	1, 4, 9, 16, 25, 49, 81, 196, 784
$[95.23, 101.87)$	1-13,15-25	1, 4, 9, 16, 25, 49, 81, 196, 784
$[90.01, 95.23)$	1-13,15-24	1, 4, 9, 16, 25, 49, 81, 196, 784
$[89.37, 90.01)$	1-13, 15-23	1, 4, 9, 16, 25, 49, 81, 196, 784
$[84.36, 89.37)$	1-12, 15-23	1, 4, 9, 16, 25, 49, 81, 196, 784
$[78.36, 84.36)$	1-12, 15-22	1, 4, 9, 16, 25, 49, 81, 196, 784
$[73.75, 78.36)$	1-12, 15-21	1, 4, 9, 16, 25, 49, 81, 196, 784
$[72.31, 73.75)$	1-8, 10-12, 15-21	1, 4, 9, 16, 25, 49, 81, 196, 784
$[71.85, 72.31)$	1-8, 10-12, 15-20	1, 4, 9, 16, 25, 49, 81, 196, 784
$[71.68, 71.85)$	1-8, 10-11, 15-20	1, 4, 9, 16, 25, 49, 81, 196, 784
$[67.29, 71.68)$	1-8. 10-11, 16-20	1, 4, 9, 16, 25, 49, 81, 196, 784
$[67.05, 67.29)$	1-8, 10, 16-20	1, 4, 9, 16, 25, 49, 81, 196, 784
$[66.82, 67.05)$	1-6, 8, 10, 16-20	1, 4, 9, 16, 25, 49, 81, 196, 784
$[65.40, 66.82)$	1-6, 8, 10, 16-19	1, 4, 9, 16, 25, 49, 81, 196, 784
$[62.42, 65.40)$	1-6, 8, 10, 17-19	1, 4, 9, 16, 25, 49, 81, 196, 784
$[59.20, 62.42)$	1-6, 8, 10, 17-18	1, 4, 9, 16, 25, 49, 81, 196, 784
$[58.43, 59.20)$	1-6, 8, 10, 17	1, 4, 9, 16, 25, 49, 81, 196, 784
$[57.27, 58.43)$	1-6, 10	4, 9, 16, 25, 49, 81, 196, 784
$[54.71, 57.27)$	1-6, 10	4, 16, 25, 49, 81, 196, 784
$[46.87, 54.71)$	1-6	16, 25, 49, 81, 196, 784
$[43.34, 46.87)$	1-5	25, 49, 81, 196, 784
$[40.76, 43.34)$	1-4	49, 81, 196, 784
$[26.69, 40.76)$	1-3	81, 196, 784
$[16.37, 26.69)$	1,2	196, 784
$[0, 16.37)$	1	784

Bibliography

[1] L. Ambrosio and T. Paolo, *Topics on Analysis in Metric Spaces*, Oxford lecture series in mathematics and its applications, Oxford University Press, 2004.

[2] V. I. Bogachev, *Measure Theory*, vol. 1 of Graduate Texts in Mathematics, Springer, Berlin, 2007.

[3] L. C. Evans, *Measure Theory and Fine Properties of Functions*, Studies in Advanced Mathematics, American Mathematical Society, 1992.

[4] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge, MA, USA, 2016. http://www.deeplearningbook.org.

[5] J. T. Henrikson, *Completeness and total boundedness of the hausdorff metric*, 1999.

[6] J. L. Kelly, *General Topology*, Graduate Texts in Mathematics, Springer New York, NY, 1975.

[7] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Copyright Cambridge University Press, 2003.

[8] Y. Moschovakis, *Notes on Set Theory*, SpringerBriefs in Mathematics, Springer International Publishing, 2016.

[9] R. Remmert, *Theory of Complex Functions*, vol. 122 of Graduate Texts in Mathematics, Springer-Verlag, 1991.

[10] F. Santambrogio, *Optimal transport for applied mathematicians. calculus of variations, pdes and modeling*, (2015).

[11] T. Tao, *Metric entropy analogues of sum set theory*, 2016. Accessed: 2024-11-19.

[12] C. Villani, *Topics in Optimal Transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, 2003.