Title
On coverings of ellipsoids in Euclidean spaces

Journal
IEEE Transactions on Information Theory, 50(10)

ISSN
0018-9448

Authors
Dumer, Ilya
Pinsker, M S
Prelov, V V

Publication Date
2004-10-01

Peer reviewed

On Coverings of Ellipsoids in Euclidean Spaces

Ilya Dumer, *Senior Member, IEEE*, Mark S. Pinsker, and Viacheslav V. Prelov

***Abstract*—The thinnest coverings of ellipsoids are studied in the Euclidean spaces of an arbitrary dimension n. Given any ellipsoid, the main goal is to find its ε-entropy, which is the logarithm of the minimum number of the balls of radius ε needed to cover this ellipsoid. A tight asymptotic bound on the ε-entropy is obtained for all but the most oblong ellipsoids, which have very high eccentricity. This bound depends only on the volume of the sub-ellipsoid spanned over all the axes of the original ellipsoid, whose length (diameter) exceeds 2ε.**

The results can be applied to vector quantization performed when data streams from different sources are bundled together in one block.

***Index Terms*—Covering, ellipsoid, entropy, Euclidean space, unit ball.**

I. Introduction

A. Spherical ε-Coverings of Ellipsoids

LET A be a subset of an n-dimensional Euclidean space $\boldsymbol{R}^n$ and

$$B_\varepsilon^n(y) \stackrel{\text{def}}{=} \left\{ x=(x_1,\ldots,x_n)\in \boldsymbol{R}^n \mid \sum_{i=1}^n (x_i-y_i)^2 \le \varepsilon^2 \right\}$$

be the ball of radius ε centered at some point $y \in \boldsymbol{R}^n$. Consider any subset $\mathcal{M}_\varepsilon(A) \subseteq \boldsymbol{R}^n$ and the union of the balls

$$\mathbb{M}_\varepsilon(A) = \cup_{y\in\mathcal{M}_\varepsilon(A)} B_\varepsilon^n(y)$$

centered at points $y \in \mathcal{M}_\varepsilon(A)$. We say that $\mathcal{M}_\varepsilon(A)$ is an ε-*covering*[1] of A if

$$A \subseteq \mathbb{M}_\varepsilon(A).$$

The ε-*entropy* [1] $H_\varepsilon(A)$ of a set A is the logarithm of the size of its minimal covering

$$H_\varepsilon(A) = \log \min |\mathcal{M}_\varepsilon(A)|$$

Manuscript received May 16, 2003; revised May 5, 2004. The work of I. Dumer was supported by the National Science Foundation under Grant CCR-0097125. The work of M. S. Pinsker and V. V. Prelov was supported in part by the National Science Foundation under Grant CCR-0097125, by INTAS under Grant 00-738, and by the Russian Foundation for Basic Research, Projects 03-01-00592 and 03-01-00098.

I. Dumer is with the College of Engineering, University of California, Riverside, CA 92521 USA (e-mail: dumer@ee.ucr.edu).

M. S. Pinsker, deceased, was with the Institute for Problems of Information Transmission, Russian Academy of Sciences, Moscow 127994, Russia.

V. V. Prelov is with the Institute for Problems of Information Transmission, Russian Academy of Sciences, Moscow 127994, Russia.

Communicated by S. Litsyn, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2004.834759

[1]Here we mostly follow coding terminology. In information theory, it is customary to say that $\mathcal{M}(A)$ is an ε-*net*.

where minimum is taken over all ε-coverings $\mathcal{M}_\varepsilon(A)$ and $\log(\cdot)$ is a natural logarithm.

In what follows, we study the ε-entropy of an arbitrary ellipsoid

$$E_a^n \stackrel{\text{def}}{=} \left\{ x=(x_1,\ldots,x_n)\in \boldsymbol{R}^n \mid \sum_{i=1}^n \frac{x_i^2}{a_i^2} \le 1 \right\} \tag{1}$$

where $a=(a_1,\ldots,a_n)$ is a vector with n positive symbols. Each symbol a_i gives half the length of the i-axis in E_a^n. Without loss of generality, we assume that symbols a_i form a nondecreasing sequence so that

$$0 < a_1 \le a_2 \le \cdots \le a_n.$$

Our main goal is to find the asymptotic ε-entropy $H_\varepsilon(E_a^n)$ as $n\to\infty$. More generally, we will base our study on a single requirement that $H_\varepsilon(E_a^n)\to\infty$ for all the sets of parameters n, ε, and a. Precise statements of our results are given in Section II and the corresponding proofs are relegated to Section III.

B. Preliminaries

Optimal ε-coverings have been long studied for an Euclidean ball $B_\rho^n \stackrel{\text{def}}{=} B_\rho^n(0)$. Various bounds on its minimum covering size are obtained in papers [2] and [3]. By linear transformation of $\boldsymbol{R}^n$, one can always replace any ε-covering of B_ρ^n using the *unit balls* to cover a ball $B_{\rho/\varepsilon}^n$. For this reason, we will mostly consider the unit balls and remove the index ε from the notation $\mathcal{M}_\varepsilon(A)$. One particularly important result is obtained by Rogers [2] who proved that for $n\ge 9$, the thinnest covering with unit balls has size

$$\left|\mathcal{M}\left(B_\rho^n\right)\right| \le \begin{cases} Cn(\log n)\rho^n, & \text{if } \rho\ge n\\ Cn^{5/2}\rho^n, & \text{if } \rho<n \end{cases} \tag{2}$$

where C is an absolute constant.

Note that any set A of infinite volume $V(A)=\infty$ also produces infinite coverings $\mathcal{M}(A)$. Therefore, for coverings of the whole space $A=\boldsymbol{R}^n$, it is customary [4] to first consider a sequence of balls B_ρ^n of growing radius ρ. Given a sequence of coverings $\mathcal{M}_\rho^n = \mathcal{M}\left(B_\rho^n\right)$, the *lower density* for this sequence is defined as the asymptotic infimum of the mean number of balls covering a point in B_ρ^n

$$\delta\left(\mathcal{M}_\rho^n\right) = \lim_{\rho\to\infty}\inf \frac{\sum_{y\in\mathcal{M}_\rho^n} V\left(B^n(y)\cap B_\rho^n\right)}{V\left(B_\rho^n\right)}.$$

The main problem is to define the minimum density $\delta = \inf\delta\left(\mathcal{M}_\rho^n\right)$ obtained in $\boldsymbol{R}^n$ over all coverings $\mathcal{M}_\rho^n$. Here, we refer to monograph [4], which gives a detailed account of the lower and upper bounds on δ along with an extensive bibliography on this subject. Detailed surveys are also presented in [5] and [6].

Coverings of other sets—different from the balls B_ρ^n and spaces $\boldsymbol{R}^n$—have also been studied for general convex bodies in Banach spaces (see [7], [8], and references therein). For the Hamming spaces, various coverings associated with codes are studied in the monograph [9], while coverings of the Hamming ellipsoids are considered in [10]–[12]. Later, in Sections II and III, we address a similar problem in the Euclidean spaces. We also compare our results with those obtained in [8] for convex bodies.

C. General Ellipsoidal Θ-Covering

Given a vector $\Theta = (\theta_1, \ldots, \theta_n)$ with n positive symbols, consider the ellipsoid

$$E_\Theta^n(y) \stackrel{\text{def}}{=} \left\{ x = (x_1, \ldots, x_n) \in \boldsymbol{R}^n \mid \sum_{i=1}^n \frac{(x_i - y_i)^2}{\theta_i^2} \leq 1 \right\}$$

centered at $y \in \boldsymbol{R}^n$. Given any (other) ellipsoid E_a^n, we say that a subset $\mathcal{M}_\Theta(E_a^n) \subseteq \boldsymbol{R}^n$ forms its *ellipsoidal* covering if E_a^n belongs to the union

$$\mathbb{M}_\Theta(E_a^n) = \cup \{ E_\Theta^n(y) \mid y \in \mathcal{M}_\Theta(E_a^n) \}$$

of ellipsoids $E_\Theta^n(y)$ with centers y running through $\mathcal{M}_\Theta(E_a^n)$. Note, however, that this setting is readily converted into the former spherical covering by linear transformation $x_i' = x_i/\theta_i$ for all points $x \in \boldsymbol{R}^n$. It is clear that in this case $E_\Theta^n(y)$ becomes the unit ball $B^n(y')$, while ellipsoid E_a^n is transformed into the ellipsoid E_b^n with the new set of axial coefficients $a_1/\theta_1, \ldots, a_n/\theta_n$. Thus, for generic ellipsoids E_a^n, we have three equivalent problems:

1) covering with unit balls;
2) ε-covering, with balls of radius ε;
3) Θ-covering, with ellipsoids $E_\Theta^n(y)$.

Due to this equivalence, we will mostly address the first problem and study the unit entropy of an ellipsoid E_a^n

$$\aleph(E_a^n) \stackrel{\text{def}}{=} H_1(E_a^n), \text{ which is the logarithm of the minimum number of unit balls needed to cover } E_a^n.$$

D. Possible Applications

It is well known [4] that sphere coverings often arise in multidimensional (vector) quantizers. In particular, suppose that n-dimensional data points $x \in \boldsymbol{R}^n$ have limited maximum energy ρ^2 and therefore fall into the ball B_ρ^n with some probability distribution $p(x)$. A typical quantizer/compressor is then supposed to establish a thin covering $\{y\}$ of the ball B_ρ^n while limiting the *mean* (squared) *rounding error* $(y-x)^2$. Zador's theorem [4] shows that this error can be reduced per dimension by using quantization in higher dimensions n. Thus, combining the data in longer blocks improves the quality of an overall quantization.

To simplify the problem, it is also customary to consider the uniform distribution of original data points x in the ball B_ρ^n. In this case, given the maximum rounding error ε caused by a quantizer, we need to find the thinnest ε-covering of the ball B_ρ^n. Thus, optimal quantization becomes closely related to the problem of a minimum sphere covering.

Suppose now that quantization is performed after a few different streams of data are mixed together in a block of length n. The above spherical framework is kept intact if the combined system still operates with a limited total energy. However, different sources S_i can incur different (power) costs a_i. Given the total cost, different sources S_i are accumulated with different factors a_i in this case, and the former ball B_ρ^n is replaced by some ellipsoid E_a^n.

Another example arises when some sources S_i generate more valuable data that have higher priorities θ_i. Given that the total energy is limited by ρ^2, the combined data still belong to the ball B_ρ^n. However, a quantizer may take into account different priorities θ_i, in which case rounding errors are also weighed differently. In this case, we have an inverse setting, when a Θ-covering of the ball B_ρ^n by small ellipsoids $E_\Theta^n(y)$ can become the model of choice.

Finally, suppose that a data block of length n is split and then transmitted over a few independent memoryless Gaussian channels S_i, which have different noise powers a_i^2. The received blocks represent Gaussian vectors in $\boldsymbol{R}^n$ that have different variances a_i^2 in different positions i. General ellipsoidal setting is again more applicable in this scenario, for both the received vector and its quantized version.

The above examples show that ellipsoidal ε-coverings can be used whenever data signals incur different costs or carry different priorities or get disturbed by a different amount of noise. The minimum covering size required for this ellipsoidal setting is the main subject of this paper.

II. Summary of the Results

Given an ellipsoid E_a^n, define the quantity $K = K_a^n$ as

$$K \stackrel{\text{def}}{=} \sum_{i=1}^n \log^+(a_i) = \sum_{i:a_i>1}^n \log a_i. \tag{3}$$

We begin with a lower bound on the unit entropy of an ellipsoid E_a^n, which holds for all dimensions n and vectors a. The proof is given in Section III.

Theorem 1: (Generalized packing bound, see also [8]). For any ellipsoid E_a^n, its unit entropy satisfies the inequality

$$\aleph(E_a^n) \geq K_a^n. \tag{4}$$

In the sequel, we assume that $n \geq 2$, for the case $n = 1$ gives the immediate answer $\aleph\left(E_a^1\right) = \log\lceil a \rceil$. In the following theorem, we derive an asymptotic upper bound on the unit entropy $\aleph(E_a^n)$. Here we only assume that parameters n and a vary in such a way that $K_a^n \to \infty$. For example, n can be fixed while components a_i grow. Therefore, our asymptotic setting $K_a^n \to \infty$ (or, briefly, $K \to \infty$) will also serve as a *limiting* condition for all other conditions described in the following theorem. This theorem will be proved in Section III.

Theorem 2: The unit entropy of an ellipsoid E_a^n satisfies asymptotic equality

$$\aleph(E_a^n) = K(1+o(1)), \qquad K \to \infty \tag{5}$$

provided that

$$\log a_n = o\left(\frac{K^2}{m \log n}\right), \quad K \to \infty \tag{6}$$

where K is defined in (3) and m is the number of half-axes a_i of length greater than one

$$m = |\{i : a_i > 1\}|. \tag{7}$$

Geometric Interpretation: Consider the sub-ellipsoid

$$\widehat{E}_a^m \overset{\text{def}}{=} \left\{ (x_{n-m+1}, \ldots, x_n) \mid \frac{x_{n-m+1}^2}{a_{n-m+1}^2} + \cdots + \frac{x_n^2}{a_n^2} \le 1 \right\}$$

obtained by projecting an ellipsoid E_a^n into the subspace $\boldsymbol{R}^m$ spanned over those dimensions that have half-axes $a_i > 1$. Note that $\widehat{E}_a^m$ has the largest volume $V\left(\widehat{E}_a^m\right)$ among all m-dimensional sub-ellipsoids obtained from E_a^n. Then our bound (5) can be rewritten as

$$\aleph(E_a^n) = \left(\log \frac{V\left(\widehat{E}_a^m\right)}{V(B^m)}\right)(1+o(1)).$$

Also, define the ball B_ρ^m of the same volume $V(B_\rho^m) = V(\widehat{E}_a^m)$. Then $\log \rho$ is the mean of $\log a_i$, which in turn is equal to K/m. Now condition (6) can be rewritten as

$$\frac{\log a_n}{\log \rho} = o\left(\frac{K}{\log n}\right), \qquad K \to \infty. \tag{8}$$

In our proof of Theorem 2, we will also show that condition (6) is equivalent to the combination of the following conditions:

$$\lim_{K\to\infty} \frac{K}{\log n} = \infty \tag{9}$$

and

$$\log \frac{a_n}{a_{n-m+1}} = o\left(\frac{K^2}{m \log n}\right), \qquad K \to \infty. \tag{10}$$

Now we see that condition (8) shows that our tight asymptotic bound (5) can fail only on the most oblong ellipsoids, for which $\log \rho = o(\log a_n)$ as $K \to \infty$. In other words, condition (8) fails if the longest half-axis a_n is lower-bounded as a polynomial ρ^s of increasing degree $s \to \infty$. In this case, the eccentricity a_n/a_{n-m+1} of sub-ellipsoid $\widehat{E}_a^m$ also undergoes an increasingly rapid growth.

Remark 1: Recall that an ε-covering of an ellipsoid E_a^n is equivalent to a unit covering of the ellipsoid $E_{a/\varepsilon}^n$. Therefore, the ε-entropy $H_\varepsilon(E_a^n)$ is readily obtained from formulas (3), (6), and (7), by replacing symbols a_i with rescaled quantities a_i/ε. In this case, the bound of Theorem 2 reads

$$H_\varepsilon(E_a^n) = K_\varepsilon(1+o(1))$$

where

$$K_\varepsilon \overset{\text{def}}{=} \sum_{i=1}^n \log^+(a_i/\varepsilon). \tag{11}$$

Remark 2: Consider the case when ellipsoid E_a^n has n equal coefficients $a_i = \rho$ and, therefore, forms the ball B_ρ^n. In this case, Theorem 2 gives $\aleph(E_a^n) \sim n \log \rho$, provided that condition (9) holds

$$n(\log \rho)/(\log n) \to \infty.$$

Note that the above result of Rogers also gives the same asymptotics $\aleph(E_a^n) \sim n \log \rho$, since

$$\aleph(E_a^n) \le n \log \rho + 5(\log n)/2 + \log C$$

according to (2).

Remark 3: Note also that expression (11) coincides with the expression obtained in [13], [14] for the ε-entropy of a random Gaussian vector $(\xi_1, \ldots, \xi_n)$ whose n independent components ξ_i have (possibly different) variances $a_i^2 \ge \varepsilon^2$ for all $i = 1, \ldots, n$.

Examples:

1) Let $a_i = \rho(i/n)^\gamma$ for all $i = 1, \ldots, n$, where ρ and γ are positive constants. In other words, here coefficients a_i form a power series bounded by ρ. Then direct calculations show that Theorem 2 holds and

$$K = \gamma[c - \log c - 1]n(1+o(1)), \qquad n \to \infty$$

where $c = \rho^{-1/\gamma}$. Thus, the unit entropy grows linearly in n.

2) Let $a_i = \rho(i^\gamma/n^\kappa)$ for all $i = 1, \ldots, n$, where $\rho > 0$ and $0 < \kappa < \gamma$ (for $\kappa = \gamma$, we have Example 1). Now coefficients a_i form an unbounded power series as n grows. Similar calculations again show that Theorem 2 holds but the unit entropy is quasi-linear in n, due to the asymptotic equality

$$K = (\gamma - \kappa) n \log n(1+o(1)), \qquad n \to \infty.$$

Note also that the main asymptotic term in K does not depend on ρ in this case.

Finally, we consider some subclasses of ellipsoids, for which condition (6) can be removed. Note that condition (6) and Theorem 2 always hold if

$$\lim_{K\to\infty} \frac{K}{m \log n} = \infty \tag{12}$$

due to the fact that $\log a_n \le K$ by definition (3). In particular, Theorem 2 holds if a_n or some other coefficients grow for fixed n.

Another similar condition is defined as follows. Given an ellipsoid E_a^n and some $\tau \in (0,1)$, consider the number

$$m_\tau = |\{i : a_i > \tau\}|. \tag{13}$$

We will show that the proof of Theorem 2 can be modified to obtain the following.

Theorem 3: The unit entropy of an ellipsoid E_a^n satisfies asymptotic equality (5) if there exists $\tau \in (0,1)$ such that

$$\lim_{K\to\infty} K/m_\tau = \infty. \tag{14}$$

Theorem 3 and condition (14) are closely related to the results known for general convex bodies [8]. Note, however, that the new condition (14) can still be much more restrictive for generic ellipsoids than the former condition (6). Namely, recall that $m_\tau \geq m$ and therefore condition (14) holds only for *expanding* ellipsoids, such that $\rho \to \infty$. In particular, condition (14) is not valid for a ball B_ρ^m of any given radius, or any ellipsoid, whose axes a_i fall within some interval $(0, a)$.

III. Proofs

A. Proof of Theorem 1

We present a short proof to keep the paper self-contained (see also [8] for a more general setting). The lower bound (4) can be obtained almost immediately, by slightly detailing the arguments used in the geometric interpretation of Section II. Indeed, consider the projection of an ellipsoid E_a^n into the subspace $\boldsymbol{R}^k$ spanned over the last k coordinates. Then we obtain the sub-ellipsoid

$$\widehat{E}_a^k \stackrel{\text{def}}{=} \left\{(x_{n-k+1}, \ldots, x_n) \mid \frac{x_{n-k+1}^2}{a_{n-k+1}^2} + \cdots + \frac{x_n^2}{a_n^2} \leq 1\right\}.$$

Also, the unit ball B^n becomes B^k. By dividing the volume of $\widehat{E}_a^k$ by the volume of B^k, we define the minimum number of covering balls in $\boldsymbol{R}^k$. Thus,

$$\aleph(E_a^n) \geq \max_{1 \leq k \leq n} \log \frac{V\left(\widehat{E}_a^k\right)}{V(B^k)}.$$

It is well known that the unit ball B^k has the volume

$$A_k = \int_{x \in B^k} dx_1 \ldots dx_k = \frac{\pi^{k/2}}{\Gamma(1 + k/2)}$$

whereas the ellipsoid $\widehat{E}_a^k$ has the volume

$$V\left(\widehat{E}_a^k\right) = A_k a_{n-k+1} \cdots a_n.$$

Thus,

$$\aleph(E_a^n) \geq \max_{1 \leq k \leq n} \sum_{i=n-k+1}^{n} \log a_i = \sum_{i=1}^{n} \log^+(a_i)$$

where the last equality simply reflects the fact that $\log a_i \leq 0$ if $a_i \leq 1$. □

Note that in case $k = n$, both sets—E_a^n and B^n—are taken in the original space $\boldsymbol{R}^n$, and the above arguments yield the (Hamming) packing bound. Obviously, the new condition (4), which we call the generalized packing bound, is stronger whenever $a_i < 1$ for some i.

To prove Theorem 2, we also need the following definitions and an intermediate statement—formulated later as Theorem 4. First, note that an ellipsoid E_a^n is contained in the unit ball B^n if inequalities $a_i \leq 1$ hold for all n coefficients a_i. Therefore, in the following we take vectors a such that $a_i > 1$ for some $i \in [1, n]$. Consider the subset $\{R\}$ of vectors $R = (r_1, \ldots, r_n)$ that satisfy restrictions

$$\begin{aligned} &r_i \geq 0, \qquad i = 1, \ldots, n \\ &\sum_{i=1}^{n} r_i \leq 1. \end{aligned} \tag{15}$$

Note that $\{R\}$ is a closed set. Given a vector a, we also consider the subset $\{\mathcal{E}(R)\}$ of vectors $\mathcal{E} = (\varepsilon_1, \ldots, \varepsilon_n)$ that satisfy restrictions

$$\begin{aligned} &0 \leq \varepsilon_i \leq r_i a_i, \qquad i = 1, \ldots, n \\ &\sum_{i=1}^{n} \varepsilon_i \leq 1. \end{aligned} \tag{16}$$

Then for any a, we define the function

$$f(R, \mathcal{E}) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \log \frac{r_i a_i}{\varepsilon_i} \tag{17}$$

where we assume that $\log \frac{0}{0} = 0$. Given R, we say that $\mathcal{E}^*(R)$ is an R-*optimal* vector if $f(R, \mathcal{E})$ achieves its minimum

$$f^*(R) \stackrel{\text{def}}{=} f(R, \mathcal{E}^*(R)) = \min_{\mathcal{E} \in \{\mathcal{E}(R)\}} f(R, \mathcal{E}) \tag{18}$$

on $\mathcal{E}^*(R)$. Note that $f^*(R)$ exists for any vector R since $f(R, \mathcal{E})$ is a continuous function on the closed bounded subset $\{\mathcal{E}(R)\}$. In the sequel, we also prove that $f^*(R)$ is a continuous function on the closed bounded subset $\{R\}$, in which case there exists an optimum pair $(R^*, \mathcal{E}^*)$ that achieves the (global) optimum

$$\begin{aligned} f^* &\stackrel{\text{def}}{=} f(R^*, \mathcal{E}^*(R^*)) = \max_{R \in \{R\}} f^*(R) \\ &= \max_{R \in \{R\}} \min_{\mathcal{E} \in \{\mathcal{E}(R)\}} f(R, \mathcal{E}). \end{aligned} \tag{19}$$

The following theorem—proven in the Appendix—will play a key role in deriving the upper bound on $\aleph(E_a^n)$.

Theorem 4: For all dimensions n and vectors a

$$f^* = K_a^n. \tag{20}$$

As above, let B_ρ^n be an n-dimensional ball of radius ρ centered at the origin. We start the proof of Theorem 2 with the following lemmas.

Lemma 5: For any dimension n, a ball B_ρ^n of radius $\rho > 1$ has unit entropy

$$\aleph\left(B_\rho^n\right) \leq n \log \rho + c \log(n+1) \tag{21}$$

where the constant c does not depend on the dimension n nor on the radius ρ.

Proof: Note that for $n \geq 9$, inequality (21) immediately follows from the Rogers bound (2). Next, we prove that inequality (21) also holds for $n \leq 8$. Indeed, for any n we can use the inequality

$$\begin{aligned} \aleph\left(B_\rho^n\right) &\leq \log(\rho\sqrt{n} + 1)^n \\ &\leq n \log \rho + (n \log n)/2 + \sqrt{n}/\rho. \end{aligned} \tag{22}$$

To prove this inequality, note that the ball B_ρ^n can be enclosed in a cube with a side 2ρ, centered at the origin. In turn, this cube can be cut into smaller cubes with a side $2/\sqrt{n}$, having at most $\rho\sqrt{n} + 1$ smaller cubes at each side (rounding off to the closest integer from above). Finally, each small cube can be enclosed in a unit ball. Thus, (22) directly follows from the bound $(\rho\sqrt{n} + 1)^n$ on the number of unit balls used in this enclosure. □

Lemma 6: Condition (6) is equivalent to the combination of conditions (9) and (10).

Proof: First, note that the former gives the combination of the latter. Indeed, (10) immediately follows from (6). Also, $\log a_n \geq K/m$, in which case (6) gives the condition $1 = o(K/\log n)$, which is equivalent to (9).

In turn, let us show that condition (6) also follows from (9) and (10). Indeed, the radius ρ satisfies both the equality $\rho = K/m$ and the inequality $\rho \geq a_{n-m+1}$. Thus, we can replace (10) as

$$\frac{\log(a_n/\rho)}{\log \rho} = o\left(\frac{K}{\log n}\right).$$

According to (9), $K/\log n \to \infty$ and we have condition (8)

$$\frac{\log a_n}{\log \rho} = o\left(\frac{K}{\log n}\right)$$

which is equivalent to (6). This completes the proof. □

B. Proof of Theorem 2

The proof will include three steps. The main idea of the first step is to cover an ellipsoid E_a^n with a finite number of subsets each of which is a direct product of the balls (of lesser dimensions). In the second step, we obtain a general upper bound on the unit entropy $\aleph(E_a^n)$, which depends on parameters introduced in the first step. Here we will use Lemma 5 and Theorem 4. Finally, in the third step we estimate the asymptotic behavior of the bound and employ the two asymptotic conditions of Lemma 6.

Step 1: Divide the set of integers $j = 1, \ldots, n$ into a set $\mathcal{I}$ of t subintervals $I_i = [n_{i-1}+1, n_i]$ of length $s_i = n_i - n_{i-1}$, where $i = 1, \ldots, t$ and

$$0 = n_0 < n_1 < \cdots < n_t = n.$$

From now on, we assume that the lengths s_i are fixed. These lengths will be optimized later, in Step 3 of our proof. We will also use the parameter $h \in (0,1)$ and the set of numbers

$$\mathcal{H} = \{ih, i = 1, \ldots, \lfloor h^{-1} \rfloor + 1\}$$

with increments equal to h. For any $w \in [0,1]$, let $\overline{w} \in (w, w+h]$ be the closest point in $\mathcal{H}$ exceeding w. Finally, consider any vector $u = (u_1, \ldots, u_t) \in \mathcal{H}^t$ (with all t components from $\mathcal{H}$) such that

$$\sum_{i=1}^{t} u_i \leq 1 + th. \tag{23}$$

In what follows, $U \subset \mathcal{H}^t$ denotes the subset of all such vectors u.

For each point $x \in \boldsymbol{R}^n$, let $x_{I_i} = (x_{n_{i-1}+1}, \ldots, x_{n_i})$ for any $i = 1, \ldots, t$. Also, consider a ball

$$B_{\rho_i}^{s_i} = \left\{ x_{I_i} : \sum_{j \in I_i} x_j^2 \leq \rho_i^2 \right\}, \qquad \rho_i = u_i^{1/2} a_{n_i}$$

of dimension s_i. Here the radius ρ_i is defined by the parameter u_i. Finally, define the direct products of all t balls

$$D_u^n = \prod_{i=1}^{t} B_{\rho_i}^{s_i}.$$

Equivalently

$$D_u^n = \left\{ x \in \boldsymbol{R}^n | \sum_{j \in I_i} \frac{x_j^2}{a_{n_i}^2} \leq u_i,\ i = 1, \ldots, t \right\}.$$

Our next goal is to prove an important property

$$E_a^n \subseteq \bigcup_{u \in U} D_u^n. \tag{24}$$

Indeed, consider any point $x \in E_a^n$ and any set $\mathcal{I}$ of t subintervals. Let

$$w_i = \sum_{j \in I_i} x_j^2 / a_j^2$$

denote the contribution of the point x_{I_i} to the overall squared "weight" in (1). Then we consider the ball $B_{\rho_i}^{s_i}$ defined by parameter $u_i = \overline{w}_i$. Recall that $a_{n_i} = \max\{a_j,\ j \in I_i\}$. Therefore, the inclusion $x_{I_i} \in B_{\rho_i}^{s_i}$ holds according to the inequalities

$$\sum_{j \in I_i} x_j^2 / a_{n_i}^2 \leq \sum_{j \in I_i} x_j^2 / a_j^2 = w_i \leq u_i.$$

Second, for any point $x \in E_a^n$ and any set $\mathcal{I}$, parameters $u_i = \overline{w}_i$ satisfy inequality (23)

$$\sum_{i=1}^{t} u_i \leq \sum_{i=1}^{t} (w_i + h) \leq 1 + th.$$

Thus, by considering all possible vectors $u \in U$ for any point $x \in E_a^n$, we find a subset D_u^n that covers this point. As a result, inclusion (24) holds.

Step 2: Given a subset D_u^n, consider its covering by unit balls. Let us cover each ball $B_{\rho_i}^{s_i}$, $i = 1, \ldots, t$, with the balls $B_{e_i}^{s_i}$ of some radius e_i. Then the direct product D_u^n of the balls is completely covered by the direct product of their coverings. Correspondingly, this set D_u^n has unit entropy

$$\aleph(D_u^n) \leq \inf_e \sum_{i=1}^{t} H_{e_i}\left(B_{\rho_i}^{s_i}\right) \tag{25}$$

where different radii e_i form a vector $e = (e_1, \ldots, e_t)$ such that

$$\sum_{i=1}^{t} e_i^2 \leq 1$$
$$0 \leq e_i \leq \rho_i, \qquad i = 1, \ldots, t. \tag{26}$$

The first restriction limits the overall radius of our covering by 1, while t other restrictions $e_i \leq \rho_i$ reflect the fact that $H_{e_i}\left(B_{\rho_i}^{s_i}\right) = 0$ if $e_i > \rho_i$. Obviously, the infimum in (25) cannot be achieved in the latter case.

Note that inequality (23) can be rewritten as

$$\sum_{i=1}^{t} u_i h^{-1} \le t + h^{-1}.$$

Also, all numbers $u_i h^{-1}$ are positive integers. Therefore, the number of vectors $u \in \mathcal{H}^t$ that satisfy inequality (23) is equal to $\mathcal{N}(t,h)$, where

$$\mathcal{N}(t,h) = \binom{t+\lfloor h^{-1}\rfloor}{t} \le \frac{(t+\lfloor h^{-1}\rfloor)^t}{t!}. \tag{27}$$

From (24) and (25)

$$\aleph\left(E_a^n\right) \le \max_u \inf_e \sum_{i=1}^{t} H_{e_i}\left(B_{\rho_i}^{s_i}\right) + \log \mathcal{N}(t,h). \tag{28}$$

Now we use Lemma 5 to estimate quantities $H_{e_i}\left(B_{\rho_i}^{s_i}\right)$ in (28). This yields the estimate

$$\aleph\left(E_a^n\right) \le \max_u \min_e \sum_{i=1}^{t} s_i \log \frac{\rho_i}{e_i} + c\sum_{i=1}^{t} \log(s_i+1) + \log \mathcal{N}(t,h) \tag{29}$$

where constant c does not depend on all other parameters.

Next, we rewrite the first sum in (29) to employ Theorem 4. Obviously, this sum can be rewritten without coefficients s_i, by taking n terms $\log(\rho_i/e_i)$ and using s_i times our parameters u_i, a_{n_i}, and e_i. Secondly, for all $i = 1,\ldots,t$, let us temporarily replace variables u_i, e_i, and coefficients a_{n_i} with

$$r_i = u_i/\gamma^2, \quad \varepsilon_i = e_i^2, \quad \alpha_{n_i} = (a_{n_i}\gamma)^2$$

where

$$\gamma = \sqrt{1+th}. \tag{30}$$

Note that $\varepsilon_i \le r_i \alpha_{n_i}$ for all i, according to (26). Moreover, variables r_i and ε_i satisfy all other restrictions (15) and (16). Therefore, we can use Theorem 4 with coefficients α_{n_i} as follows:

$$\begin{aligned}\max_u \min_e &\sum_{i=1}^{t} s_i \log \frac{\rho_i}{e_i} \\ &= \max_u \min_e \sum_{i=1}^{t} \frac{s_i}{2} \log \frac{r_i \alpha_{n_i}}{\varepsilon_i} \\ &\le \sum_{i=1}^{t} \frac{s_i}{2} \log^+ \alpha_{n_i} = \sum_{i\gamma a_{n_i}>1} s_i \log(\gamma a_{n_i}) \\ &\le \sum_{i=1}^{t} s_i \log^+ a_{n_i} + \log\gamma \sum_{i\gamma a_{n_i}>1} s_i. \end{aligned} \tag{31}$$

Now consider vector $\widehat{a} = (\widehat{a}_1,\ldots,\widehat{a}_n)$ and ellipsoid $E_{\widehat{a}}^n$ with coefficients

$$\widehat{a}_j = a_{n_i}, \qquad j \in I_i, \;\; i = 1,\ldots,t.$$

Then

$$\sum_{i=1}^{t} s_i \log^+(a_{n_i}) = K_{\widehat{a}}^n. \tag{32}$$

Our next goal is to estimate the difference $K_{\widehat{a}}^n - K_a^n$. We choose $n_1 = n - m$, where m is defined in (7). Then it is readily verified that

$$0 \le K_{\widehat{a}}^n - K_a^n \le \sum_{i=2}^{t} (s_i - 1) \log \frac{a_{n_i}}{a_{n_{i-1}+1}}. \tag{33}$$

Now from (29) and (31)–(33), we obtain the upper bound

$$\aleph\left(E_a^n\right) \le K_a^n + c\sum_{i=1}^{t} \log(s_i+1) + \log\mathcal{N}(t,h) + n\log\gamma + \sum_{i=2}^{t}(s_i-1)\log\frac{a_{n_i}}{a_{n_{i-1}+1}} \tag{34}$$

which is used in Step 3 to derive tight asymptotic bound (5).

Step 3: To prove asymptotic equality (5), we use conditions (9) and (10), which are equivalent to the original condition (6) according to Lemma 6. For the asymptotic setting $K \to \infty$, our goal is to optimize the set of subintervals $\mathcal{I}$ and the quantization step h, so that all terms in the right-hand side of (34) fall to the order of $o(K)$, with the exception of the first term K_a^n. In so doing, we will also employ a positive vanishing function $\eta = \eta(K)$ such that for $K \to \infty$

$$\lim_{K\to\infty} \eta = 0, \quad \log \frac{a_n}{a_{n-m+1}} = o\left(\frac{\eta K^2}{m \log n}\right). \tag{35}$$

Obviously, our original conditions (9) and (10) can be modified into the latter conditions if this function η approaches 0 slowly enough.

We take $n_1 = n - m$ and choose all other intervals I_i, $i = 2,\ldots,t-1$, of equal length

$$\begin{aligned} s_i &\equiv s = \lceil (m\log n)/(\eta K) \rceil \\ s_t &= n - s \le s. \end{aligned} \tag{36}$$

From (36), we derive the upper bound

$$t = \lceil m/s \rceil + 1 \le \left\lceil \frac{\eta K}{\log n} \right\rceil + 1. \tag{37}$$

Now we can estimate the terms in the right-hand side of (34). We rewrite the second term using convexity of the logarithmic function. Then we use (9) and (37) to obtain the following estimates:

$$\begin{aligned}\sum_{i=1}^{t}\log(s_i+1) &\le \log(n-m+1)+(t-1)\log\left(\frac{m}{t-1}+1\right)\\ &\le \log(n-m+1)+(t-1)\log(m+1)\\ &\le \log(n-m+1)+o(K)=o(K).\end{aligned}\tag{38}$$

Next, we use asymptotic conditions (35), which show that the last term in (34) also has the order of $o(K)$

$$\begin{aligned}\sum_{i=2}^{t}(s_i-1)\log\frac{a_{n_i}}{a_{n_{i-1}+1}} &\le \frac{m\log n}{\eta K}\sum_{i=2}^{t}\log\frac{a_{n_i}}{a_{n_{i-1}+1}}\\ &= \frac{m\log n}{\eta K}\log\frac{a_n}{a_{n-m+1}}.\end{aligned}\tag{39}$$

Finally, we choose the quantization step $h=1/n$, and verify that the two remaining terms in (34) have the same order $o(K)$. Indeed, we use Stirling formula for $t!$ in (27) and obtain

$$\begin{aligned}\log\mathcal{N}(t,h) &\le \log(e(1+t^{-1}h^{-1}))^t\\ &\le t(1+\log(n+1))\\ &\le \eta K+t+t/n.\end{aligned}\tag{40}$$

Also, from (30)

$$n\log\gamma\le nth/2=t/2.\tag{41}$$

Now combining our estimates (38)–(41) with the upper bound (34), we obtain

$$\aleph\left(E_a^n\right)\le K+o(K),\quad K\to\infty.$$

The latter bound has the same order as the lower bound of Theorem 1, and the proof of Theorem 2 is completed. □

Our final goal is to prove Theorem 3. Here we extensively use the proof of Theorem 2.

C. Proof of Theorem 3

Given condition (14) for some $\tau\in(0,1)$, we choose $t=m_\tau+1$ intervals I_i of length

$$\begin{aligned}s_1&=n-m_\tau\\ s_i&=1,\qquad i=2,\dots,t.\end{aligned}\tag{42}$$

Also, we choose

$$h=c_\tau t^{-1}\tag{43}$$

where

$$c_\tau=\min(1,\tau^{-2}-1).\tag{44}$$

Finally, given any vector $(u_1,\dots,u_t)\in\mathcal{H}^t$, we *completely* cover the first ball $B_{\rho_1}^{s_1}$ of radius ρ_1 taking the *identical* covering radius

$$e_1=\rho_1=u_1^{1/2}a_{n_1}.$$

First, we prove that $e_1<1$, so that such a choice is possible. Indeed, note that $a_{n_1}\le\tau$ and $u_1\le 1+h$. Also, $c_\tau\tau^2\le 1-\tau^2$ for all $\tau\in(0,1)$. Therefore,

$$e_1\le\tau(1+h)^{1/2}\le\left(\tau^2+c_\tau\tau^2t^{-1}\right)^{1/2}<1.\tag{45}$$

Now we follow Steps 2 and 3 of the Proof of Theorem 2, with all modifications resulting from (42)–(44). Since $s_i=1$ for all $i=2,\dots,t$, each ball $B_{\rho_i}^{s_i}$ is now replaced with the one-dimensional *interval* $B_{\rho_i}^1$ of radius $\rho_i=u_i^{1/2}a_{n_i}$. To cover these intervals we can use any set of nonnegative numbers e_i and (h-incremental) numbers $u_i\in\mathcal{H}$ provided that

$$\begin{aligned}&\textstyle\sum_{i=2}^{t}e_i^2\le 1-u_1\tau^2\\ &\textstyle\sum_{i=2}^{t}u_i\le 1+th-u_1,\quad 0\le e_i\le\rho_i,\ i=1,\dots,t.\end{aligned}\tag{46}$$

Now we can rewrite the estimate (29) as

$$\aleph\left(E_a^n\right)\le\max_u\min_e\sum_{i=2}^{t}\log(\rho_i/e_i)+c(t-1)+\log\mathcal{N}(t,h).\tag{47}$$

(Here the universal constant c is taken from Lemma 5). Next, we rewrite the first sum in (47) to employ Theorem 4. We replace variables u_i, e_i, and coefficients a_{n_i} with the new variables

$$r_i=u_i/\gamma^2,\quad \varepsilon_i=(e_i/\beta)^2,\quad \alpha_{n_i}=(a_{n_i}\gamma)^2$$

using two scaling factors

$$\gamma=(1+th-u_1)^{1/2},\qquad \beta=\left(1-u_1\tau^2\right)^{1/2}.\tag{48}$$

Now the new variables r_i and ε_i satisfy all earlier restrictions (15) and (16). Therefore, we can use Theorem 4 with coefficients α_{n_i} and estimate the first term in (47) as follows:

$$\begin{aligned}\max_u\min_e\sum_{i=1}^{t}\log(\rho_i/e_i) &= \max_u\min_e\sum_{i=2}^{t}\left(\log\frac{r_i\alpha_{n_i}}{\varepsilon_i\beta^2}\right)/2\\ &\le\sum_{i=2}^{t}\log^+(\gamma a_{n_i}/\beta)\\ &\le\sum_{i=2}^{t}\log^+a_{n_i}+(t-1)\log(\gamma/\beta).\end{aligned}$$

Finally, recall that $s_i=1$ for $i\ge 2$, in which case the numbers a_{n_i} entirely fill the set of integers $[n-m_\tau,n]$. In this case, we have equalities

$$\sum_{i=2}^{t}\log^+(a_{n_i})=\sum_{j=n-m_\tau+1}^{n}\log^+(a_j)=K.$$

Therefore, we arrive at the estimate

$$\aleph(E_a^n) \le K + c(t-1) + \log \mathcal{N}(t,h) + (t-1)\log(\gamma/\beta) \quad (49)$$

which only depends on the ratio γ/β from (48) defined by u_1. Given our choice of h, γ, and β from (43), (44), and (48), it is easy to see that for any u_1

$$\gamma/\beta = \tau^{-1}\left(\frac{\min(2,\tau^{-2}) - u_1}{\tau^{-2} - u_1}\right)^{1/2} \le \min(\sqrt{2}, \tau^{-1})$$

with maximum achieved at $u_1 = 0$. Also, similarly to (40)

$$\begin{aligned}\log \mathcal{N}(t,h) &\le \log(e(1+t^{-1}h^{-1}))^t \\ &\le t\left(1 + \log\left(1 + c_\tau^{-1}\right)\right).\end{aligned}$$

Summarizing, we see that the last three terms in (49) are linear in $m_\tau = t-1$

$$\aleph(E_a^n) \le K + t\left(c + 1 + \log\left(1 + c_\tau^{-1}\right) + \min(\sqrt{2}, \tau^{-1})\right).$$

Now we see that restriction $K/m_\tau \to \infty$ gives tight bound $\aleph(E_a^n) \le K + o(K)$. □

APPENDIX
PROOF OF THEOREM 4

First, note that for any a

$$f^* \ge K_a^n.$$

Indeed, we can always take the vector R with components

$$r_i = \begin{cases} 0, & \text{if } i \le n-m \\ 1/m, & \text{if } i > n-m \end{cases}$$

and obtain the lower bound

$$f^* \ge \min_{\mathcal{E}} \sum_{i=n-m+1}^{n} \log \frac{a_i}{m\varepsilon_i} = \sum_{i=n-m+1}^{n} \log a_i = K_a^n.$$

Here $\mathcal{E}^*(R)$ has m equal components $1/m$, due to convexity of the logarithmic function. Also, we see that $f^* > 0$, since $a_i > 1$ for some positions i.

Second, let $\{\overline{R}\}$ be the subset of vectors $R \in \{R\}$ that satisfy inequality

$$\sum_{i=1}^{n} r_i a_i \ge 1. \quad (50)$$

Note that optimum $f^* > 0$ can be sought only in the (closed) subset $\{\overline{R}\}$. Indeed, if inequality (50) does not hold, we can always take $\varepsilon_i = a_i r_i$ for all i and obtain $f^*(R) = 0$. For this reason, in the following we replace our former set $\{R\}$ by $\{\overline{R}\}$. Note that $\{\overline{R}\}$ is nonempty, since otherwise $f^* = 0$, which contradicts the preceding arguments. Also, f^* is achieved inside $\{\overline{R}\}$ since $f^*(R) = 0$ if equality holds in (50).

Now we need to prove that equality (20) holds. Recall that $f(R,\mathcal{E})$ is the *difference* between the two convex functions $\sum \log a_i r_i$ and $\sum \log \varepsilon_i$. In this case, our arguments become slightly more convoluted than those used in direct convex optimization. However, below we also make an extensive use of convexity of the logarithmic function. Our proof includes the following lemmas.

Lemma 7: For any vector $R \in \{\overline{R}\}$, any R-optimal vector $\mathcal{E}^*(R)$ has components

$$\varepsilon_i = \min\{r_i a_i, \nu\} \quad (51)$$

where $i = 1, \ldots, n$, and $\nu = \nu(R)$ is the root of the equation

$$\sum_{i=1}^{n} \min\{r_i a_i, \nu\} = 1. \quad (52)$$

Proof: Given any vector $R \in \{\overline{R}\}$, consider any R-optimal vector $\mathcal{E}^*(R)$. We first prove that equality $\varepsilon_i = \varepsilon_j$ holds for any pair of its symbols such that $\varepsilon_i < r_i a_i$ and $\varepsilon_j < r_j a_j$. Indeed, suppose that the latter two inequalities hold but $\varepsilon_i < \varepsilon_j$. Obviously, all three inequalities are satisfied if we choose sufficiently small $\delta > 0$, and use new symbols $\varepsilon_i' = \varepsilon_i + \delta$ and $\varepsilon_j' = \varepsilon_j - \delta$. Let $\mathcal{E}'$ be the new vector obtained from $\mathcal{E}^*(R)$ by this replacement. Due to convexity of the logarithmic function, we have the inequality

$$\log(\varepsilon_i + \delta) + \log(\varepsilon_j - \delta) > \log \varepsilon_i + \log \varepsilon_j.$$

Therefore, $f(R,\mathcal{E}') < f(R,\mathcal{E}^*(R))$, and vector $\mathcal{E}^*(R)$ is not R-optimal. This contradiction shows that there exists some parameter ν such that $\varepsilon_i = \nu$ whenever $\varepsilon_i < r_i a_i$. Let I be the subset of such positions i. Obviously, I is nonempty, since otherwise we have n equalities $\varepsilon_i = r_i a_i$, which is a contradiction to inequality (50).

Next, consider the subset J of remaining positions j, for which $\varepsilon_j = r_j a_j$. Then $\nu \ge r_j a_j$ for any such j. Indeed, given the opposite inequality $v < r_j a_j$ for some j, we take any $i \in I$ and replace the former symbols ε_i and ε_j by the new symbols $\varepsilon_i + \delta$ and $\varepsilon_j - \delta$, where δ is sufficiently small. Now the proof completely repeats the first part and shows that the vector $\mathcal{E}^*(R)$ is not R-optimal. Thus, any R-optimal vector $\mathcal{E}^*(R)$ has components (51).

Finally, we use restrictions (50) valid for any $R \in \{\overline{R}\}$. In this case, any R-optimal vector $\mathcal{E}^*(R)$ with components (51) must satisfy the equality $\sum_{i=1}^{n} \varepsilon_i = 1$, since some components ε_i can be increased otherwise. This gives a unique solution (52) and completes the proof. □

The preceding lemma also shows that $f^*(R)$ is a continuous function of R on the bounded closed subset $\{\overline{R}\}$. Thus, there exists some (not necessarily unique) optimal vector $R^* = (r_1, \ldots, r_n)$. Also, the proof of Lemma 7 shows that for any optimal pair $(R^*, \mathcal{E}^*)$ there exists ν such that positions $i = 1, \ldots, n$ form two complementary subsets

$$\begin{aligned} I &= \{i : \varepsilon_i = \nu, \quad a_i r_i > \nu\}, \qquad |I| > 0 \\ J &= \{i : \varepsilon_i = a_i r_i, \quad a_i r_i \le \nu\}. \end{aligned} \quad (53)$$

Lemma 8: Any optimal vector R^* satisfies equality

$$\sum_{i=1}^{n} r_i = 1$$

and has equal components on the subset I defined in (53)

$$r_i = r, \qquad \text{for any } i \in I. \quad (54)$$

Proof: The proof is similar to that of Lemma 7. First, we can always increase some components r_i and f^* if

$$\sum_{i=1}^{n} r_i < 1.$$

Second, suppose that $r_i < r_j$ in vector R^* for some indices i, $j \in I$. Then we can increase $f^*(R^*)$, by using δ-rebalancing and taking $r_i + \delta$ and $r_j - \delta$ for sufficiently small δ. Note that both restrictions (15) and (16) are satisfied for the new vector R'. Also, $R' \in \{\overline{R}\}$, since $f^*(R') > 0$. Thus, we have a contradiction, for $f^*(R') > f^*(R^*)$. □

Lemma 9: There exists an optimal vector R^* with zero components on the subset J defined in (53)

$$r_i = 0, \qquad \text{for any } i \in J.$$

Proof: Rephrasing Lemmas 7 and 8, we see that for any optimal pair $(R^*, \mathcal{E}^*(R^*))$, there exist two subsets I and J such that

$$\begin{cases} r_i = r, & \text{for any } i \in I \\ \varepsilon_i = \nu, & \text{for any } i \in I \\ ra_i > \nu, & \text{for any } i \in I \\ \varepsilon_i = r_i a_i, & \text{for any } i \in J \end{cases} \tag{55}$$

where

$$\begin{cases} r = \left(1 - \sum_{i \in J} r_i\right) / (n-l) \\ \nu = \left(1 - \sum_{i \in J} r_i a_i\right) / (n-l) \end{cases} \tag{56}$$

and $|J| = l$ is the size of J. Also, $|I| = n - l$.

Note that conditions (55) and (56) always give vectors $R \in \{\overline{R}\}$ and $\mathcal{E} \in \{\mathcal{E}(R)\}$. Now consider all such pairs $(R, \mathcal{E})$ given some "boundary" vector R_J with components r_s, $s \in J$. This vector defines both parameters r and ν. Therefore, $f^*(R)$ also becomes the function

$$f^*(R_J) = \sum_{i \in I} \log \frac{ra_i}{\nu}$$

of vector R_J. Next, we study how this function $f^*(R_J)$ changes with any argument r_s, $s \in J$. By changing r_s, we also replace r and ν, according to (55) and (56). Let

$$\Delta_i = ra_i - \nu, \qquad i = 1, \ldots, n. \tag{57}$$

Using (56), one can readily obtain equality

$$\frac{\partial f^*(R_J)}{\partial r_s} = \frac{\Delta_s \cdot |I|}{r\nu(n-l)} = \frac{\Delta_s}{r\nu}. \tag{58}$$

When combined, (56) and (57) also show that

$$\frac{\partial \Delta_i}{\partial r_s} = \frac{a_s - a_i}{n - l}.$$

In particular,

$$\frac{\partial \Delta_s}{\partial r_s} = 0. \tag{59}$$

Equality (58) shows that an optimal vector R^* can be found among those vectors $R \in \{\overline{R}\}$, which either satisfy the equality $\Delta_s = 0$ or belong to the boundary $r_s = 0, 1$ for some $s \in J$.

Consider the first case with $\Delta_s = 0$, and let $r_s > 0$ for some s. First, note that Δ_i is a nondecreasing function of $i \in I$ by definition (57), and is positive according to inequalities in (55). Since $\Delta_s = 0$, the inequality $a_i > a_s$ holds for any $i \in I$. According to formulas (58) and (59), both quantities Δ_s and $f^*(R_J)$ do not change when r_s is being reduced. In this process, we also redefine all parameters using equalities in (55) and (56).

Finally, note that inequalities $ra_i > \nu$ in (55) also hold after reduction. Indeed, the original value $\Delta_s = 0$ is left unchanged, and therefore, $\Delta_i > 0$ for $a_i > a_s$. Thus, by taking $r_s = 0$, we keep the function $f^*(R_J)$ unchanged and obtain new valid vectors R^* and $\mathcal{E}^*(R^*)$. Repeating this reduction, we obtain components $r_s = 0$ for all $s \notin I$. This completes the proof. □

Now we see that Theorem 4 directly follows from Lemma 9. □

Acknowledgment

The authors thank Shiri Artstein for her insightful comments on the content of the paper and helpful discussion of the results of [7] and [8].

References

[1] A. N. Kolmogorov and V. M. Tikhomirov, "ε-entropy and ε-capacity," *Usp. Mat. Nauk*, vol. 14, pp. 3–86, 1959.
[2] C. A. Rogers, "Covering a sphere with spheres," *Mathematika*, vol. 10, pp. 157–164, 1963.
[3] A. D. Wyner, "Random packings and coverings of the unit N-sphere," *Bell Syst. Tech. J.*, vol. 40, pp. 2111–2118, 1967.
[4] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. New York: Springer-Verlag, 1988.
[5] D. L. Donoho, "Counting bits with Kolmogorov and Shannon," Stanford Univ., Stanford, CA, Tech. Rep. 2000-38, 2000.
[6] G. F. T'oth, "New results in the theory of packing and covering," in *Convexity and its Applications*, P. M. Gruber and J. M. Wills, Eds. Basel, Switzerland: Birkhauser-Verlag, 1983, pp. 318–359.
[7] Y. Gordon, H. König, and C. Schütt, "Geometric and probabilistic estimates for entropy and approximation numbers of operators," *J. Approx. Theory*, vol. 49, no. 3, pp. 219–239, 1987.
[8] G. Pisier, "The volume of convex bodies and banach space geometry," in *Cambridge Tracts in Mathematics, 94*. Cambridge, U.K.: Cambridge Univ. Press, 1989.
[9] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein, *Covering Codes*. Amsterdam, The Netherlands: Elsevier, 1997.
[10] I. I. Dumer, M. S. Pinsker, and V. V. Prelov, "Epsilon-entropy of an ellipsoid in a Hamming space," *Probl. Inform. Transm.*, vol. 38, no. 1, pp. 1–15, 2002.
[11] ——, "An optimization problem related to the computation of the epsilon entropy of an ellipsoid in a Hamming space," *Probl. Inform. Transm.*, vol. 38, no. 2, pp. 99–112, 2002.
[12] V. V. Prelov and E. C. van der Meulen, "On epsilon-entropy of a certain class of ellipsoids in a Hamming space," *Probl. Inform. Transm.*, vol. 38, no. 2, pp. 113–122, 2002.
[13] A. N. Kolmogorov, "Theory of information transmission," in *Proc. Session on Scientific Problems of Automation*, ser. 2. Moscow, U.S.S.R.: Izd. Akad. Nauk. SSSR, 1956, vol. 33, pp. 291–321. Translation in *AMS Trans.* ser. 2, vol. 33, pp. 291–321, 1956.
[14] M. S. Pinsker, "Gaussian sources," *Probl. Inform. Transm.*, vol. 14, pp. 59–100, 1963.