Universal Class-Based Block Source Coding for Discrete Memoryless Sources

Mehdi Dabirnia
CTTC
mehdi.dabirnia@cttc.es

Albert Guillén i Fàbregas
University of Cambridge
Universitat Politècnica de Catalunya
guillen@ieee.org

Alfonso Martinez
Universitat Pompeu Fabra
alfonso.martinez@ieee.org

***Abstract*—We introduce a general class-based block source coding framework which, for the case of discrete memoryless sources, is specialized to four different constructions with varying levels of underlying structure in the class definitions. For all the constructions, we show that the matched random coding error exponent is achieved universally with a corresponding universal decoder. Increasing the structure in the classes results in a simpler corresponding universal decoder.**

I. Introduction

Lossless source coding has been studied in three main different settings. The setting that has been studied most extensively is strictly-lossless variable-length prefix data compression. In this setting, a source code is an injective mapping of source sequences of fixed-length n to variable-length binary codewords where no codeword is a prefix of another (see e.g. [1]). In the second setting, the prefix constraint is dropped when a whole file is to be compressed at once, while retaining strictly-lossless, variable-length codes. An optimal code in this setting is a deterministic mapping of length-n source sequences ordered with decreasing probability to binary sequences of increasing length. A rigorous treatment of the optimal code in this setting for the known distribution case and universal case can be found in [2], [3]. The third and last setting is almost-lossless, fixed-length to fixed-length data compression which we refer to in this paper as *block source coding*. The optimal n-to-k code in this setting maps $2^k - 1$ most likely sequences to distinct codewords and maps all the remaining ones to the last codeword (error index). The analysis of optimal code for a discrete memoryless source (DMS) shows the exponential decay of the error probability with the optimal exponent given by [4]

$$E(R) = \max_{\rho \geq 0} \rho R - E_{\mathrm{s}}(\rho) \tag{1}$$

where

$$E_{\mathrm{s}}(\rho) = \log \left(\sum_{v \in \mathcal{V}} P_V(v)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \tag{2}$$

When the distribution of a DMS is unknown, an optimal universal code can be constructed by one-to-one mapping the lowest empirical entropy sequences instead of the most likely ones [1]. Using this construction, the optimal exponent can be achieved universally [5], [6]. The literature on variable-length universal source coding is extensive (see e.g. [7] and references therein). Practical block-code designs have been considered in a number of works, see e.g. [8], [9]. In this paper, we introduce a general class-based block source coding framework, in analogy with channel coding, where source sequences are partitioned into disjoint classes, and each class is encoded independently. For DMSs, we specialize the framework to four code constructions, depending on the structure imposed on the code. We show that the random coding error exponent is universally achievable in all four cases.

This work was supported in part by the European Research Council under Grants 101142747 and 101158232, and in part by the Spanish Government under Grants PID2020-116683GB-C22, PID2021-128373OB-I00, and PID2024-159557OB-C22, and partially by the WAVE-XL project (PID2024-160457OB-I00) funded by MCIN/AEI/10.13039/501100011033 and by FEDER, UE.

A. Class-Based Block Source Coding

Consider a discrete source with finite alphabet $\mathcal{V}$ that outputs sequences of length n, $\boldsymbol{V} = V_1, \ldots, V_n$, with probability distribution $P_{\boldsymbol{V}}$, and realization denoted by $\boldsymbol{v} = v_1, \ldots, v_n$.

We partition the source-message set $\mathcal{V}^n$ into N_n disjoint subsets (classes) $\mathcal{A}_i^n$, $i = 1, \ldots, N_n$, with $\bigcup_{i=1}^{N_n} \mathcal{A}_i^n = \mathcal{V}^n$. We let the partitions be such that, for each n, source sequences are assigned to classes depending on a common property: for example having same empirical type.

The codeword set $\mathcal{M} = \{1, \cdots, M\}$ is also partitioned into N_n disjoint subsets as $\{\mathcal{M}_1, \cdots, \mathcal{M}_{N_n}\}$ where $\mathcal{M}_i$ is the codeword set for the class $\mathcal{A}_i^n$. A class-based block source code $\boldsymbol{\mathcal{C}} = \{\mathcal{C}_1, \cdots, \mathcal{C}_{N_n}\}$ is a union of N_n codes where each $\mathcal{C}_i$ is an (n, R_i) block code for source class $\mathcal{A}_i^n$ with mapping function $\phi_i : \mathcal{A}_i^n \to \mathcal{M}_i$ and rate $R_i = \frac{\log |\mathcal{M}_i|}{n}$. Each code $\mathcal{C}_i$ can be represented as a partitioning of the corresponding class $\mathcal{A}_i^n$ into $M_i = |\mathcal{M}_i|$ subsets as $\{\mathcal{A}_{i1}, \cdots, \mathcal{A}_{iM_i}\}$.

The decoder for a class-based code $\boldsymbol{\mathcal{C}}$ is a set of mappings $\psi_i : \mathcal{M}_i \to \mathcal{A}_i^n$ for every $i \in \{1, \cdots, N_n\}$. We consider using a maximum metric decoder as follows: for every $m \in \mathcal{M}_i$

$$\psi_i(m) = \underset{\boldsymbol{v} \in \mathcal{A}_i^n : \phi_i(\boldsymbol{v}) = m}{\arg\max} \; q(\boldsymbol{v}), \tag{3}$$

where $q(\boldsymbol{v})$ is a positive decoding metric. For a sequence $\boldsymbol{v} \in \mathcal{A}_i^n$, the decoder makes an error whenever $\psi_i(\phi_i(\boldsymbol{v})) \neq \boldsymbol{v}$. For a class-based code, an error can only happen between the source sequences of the same class. Here we consider the average probability of error as the metric of interest.

Definition 1. *A Class-based random coding ensemble is the set of all (n, R) block codes for the source sequences $\mathcal{V}^n$ with a probability measure over the codes having the following property: For every source class $\mathcal{A}_i^n$, $i \in \{1, \cdots, N_n\}$, each source sequence $\boldsymbol{v} \in \mathcal{A}_i^n$ is independently assigned with equal probability $\frac{1}{M_i}$ to each of the codewords in $\mathcal{M}_i$.*

B. Discrete Memoryless Sources

Following conventional definitions, the type of a sequence $\boldsymbol{v}$ generated by a DMS is the probability distribution $\hat{P}_{\boldsymbol{v}}$ with $\hat{P}_{\boldsymbol{v}}(v) = \frac{n_{\boldsymbol{v}}(v)}{n}$ for any $v \in \mathcal{V}$ where $n_{\boldsymbol{v}}(v)$ is the number of appearances of symbol v in $\boldsymbol{v}$. The type class $\{\bar{\boldsymbol{v}} \in \mathcal{V}^n : \hat{P}_{\bar{\boldsymbol{v}}} = \hat{P}_{\boldsymbol{v}}\}$ is denoted by $\mathcal{T}_n(\hat{P}_{\boldsymbol{v}})$. The set of all possible types of sequences $\boldsymbol{v} \in \mathcal{V}^n$ is denoted by $\mathcal{P}_n(\mathcal{V})$.

We consider and study four different partitionings of the source n-tuples $\mathcal{V}^n$ to classes as follows:

1) Type class: $\mathcal{A}_i^n = \mathcal{T}_n(\hat{P}_i)$ for $i \in \{1, \cdots, |\mathcal{P}_n(\mathcal{V})|\}$.
2) Empirical entropy class: $\mathcal{A}_i^n = \bigcup_j \mathcal{T}_n(\hat{P}_j) : H(\hat{P}_j) = \hat{H}_i$ for $i \in \{1, \cdots, N_n\}$, where $\hat{H}_i$'s are distinct empirical entropy levels for source n-tuples.
3) Tilted family class: $\mathcal{A}_i^n = \bigcup_j \mathcal{T}_n(\hat{P}_j) : \hat{P}_j \in \mathcal{E}_{Q_i}$, where $\mathcal{E}_{Q_i}$ are distinct tilted families (Definition 2) for empirical distributions of source n-tuples.
4) Single class: $\mathcal{A}^n = \mathcal{V}^n$.

For each of the four cases, we consider class-based block source codes with corresponding universal decoding metrics:

- For 1) and 2), $q(\boldsymbol{v})$ is an arbitrary decoding metric,
- For 3), we consider $q(\boldsymbol{v}) = \bar{Q}_i$ for $\boldsymbol{v} \in \mathcal{A}_i^n$ where $\bar{Q}_i$ is an arbitrary distribution from the tilted family $\mathcal{E}_{Q_i}$,
- For 4), we consider the empirical probability metric $q(\boldsymbol{v}) = e^{-nH(\hat{P}_{\boldsymbol{v}})} = \hat{P}_{\boldsymbol{v}}(\boldsymbol{v})$. Such a maximum metric decoder is equivalent to a minimum empirical entropy decoder, the source coding counterpart of the maximum mutual information decoder for channel coding [6].

II. Main result

Our main result is an achievable error exponent for class-based ensembles. First, we give the random coding error exponent for the ensemble of standard random block codes, a special case of class-based random coding with a single class. Considering a matched maximum likelihood (ML) decoder, which minimizes the average error probability, the random coding exponent for standard ensemble has a similar form to (1) where the maximization is over $\rho \in [0, 1]$,

$$E_r(R) = \max_{\rho \in [0,1]} \rho R - E_{\mathrm{s}}(\rho). \tag{4}$$

Theorem 1. *For discrete memoryless sources and each of the four above-mentioned class-based partitionings, there exists a class-based (n, R) block source code with $M = \lceil e^{nR} \rceil$ codewords such that a maximum metric decoder with universal metric $q(\boldsymbol{v})$ achieves the random coding exponent in* (4).

Next, we give bounds on the average error probability of a random code with generic decoding metrics that we later particularize to the four ensemble-decoder pairs above.

A. Random Coding Error Probability Analysis

A random code from the class-based ensemble is denoted by $\mathsf{C} = \{\mathsf{C}_1, \cdots, \mathsf{C}_{N_n}\}$ with encoding functions $\{\Phi_i\}_{i=1}^{N_n}$. Over the ensemble of random class-based codes, the average error probability of a source sequence $\boldsymbol{v} \in \mathcal{A}_i^n$ can be written as

$$\bar{p}_e(\boldsymbol{v}, \mathsf{C}_i) = \mathbb{E}\left[\mathbb{1}\left[\bigcup_{\substack{\bar{\boldsymbol{v}} \neq \boldsymbol{v} \\ \bar{\boldsymbol{v}} \in \mathcal{A}_i^n}} \{q(\bar{\boldsymbol{v}}) \geq q(\boldsymbol{v}), \Phi_i(\bar{\boldsymbol{v}}) = \Phi_i(\boldsymbol{v})\}\right]\right] \tag{5}$$

$$\leq \sum_{\bar{\boldsymbol{v}} \in \mathcal{A}_i^n} \mathbb{1}\left[q(\bar{\boldsymbol{v}}) \geq q(\boldsymbol{v})\right] \mathbb{P}\left[\Phi_i(\bar{\boldsymbol{v}}) = \Phi_i(\boldsymbol{v})\right], \tag{6}$$

$$\leq \frac{1}{M_i} \sum_{\bar{\boldsymbol{v}} \in \mathcal{A}_i^n} \mathbb{1}\left[q(\bar{\boldsymbol{v}}) e^{f(\bar{\boldsymbol{v}})} \geq q(\boldsymbol{v}) e^{f(\boldsymbol{v})}\right], \tag{7}$$

$$\leq \frac{1}{M_i} \sum_{\bar{v}^n \in \mathcal{A}_i^n} \left(\frac{q(\bar{\boldsymbol{v}}) e^{f(\bar{\boldsymbol{v}})}}{q(\boldsymbol{v}) e^{f(\boldsymbol{v})}}\right)^{s_i}, \tag{8}$$

where we use the union bound in (6), in (7) $f(\cdot)$ is any arbitrary function that has same value for all the sequences in the same class and possibly different values for the sequences of different classes, and (8) holds for any $s_i \geq 0$. We note that the ratio $\frac{e^{f(\bar{\boldsymbol{v}})}}{e^{f(\boldsymbol{v})}} = 1$ in (8), however the function $f(\cdot)$ will be an extra degree of optimization after simplifying the bound.

Averaging over all source sequences yields an upper bound on the ensemble average error probability,

$$p_e(\mathsf{C}) = \sum_{i=1}^{N_n} p_e(\mathsf{C}_i), \tag{9}$$

where

$$p_e(\mathsf{C}_i) \leq \sum_{\boldsymbol{v} \in \mathcal{A}_i^n} P_{\boldsymbol{V}}(\boldsymbol{v}) \left(\frac{1}{M_i} \sum_{\bar{\boldsymbol{v}} \in \mathcal{A}_i^n} \left(\frac{q(\bar{\boldsymbol{v}}) e^{f(\bar{\boldsymbol{v}})}}{q(\boldsymbol{v}) e^{f(\boldsymbol{v})}}\right)^{s_i}\right)^{\rho_i}, \tag{10}$$

and (10) holds for any $\rho_i \in [0, 1]$.

The average error probability $p_e(\mathsf{C})$ in (9) can be further upper bounded as

$$p_e(\mathsf{C}) \leq N_n \max_i p_e(\mathsf{C}_i), \tag{11}$$

which implies that the corresponding error exponent is dominated by the exponent of the worst class as long as N_n grows sub-exponentially with n.

In order to find a simpler bound on $p_e(\mathsf{C})$ that does not require maximization over the classes, we first weaken the bound in (10) by including all $\bar{\boldsymbol{v}}$ in the inner sum and choose $M_i = \frac{M}{N_n}$ and fix $s_i = s$ and $\rho_i = \rho$, hence we obtain

$$p_e(\mathsf{C}) \leq \sum_{\boldsymbol{v} \in \mathcal{V}^n} P_{\boldsymbol{V}}(\boldsymbol{v}) \left(\frac{N_n}{M} \sum_{\bar{\boldsymbol{v}} \in \mathcal{V}^n} \left(\frac{q(\bar{\boldsymbol{v}}) e^{f(\bar{\boldsymbol{v}})}}{q(\boldsymbol{v}) e^{f(\boldsymbol{v})}}\right)^{s}\right)^{\rho}, \tag{12}$$

for any $s \geq 0$ and $\rho \in [0, 1]$. The bound in (12) can be optimized over the choice of $f(\cdot)$ which only needs to satisfy $f(\boldsymbol{v}) = \gamma_i$ for all $\boldsymbol{v} \in \mathcal{A}_i^n$ where γ_i is a constant for every i.

III. Proof of Theorem 1

We now prove the achievability of the random coding exponent with matched decoding for discrete memoryless sources with each of the four class-based block source codings introduced in section I-B with a corresponding universal decoder for each case.

A. Type Class

In this case, each source type class is considered as a separate class, hence, $N_n = |\mathcal{P}_n(\mathcal{V})|$ and $f(\boldsymbol{v})$ in (12) can be any arbitrary function that depend on the source sequence $\boldsymbol{v}$ only through its type.

Now setting $s = \frac{1}{1+\rho}$, $f(\boldsymbol{v}) = \log \frac{P_{\boldsymbol{V}}(\boldsymbol{v})}{q(\boldsymbol{v})}$ in (12) and using memoryless property of the source we obtain

$$p_e(\mathsf{C}) \leq \left(\frac{|\mathcal{P}_n(\mathcal{V})|}{M}\right)^{\rho} \left(\sum_{v\in\mathcal{V}} P_V(v)^{\frac{1}{1+\rho}}\right)^{1+\rho} \tag{13}$$

$$= e^{-n(\rho(R-\delta_n)-E_s(\rho))} \tag{14}$$

where $\delta_n = \frac{\log|\mathcal{P}_n(\mathcal{V})|}{n} \to 0$ as $n \to \infty$ and $E_{\mathrm{s}}(\rho)$ is given in (2). This shows that for discrete memoryless sources, type class-based source coding achieves the matched random coding error exponent universally irrespective of the decoding metric since the choice of function $f(\boldsymbol{v}) = \log \frac{P_{\boldsymbol{V}}(\boldsymbol{v})}{q(\boldsymbol{v})}$ cancels the effect of decoding metric.

B. Empirical Entropy Class

In this case, all the source type classes with equal empirical entropy are partitioned together into a single class, hence, the number of separate classes N_n is equal to the number of distinct empirical entropy levels. Denoting the maximum number of types within an equal empirical entropy class by K_n, we consider simple upper bounds of $K_n \leq |\mathcal{P}_n(\mathcal{V})|$ and $N_n \leq |\mathcal{P}_n(\mathcal{V})|$ in the following.

For a source sequence $\boldsymbol{v}$ and probability distribution $P_{\boldsymbol{V}}(\cdot)$ define

$$\mathcal{B}(\boldsymbol{v}) = \{\bar{\boldsymbol{v}} \in \mathcal{V}^n : P_{\boldsymbol{V}}(\bar{\boldsymbol{v}}) \geq P_{\boldsymbol{V}}(\boldsymbol{v})\}. \tag{15}$$

For any source sequence $\boldsymbol{v} \in \mathcal{A}_i^n$ we define the set

$$\tilde{\mathcal{A}}(\boldsymbol{v}) = \{\bar{\boldsymbol{v}} \in \mathcal{A}_i^n : q(\bar{\boldsymbol{v}}) \geq q(\boldsymbol{v})\}, \tag{16}$$

which are the set of sequences that would cause error if they were assigned to the same codeword. Noting that for any arbitrary metric $q(\cdot)$ we have $|\tilde{\mathcal{A}}(\boldsymbol{v})| \leq |\mathcal{A}_i^n|$ for $\boldsymbol{v} \in \mathcal{A}_i^n$ and since types with equal empirical entropy have type classes with equal size and also since $\mathcal{B}(\boldsymbol{v})$ includes type class $\mathcal{T}_n(\hat{P}_{\boldsymbol{v}})$, we upper bound the size of $\tilde{\mathcal{A}}(\boldsymbol{v})$ as

$$|\tilde{\mathcal{A}}(\boldsymbol{v})| \leq K_n|\mathcal{B}(\boldsymbol{v})|. \tag{17}$$

Now we upper bound the average error probability of a source sequence $\boldsymbol{v} \in \mathcal{A}_i^n$ using (6) as

$$\bar{p}_e(\boldsymbol{v}, \mathsf{C}_i) \leq \frac{|\tilde{\mathcal{A}}(\boldsymbol{v})|}{M_i}, \tag{18}$$

$$\leq \frac{K_n N_n}{M} \sum_{\bar{\boldsymbol{v}}\in\mathcal{V}^n} \left(\frac{P_{\boldsymbol{V}}(\bar{\boldsymbol{v}})}{P_{\boldsymbol{V}}(\boldsymbol{v})}\right)^s, \tag{19}$$

where we use (17) and the bound $\mathbb{1}[P_{\boldsymbol{V}}(\bar{\boldsymbol{v}}) \geq P_{\boldsymbol{V}}(\boldsymbol{v})] \leq \left(\frac{P_{\boldsymbol{V}}(\bar{\boldsymbol{v}})}{P_{\boldsymbol{V}}(\boldsymbol{v})}\right)^s$ for $s \geq 0$.

Now averaging over all source sequences we obtain

$$p_e(\mathsf{C}) \leq \sum_{\boldsymbol{v}\in\mathcal{V}^n} P_{\boldsymbol{V}}(\boldsymbol{v}) \left(\frac{K_n N_n}{M} \sum_{\bar{\boldsymbol{v}}\in\mathcal{V}^n} \left(\frac{P_{\boldsymbol{V}}(\bar{\boldsymbol{v}})}{P_{\boldsymbol{V}}(\boldsymbol{v})}\right)^s\right)^{\rho}, \tag{20}$$

which holds for any $\rho \in [0, 1]$.

Now setting $s = \frac{1}{1+\rho}$ in (20) and using memoryless property of the source we obtain a similar bound as (14) with $\delta_n = \frac{\log K_n N_n}{n} \leq \frac{2\log|\mathcal{P}_n(\mathcal{V})|}{n} \to 0$ as $n \to \infty$. This shows that for discrete memoryless sources, empirical entropy class-based source coding achieves the matched exponent universally irrespective of the decoding metric. This is because the bound in (17) is valid regardless of the decoding metric.

C. Tilted Family Class

We start by defining the tilted family and state a lemma that will be used to prove the result for tilted family class.

Definition 2. *For a probability distribution Q with support $\mathcal{V}$, the tilted family of Q is defined as*

$$\mathcal{E}_Q = \{Q^{(s)} : s \in \mathbb{R}^+\},$$

where $Q^{(s)}$ is the tilted distribution of order s of Q and for every $v \in \mathcal{V}$ it is given by

$$Q_V^{(s)}(v) = \frac{Q_V(v)^s}{\sum_{\bar{v}\in\mathcal{V}} Q_V(\bar{v})^s}.$$

The tilted family $\mathcal{E}_Q$ is an exponential family of probability distributions written as

$$Q_V^{(s)}(v) = ce^{sf(v)}$$

where $f(v) = \log Q_V(v)$ and $c = \big(\sum_{v\in\mathcal{V}} e^{sf(v)}\big)^{-1}$. The tilted family $\mathcal{E}_Q$ depends on Q only in a weak manner, since any element of $\mathcal{E}_Q$ could play the role of Q.

For tilted family class case, all the source type classes having same tilted family are grouped together into a single class, hence, the number of separate classes N_n is equal to the number of distinct tilted families of empirical distributions on source n-tuples. We consider a simple upper bound of $N_n \leq |\mathcal{P}_n(\mathcal{V})|$ in the following.

For the i-th tilted family class $\mathcal{A}_i^n$, we consider the universal decoding metric $q(\boldsymbol{v}) = \bar{Q}_i$ where $\bar{Q}_i$ is an arbitrary distribution from the corresponding tilted family $\mathcal{E}_{Q_i}$. For any source sequence $\boldsymbol{v} \in \mathcal{A}_i^n$ we define the set

$$\tilde{\mathcal{A}}(\boldsymbol{v}) = \{\bar{\boldsymbol{v}} \in \mathcal{A}_i^n : \bar{Q}_i(\bar{\boldsymbol{v}}) \geq \bar{Q}_i(\boldsymbol{v})\}. \tag{21}$$

Considering the definition of $\mathcal{B}(\boldsymbol{v})$ in (15), we have the following inequality on the size of the sets $\tilde{\mathcal{A}}(\boldsymbol{v})$ and $\mathcal{B}(\boldsymbol{v})$.

Lemma 1. *For any source sequence $\boldsymbol{v}$ from tilted family class $\mathcal{A}_i^n$, we have*

$$|\tilde{\mathcal{A}}(\boldsymbol{v})| \leq |\mathcal{B}(\boldsymbol{v})| \tag{22}$$

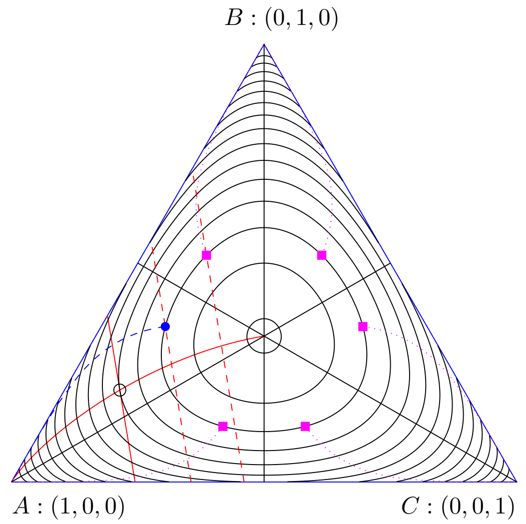

Fig. 1: Ternary simplex and illustration of permuted tilted families, iso-cross-entropy families and iso-entropy contours.

for any arbitrary probability distribution P_V in definition of $\mathcal{B}(\boldsymbol{v})$ in (15).

The full proof of Lemma 1 is involved and omitted here. We only provide an sketch of the proof through an illustrative ternary example in the following.

Fig. 1 shows the ternary simplex with iso-entropy contours. Assume that solid red curve connecting the corner point A to the center of the simplex U (uniform distribution) shows the tilted family $\mathcal{E}_{P_V}$ of the (arbitrary) distribution P_V in the definition of $\mathcal{B}(\boldsymbol{v})$.

As first case consider a sequence $\boldsymbol{v}$ such that the empirical type $\hat{P}_{\boldsymbol{v}}$ belongs to the tilted family $\mathcal{E}_{P_V}$. The black circle on the figure illustrates one such type. The set $\tilde{\mathcal{A}}(\boldsymbol{v})$ is union of all the type classes with empirical type on the portion of titled family curve between the corner point A and $\hat{P}_{\boldsymbol{v}}$. The solid red line passing through $\hat{P}_{\boldsymbol{v}}$ shows the iso-cross-entropy family $\mathcal{M}_{\hat{P}_{\boldsymbol{v}}} = \{P : H(P\|\hat{P}_{\boldsymbol{v}}) = H(\hat{P}_{\boldsymbol{v}})\}$. The iso-cross-entropy families corresponding to other distributions from tilted family $\mathcal{E}_{P_V}$ are parallel lines to $\mathcal{M}_{\hat{P}_{\boldsymbol{v}}}$. The set $\mathcal{B}(\boldsymbol{v})$ is union of all the type classes with empirical type over the triangle formed by the vertex A and the solid red line as opposite side which includes the $\tilde{\mathcal{A}}(\boldsymbol{v})$ as a subset. Notice that in this case we can also replace $\bar{Q}_i$ with P_V in (21) and show that $\tilde{\mathcal{A}}(\boldsymbol{v}) \subset \mathcal{B}(\boldsymbol{v})$.

As second case consider a sequence $\boldsymbol{v}$ such that the empirical type $\hat{P}_{\boldsymbol{v}}$ belongs to the same region of the simplex as P_V according to ordering of symbol probabilities. The blue dot on the figure illustrates one such type. The set $\tilde{\mathcal{A}}(\boldsymbol{v})$ is union of all the type classes on the portion of titled family curve (dashed blue curve) between the corner point A and the blue dot. The iso-cross-entropy family for a tilted distribution $P_V^{(s)}$ passing through $\hat{P}_{\boldsymbol{v}}$ (blue dot), namely $\mathcal{M}_{P_V^{(s)}}$, is shown by dashed red line. The set $\mathcal{B}(\boldsymbol{v})$ is union of all the type classes with empirical type over the triangle formed by the vertex A and the dashed red line (passing through blue dot) as opposite side which includes the $\tilde{\mathcal{A}}(\boldsymbol{v})$ as a subset.

For the last case, consider a sequence $\boldsymbol{v}$ such that the empirical type $\hat{P}_{\boldsymbol{v}}$ belongs to a different region of the simplex than P_V. The magenta squares on the figure illustrate such types. The set $\tilde{\mathcal{A}}(\boldsymbol{v})$ is union of all the type classes on the portion of titled family curve (dotted magenta curve) between the corresponding corner point and the magenta square. For any such $\boldsymbol{v}$ there is a permutation of $\hat{P}_{\boldsymbol{v}}$ and corresponding tilted family $\mathcal{E}_{\hat{P}_{\boldsymbol{v}}}$ denoted by $\pi(\hat{P}_{\boldsymbol{v}})$ and $\mathcal{E}_{\pi(\hat{P}_{\boldsymbol{v}})}$, respectively, that are on the same region of the simplex as P_V. The dashed blue curve is the corresponding permutation of dotted magenta curves. We first note that since any type class lying on a dotted magenta curve has a corresponding permuted type class on the dashed blue curve, therefore, $|\tilde{\mathcal{A}}(\boldsymbol{v})| = |\tilde{\mathcal{A}}(\bar{\boldsymbol{v}})|$ for any $\bar{\boldsymbol{v}} \in \mathcal{T}_n(\pi(\hat{P}_{\boldsymbol{v}}))$. Therefore using the result of the second case we have $|\tilde{\mathcal{A}}(\boldsymbol{v})| \leq |\mathcal{B}(\bar{\boldsymbol{v}})|$. The set $\mathcal{B}(\boldsymbol{v})$ is union of all the type classes with empirical type over the triangle formed by the vertex A and the dashed red line (the iso-cross-entropy family passing through $\hat{P}_{\boldsymbol{v}}$) as opposite side. Also note that $P_V(\bar{\boldsymbol{v}}) \geq P_V(\boldsymbol{v})$ since $\pi(\hat{P}_{\boldsymbol{v}})$ has similar ordering to P_V whilst $\hat{P}_{\boldsymbol{v}}$ has different ordering, therefore $\mathcal{B}(\bar{\boldsymbol{v}}) \subseteq \mathcal{B}(\boldsymbol{v})$.

Using Lemma 1 we upper bound the average error probability of a source sequence $\boldsymbol{v} \in \mathcal{A}_i^n$ from (6) as

$$\bar{p}_e(\boldsymbol{v}, \mathsf{C}_i) \leq \frac{|\tilde{\mathcal{A}}(\boldsymbol{v})|}{M_i}, \tag{23}$$

$$\leq \frac{N_n}{M} \sum_{\bar{\boldsymbol{v}} \in \mathcal{V}^n} \left(\frac{P_{\boldsymbol{V}}(\bar{\boldsymbol{v}})}{P_{\boldsymbol{V}}(\boldsymbol{v})}\right)^s, \tag{24}$$

where the (24) holds for any arbitrary probability distribution, especially for the matched distribution P_V.

Averaging over all source sequences and using memoryless property of the source we obtain a similar bound as (14) with $\delta_n = \frac{\log N_n}{n} \leq \frac{\log|\mathcal{P}_n(\mathcal{V})|}{n} \to 0$ as $n \to \infty$.

This shows that for discrete memoryless sources, tilted family class-based random coding achieves the matched random coding error exponent universally with corresponding universal decoding metric.

D. Single Class

In this case we consider a universal decoding metric given by $q(\boldsymbol{v}) = 2^{-nH(\hat{P}_{\boldsymbol{v}})} = \hat{P}_{\boldsymbol{v}}(\boldsymbol{v})$ which is the empirical probability of the sequence $\boldsymbol{v}$. For any sequence $\boldsymbol{v}$, we define the set $\tilde{\mathcal{A}}(\boldsymbol{v})$ as the set of sequences $\bar{\boldsymbol{v}} \in \mathcal{V}^n$ that would cause error if they were assigned to the same codeword, namely for such sequences we have $\hat{P}_{\bar{\boldsymbol{v}}}(\bar{\boldsymbol{v}}) \geq \hat{P}_{\boldsymbol{v}}(\boldsymbol{v})$, or equivalently $H(\hat{P}_{\bar{\boldsymbol{v}}}) \leq H(\hat{P}_{\boldsymbol{v}})$, i.e.,

$$\tilde{\mathcal{A}}(\boldsymbol{v}) = \{\bar{\boldsymbol{v}} \in \mathcal{V}^n : H(\hat{P}_{\bar{\boldsymbol{v}}}) \leq H(\hat{P}_{\boldsymbol{v}})\}. \tag{25}$$

Noting that a type with lower empirical entropy, has a smaller type class and also since $\mathcal{B}(\boldsymbol{v})$ includes type class $\mathcal{T}_n(\hat{P}_{\boldsymbol{v}})$, we upper bound the cardinality of the set $\tilde{\mathcal{A}}(\boldsymbol{v})$ as

$$|\tilde{\mathcal{A}}(\boldsymbol{v})| \leq |\mathcal{P}_n(\mathcal{V})||\mathcal{B}(\boldsymbol{v})|. \tag{26}$$

Using (26) we upper bound the average error probability of a source sequence $\boldsymbol{v} \in \mathcal{V}^n$ from (6) as

$$\bar{p}_e(\boldsymbol{v}, \mathsf{C}) \leq \frac{|\tilde{\mathcal{A}}(\boldsymbol{v})|}{M}, \tag{27}$$

$$\leq \frac{|\mathcal{P}_n(\mathcal{V})|}{M} \sum_{\bar{\boldsymbol{v}} \in \mathcal{V}^n} \left(\frac{P_{\boldsymbol{V}}(\bar{\boldsymbol{v}})}{P_{\boldsymbol{V}}(\boldsymbol{v})} \right)^s \tag{28}$$

Averaging over all source sequences and using memoryless property of the source we obtain a similar bound as (14) with $\delta_n = \frac{\log |\mathcal{P}_n(\mathcal{V})|}{n} \to 0$ as $n \to \infty$.

This shows that for discrete memoryless sources, single class block source coding with universal maximum empirical probability decoding (minimum empirical entropy decoding) achieves the matched random coding error exponent.

References

[1] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2006.

[2] I. Kontoyiannis and S. Verdú, "Optimal lossless data compression: Non-asymptotics and asymptotics," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 777–795, 2014.

[3] O. Kosut and L. Sankar, "Asymptotics and non-asymptotics for universal fixed-to-variable source coding," *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 3757–3772, 2017.

[4] F. Jelinek, *Probabilistic information theory*. McGraw-Hill, 1968.

[5] G. Longo and A. Sgarro, "The source coding theorem revisited: A combinatorial approach," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 544–548, 1979.

[6] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[7] S. Verdú, "Fifty years of Shannon theory," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2057–2078, 1998.

[8] T. Ancheta, "Syndrome-source-coding and its universal generalization," *IEEE Trans. Inf. Theory*, vol. 22, no. 4, pp. 432–436, 1976.

[9] G. Caire, S. Shamai, and S. Verdú, "Noiseless data compression with low-density parity-check codes," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 66, pp. 263–284, 2004.