



---

Soviet-era science, translated into English

# CAPACITY OF A SET IN SIGNAL SPACE AND A RIEMANNIAN METRIC

MATHEMATICS

1966

SovietRxiv

---

View the original and related papers at <https://sovietrxiv.org/items/ru-196601.22692>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

**Abstract**

**Full Text**

UDC 519.25

*MATHEMATICS*

**V. P. KOZLOV**

## CAPACITY OF A SET IN SIGNAL SPACE AND A RIEMANNIAN METRIC

*(Presented by Academician Yu. V. Linnik on 24 V 1965)*

1. Let  $\Omega$  be a measurable space on which a family of probability measures  $P_s(\cdot)$ ,  $s \in S$ , is given, where  $S$  is some set. For fixed  $(\Omega, S)$ , the family  $P_s(\cdot)$  defines a channel for which the set  $S$  is the space of input signals and  $\Omega$  is the space of output signals. Let, further,  $D \subset S$  be some subset of the set  $S$ .

**Definition 1.** The **capacity** of the set  $D$  relative to the channel  $P_s(\cdot)$  will be called the number

$$M(D) = \sup \sum_k \sup_{s \in D} P_s(E_k), \quad (1)$$

where the least upper bound is taken over all finite partitions of the space  $\Omega$ .

It is not difficult to verify that the capacity of a set is a nondecreasing subadditive function of the set.

**Definition 2.** We shall say that the set  $D$  **decomposes into a product of independent sets**  $D'$  and  $D''$ :  $D = D' \times D''$ , if: 1)  $D$  is the set of ordered pairs  $s = (s', s'')$ ,  $s' \in D'$ ,  $s'' \in D''$ ; 2) the measurable space  $\Omega$  can be decomposed into the direct product of measurable spaces  $\Omega' \times \Omega''$  in such a way that, for any  $s' \in D'$  and  $s'' \in D''$ , the probability measure  $P_s(\cdot)$  on  $\Omega$  is the product of the probability measures  $P_{s'}(\cdot)$  and  $P_{s''}(\cdot)$ , given on  $\Omega'$  and  $\Omega''$ , respectively.

Under the conditions specified in Definition 2, the relation

$$M(D' \times D'') = M(D')M(D'') \quad (2)$$

is valid.

- 2) The quantity  $M(D)$  is directly connected with other quantities characterizing the possibility of obtaining, with the aid of the channel  $P_s(\cdot)$ , one or another information concerning the elements of the set  $D$ . Choose from the set  $D$  some number of elements  $s_1, s_2, \dots, s_N$  (the set  $D$  itself

may be finite, countable, or have the cardinality of the continuum) and construct at the output of the channel a decision scheme for identifying the transmitted symbols, i.e., fix a partition of the space  $\Omega$  into disjoint sets  $E_1, E_2, \dots, E_N$  put in one-to-one correspondence with the selected elements. The quantity  $P_{s_j}(E_j)_{j=1, \dots, N}$  determines in this case the probability of correct identification under the condition that the symbol  $s_j$  has been transmitted. A natural measure of the reliability of transmission is the sum of the conditional probabilities of correct identification, which should be as close as possible to the number  $N$  of selected elements. This, however, is not always possible; namely, for every partition  $\{E_1, \dots, E_N\}$  the inequality

$$\sum_{j=1}^N P_{s_j}(E_j) \leq M(D), \quad (3)$$

holds.

which is evidently valid for any choice of elements  $s_1, \dots, s_N$  from the set  $D$  and for all  $N$ . In fact, inequality (3) limits the number of mutually distinguishable elements that can be chosen from the set  $D$ .

Another important property of the quantity  $M(D)$  is its relation to the amount of information  $I(D, \Omega)$  (in Shannon's sense) contained in  $\Omega$  with respect to the elements of the set  $D$ . Shannon's definition of information requires the specification of an "a priori" probability distribution on the set of possible signals. Let this distribution be given by a probability measure  $\mu$ , defined on some algebra of subsets of the set  $D$  and normalized so that  $\mu(D) = 1$ . If, for any  $E \subset \Omega$ , the measure  $P_s(E)$ , as a function of the argument  $s$ , is measurable with respect to the same algebra, then one may define, according to <sup>(1)</sup>, the information in Shannon's sense by the relation

$$I(D, \Omega) = \sup \sum_{i,k} P(E_i; \Delta_k) \log \frac{P(E_i; \Delta_k)}{P(E_i, D)\mu(\Delta_k)}, \quad (4)$$

where  $P(E, \Delta) = \int_{\Delta} P_s(E) d\mu$ ,  $\Delta \subset D$ ,  $E \subset \Omega$ , and the least upper bound is taken over all finite measurable partitions  $\{E_i\}$  and  $\{\Delta_k\}$  of the spaces  $\Omega$  and  $D$ , respectively. In this case the inequality

$$2^{I(D, \Omega)} \leq M(D) \quad (5)$$

holds.

Inequality (5) estimates the information  $I(D, \Omega)$  for any choice of the probability distribution at the input of the channel, i.e., it is an estimate of the capacity <sup>(2)</sup> of a channel with input alphabet from the set  $D$ . It is easy to see that equality

in (5) is attained, for example, if  $D$  is a finite set of order  $N$  and  $M(D) = N$  (all elements of the set are absolutely distinguishable), i.e., estimate (5) in the general case cannot be improved.

4. Let now  $S = S^n$  be a continuous  $n$ -dimensional manifold <sup>(3)</sup>, i.e., for any measurable  $E \subset \Omega$ ,  $P_s(E) = P(E; x^1, \dots, x^n)$  is a continuous twice differentiable function of the variables  $x^1, \dots, x^n$ —the coordinates of the point  $s$  in the manifold  $S^n$ ; moreover, the coordinate system  $\{x^1, \dots, x^n\}$  is specified up to an arbitrary nondegenerate twice differentiable change of variables. A Riemannian metric in the manifold  $S^n$  is naturally introduced by means of the Fisher information matrix <sup>(4–6)</sup>. We define the metric quadratic form by the relation

$$d\rho^2 = g_{ik} dx^i dx^k = \sup \left[ dx^i dx^k \sum_{\alpha} \frac{1}{P_s(E_{\alpha})} \frac{\partial P_s(E_{\alpha})}{\partial x^i} \frac{\partial P_s(E_{\alpha})}{\partial x^k} \right], \quad (6)$$

where the least upper bound is taken over all finite partitions of the space  $\Omega$ .

The definition given for the system of quantities  $g_{ik}$ ,  $i, k = 1, \dots, n$ , coincides with the usual definition of the Fisher information matrix <sup>(4)</sup>, if the functions  $\partial P_s(\cdot)/\partial x^i$ ,  $i = 1, \dots, n$ , considered as functions of a set in the space  $\Omega$ , are absolutely continuous with respect to the measure  $P_s(\cdot)$ . The geometry generated in  $S^n$  by the quadratic form (6) does not depend on the chosen coordinate system  $\{x^1, \dots, x^n\}$  and is determined only by the original family of probability measures  $P_s(\cdot)$ . Below we shall clarify the informational meaning of the geometric invariants of certain sets from  $S^n$ , establishing a connection between these invariants and the capacities of the corresponding sets.

The simplest nontrivial set is a pair of points  $(s_1, s_2)$ , and its only geometric invariant is the distance between the points—

families, defined as usual by (3):

$$\rho(s_1, s_2) = \inf \int_{s_1 s_2} \sqrt{g_{ik} dx^i dx^k}, \quad (7)$$

where the greatest lower bound is taken over all continuously differentiable curves connecting the points  $s_1$  and  $s_2$  in the manifold  $S^n$ . Between the capacity  $M(s_1, s_2)$  of a pair of points and the distance  $\rho(s_1, s_2)$  there is the inequality

$$M(s_1, s_2) \leq 1 + \frac{1}{2}\rho(s_1, s_2). \quad (8)$$

Inequality (8) limits the possibility of distinguishing elements if  $0 \leq \rho \leq 2$  (for  $M(s_1, s_2) = 1$  the elements are absolutely indistinguishable). The maximum

probability of correct identification under equiprobable presentation of the elements is simply  $\frac{1}{2}M(s_1, s_2)$ , i.e., the value  $\rho = 1$  practically corresponds to the threshold of distinguishability.\*

The capacity of a one-dimensional set—a simple arc  $\widehat{AB}$ —in the space  $S^n$  is estimated by the inequality

$$M(\widehat{AB}) \leq 1 + \int_{\widehat{AB}} \sqrt{g_{ik} dx^i dx^k}, \quad (9)$$

where the integral on the right-hand side determines the length of the arc  $L(\widehat{AB})$  in the metric (6). Since the capacity  $M(\widehat{AB})$ , in turn, estimates the Shannon information for a channel with a one-dimensional input, inequalities (5) and (9) establish a relation between the information quantities of Shannon and Fisher in the one-dimensional case.

The following inequality refines the semiadditivity of the capacity of a set  $D$  when an isolated point  $A \in S^n$  is adjoined:

$$M(D \cup A) \leq M(D) + \rho(D; A), \quad (10)$$

where  $\rho(D; A) = \inf_{B \in D} \rho(A, B)$  is the distance from the point  $A$  to the set  $D$ .

5. In the general case of an arbitrary domain  $G \subset S^n$ , it is not possible to estimate the quantity  $M(G)$  in terms of the geometric invariants of the domain  $G$ . However, the following simple example shows that, in principle, this apparently is possible. Let the domain  $G$  decompose (in the sense of definition 2) into a product of independent intervals  $L_1 \times L_2 \times \dots \times L_n$ , i.e., in the space  $S^n$  (which in this case is Euclidean) it represents a rectangular parallelepiped; then

$$M(L_1 \times L_2 \times \dots \times L_n) \leq (1 + L_1)(1 + L_2) \dots (1 + L_n). \quad (11)$$

Expanding the brackets on the right-hand side, we obtain the sum of the geometric invariants of the parallelepiped, beginning with the sum of the lengths of its edges and ending with its surface area and volume.\*\* The latter will play the chief role if all  $L_i \gg 1$ .

6. In conclusion, let us give results connected with the normal law of distribution. Let  $\Omega$  be a Borel space of dimension  $n$ , and let  $P_s(\cdot)$  be a family of normal probability measures with a fixed correlation matrix  $B$ , and  $s = (x^1, \dots, x^n)$  the vector of mean values. Then (5) the space  $S^n$  is Euclidean with metric tensor  $\|g_{ik}\| = B^{-1}$ . The capacity of a rectangular parallelepiped in this

\* Comparison with the Cramér-Rao inequality (5) shows that the scale of the metric (6) is determined by the variance of an effective estimate of the coordinate in the corresponding direction.

\*\* All geometric invariants (volume, area, length, etc.) in the metric (6) are expressed by dimensionless numbers.

in the space can be computed exactly:

$$M(L_1 \times L_2 \times \dots \times L_n) = \left(1 + \frac{L_1}{\sqrt{2\pi}}\right) \dots \left(1 + \frac{L_n}{\sqrt{2\pi}}\right). \quad (11')$$

For any simply connected convex domain  $G \subset S^n$ ,

$$M(G) = (2\pi)^{-n/2} V_G + O(\Sigma_G), \quad (12)$$

where

$$V_G = \int_G \dots \int \sqrt{g} dx^2 \dots dx^n$$

is the volume of the domain  $G$  in the metric (6), and  $\Sigma_G$  is the area of the hypersurface bounding the domain  $G$  ( $g$  is the determinant of the matrix  $\|g_{ik}\|$ ). For a convex domain  $G \subset S^2$ , the exact result has the form

$$M(G) = \frac{1}{2\pi} \Sigma_G + \frac{1}{2\sqrt{2\pi}} L_G + 1, \quad (13)$$

where  $\Sigma_G$  is the area, and  $L_G$  the perimeter, of the domain  $G$ .

The results presented above show that the geometry, consistent with the channel  $P_s(\cdot)$ , which arises in the manifold  $S^n$  upon introducing the Riemannian metric (6), largely determines the informational properties of the set of input signals and of its subsets.

The author expresses his gratitude to V. N. Sudakov and R. A. Zaidman for discussion of the results and valuable comments.

Received  
13 V 1965

## REFERENCES

1. I. M. Gelfand, A. N. Kolmogorov, A. M. Yaglom, DAN, **111**, 745 (1956).
2. A. Feinstein, *Foundations of Information Theory*, Moscow, 1960.

3. P. K. Rashevsky, *Riemannian Geometry and Tensor Analysis*, Moscow, 1953.
4. S. Kullback, *Information Theory and Statistics*, N. Y., 1959.
5. C. R. Rao, *Advanced Statistical Methods in Biometric Research*, N. Y., 1952.
6. H. Jeffreys, *Theory of Probability*, Oxford, 1948.

*Note: Figure translations are in progress. See original paper for figures.*

*Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.*