Entropy **2015**, *xx*, 1-x; doi:10.3390/——

OPEN ACCESS
entropy
ISSN 1099-4300
www.mdpi.com/journal/entropy

Article

Entropy approximation

Marek Śmieja [1,*] and Jacek Tabor [1]

[1] Department of Mathematics and Computer Science, Jagiellonian University, Lojasiewicza 6, 30-348 Kraków, Poland

* Author to whom correspondence should be addressed; marek.smieja@ii.uj.edu.pl.

Received: xx / Accepted: xx / Published: xx

Abstract: In this paper we investigate a lossy source coding problem, where an upper limit on the permitted distortion is defined for every data set element. It can be seen as an alternative approach to rate distortion theory where a bound on the allowed average error is specified. In order to find the entropy, which gives a statistical length of source code compatible with a fixed distortion bound, a corresponding optimization problem has to be solved. First, we show how to simplify this general optimization by reducing the number of coding partitions which are irrelevant for the entropy calculation. In our main result we present a fast and feasible for implementation greedy algorithm which allows to approximate the entropy within an additive error term of $\log_2 e$. The proof is based on the minimum entropy set cover problem, for which a similar bound was obtained.

Keywords: Shannon entropy, entropy approximation, minimum entropy set cover, lossy compression, source coding.

1. Introduction

The lossy source coding transforms a possibly continuously distributed information into a finite number of codewords [1,5]. Although this allows to encode data efficiently, such operation is irreversible and once modified information cannot be restored accurately. One of the fundamental questions in lossy coding is the following: *What is the lowest achievable statistical code length given a maximal coding error?* To answer to this question, the precise formulation of the coding error and related definition of the entropy needs to be given. In this paper we present how to approximate the value of the entropy in the case when every entry element has a fixed upper limit on the permitted error.

1.1. Motivation

In order to explain our results we give more precise problem formulation. Suppose that a random source represented by a probability measure μ produces the information from space X. We fix a partition $\mathcal{P}$ of X and encode an arbitrary element $x \in X$ by a unique $P \in \mathcal{P}$ such that $x \in P$. The statistical code length is described by the Shannon entropy of μ with respect to $\mathcal{P}$ [16]:

$$h(\mu;\mathcal{P}) = \sum_{P \in \mathcal{P}} \mu(P) \log_2 \mu(P).$$

Example 1. *Suppose that we want to encode numbers picked randomly from* $[0, n)$*. One can use a coding partition with equally-sized sets, e.g.* $\mathcal{P}_\delta = \{[k, k+\delta) : k = 0, \delta, \dots, n-\delta\}$*. When the source elements are encoded by the centers of these sets, then (de)coding error does not exceed* $\delta/2$*. Clearly, one may construct partitions which contain different types of sets. Roughly speaking, highly probable elements should be coded with high accuracy (smaller sets) while the rare numbers can be coded with low precision (larger sets). DjVu is an example of a file format which uses different precision for various image elements - it compresses the text layer and the background separately [6]. Proposed approach allows to define a maximal coding error for every data set element separately.*

To control the maximal coding error in the above formulation we propose to use an additional family $\mathcal{Q}$ of subsets of X which we call an error-control family. The error-control family is a kind of a fidelity criterion [9]. We accept only these coding partitions $\mathcal{P}$ where every element is a subset of some element of $\mathcal{Q}$ (we say that $\mathcal{P}$ is $\mathcal{Q}$-acceptable). The optimal $\mathcal{Q}$-acceptable coding partition is the one with the minimal entropy. Thus we define the entropy of an error-control family $\mathcal{Q}$ by [17–19]:

$$H(\mu;\mathcal{Q}) := \inf\{h(\mu;\mathcal{P}) \in [0,\infty] : \mathcal{P} \text{ is } \mathcal{Q}\text{-acceptable partition}\}. \tag{1}$$

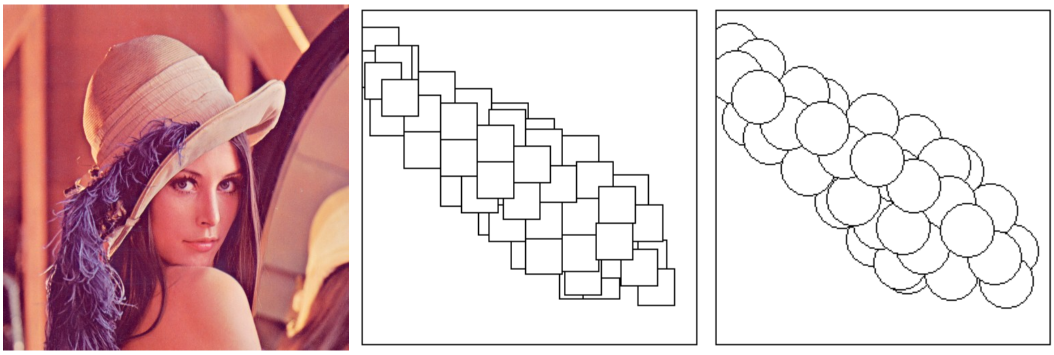
The above problem differs from the rate distortion theory [2,12] which is the most common approach to lossy source coding. Instead of specifying an upper limit on the allowed average distortion rate, an upper limit on the permitted distortion for any symbol is considered. The above formulation can be seen as a kind of vector quantization [4,11] and was partially motivated by the notion of epsilon-entropy proposed by Posner [13,14]. The entropy of specific error-control families in the case of metric spaces appears also in the definition of Rényi entropy dimension [15]. It is worth to mention that our idea is partially connected with perceptual source coding considered by Jayant et al. [8].

The answer to the question raised at the beginning of the paper concerning the lowest code length given a maximal coding error is equivalent to the calculation of the entropy of an error-control family (1). In this paper we focus on methods which allow to approximate this quantity.

1.2. Main results

In our main result, Theorem 12, we present a method which allows to approximate the entropy of an error-control family within an additive term. More precisely, we propose fast and easy to implement algorithm, Greedy Entropy Algorithm, which for a given finite error-control family $\mathcal{Q}$ produces a $\mathcal{Q}$-acceptable partition $\mathcal{P}$ satisfying:

$$h(\mu;\mathcal{P}) \le H(\mu;\mathcal{Q}) + \log_2 e. \tag{2}$$



(a) **(b)** **(c)**

Figure 1. Input image for compression (1a) and partitions produced by Greedy Entropy Algorithm for two cases of error-control families: the first one (1b) consists of cubes with a given side length while the second (1c) contains balls with a given radius.

Obtained bound is sharp and cannot be improved. Moreover, it is independent from a coding problem instance characterized by a probability measure μ and an error-control family $\mathcal{Q}$.

Our method is a reminiscent of the procedure used to approximate the solution of the minimum entropy set cover problem (MESC) [3,7], where a similar bound was derived. Roughly speaking, MESC focuses on an optimization problem where one seeks for a partition[1] compatible with a fixed cover of a finite data set X with minimal entropy. A similar greedy algorithm was used for producing a partition with the entropy approximating the minimal entropy value. To be able to apply the results obtained for MESC, the precise relationships between these two problems were established. In particular, in Theorem 10 we show that the entropy of an error-control family equals the minimal entropy for a set cover.

Let us observe that our main minimization problem (1) is very complex since for most examples of error-control families there exists uncountable number of acceptable partitions (see Section 2.2). In consequence, an exhaustive search through all acceptable partitions cannot be done in practice. Second key part for establishing the inequality (2) was to show that the number of partitions relevant for the entropy calculation may be drastically reduced . In Theorem 3 we derive that to find the entropy of an error-control family, it is sufficient to consider only partitions constructed from the elements of σ-algebra generated by an error-control family.

1.3. Discussion

In order to demonstrate the effects of Greedy Entropy Algorithm we apply this procedure for image segmentation. For simplicity we assume that every pixel is represented by a 3-dimensional feature vector, i.e. the intensity of each color coordinate ranges between 0 and 255.

Let the error-control family $\mathcal{Q}_\delta^1$ consists of all cubes with a side length δ, i.e. $\mathcal{Q}_\delta^1 = \{[k, k+\delta)^3 : k = 0, 1, \ldots, 255 - k\}$. Greedy Entropy Algorithm selects sequentially the most probable cubes from the image histogram. The final partition $\mathcal{P}_\delta^1$, projected on two dimensions, is shown in the Figure 1b. Table

[1] In fact, it looks for an assignment $f : X \to \mathcal{Q}$ which can be seen as a special case of a partition.

Table 1. Entropies calculated for partitions $\mathcal{P}^1_\delta(1a), \mathcal{P}^2_\delta$ (1b) returned by Greedy Entropy Algorithm for two cases of error control families. The first one consists of cubes with a given side length while the second contains balls with a given radius. In each case the results are compared with the entropy of acceptable partition consisting of maximally-sized pairwise disjoint cubes $\mathcal{P}_\delta, \mathcal{P}_{\lfloor \frac{2\sqrt{3}}{3}\delta \rfloor}$ respectively.

(a)

δ	$h(\mu;\mathcal{P}_\delta)$	$h(\mu;\mathcal{P}^1_\delta)$
3	12.73	12.59
5	10.81	10.83
9	8.62	8.62
15	6.95	6.72

(b)

δ	$h(\mu;\mathcal{P}_{\lfloor \frac{2\sqrt{3}}{3}\delta \rfloor})$	$h(\mu;\mathcal{P}^2_\delta)$
5	14.11	12.25
9	10.81	9.87
17	8.62	7.23
25	7.15	5.82

1a presents the comparison between the entropies calculated for partition $\mathcal{P}^1_\delta$ and another $\mathcal{Q}^1_\delta$-acceptable partition $\mathcal{P}_\delta$ including all pairwise disjoint cubes with the side length $\delta > 0$. Surprisingly, both partitions gave similar entropies. This might be explained by the fact that $\mathcal{P}_\delta$ contains only the elements of the error-control family $\mathcal{Q}^1_\delta$.

Clearly, this is not always the case. To observe this, let the error-control family $\mathcal{Q}^2_\delta$ consists of all balls with radius δ. Figure 1c shows the partition $\mathcal{P}^2_\delta$ returned by Greedy Entropy Algorithm projected on two dimensional space. As in the previous case, we consider also a $\mathcal{Q}^2_\delta$-acceptable partition $\mathcal{P}_{\lfloor \frac{2\sqrt{3}}{3}\delta \rfloor}$ of maximally-sized pairwise disjoint cubes with fixed side length. We can see from the results placed in Table 1b that a greedy selection provided significantly lower entropy values.

1.4. Paper organization

The paper is organized as follows. In Section 2 we formulate our lossy source coding problem and show that the entropy optimization problem can be simplified by reducing the number of partitions which are irrelevant for entropy calculation. Section 3 contains our main result. First, the minimum entropy set cover problem is recalled and its relationship with our notion of the entropy is established. Next, we define a Greedy Entropy Algorithm and derive that it constructs a partition with the entropy close to the optimal one.

2. Entropy calculation

We start this section with establishing basic notations and definitions. Then, we present the main problem of this paper concerning the entropy calculation and show how to reduce its complexity by eliminating irrelevant coding partitions.

2.1. Lossy source coding and error-control families

Let us assume that (X, Σ, μ) is a probability space. In our formulation of lossy source coding we are interested in encoding elements of X produced by a probability measure μ by a countable number of symbols. The source code is determined by a partition of X which is a countable family of measurable, pairwise disjoint subsets of X such that

$$\mu(X \setminus \bigcup_{P \in \mathcal{P}} P) = 0.$$

More precisely, every element $x \in X$ is transformed into a code related with a unique $P \in \mathcal{P}$ such that $x \in P$. The statistical code length of arbitrary element of X in optimal coding scheme can be calculated by the entropy of $\mathcal{P}$ [16]:

Definition 1. *The entropy of a partition $\mathcal{P}$ is defined by:*

$$h(\mu; \mathcal{P}) := \sum_{P \in \mathcal{P}} \mathrm{sh}(\mu(P)),$$

where $\mathrm{sh} : [0, 1] \to [0, \infty)$ *is the Shannon function, i.e.*

$$\mathrm{sh}(x) := \begin{cases} -x \cdot \log_2(x) & \text{for } x \in (0, 1], \\ 0 & \text{for } x = 0. \end{cases}$$

Although the entropy and the partition depend strictly on the probability measure, we consequently omit the symbol μ in their definitions to simplify notations.

The use of partition causes a coding error (distortion). To be able to control the maximal coding error (upper limit of permitted distortion) an additional family $\mathcal{Q}$ of subsets of X is introduced which we call an error-control family [18]. The error-control family restricts the number of permissible partitions which can be used for encoding. More precisely, a partition $\mathcal{P}$ is said to be $\mathcal{Q}$-acceptable iff for every $P \in \mathcal{P}$ there exists $Q \in \mathcal{Q}$ such that $P \subset Q$ (which we write $\mathcal{P} \prec \mathcal{Q}$).

The partitions which are allowed to be used for encoding are limited by a fixed error-control family. The optimal lossy coding scheme (determined by a partition) is the one which minimizes the entropy and does not violate the upper limit of permitted distortion, i.e. $\mathcal{P} \prec \mathcal{Q}$. This leads to the following definition of the entropy of an error-control family [18].

Definition 2. *Let $\mathcal{Q} \subset \Sigma$ be an error-control family. The entropy of $\mathcal{Q}$ is defined by:*

$$H(\mu; \mathcal{Q}) := \inf\{h(\mu; \mathcal{P}) \in [0, \infty] \, : \, \mathcal{P} \text{ is a partition and } \mathcal{P} \prec \mathcal{Q}\}. \tag{3}$$

In this paper we focus on the computational methods for finding the value of $H(\mu; \mathcal{Q})$ and possibly a partition $\mathcal{P} \prec \mathcal{Q}$ which satisfies $h(\mu; \mathcal{P}) \approx H(\mu; \mathcal{Q})$[2].

2.2. Partition reduction

In order to find the entropy of an error-control family a minimization problem (3) has to be solved. Let us first observe that for very simple error-control families the number of acceptable

[2] One should look for a partition $\mathcal{P}$ satisfying $h(\mu; \mathcal{P}) = H(\mu; \mathcal{Q})$. However, Example [18, II.1] shows that the value of entropy does not have to be attained on any partition in general.

partitions can be uncountable. Given a family $\mathcal{Q} = \{(-\infty, 1], [0, +\infty)\}$ any partition of the form $\mathcal{P} = \{(-\infty, a], (a, \infty)\}$, where $a \in (0, 1)$, is $\mathcal{Q}$-acceptable. Clearly, some of them do not lead to the optimal solution. In consequence, it is extremely important to eliminate partitions which are irrelevant for entropy calculation (3).

The main result of this section shows that to find the entropy it is sufficient to consider only partitions constructed from the sets of $\Sigma_{\mathcal{Q}}$ – the σ-algebra generated by the error-control family $\mathcal{Q}$. In the aforementioned example there are only 3 such partitions: $\mathcal{P}_1 = \{(-\infty, 0], (0, +\infty)\}$, $\mathcal{P}_2 = \{(-\infty, 1], (1, +\infty)\}$ and $\mathcal{P}_3 = \{(-\infty, 0], (0, 1], (1, +\infty)\}$.

Theorem 3. *Let (X, Σ, μ) be a probability space and let $\mathcal{Q}$ be an error-control family. Then, we have:*

$$H(\mu; \mathcal{Q}) := \inf\{h(\mu; \mathcal{P}) \in [0, \infty] : \mathcal{P} \text{ is a partition, } \mathcal{P} \prec \mathcal{Q} \text{ and } \mathcal{P} \subset \Sigma_{\mathcal{Q}}\}.$$

To derive this fact, for any $\mathcal{Q}$-acceptable partition $\mathcal{P}$, we will construct a partition $\mathcal{R} \subset \Sigma_{\mathcal{Q}}$ with the entropy not greater than $h(\mu; \mathcal{P})$. To describe the process of construction of such a partition let us first establish the notation: for a given family $\mathcal{Q}$ of subsets of X and a set $A \subset X$, we denote:

$$\mathcal{Q}_A = \{Q \cap A : Q \in \mathcal{Q}\}. \tag{4}$$

Then, for a $\mathcal{Q}$-acceptable partition $\mathcal{P}$, a family $\mathcal{R}$ is built by the following algorithm:

Partition Reduction Algorithm

---

initialization
  $i := 0$
  $X_0 := X$
  $\mathcal{R} := \emptyset$
while $\mu(X_i) > 0$ **do**
  Let $P^i_{\max} \in \mathcal{P}_{X_i}$ *be such that*
    $\mu(P^i_{\max}) = \max\{\mu(P) : P \in \mathcal{P}_{X_i}\}$
  Let $R_i \in \mathcal{Q}_{X_i}$ *be an arbitrary set*
    which satisfies $P^i_{\max} \subset R_i$
  $\mathcal{R} = \mathcal{R} \cup \{R_i\}$
  $X_{i+1} := X \setminus (R_1 \cup \ldots \cup R_i)$
  $i := i + 1$
end while
return $\mathcal{R}$

Our goal is to show that Partition Reduction Algorithm produces a partition $\mathcal{R}$ such that

$$h(\mu; \mathcal{R}) \leq h(\mu; \mathcal{P}).$$

We need to observe that the following property of the Shannon function holds.

Proposition 4. *Given numbers $p \geq q \geq 0$ and $r > 0$ such that $p, q, p + r, q - r \in [0, 1]$, we have:*

$$\mathrm{sh}(p) + \mathrm{sh}(q) \geq \mathrm{sh}(p + r) + \mathrm{sh}(q - r).$$

Let us focus on a single iteration of the Partition Reduction Algorithm.

Lemma 5. *Let (X, Σ, μ) be a subprobability space*[3]*. We consider an error-control family $\mathcal{Q}$ and a $\mathcal{Q}$-acceptable partition $\mathcal{P}$ of X. Let $P_{\max} \in \mathcal{P}$ be such that:*

$$\mu(P_{\max}) = \max\{\mu(P) : P \in \mathcal{P}\}.$$

If $Q \in \mathcal{Q}$ is chosen so that $P_{\max} \subset Q$ then

$$h(\mu; \{Q\} \cup \mathcal{P}_{X \setminus Q}) \leq h(\mu; \mathcal{P}). \tag{5}$$

Proof. Clearly, if $h(\mu; \mathcal{P}) = \infty$ then the inequality (5) holds trivially. Thus we assume that $h(\mu; \mathcal{P}) < \infty$.

Let us observe that it is enough to consider only elements of $\mathcal{P}$ with non zero measures – the number of such sets can be at most countable. Thus, let us assume that $\mathcal{P} = \{P_i\}_{i=1}^{\infty}$ (the case when $\mathcal{P}$ is finite can be treated in similar manner).

For a simplicity we put $P_1 := P_{\max}$. For every $k \in \mathbb{N}$, we consider the sequence of sets, defined by

$$Q_k := \bigcup_{i=1}^{k} (P_i \cap Q).$$

Clearly, for $k \in \mathbb{N}$, we have

$$Q_1 = P_1,$$

$$Q_k \subset Q_{k+1},$$

$$P_i \cap Q_k = P_i \cap Q, \text{ for } i \leq k, \tag{6}$$

$$P_i \cap Q_k = \emptyset, \text{ for } i > k, \tag{7}$$

$$\lim_{n \to \infty} \mu(Q_n) = \mu(Q). \tag{8}$$

To complete the proof it is sufficient to derive that for every $k \in \mathbb{N}$, we have:

$$h(\mu; \{Q_k\} \cup \mathcal{P}_{X \setminus Q_k}) \geq h(\mu; \{Q_{k+1}\} \cup \mathcal{P}_{X \setminus Q_{k+1}}) \tag{9}$$

and

$$h(\mu; \{Q_k\} \cup \mathcal{P}_{X \setminus Q_k}) \geq h(\mu; \{Q\} \cup \mathcal{P}_{X \setminus Q}). \tag{10}$$

Let $k \in \mathbb{N}$ be arbitrary. Then from (6) and (7), we get

$$h(\mu; \{Q_k\} \cup \mathcal{P}_{X \setminus Q_k}) = \mathrm{sh}(\mu(Q_k)) + \sum_{i=2}^{\infty} \mathrm{sh}(\mu(P_i \setminus Q_k))$$

$$= \mathrm{sh}(\mu(Q_k)) + \sum_{i=2}^{k} \mathrm{sh}(\mu(P_i \setminus Q)) + \sum_{i=k+1}^{\infty} \mathrm{sh}(\mu(P_i))$$

---

[3] (X, Σ) is measurable space and μ is a non-negative measure on (X, Σ) such that $\mu(X) \leq 1$

$$= h(\mu; \{Q_{k+1}\} \cup \mathcal{P}_{X\setminus Q_{k+1}}) + \mathrm{sh}(\mu(Q_k)) - \mathrm{sh}(\mu(Q_{k+1}))$$
$$+\mathrm{sh}(\mu(P_{k+1})) - \mathrm{sh}(\mu(P_{k+1} \setminus Q)).$$

Making use of Observation 4, we obtain

$$\mathrm{sh}(\mu(Q_k)) + \mathrm{sh}(\mu(P_{k+1}))$$
$$\geq \mathrm{sh}(\mu(Q_{k+1})) + \mathrm{sh}(\mu(P_{k+1} \setminus Q)),$$

which proves (9).

To derive (10), we first use inequality (9). Then

$$h(\mu; \{Q_k\} \cup \mathcal{P}_{X\setminus Q_k}) = \mathrm{sh}(\mu(Q_k)) + \sum_{i=1}^{\infty} \mathrm{sh}(\mu(P_i \setminus Q_k))$$
$$\geq \lim_{n\to\infty} [\mathrm{sh}(\mu(Q_n)) + \sum_{i=1}^{\infty} \mathrm{sh}(\mu(P_i \setminus Q_n))].$$

By (8),

$$\lim_{n\to\infty} \mathrm{sh}(\mu(Q_n)) = \mathrm{sh}(\mu(Q)) < \infty.$$

To calculate $\lim\limits_{n\to\infty} \sum_{i=1}^{\infty} \mathrm{sh}(\mu(P_i \setminus Q_n))$, we will use the Lebesgue Theorem [10]. We consider a sequence of functions

$$f_n : \mathcal{P} \ni P \to \mathrm{sh}(\mu(P \setminus Q_n)) \in \mathbb{R}, \text{ for } n \in \mathbb{N}.$$

Let us observe that the Shannon function sh is increasing on $[0, 2^{-\frac{1}{\ln 2}}]$ and decreasing on $(2^{-\frac{1}{\ln 2}}, 1]$. Thus for a certain $m \in \mathbb{N}$,

$$\mathrm{sh}(\mu(P_i \setminus Q_n)) \leq 1, \text{ for } i \leq m$$

and

$$\mathrm{sh}(\mu(P_i \setminus Q_n)) \leq \mathrm{sh}(\mu(P_i)), \text{ for } i > m,$$

for every $n \in \mathbb{N}$. Since $h(\mu; \mathcal{P}) < \infty$ then

$$\sum_{i=1}^{\infty} \mathrm{sh}(\mu(P_i \setminus Q_n)) \leq m + \sum_{i=m+1}^{\infty} \mathrm{sh}(\mu(P_i)) < \infty.$$

Moreover,

$$\lim_{n\to\infty} \mathrm{sh}(\mu(P \setminus Q_n)) = \mathrm{sh}(\mu(P \setminus Q)),$$

for every $P \in \mathcal{P}$.

As the sequence of functions $\{\mathrm{sh}(\mu(P \setminus Q_n))\}_{n\in\mathbb{N}}$ satisfies the assumptions of the Lebesgue Theorem [10] then, we get

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} \mathrm{sh}(\mu(P_i \setminus Q_n)) = \sum_{i=1}^{\infty} \lim_{n\to\infty} \mathrm{sh}(\mu(P_i \setminus Q_n))$$
$$= \sum_{i=1}^{\infty} \mathrm{sh}(\mu(P_i \setminus Q)) < \infty.$$

Consequently, we have

$$h(\mu; \{Q_k\} \cup \mathcal{P}_{X \setminus Q_k}) \geq \lim_{n\to\infty} [\mathrm{sh}(\mu(Q_n)) + \sum_{i=1}^{\infty} \mathrm{sh}(\mu(P_i \setminus Q_n))]$$

$$= \mathrm{sh}(\mu(Q)) + \sum_{i=1}^{\infty} \mathrm{sh}(\mu(P_i \setminus Q)) = h(\mu; \{Q\} \cup \mathcal{P}_{X \setminus Q}),$$

which completes the proof. □

Making use of the above lemma we summarize the analysis of Partition Reduction Algorithm in the following theorem.

Theorem 6. *We assume that (X, Σ, μ) is a subprobability space. Let $\mathcal{Q}$ be an error-control family on X and let $\mathcal{P}$ be a $\mathcal{Q}$-acceptable partition of X. A family $\mathcal{R}$ constructed by the Partition Reduction Algorithm is a partition of X and satisfies:*

$$h(\mu; \mathcal{R}) \leq h(\mu; \mathcal{P}). \tag{11}$$

Proof. Directly from the Partition Reduction Algorithm, we get that $\mathcal{R}$ is countable family of pairwise disjoint sets. The fact that

$$\mu(X \setminus \bigcup_{R \in R} R) = 0,$$

follows from the Lebesgue Theorem [10] applied to a sequence of functions $f_n : \mathcal{P} \to \mathbb{R}$ defined by

$$f_n(P) := \mu(P \setminus \bigcup_{i=1}^{n} R_i), \text{ for } P \in \mathcal{P}.$$

We prove the inequality (11). If $h(\mu; \mathcal{P}) = \infty$ then the inequality (11) is straightforward. Thus, let us discuss the case when $h(\mu; \mathcal{P}) < \infty$.

We denote $\mathcal{P} = \{P_i\}_{i=1}^{\infty}$, since at most countable number of elements of partition can have positive measure (the case when $\mathcal{P}$ is finite follows similarly). We will use the notation introduced in Partition Reduction Algorithm.

Directly from Lemma 5, we obtain

$$h(\mu; \mathcal{P}_{X_k}) \geq h(\mu; \mathcal{P}_{X_{k+1}} \cup \{R_k\}), \text{ for } k \in \mathbb{N}.$$

Consequently, for every $k \in \mathbb{N}$, we get

$$h(\mu; \bigcup_{i=1}^{k} \{R_i\} \cup \mathcal{P}_{X_k}) \geq h(\mu; \bigcup_{i=1}^{k+1} \{R_i\} \cup \mathcal{P}_{X_{k+1}}). \tag{12}$$

Our goal is to show that

$$h(\mu; \bigcup_{i=1}^{k} \{R_i\} \cup \mathcal{P}_{X_k}) \geq h(\mu; \mathcal{R}),$$

for every $k \in \mathbb{N}$.

Making use of (12), we have

$$h(\mu; \bigcup_{i=1}^{k}\{R_i\} \cup \mathcal{P}_{X_k})$$

$$= \sum_{i=1}^{k} \mathrm{sh}(\mu(R_i)) + \sum_{i=1}^{\infty} \mathrm{sh}(\mu(P_i \setminus \bigcup_{j=1}^{k} R_j))$$

$$\geq \lim_{n\to\infty} [\sum_{i=1}^{n} \mathrm{sh}(\mu(R_i)) + \sum_{i=1}^{\infty} \mathrm{sh}(\mu(P_i \setminus \bigcup_{j=1}^{n} R_j))],$$

for every $k \in \mathbb{N}$.

We will calculate $\lim\limits_{n\to\infty} \sum\limits_{i=1}^{\infty} \mathrm{sh}(\mu(P_i \setminus \bigcup\limits_{j=1}^{n} R_j))$ using the Lebesgue Theorem [10] for a sequence of functions $\{f_n\}_{n=1}^{\infty}$, defined by

$$f_n : \mathcal{P} \ni P \to \mathrm{sh}(\mu(P \setminus \bigcup_{j=1}^{n} R_j)) \in \mathbb{R}, \text{ for } n \in \mathbb{N}.$$

Similarly to the proof of Lemma 5, we may assume that there exists $m \in \mathbb{N}$ such that

$$\mathrm{sh}(\mu(P_i \setminus \bigcup_{j=1}^{n} R_j)) < 1, \text{ for } i \leq m$$

and

$$\mathrm{sh}(\mu(P_i \setminus \bigcup_{j=1}^{n} R_j)) < \mathrm{sh}(\mu(P_i)), \text{ for } i > m,$$

for every $n \in \mathbb{N}$. Moreover,

$$\lim_{n\to\infty} \mathrm{sh}(\mu(P \setminus \bigcup_{j=1}^{n} R_j)) = \mathrm{sh}(\mu(P \setminus \bigcup_{j=1}^{\infty} R_j)) = 0,$$

for every $P \in \mathcal{P}$ since $\mathcal{R}$ is a partition of X.

Making use of the Lebesgue Theorem [10], we get

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} \mathrm{sh}(\mu(P_i \setminus \bigcup_{j=1}^{n} R_j)) = \sum_{i=1}^{\infty} \mathrm{sh}(\mu(P_i \setminus \bigcup_{j=1}^{\infty} R_j)) = 0.$$

Consequently, for every $k \in \mathbb{N}$, we have

$$h(\mu; \bigcup_{i=1}^{k}\{R_i\} \cup \mathcal{P}_{X_k})$$

$$\geq \lim_{n\to\infty} [\sum_{i=1}^{n} \mathrm{sh}(\mu(R_i)) + \sum_{i=1}^{\infty} \mathrm{sh}(\mu(P_i \setminus \bigcup_{j=1}^{n} R_j))]$$

$$= \sum_{i=1}^{\infty} \mathrm{sh}(\mu(R_i)) = h(\mu; \mathcal{R}),$$

which completes the proof. □

As a consequence of Theorem 6, we directly get that Theorem 3 holds. In the case of finite error-control families we get that there exists an acceptable partition with minimal entropy.

Corollary 7. *Let (X, Σ, μ) be a probability space and let $\mathcal{Q}$ be a finite error-control family. Then, there exists a $\mathcal{Q}$-acceptable partition $\mathcal{P} \subset \Sigma_{\mathcal{Q}}$ such that:*

$$H(\mu; \mathcal{Q}) = h(\mu; \mathcal{P}).$$

3. Entropy approximation

In previous section we simplified the problem of entropy calculation by reducing the number of partitions which are necessary to consider to find the entropy of an error-control family. Since the number of acceptable partitions grows exponentially with the cardinality of error-control family, it might be impossible to test all of them for entropy calculation. In this section we show an algorithm which allows to approximate the entropy within an additive term.

Presented formulation of lossy source coding is closely related with the minimum entropy set cover problem (MESC) [3,7], where one focuses on a similar optimization problem. There exists an algorithm for approximation the solution of MESC within an additive term of $\log_2 e$. First, we present a description of MESC and its relationship with introduced coding problem. Next, we use these facts to apply a similar technique for approximating the entropy of an error-control family.

3.1. Relationship with the minimum entropy set cover problem

In order to define MESC let $X = \{x_1, \ldots, x_n\}$ be a finite data set, where every observation $x \in X$ appears with a probability p_x. A random source produces a signal from a probability distribution $\{p_{x_1}, \ldots, p_{x_n}\}$ and passes it through the noisy channel. Each observation has a type but due to the noise we only know that it is one of a given set of types defined by a finite cover $\mathcal{Q} = \{Q_1, \ldots, Q_k\}$ of data space X. We map an observation to a type by defining an assignment $f^{\mathcal{Q}} : X \to \mathcal{Q}$ which is compatible with $\mathcal{Q}$, i.e. $x \in f^{\mathcal{Q}}(x)$ for all $x \in X$.

Let us denote by q_i the probability that the random point is assigned to Q_i:

$$q_i := \sum_{x \in f^{-1}(Q_i)} p_x, \text{ for } i = 1, \ldots, k.$$

The goal is to find such an assignment which minimizes the entropy of the distribution of the types, i.e.

$$h(f^{\mathcal{Q}}) := \sum_{i=1}^{k} \mathrm{sh}(q_i).$$

Such an optimal assignment is denoted by $f^{\mathcal{Q}}_{opt}$.

MESC is NP-hard problem [7, Theorem 1]. To find an assignment which approximates efficiently the minimal entropy value, a simple greedy algorithm can be used (which we call Greedy MESC Algorithm). It relies on iterative execution of the following steps:

- choose the most probable type $Q^i_{\max} \in \mathcal{Q}$,

- if $x \in Q^i_{\max}$, then assign x to $Q^i_{\max}$, i.e. put $f^{\mathcal{Q}}(x) := Q^i_{\max}$,
- remove from X (and from all $Q \in \mathcal{Q}$) the elements of $Q^i_{\max}$,

until there exists $Q \in \mathcal{Q}$ with positive probability [7].

J. Cardinal et al. proved a sharp bound for the entropy of assignment constructed with use of Greedy MESC Algorithm:

Greedy MESC Approximation (see [3, Theorem 1]) *If $f^{\mathcal{Q}}_g$ is an assignment produced by the greedy algorithm then*

$$h(f^{\mathcal{Q}}_g) \leq h(f^{\mathcal{Q}}_{opt}) + \log_2 e.$$

To be able to obtain a similar approximation of the entropy of an error-control family, the relationship between MESC and our formulation of entropy has to be established. For this purpose our lossy source coding problem will be considered on a discrete probability space (X, Σ, μ), where X is a finite set, Σ is a σ-algebra generated by all singletons of X and $\mu := \sum\limits_{x \in X} p_x \delta_x$ is an atomic measure on (X, Σ). A cover $\mathcal{Q} = \{Q_1, \ldots, Q_k\}$ plays a role of an error-control family.

We start our analysis with showing that given an assignment $f^{\mathcal{Q}}$ compatible with $\mathcal{Q}$ one can construct a $\mathcal{Q}$-acceptable partition with equal entropy:

Lemma 8. *Let $f^{\mathcal{Q}}$ be an assignment compatible with $\mathcal{Q}$. Then, the family $\mathcal{P} = \{P_i\}_{i=1}^k$, where $P_i := (f^{\mathcal{Q}})^{-1}(Q_i)$, is a $\mathcal{Q}$-acceptable partition, and*

$$h(f^{\mathcal{Q}}) = h(\mu; \mathcal{P}).$$

Proof. Directly from the definition of compatible assignment, we get that $\mathcal{P}$ is a $\mathcal{Q}$-acceptable partition of X.

Moreover, let us observe that

$$h(f^{\mathcal{Q}}) = \sum_{i=1}^{k} \mathrm{sh}(q_i) = \sum_{i=1}^{k} \mathrm{sh}(\sum_{x \in f^{-1}(Q_i)} p_x)$$

$$= \sum_{i=1}^{k} \mathrm{sh}(\sum_{x \in P_i} p_x) = \sum_{i=1}^{k} \mathrm{sh}(\mu(P_i)) = h(\mu; \mathcal{P}).$$

□

The following example illustrates that the natural inverse construction is not possible.

Example 2. *Let (X, Σ, μ) be a probability space with an error-control family $\mathcal{Q}$, where:*

$$X = \{0, 1\},$$

$$\Sigma = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\},$$

$$\mu(\{0\}) = \mu(\{1\}) = \frac{1}{2},$$
$$\mathcal{Q} = \{\{0,1\}\}.$$

Then

$$\mathcal{P} = \{\{0\},\{1\}\}$$

is a $\mathcal{Q}$-acceptable partition with the entropy equals 1*. However, the only assignment $f^{\mathcal{Q}} : X \to \mathcal{Q}$ compatible with $\mathcal{Q}$ is defined by*

$$f^{\mathcal{Q}}(0) = f^{\mathcal{Q}}(1) = \{0,1\},$$

which entropy equals 0*.*

The following result demonstrates that given a partition one can find an assignment with not greater entropy:

Lemma 9. *Let $\mathcal{P} \prec \mathcal{Q}$ be a partition. Then, there exists an assignment $f^{\mathcal{Q}}$ compatible with $\mathcal{Q}$ such that*

$$h(f^{\mathcal{Q}}) \leq h(\mu;\mathcal{P}).$$

Proof. Since $\mathcal{P}$ is a partition, the function

$$g : X \ni x \to P_x \in \mathcal{P}, \text{ for } x \in P_x,$$

is well defined. Moreover, as $\mathcal{P}$ is $\mathcal{Q}$-acceptable, we find a mapping $h : \mathcal{P} \to \mathcal{Q}$, such that

$$h(P) = Q, \text{ if } P \subset Q.$$

Finally, we put an assignment $f^{\mathcal{Q}} : X \to \mathcal{Q}$, by

$$f^{\mathcal{Q}} = h \circ g.$$

Clearly, $f^{\mathcal{Q}}$ is an assignment compatible with $\mathcal{Q}$. Let us calculate the entropy of $f^{\mathcal{Q}}$. We have

$$h(f^{\mathcal{Q}}) = \sum_{Q\in\mathcal{Q}} \mathrm{sh}(\sum_{x\in(f^{\mathcal{Q}})^{-1}(Q)} p_x)$$

$$= \sum_{Q\in\mathcal{Q}} \mathrm{sh}(\sum_{x\in g^{-1}(h^{-1}(Q))} p_x) = \sum_{Q\in\mathcal{Q}} \mathrm{sh}(\sum_{P\in h^{-1}(Q)} \sum_{x\in g^{-1}(P)} p_x).$$

Making use of the subadditivity of Shannon function, we get

$$\sum_{Q\in\mathcal{Q}} \mathrm{sh}(\sum_{P\in h^{-1}(Q)} \sum_{x\in g^{-1}(P)} p_x) \leq \sum_{Q\in\mathcal{Q}} \sum_{P\in h^{-1}(Q)} \mathrm{sh}(\sum_{x\in g^{-1}(P)} p_x)$$

$$= \sum_{P\in\mathcal{P}} \mathrm{sh}(\sum_{x\in g^{-1}(P)} p_x) = \sum_{P\in\mathcal{P}} \mathrm{sh}(\sum_{x\in P} p_x) = h(\mu;\mathcal{P}).$$

□

As a consequence, we get that the entropy of optimal assignment equals the entropy of an error-control family.

Theorem 10. *We have*

$$h(f^{\mathcal{Q}}_{opt}) = H(\mu;\mathcal{Q}).$$

Proof. If $f^{\mathcal{Q}}_{opt}$ is an optimal assignment, then making use of Lemma 8, we construct a $\mathcal{Q}$-acceptable partition $\mathcal{P}$ which satisfies:

$$h(f^{\mathcal{Q}}_{opt}) = h(\mu;\mathcal{P}).$$

On the other hand, since $\mathcal{Q}$ is a finite error-control family, then from Theorem 3 we get

$$H(\mu;\mathcal{Q}) = h(\mu;\mathcal{P}),$$

for a specific $\mathcal{Q}$-acceptable partition $\mathcal{P}$. Using Lemma 9, we find an assignment $f^{\mathcal{Q}}$ compatible with $\mathcal{Q}$ such that

$$h(\mu;\mathcal{P}) \geq h(f^{\mathcal{Q}}),$$

which completes the proof. □

3.2. Greedy approximation

In this section we show that the analogue of Greedy MESC Algorithm can be applied for the case of our formulation of lossy source coding. Furthermore similar bounds can be established.

Let us start with a modified version of approximation algorithm which we call Greedy Entropy Algorithm. We assume that $\mathcal{Q}$ is a finite error-control family.

Greedy Entropy Algorithm

---

initialization
 $i := 0$
 $\mathcal{P} := \emptyset$
 $\mathcal{Q}_0 := \mathcal{Q}$
while $\mu(\bigcup_{Q\in\mathcal{Q}_i} Q) > 0$ **do**
 Let $Q^i_{\max} \in \mathcal{Q}_i$ *be such that*
  $\mu(Q^i_{\max}) = \max\{\mu(Q) : Q \in \mathcal{Q}_i\}$
 $\mathcal{P} = \mathcal{P} \cup \{Q^i_{\max}\}$
 $\mathcal{Q}_{i+1} := \{Q \setminus Q^i_{\max} : Q \in \mathcal{Q}_i\}$
 $i := i+1$
end while
return $\mathcal{P}$

To see that Greedy Entropy Algorithm is not well defined for infinite error-control families, let us consider the example:

Example 3. *Let us consider an open segment* $(0,1)$ *with* σ*-algebra generated by all Borel subsets of* $(0,1)$*, Lebesgue measure* λ *and an error control family, defined by*

$$\mathcal{Q} = \{[a,b] : 0 < a < b < 1\}.$$

There does not exist the set of maximal measure from family $\mathcal{Q}$ – hence the Greedy Entropy Algorithm cannot be applied directly in a such case.

Let us observe that both greedy algorithms create partitions with the same entropies. For this purpose we denote by $\mathrm{Greedy}^f_{\mathcal{Q}}$ a set of all assignments produced by Greedy MESC Algorithm while by $\mathrm{Greedy}_{\mathcal{Q}}$ we denote a set of all partitions returned by Greedy Entropy Algorithm:

Proposition 11. *We have:*

- *For every $f^{\mathcal{Q}}_g \in \mathrm{Greedy}^f_{\mathcal{Q}}$, there exists $\mathcal{P}_g \in \mathrm{Greedy}_{\mathcal{Q}}$ such that:*

$$h(f^{\mathcal{Q}}_g) = h(\mu; \mathcal{P}_g).$$

- *For every $\mathcal{P}_g \in \mathrm{Greedy}_{\mathcal{Q}}$, there exists $f^{\mathcal{Q}}_g \in \mathrm{Greedy}^f_{\mathcal{Q}}$ such that:*

$$h(\mu; \mathcal{P}_g) = h(f^{\mathcal{Q}}_g).$$

The main result of this section shows that Greedy Entropy Algorithm produces a partition with the entropy not grater that the entropy of an error-control family.

Theorem 12. *Let (X, Σ, μ) be a probability space and let $\mathcal{Q}$ be a finite error-control family. Then:*

$$h(\mu; \mathcal{P}) \leq H(\mu; \mathcal{Q}) + \log_2 e \text{, for } \mathcal{P} \in \mathrm{Greedy}_{\mathcal{Q}}.$$

The proof of Theorem 12 involves two facts:

- To calculate the entropy it is sufficient to consider only partitions constructed from the elements of σ-algebra generated by $\mathcal{Q}$ (see Corollary 7).
- The calculation of the entropy of an error-control family is closely related with MESC optimization problem (see Theorem 10 and Proposition 11).

To be able to apply these facts we need an additional lemma:

Lemma 13. *Let (X, Σ, μ) be an arbitrary (not necessarily discrete) probability space and let $\mathcal{Q}$ be a finite error-control family. Then there exists a probability space $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu})$ with an error-control family $\tilde{\mathcal{Q}}$ such that $\tilde{X}$, $\tilde{\Sigma}$ and $\tilde{\mathcal{Q}}$ are finite, $\tilde{\mu}$ is an atomic measure,*

$$H(\mu; \mathcal{Q}) = H(\tilde{\mu}; \tilde{\mathcal{Q}})$$

and for every $\mathcal{P}_g \in \mathrm{Greedy}_{\mathcal{Q}}$ there exists $\tilde{\mathcal{P}}_g \in \mathrm{Greedy}_{\tilde{\mathcal{Q}}}$ satisfying:

$$h(\mu; \mathcal{P}_g) = h(\tilde{\mu}; \tilde{\mathcal{P}}_g).$$

Proof. We restrict our consideration to partitions $\mathcal{P} \subset \Sigma_{\mathcal{Q}}$ since, by Corollary 7, the entropy $H(\mu; \mathcal{Q})$ is attained on some partition generated by $\Sigma_{\mathcal{Q}}$. Let us denote the set of generators of $\Sigma_{\mathcal{Q}}$:

$$\begin{aligned}\mathrm{Gen}(\Sigma_{\mathcal{Q}}) := \ & \{G \in \Sigma_{\mathcal{Q}} : G \neq \emptyset, \text{ and for all} \\ & Q' \in \mathcal{Q} : G \subset Q' \text{ or } G \cap Q' = \emptyset\}.\end{aligned}$$

Next, for every set $G \in \mathrm{Gen}(\Sigma_{\mathcal{Q}})$, we fix exactly one point $x_G \in G$.

Then, we obtain a probability space $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu})$ and an error-control family $\tilde{\mathcal{Q}}$ by:

$$\tilde{X} = \{x_G\}_{G \in \mathrm{Gen}(\Sigma_{\mathcal{Q}})},$$

$$\tilde{\mathcal{Q}} := \{ \bigcup_{G \in \mathrm{Gen}(\Sigma_{\mathcal{Q}}), G \subset Q} \{x_G\}\}_{Q \in \mathcal{Q}}$$

$$\tilde{\Sigma} = \Sigma_{\tilde{\mathcal{Q}}}$$

$$\tilde{\mu} = \sum_{G \in \mathrm{Gen}(\Sigma_{\mathcal{Q}})} \mu(G) \delta_{x_G}.$$

It is easy to see that every $\tilde{\mathcal{Q}}$-acceptable $\tilde{\mu}$-partition $\tilde{P} \subset \Sigma_{\tilde{\mathcal{Q}}}$ corresponds naturally to a specific $\mathcal{Q}$-acceptable partition $\mathcal{P} \subset \Sigma_{\mathcal{Q}}$ and conversely. The measures of corresponding sets are equal.

Thus

$$H(\mu; \mathcal{Q}) = H(\tilde{\mu}; \tilde{\mathcal{Q}}).$$

Moreover, for every $\mathcal{P}_g \in \mathrm{Greedy}_{\mathcal{Q}}$ there exists $\tilde{\mathcal{P}}_g \in \mathrm{Greedy}_{\tilde{\mathcal{Q}}}$ which satisfies:

$$h(\mu; \mathcal{P}_g) = h(\tilde{\mu}; \tilde{\mathcal{P}}_g).$$

□

Finally, the proof of our main result is as follows:

Proof. (of Theorem 12) Making use of Lemma 13, we find a probability space $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu})$ with the error-control family $\tilde{\mathcal{Q}}$ such that $\tilde{X}$, $\tilde{\Sigma}$, $\tilde{\mathcal{Q}}$ are finite, $\tilde{\mu}$ is an atomic measure and

$$H(\mu; \mathcal{Q}) = H(\tilde{\mu}; \tilde{\mathcal{Q}}).$$

By Theorem 10, we get that if $f^{\tilde{\mathcal{Q}}}_{opt}$ is an optimal assignment compatible with $\tilde{\mathcal{Q}}$ then

$$h(f^{\tilde{\mathcal{Q}}}_{opt}) = H(\tilde{\mu}, \tilde{\mathcal{Q}}).$$

Moreover, making use of Proposition 11, for every $\tilde{\mathcal{P}}_g \in \mathrm{Greedy}_{\tilde{\mathcal{Q}}}$, we find $f^{\tilde{\mathcal{Q}}}_g \in \mathrm{Greedy}^f_{\tilde{\mathcal{Q}}}$ such that

$$h(f^{\tilde{\mathcal{Q}}}_g) = h(\tilde{\mu}, \tilde{\mathcal{P}}_g).$$

Thus by Greedy MESC Approximation, we have:

$$h(\tilde{\mu}; \tilde{\mathcal{P}}_g) \leq H(\tilde{\mu}; \tilde{\mathcal{Q}}) + \log_2 e.$$

Consequently, using Lemma 13, we get

$$h(\mu; \mathcal{P}_g) \leq H(\mu; \mathcal{Q}) + \log_2 e.$$

□

The above approximation cannot be improved. To see this, we use the example from [3, Section 2] adopted to our situation. Moreover, it is also NP-hard to approximate the entropy within an additive term lower than $\log_2 e$ (see [3, Theorem 2]).

4. Conclusion

The paper focused on a non standard type of lossy source coding. In contrast to rate distortion theory, a cover of the source alphabet, which defines a maximal distortion permitted on every element, was introduced. The calculation of the entropy in such a formulation of lossy coding is equivalent to solving the minimum entropy optimization problem, where one would like to find a coding partition compatible with a fixed distortion with minimal entropy. Our results show how to simplify this optimization problem and find the approximated entropy value. Proposed algorithm is fast, feasible for implementation and produces a partition which has a proven upper bound on accuracy, i.e. the entropy of returned partition is not higher than $\log_2 e$ than the true entropy value.

Acknowledgments

This research was supported by the National Center of Science (Poland) Grants No. 2014/13/N/ST6/01832 and 2014/13/B/ST6/01792.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Berger, T. Lossy source coding. *IEEE Trans. Inf. Theory* **1998**, *44/6*, 2693–2723.
2. Berger, T. Rate-distortion theory; Encyclopedia of Telecommunications, Wiley Online Library, 1971.
3. Cardinal, J.; Fiorini, S.; Joret, G. Tight results on minimum entropy set cover. *Algorithmica* **2008**, *51/1*, 49–60.
4. Gray, R.M. Vector quantization. *IEEE ASSP Magazine* **1984**, *1/2*, 4–29.
5. Gray, R.M.; Neuhoff, D.L. Quantization. *IEEE Trans. Inf. Theory* **1998**, *44/6*, 2325–2383.
6. Haffner, P.; Howard, P.G.; Simard, P.; Bengio, Y.; Lecun, Y. and others, High quality document image compression with "DjVu". *Journal of Electronic Imaging* **1998**, *7/3*, 410–425.
7. Halperin, E.; Karp, R.M. The minimum-entropy set cover problem. *Thoer. Comput. Sci.* **2005**, *348/2*, 240–250.
8. Jayant, N.; Johnston, J.;d Safranek, R. Signal compression based on models of human perception. *Proceedings of IEEE* **1993**, *81/10*, 1385–1422.
9. Kieffer, J.C. A survey of the theory of source coding with a fidelity criterion. *IEEE Trans. Inf. Theory* **1993**, *39/5*, 1473–1490.
10. Kingman, J.F.C.; Taylor, S.J. Introduction to measures and probability; Cambridge University Press, 1966.
11. Nasrabadi, N.M.; King, R.A. Image coding using vector quantization: A review. *IEEE Trans. Commun.* **1998**, *36/8*, 957–971.
12. Ortega, A.; Ramchandran, K. Rate-distortion methods for image and video compression. *IEEE Signal Process Mag.* **1998**, *15/6*, 23–50.

13. Posner, E.C.; Rodemich, E.R. Epsilon entropy and data compression. *Ann. Math. Stat.* **1971**, 2079–2125.
14. Posner, E.C.; Rodemich, E.R.; Rumsey J.H. Epsilon entropy of stochastic processes. *Ann. Math. Stat.* **1967**, 1000-1020.
15. Rényi A. On the dimension and entropy of probability distributions. *Acta Math. Hungar.* **1959**, *10/1-2*, 193–215.
16. Shannon, C.E. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review* **2001**, *5/1*, 3–55.
17. Śmieja, M. Weighted approach to general entropy function. *IMA J. Math. Control Inf.* **2014**, *doi: 10.1093/imamci/dnt044*, 13.
18. Śmieja, M.; Tabor J. Entropy of the mixture of sources and entropy dimension. *IEEE Trans. Inf. Theory* **2012**, *58/5*, 2719–2728.
19. Śmieja, M.; Tabor, J. Rényi entropy dimension of the mixture of measures. *Science and Information Conference (SAI)* **2014**, 685–689.