Embedding Distances into the Hamming Cube

Rafael D'Oliveira, Marcelo Firer

Embedding Distances into the Hamming Cube

Rafael Gregorio Lucas D'Oliveira and Marcelo Firer

Imecc - Unicamp
rgldoliveira@gmail.com, mfirer@gmail.com

Abstract. In Coding Theory, two different decision criteria types are mostly used: a maximal likelihood (relative to a probabilistic model of a channel) and a nearest neighbour criterion (relative to a distance model of the channel). In this work we present an algorithm that, given a maximal likelihood criterion, decides if there is a nearest neighbour criterion that matches it and, in this case, produces a metric that determines such a criterion. It is also shown that the Hamming metric is universal, in the sense that any metric, up to a decoding equivalence, can be isometrically embedded into a hypercube with the Hamming metric.

1 Introduction

The Binary Symmetric Channel (BSC) is the simplest model of a discrete channel, and is used for many reasons, including the fact it represents the worst case scenario, when there is no available information about the channel. In this context, the Hamming metric has an outstanding and prominent status, since it is matched to the BSC, in the sense that, for any code C and for any received message x, we have that $c \in C$ is the most likely codeword to be sent if, and only if, it is the codeword closest to x. Many different channels are considered in the context of Information Theory, and also many different metrics are considered in the context of Coding Theory. A recent survey can be found in [3] and in [1][Chapter 16]. However, the general question about matching a channel and a metric has been little explored in the literature. This general setting, of matching a metric to a channel (and vice-versa) was first asked by Massey, in a class note from 1967 [4]. Considering only the family of metrics that are defined over the alphabet and extended additively over the coordinates (which includes the Hamming and the Lee metrics), in 1980 Gérald Séguin [5] gave necessary and sufficient conditions to a channel to be matched to such a metric. Since then, the problem of matching a channel and a metric stayed abandoned, until recently, when Walker and Firer proved that a metric can be matched to the Z-Channel (and some other general questions on the subject)[2].

In this work we look at this problem as a problem of finding a solution to a set of linear inequalities and produce an algorithm (Section 3) that either produces a metric matched to the channel or shows the non-existence of such a metric. If the possibility of considering metrics holds for a vast family of channels (other then the BSC) and metrics (other then the Hamming), the universality of the Hamming metric is fully assessed in Section 4, where we show that any metric

on a finite set, up to a decoding equivalence, may be isometrically embedded into a hypercube with the Hamming metric (d_H), which, from now on, we refer to as the *Hamming cube*.

2 Distances, Channels and Decoders

Even though we speak about metrics in the introduction, we will consider more general distances.

Let X be a set. A *distance* is a function $d : X \times X \mapsto \mathbb{R}$ such that

1. $d(x, y) \geq 0$
2. $d(x, y) = d(y, x)$
3. $d(x, x) = 0$

If the distance also satisfies the *triangle inequality* ($d(x, z) \leq d(x, y) + d(y, z)$) and the *identity of indiscernibles* ($d(x, y) = 0$ iff $x = y$) it is a metric.

If the set X is an additive group and $d(x + z, y + z) = d(x, y)$ we say that d is *translation invariant*. In this case the distance is totally determined by a weight function $\omega : X \mapsto \mathbb{R}$, defined by $\omega(x) := d(x, 0)$, so that $d(x, y) = \omega(x - y)$.

We are concerned only with finite sets, thus nothing is lost if we assume X to be the set $[n] = \{1, 2, 3, \ldots, n\}$. We also identify the distance with its matrix representation $d_{n\times n}$ such that $d_{ij} = d(i, j)$.

We define a channel using the notation of probability theory.

Definition 1. *A channel over $[n]$ is an $n \times n$ probability matrix P such that*

$$P_{ij} = P(j \text{ received } | i \text{ sent}).$$

Thus, it is a matrix $P_{n\times n}$ with entries in $[0, 1]$ such that

$$\sum_j P_{ij} = 1.$$

As usual, in the context of information theory, the interpretation is that a transmitter sends a symbol $j \in [n]$ to a receiver and the channel determines the probability he receives $i \in [n]$. We remark that the probabilities are defined over the set of elements of X, that may be considered as a set of all possible messages and not an alphabet out of which messages are written by considering blocks of letters.

Given a code $C \subseteq [n]$ the maximum likelihood decoder (MLD) decodes j by searching which $c \in C$ maximizes $P(j \text{ recieved } | c \text{ sent})$, that is, j is decoded as an element of the set $\arg\max\{P(j|c) : c \in C\}$. This is equivalent to looking for the largest number in the jth column restricted to the rows corresponding to codewords. This means that the MLD only cares about the ordering of the elements in the columns, and not their values.

If there is a distance function d defined on $[n]$, then the minimum distance decoder (MDD) decodes j by searching which $c \in C$ minimizes $d(j, c)$.

Given a distance d and a channel over $[n]$ we say that they are *matched to each other* if

$$\arg\min\{d(j,c); c \in C\} = \arg\max\{P(j|c); c \in C\}$$

for every $C \subseteq [n]$, and $j \in [n]$. It is well known that the Hamming metric is matched to the binary symmetric channel. What about for general channels? As was said in the introduction, very little is known about this problem. We do not face this problem directly, but in the next section we will show an algorithm that enables us to match a distance for any particular channel or show that no such distance exists.

3 Finding a Distance Matched to a Channel

There are many different notions of equivalence between distances which are used all throughout mathematics. The following one has not been investigated yet (to the authors' knowledge), despite it being quite natural in the context of coding theory.

Definition 2. *Let X be a set and d_1, d_2 two distance functions on X. We say that $d_1 \sim d_2$ if they define the same minimum distance decoder, in other words, if $\arg\min\{d_1(j,c); c \in C\} = \arg\min\{d_2(j,c); c \in C\}$ for any code $C \subseteq X$ and any $i \in X$.*

It is straightforward to prove that this amounts to the distances having the same set of balls:

Proposition 1. *$d_1 \sim d_2$ iff for every $x_0 \in X$ and $r_1 \in \mathbb{R}$ there exists an $r_2 = r_2(x_0, r_1) \in \mathbb{R}$ such that*

$$B_{d_1}(x_0, r_1) = B_{d_2}(x_0, r_2),$$

where

$$B_d(x, r) = \{y \in X : d(y, x) \leq r\}.$$

In the coding theory literature most, if not all, distances used are metrics. However, there are many known basic constructions which can transform a distance satisfying the identity of indiscernibles into an equivalent metric, and thus matching a channel to a distance or to a metric is essentially the same problem. The next example is a particular case of the "squeezing" argument presented in [2].

Example 1. *Let d_1 be a distance satisfying the identity of indiscernibles. Then*

$$d_2(x, y) = 1 + \frac{d_1(x, y)}{\max_{u,v} d_1(u, v)}$$

for $x \neq y$ and zero otherwise is a metric which is equivalent to d_1.

Analogously to the distance case we have the following equivalence relation between channels.

Definition 3. *Let M and N be two channels. We say that $M \sim N$ if for any code $C \subset X$, they define the same maximum likelihood decoder.*

As previously discussed, after Definition 1, when decoding by the MLD only the order of the elements in a column are important. We therefore define a function which orders a matrix's entries in each column. Because of space constraints we give a simple definition which is made clear by an example. A more rigorous definition would require us to define other concepts, in this case related to partially ordered sets.

Definition 4. *Given a matrix M its decreasing column ordered matrix is the matrix O^-M such that $(O^-M)_{ij} = k$ iff it is the kth largest element of the jth column.*

The definition is analogous for the increasing column matrix O^+M.

An example will make the idea clear.

Example 2. *If*

$$M = \begin{pmatrix} 9\ 2\ 1 \\ 9\ 7\ 0 \\ 8\ 6\ 8 \end{pmatrix}$$

then

$$O^-M = \begin{pmatrix} 1\ 3\ 2 \\ 1\ 1\ 3 \\ 2\ 2\ 1 \end{pmatrix} \quad \text{and} \quad O^+M = \begin{pmatrix} 2\ 1\ 2 \\ 2\ 3\ 1 \\ 1\ 2\ 3 \end{pmatrix}$$

Note that $O^-M = O^-N$ iff $O^+M = O^+N$.

Directly from the definition we have the following.

Proposition 2. *$M \sim N$ iff $O^-M = O^-N$ or equivalently $O^+M = O^+N$.*

Proposition 3. *Given two distances d_1 and d_2, $d_1 \sim d_2$ iff $O^+d_1 = O^+d_2$ or equivalently $O^-d_1 = O^-d_2$. As said earlier we are identifying the distance with its matrix representation.*

Thus, determining if for some channel M there is a distance d matched to it is equivalent to determining if there exists a distance d such that $O^-M = O^+d$, in other words, we have the following question:

Question 1. *Given a matrix $M_{n\times n}$ does there exist a distance d on $[n]$ such that $O^-M = O^+d$?*

This is equivalent to solving the following inequalities on $d(i,j)$:

$$d(i,j) = d(k,j) \qquad \text{if } O^-M_{ij} = O^-M_{kj}$$

$$d(i,j) < d(k,j) \qquad \text{if } O^-M_{ij} < O^-M_{kj}$$

and, the symmetry conditions $d(i,j) = d(j,i)$.

Note that since $d(i,i) = 0$, this implies that $O^+ d_{ii} = 1$, thus we must have $O^- M_{ii} = 1$. Note also that the inequalities come from fixing the column.

We proceed as follows:

We have n chains of inequalities, each corresponding to a certain column. The smallest element of each chain must be $d(i,i)$ which we set to equal zero. Then:

1. We take the first chain and set arbitrary values for the distances with the condition that the inequalities hold true;
2. We then set the same values to their corresponding symmetric term, i.e. $d(i,j) = d(j,i)$;
3. We continue to do this until we have assigned a value to every distance and have therefore found a distance matched to our channel;
   or,

3'. Find that some distance cannot have a valued assigned to it, and thus there is no distance matched for this channel.

Example 3. *Let*

$$M = \begin{pmatrix} \frac{5}{8} & \frac{1}{8} & \frac{2}{8} \\ \frac{2}{8} & \frac{5}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{2}{8} & \frac{5}{8} \end{pmatrix}.$$

Then,

$$O^- M = \begin{pmatrix} 1 & 3 & 2 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

We have the following three chains of inequalities:

$$0 = d(1,1) < d(1,2) < d(1,3)$$

$$0 = d(2,2) < d(2,3) < d(2,1)$$

$$0 = d(3,3) < d(3,1) < d(3,2)$$

We set arbitrary values to the first column and the same ones to their symmetric counterparts.

$$0 = d(1,1) < 1 = d(1,2) < 2 = d(1,3)$$

$$0 = d(2,2) < d(2,3) < 1 = d(2,1)$$

$$0 = d(3,3) < 2 = d(3,1) < d(3,2)$$

In the next step we must set an arbitrary value to $d(2,3)$ but it is impossible to do this since it must satisfy

$$2 = d(3,1) < d(2,3) < 1 = d(2,1).$$

Therefore, no distance exists for M since it would need to satisfy

$$d(2,1) < d(3,1) < d(2,3) < d(2,1).$$

Example 4. *Let*

$$M = \begin{pmatrix} \frac{5}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{2}{8} & \frac{5}{8} & \frac{2}{8} \\ \frac{1}{8} & \frac{2}{8} & \frac{5}{8} \end{pmatrix}.$$

Then,

$$O^- M = \begin{pmatrix} 1 & 3 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

We have the following three chains of inequalities:

$$0 = d(1,1) < d(1,2) < d(1,3)$$
$$0 = d(2,2) < d(2,3) < d(2,1)$$
$$0 = d(3,3) < d(3,2) < d(3,1)$$

We set arbitrary values to the first chain and to their symmetric counterparts.

$$0 = d(1,1) < 1 = d(1,2) < 2 = d(1,3)$$
$$0 = d(2,2) < d(2,3) < 1 = d(2,1)$$
$$0 = d(3,3) < d(3,2) < d(3,1) = 2$$

We do the same for the second chain.

$$0 = d(1,1) < 1 = d(1,2) < 2 = d(1,3)$$
$$0 = d(2,2) < d(2,3) = \frac{1}{2} < 1 = d(2,1)$$
$$0 = d(3,3) < d(3,2) = \frac{1}{2} < d(3,1) = 2$$

We were able to set values to all the distances. Therefore, M has the following distance:

$$d = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & \frac{1}{2} \\ 2 & \frac{1}{2} & 0 \end{pmatrix}.$$

4 Embedding Distances into the Hamming Cube

In this section we prove that any distance satisfying the identity of indiscernibles is equivalent to a distance which is embeddable into the Hamming cube. If, in addition to this, the distance is translation invariant over $\mathbb{F}_2^n$, the embedding, f, will behave well with respect to addition, i.e. $f(x+y) = f(x) + f(y)$.

We first note that there is a natural bijection between the n-dimensional Hamming cube H^n, and the subsets of $[n]$, $2^{[n]}$ given by

$$supp : H^n \leftrightarrow 2^{[n]}$$

where $supp(x) = \{i : x_i \neq 0\}$.

This function satisfies the following properties:

1. $supp(x+y) = supp(x) \triangle supp(y)$
2. $\omega_H(x) = |supp(x)|$

where $A \triangle B = \{x \in A - B\} \cup \{x \in B - A\}$ is the symmetric difference between the sets A and B.

This is important because we will pose the problem as a problem on sets.

We first prove the translation invariant case over $\mathbb{F}_2^n$. The general case will follow as a consequence.

Theorem 1. *Let d_1 be a translation invariant distance, with weight ω_1, satisfying the identity of indiscernibles over $\mathbb{F}_2^n$. Then there exists a translation invariant distance d_2, with weight ω_2, over $\mathbb{F}_2^n$ that is equivalent to d_1 and is isometrically embeddable, preserving addition, into the Hamming cube.*

Proof. Denote by $\{e_1, e_2, \cdots e_n\}$ the canonical basis of $\mathbb{F}_2^n$.

Let

$$\delta_I = \omega_1(\sum_{i \in I} e_i)$$

for every $I \subseteq [n]$ such that $I \neq \emptyset$.

Without loss of generality we can assume that $\delta_I \in \mathbb{Q}_+$ since only the order relation between the values matters.

Note that for any given $m, k \in \mathbb{N}$ if

$$w_2(\sum_{i \in I} e_i) = m\delta_I + k$$

for every $I \subseteq [n]$ such that $I \neq \emptyset$, then $d_1 \sim d_2$.

In Theorem 3 we prove that there exists finite sets A_i and constants $m, k \in \mathbb{Z}_+$ such that for every I, the cardinality of their symmetric difference is

$$|\triangle_{i \in I} A_i| = m\delta_I + k.$$

Let $N = |\cup_i A_i|$ and $f : \mathbb{F}_2^n \to 2^N$ such that $f(e_i) = A_i$. Then it is straightforward to show that $supp^{-1} \circ f$ is a weight preserving embedding from $(\mathbb{F}_2^n, d_2)$ into the Hamming cube which is equivalent to d_1 and also that the transformation from d_1 to the Hamming cube behaves well with respect to addition. The identity of indiscernibles is needed to prove that f is injective. □

As a consequence of Theorem 1 we can now prove the result for any distance satisfying the identity of indiscernibles.

Theorem 2. *Let d_1 be a distance over X satisfying the identity of indiscernibles. Then there exists a distance d_2 such that $d_1 \sim d_2$ and d_2 is embeddable into the Hamming cube.*

Proof. If we let $X = \{x_1, x_2, \ldots, x_n\}$ and

$$\delta_I = \begin{cases} d(x_{i_1}, x_{i_2}) \text{ if } & I = \{i_1, i_2\} \quad (i_1 \neq i_2) \\ 1 & \text{otherwise} \end{cases}$$

for every $I \subset [n]$ such that $I \neq \emptyset$, we can use the same arguments as in Theorem 1. Essentially, we define an additive group structure on d_1 and end up in the same situation as Theorem 1. □

If we remove the condition of satisfying the identity of indiscernibles, as said before, we lose injectivity in our embedding. There are many ways to deal with this, but we will not explore them here.

To prove Theorems 1 and 2 we need to solve the following problem:

Given $\delta_I \in \mathbb{Q}_+$ where $I \subset [n]$, does there exist finite sets $A_1, \ldots, A_n$ and constants $m, k \in \mathbb{Z}_+$ such that

$$|\triangle_{i \in I_j} A_i| = m\delta_{I_j} + k.$$

We call these equations a system of symmetric differences.

We solve this system by trying to solve

$$|\triangle_{i \in I_j} A_i| = \delta_{I_j}$$

and working out the problems that prevent us from finding a solution.

We use the following relation, if $A = \{A_1, \ldots, A_n\}$ then

$$|\triangle A| = \sum_{l=1}^{n} (-2)^{l-1} \sum_{i_1 \neq i_2 \neq \ldots \neq i_l} |A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_l}|. \tag{1}$$

This gives a recursive formula to find all intersections in terms of the symmetric differences. With all the intersections it is possible to build the full Venn Diagram of $A_1, \ldots, A_n$.

To prove that the algorithm works, at each step, the constant m is used to ensure that $m\delta_I \in \mathbb{N}$ (since the cardinality of a set is a natural number) and the constant k is used to ensure that, at each step, the correct intersection inequalities are satisfied, like for example,

$$|A \cap B \cap C| \leq |A \cap B|.$$

Essentially, we need each component of the Venn diagram of $A_1, \ldots, A_n$ to be non-negative.

The key to prove Theorem 3 are the following lemmas which we provide without proof, both of which can be done by induction on Equation 1.

Lemma 1. *Let $A = \{A_1, \ldots, A_n\}$ and $B = \{B_1, \ldots, B_n\}$. Suppose that for $\emptyset \neq I \subseteq [n]$*

$$|\triangle_{i \in I} A_i| = \delta_I$$

and

$$|\triangle_{i \in I} B_i| = m\delta_I.$$

Then,

$$\left|\bigcap_{i\in I} B_i\right| = m\left|\bigcap_{i\in I} A_i\right|$$

considered formally in the sense that they satisfy equation 1.

Lemma 2. *Let* $A = \{A_1, \ldots, A_n\}$ *and* $B = \{B_1, \ldots, B_n\}$. *Suppose that for* $\emptyset \neq I \subseteq [n]$

$$|\triangle_{i\in I} A_i| = \delta_I$$

and

$$|\triangle_{i\in I} B_i| = \delta_I + 1.$$

Then,

$$\left|\bigcap_{i\in I} B_i\right| = \left|\bigcap_{i\in I} A_i\right| + \frac{1}{2^{|I|-1}}$$

considered formally in the sense that they satisfy equation 1.

With these two lemmas we can prove our result.

Theorem 3. *Given* $\delta_I \in \mathbb{Q}_+$ *where* $\emptyset \neq I \subset [n]$, *there exists finite sets* $A_1, \ldots, A_n$ *and constants* $m, k \in \mathbb{Z}$ *such that*

$$|\triangle_{i\in I} A_i| = m\delta_I + k.$$

Proof. Find the values of the intersections of the A_i recursively from

$$|\triangle_{i\in I} A_i| = \delta_I$$

using Equation 1. Then, use Lemma 1 to make all values integers and Lemma 2 to guarantee the intersection inequalities are valid, and therefore, that each component of the Venn diagram of $A_1, \ldots, A_n$ is non-negative.

□

We give an example that illustrates how to obtain m and k.

Example 5. *Consider the following translation invariant metric,* d *with weight* ω, *over* $\mathbb{F}_2^3$.

$$\begin{array}{lll} \omega(001) = 3 & \omega(011) = 3 & \omega(111) = 3 \\ \omega(010) = 2 & \omega(101) = 3 & \\ \omega(100) = 1 & \omega(110) = 2 & \end{array}$$

Thus we must solve the following system of symmetric difference.

$$\begin{array}{lll} |A| = 3 & |A \triangle B| = 3 & |A \triangle B \triangle C| = 3 \\ |B| = 2 & |A \triangle C| = 3 & \\ |C| = 1 & |B \triangle C| = 2 & \end{array}$$

Using Equation 1 recursively we can determine that

$$|A \cap B| = 1 \qquad |B \cap C| = \tfrac{1}{2}$$
$$|A \cap C| = \tfrac{1}{2} \qquad |A \cap B \cap C| = \tfrac{1}{4}$$

Adding 1 *to the initial conditions and then multiplying by* 2*, using the lemmas, we get*

$$|A| = 8 \qquad |A \cap B| = 3 \qquad |A \cap B \cap C| = 1$$
$$|B| = 6 \qquad |A \cap C| = 2$$
$$|C| = 4 \qquad |B \cap C| = 2$$

This gives us an embedding $f : \mathbb{F}_2^3 \mapsto H^{12}$ *such that*

$$f(001) = 111111110000$$
$$f(010) = 000001111110$$
$$f(100) = 000011000011$$

and decoding on the image of f *in* (H^{12}, d_H) *is equivalent to decoding on* $(\mathbb{F}_2^3, d)$.

As stated before, Theorem 1 and Theorem 2 follow straighforward from Theorem 3. Those two theorems together with the inequalities analysis made in Section 3, allows for a geometric-analytic approach to the problem of matching distances and metrics to channels including the possibility of approximating a channel by a metric and other analytic-like results. One of our goals is to study the packing radius under general metrics, and in particular to see if we can better understand perfect codes, since for many metrics, it is an open problem if there exists any. This geometric approach also allows us to get better embeddings. For example, we can show that the best embedding, in terms of minimizing the dimension of the Hamming cube, of Example 5 is into H^{11}.

Some results in this direction that are already proved and some that are under work, will be presented in future works.

References

1. Michel Marie Deza and Elena Deza, "Encyclopedia of distances," 3rd revised edition, *Springer-Verlag*, 2014.
2. Marcelo Firer and Judy L. Walker, "Matched Metrics and Channels", *preprint*, 2014.
3. Ernst Gabidulin, "A brief survey of metrics in coding theory," *Mathematics of Distances and Applications* (2012): 66.
4. James L. Massey, "Notes on coding theory, class notes for course 6.575 (spring)", M.I.T., Cambridge, MAA, 1967.
5. Gérald Séguin, "On metrics matched to the discrete memoryless channel", *J. Franklin Inst. 309*, no. 3, 179189, 1980.