Coding of Signals with Finite Spectrum and Sound Recording Problems

A. G. Vitushkin

We will discuss one well-known problem of information theory, the problem which at present arises in various branches of radio engineering. We mean the problem of coding signals with finite spectrum. By way of an example, we consider how such problems arise, what comprises their mathematical content, and what conclusions can be drawn from results obtained. We will present an estimate of the length of codes for signals with finite spectrum and discuss it in connection with the problems of sound recording.

1. Raising of the question. Of the sound recording techniques the most widely used method is the so-called analogue method. When using this method a signal to be retained is recorded in its natural form without any preceding transformations. This system of recording is remarkable for its simplicity. The system's disadvantage is the impossibility of defending signals from interference. All defects of recording and reproducing devices, the inhomogeneity and aging of materials and the like lead to distortions in reproducing.

Another method of recording in which we are interested, the digital one, consists of the following: The signal is transformed into a discrete code, the code of the signal is recorded and, in order to be reproduced, it is again transformed into its natural continuous form. As far as this system is concerned, there are many ways of protecting signals from various sorts of noises. But in sound reproduction this system is not used because the existing schemes of coding still remain unacceptably complex.

Successful development of a digital recording system requires the construction of a handy mathematical model of sound signals and the discovery of simple schemes.
of coding. The first question arising is: How long must the codes be for the signals to be reproduced with a desirable accuracy? The qualitative estimate obtained seems encouraging.

2. The choice of the class of functions. The concept of signals with finite spectrum is usually associated with the Bernstein class of entire functions. We shall denote by \( B_\sigma \) the class of entire functions, real-valued on the real axis bounded in modulus by the constant 1 on the whole axis, and such that their Fourier transforms vanish outside the segment \([-\sigma, \sigma]\). We will call the functions of this class signals with spectrum \( \sigma \).

By Kotelnikov's theorem [1] the informational content of a signal with spectrum \( \sigma \) is proportional to \( \sigma \). Really, representing a function \( f \in B_\sigma \) in the form

\[
f(t) = \sum_{k=\infty}^{\infty} f\left(\frac{K\pi}{\sigma'}\right) \frac{\sin \sigma'(t - K\pi/\sigma')}{\sigma'(t - K\pi/\sigma')}
\]

(this representation is valid with any \( \sigma' > \sigma \)), we see that the number of parameters (per unit time) defining the function is proportional to \( \sigma \).


By the Kolmogorov-Tichomirov theorem [3] the entropy \( H_\varepsilon(B_\sigma, T) \) of the class \( B_\sigma \) (the norm being the maximum of a function on the segment \([-T, T]\)) satisfies the following inequality:

\[
\frac{2\sigma T}{\pi} \log \frac{c_1}{\varepsilon} \leq H_\varepsilon(B_\sigma, T) \leq \frac{2\sigma T}{\pi} \log \frac{c_2}{\varepsilon}
\]

where \( c_1 \) and \( c_2 \) are absolute constants.

It should be noted that this kind of theorem, formulated in terms of the uniform metric, has rather limited applications because in practice one has as a rule to deal with more complex forms of measurements.

3. The complexity of apparatus. Now we define the notion of an apparatus and the parameters characterizing its quality and complexity. An apparatus \( P \) is a pair of transformations \( P_1 \) and \( P_2 \) possessing the following properties.

A real-valued function \( f(t) \) defined on the whole real axis is transformed by the operator \( P_1 \) into a function \( \phi = \phi(K\pi, f) \) defined for all integers \( K \). Here \( \tau \) is a positive number constant for all input functions \( f(t) \). The function \( \phi \) may take only one of two values: either 1 or 0. In other words, the operator \( P_1 \) puts, in correspondence to the input function \( f(t) \), the sequence of binary numbers \( \phi(K\pi, f) \) \((K = -\infty, \ldots, 0, \ldots, \infty)\) uniformly distributed in time with the density \( \tau^{-1} \) per unit time. This sequence is called a binary code of the function \( f(t) \).

The second operator \( P_2 \) transforms the sequence \( \phi(K\pi, f) \) into a real-valued function \( f^* = P(f) \) defined on the whole real axis and bounded in modulus by the constant 1.

It is assumed, moreover, that there exists a positive constant \( l \) such that for any function \( f(t) \) and every integer \( K \) the value \( \phi(K\pi, f) \) depends only on the values of
the function \( f(t) \) at the segment \([K\tau - l, K\tau + l]\) and for every \( t \) the value \( f^*(t) \) depends only on the values \( P_t(K\tau) \) at the segment \( t - l \leq K\tau \leq t + l \).

The constant \( l \) is called a delay of the apparatus and the number \( h = \tau^{-1} \) is called a code density of the apparatus. If the condition of the boundedness of the apparatus delay were omitted from the definition, the notion of code density would not be strict. Really, by stretching the code sequence we can turn the code density into any desirable number.

The parameters \( h \) and \( l \) characterize, in a sense, the complexity of the apparatus.

4. The quality of apparatus. To describe the quality of reproduction we shall use three parameters: \( \sigma, \varepsilon \) and \( \delta \). But first of all we must say a few words about parameters used for the same purpose in engineering. The most essential of such parameters are the following. The first is \( \sigma \). It is the maximal frequency which can be reproduced by the apparatus. The second is \( \varepsilon \). This parameter characterises relative error of reproduction. The third is \( \delta = 20 \log_{10}(M/\delta) \). It is called the dynamic range of the apparatus. Here \( M \) is the maximum of the norm of output signals and \( \delta \) is the norm of apparatus noise. The norm of a signal is defined as

\[
\|f(t)\| = \max_t \left( \frac{1}{2\pi} \int_{t-r}^{t+r} \|f(x)\| \, dx \right)^{1/2}
\]

where \( \tau \) is a positive constant comparable to \( \delta^{-1} \).

**Definition.** We fix positive constants \( \sigma, \varepsilon, \delta \) and \( r \geq \sigma^{-1} \). Let \( f(t) \) and \( f^*(t) \) be two functions defined on the whole real axis. We will say that the function \( f^*(t) \) is close to \( f(t) \) if for any real \( t \) the following inequality is valid:

\[
|f(t) - f^*(t)| \leq \varepsilon \max_{t-r \leq x \leq t+r} |f(x)| + \delta.
\]

We will say that the parameters of an apparatus are not worse than \( \sigma, \varepsilon, \delta \) if for every function \( f \in B_{\sigma} \) the corresponding function \( f^* = P(f) \) is close to \( f(t) \). To put it otherwise, the apparatus has parameters \( \sigma, \varepsilon, \delta \) if it records and reproduces signals so precisely that for any signal with spectrum \( \sigma \) the corresponding output signal is close to the input one.

For an apparatus with parameters \( \sigma, \varepsilon, \delta \) the number \( \mathcal{D} = 20 \log_{10} \delta^{-1} \) is called the dynamic range of the apparatus. If an apparatus has a wide dynamic range, it means that both large and small signals can be reproduced with the same accuracy.

Thus all necessary definitions have been given and we can formulate the result.

5. Estimate of code density. For any positive numbers \( \sigma, \varepsilon \) and \( \delta \) it is possible to construct an apparatus the parameters of which are not worse than \( \sigma, \varepsilon, \delta \), while the complexity of the apparatus is characterized by the following inequalities:

\[
h \leq \left( \frac{\sigma}{\pi} \right) \log \left( \frac{c}{\varepsilon} \right) \quad \text{and} \quad l \leq \max \{c/\varepsilon, c/\delta\},
\]

where \( c \) is an absolute constant.

It should be pointed out that the right-hand side of the first inequality does not contain the parameter \( \delta \). It means that it is possible to construct an apparatus with any desirable dynamic range, using codes with the density which is independent of dynamic range.

This is rather unexpected because in engineering another point of view prevails:
A sufficiently wide dynamic range is the most difficult thing to obtain when one constructs an apparatus with the analogue system of recording.

But we need not think that wide dynamic range can be obtained without any difficulties at all. In the digital system, obtaining wide dynamic range requires either long codes or complex schemes of coding.

It should be noted as well that it is impossible to construct an apparatus with infinite dynamic range using codes of finite density.

6. Entropy of the class $B_\sigma$. The estimate of code density consists, as usual, in counting the entropy of the corresponding functional class.

Let the numbers $\sigma$, $\varepsilon$, $\delta$ and $r$ introduced above be fixed. Let $B^*$ be a set of functions defined on a segment $[-T, T]$. This set is called a net of the class $B_\sigma$ on the segment $[-T, T]$, if for any function $f \in B_\sigma$ there exists a function $f^* \in B^*$ close to $f$ on the segment $[-T, T]$, i.e., such that for any $t \in [-T, T]$ the following inequality is valid:

$$|f(t) - f^*(t)| \leq \varepsilon \max_{t-r \leq x \leq t+r} |f(x)| + \delta.$$  

Denote by $N(T)$ the number of elements of the minimal net of the set $B_\sigma$ on the segment $[-T, T]$. The number $H(T) = \log N(T)$ is called an $(\varepsilon, \delta)$-entropy of the set $B_\sigma$ on the segment $[-T, T]$.

**Theorem.** Let $\sigma, \varepsilon \leq 1$, $\delta \leq 1$ and $r \geq \sigma^{-1}$ be positive numbers. Then for any sufficiently large $T$ the entropy is

$$H(T) = \frac{2\sigma T}{\pi} \log \frac{c}{\max\{\varepsilon, \delta\}},$$

where $c$ is a positive function of $\sigma$, $\varepsilon$, $\delta$, $r$ which satisfies the inequality $c_1 \leq c \leq c_2$, where $c_1$ and $c_2$ are absolute positive constants [4].

Denote by $H = H(\sigma, \varepsilon, \delta)$ the minimum of code density $h = h(P)$ taken for all apparatuses with parameters $\sigma$, $\varepsilon$ and $\delta$. It can be easily shown that

$$H = \lim_{T \to \infty} \frac{1}{2T} H(T),$$

because for any $T$, on the one hand, any apparatus with parameters $\sigma$, $\varepsilon$ and $\delta$ generates a net of the class $B_\sigma$ on the segment $[-T, T]$ (this net is the set of all output signals when the input ones are all functions from the class $B_\sigma$) and, on the other hand, any net can be looked upon as an apparatus which puts, in correspondence to every function from $B_\sigma$, one of the nearest elements of the net.

So the theorem just formulated implies that

$$H = \frac{\sigma}{\pi} \log \frac{c}{\max\{\varepsilon, \delta\}},$$

i.e., the code density of the most economical apparatus with parameters $\sigma$, $\varepsilon$, $\delta$ is equal to $(\sigma/\pi) \log (c/\max\{\varepsilon, \delta\})$.

7. Estimate of polynomial derivatives. Now we present a result obtained while proving the above theorem. It seems to be interesting by itself.
Let $P(t)$ be a polynomial of degree $K$. Put $M = \max_{-1 \leq t \leq 1} |P(t)|$. By the Markov-Bernstein theorem the derivative $P'(0)$ at the point 0 satisfies the inequality $|P'(0)| \leq MK$. It is well known that this estimate is the least upper bound. Buslayev has found another form of estimating derivatives.

If the polynomial $P(t)$ has real coefficients, then

$$
|P'(0)| \leq \lambda M \left( 1 + q + \sum_{i=1}^{K-q} \frac{1}{|r_i|^2} \right),
$$

$\lambda$ is an absolute constant, $q$ is the number of the roots of the polynomial located in the disk $|t| \leq 1$ and $\{r_i\}$ are the roots of the polynomial located outside the disk.

Polynomials which arise as approximations of entire functions have widely scattered roots. For this kind of polynomial this estimate turns out to be much more effective than the Markov-Bernstein theorem. For polynomials with complex coefficients this estimate, generally speaking, is not valid. A counterexample is $P(t) = (1 + t/K^{1/2})^8$.

8. Some remarks. Returning to our main subject, the estimate of code density, we would like to make some remarks.

If we put $\delta = 0$ and take $\varepsilon$ sufficiently small, then the constants $1/2^K (K$ running over all positive integers) are pairwise distant, i.e., none of these constants is close to another. Hence the entropy $H(T) = \infty$ and consequently $H = \infty$. It means that there is no apparatus with an infinite dynamic range.

It will be recalled that the definition of the closeness of signals includes the parameter $r$. We have been assuming all the time that $r \geq \sigma^{-1}$. If we put $r = 0$, then the corresponding value $H$ turns out to be equal to $(\sigma/\pi) \log (c/\min\{\varepsilon, \delta\})$. The symbol $c$ is again understood as a positive function of all parameters separated from zero and infinity. We see that in the estimate of $H$ the symbol $\min\{\varepsilon, \delta\}$ is substituted for $\max\{\varepsilon, \delta\}$, i.e., in the case when $r = 0, \delta < \varepsilon$, the code density $H$ of the most economical apparatus turns out to be equal to $(\sigma/\pi) \log (c/\delta)$. We see that $H$ turns out to be essentially dependent on the parameter $\delta$.

This circumstance shows that the conclusion, that there exists an apparatus with a wide dynamic range and relatively small code density, is correct as much as the choice of metric is reasonable.

The notion of the closeness of signals has been defined to correspond to the system of measurements which at present is used in engineering. The condition $r \geq \sigma^{-1}$ seems to be natural as well because errors of reproduction are usually related to the energy of the signal per some period of time and not to the momentary value of the signal. For sinusoidal signals, for example, the error is usually related to the energy per one period of the oscillation. So there is hope that our choice of metric is reasonable and our conclusion is correct.

Now, in conclusion, it should be noted that our article about coding has been centred around the sound recording problems only to make the discussion more concrete. The estimate presented relates to arbitrary signals with finite spectrum and therefore can be used in other applications. For example, the result may be looked upon as the estimate of the capacity of a communication channel.
Any radio communication channel uses signals with finite spectrum and hence can be interpreted as an apparatus. In this case we may use the parameters \( \sigma, \varepsilon \) and \( \delta \) to characterise the frequency range of the channel, nonlinear distortions of the channel and the level of channel noise. The entropy \( H(T) \) of the corresponding class \( B_\sigma \) characterises the information content of signals and the number \( H(\sigma, \varepsilon, \delta) \) turns out to be equal to the channel capacity.

The fact that \( H \) does not essentially depend on the parameter \( \delta \) when \( \delta \) is sufficiently small with respect to \( \varepsilon \) means that the channel capacity does not depend in fact on the level of channel noise as soon as the noise is sufficiently small with respect to distortions.

**Literature**

1. V. A. Kotel'nikov, *Materials from the First All-Union Congress on questions of reconstruction of the connection and development of light industry*, Publishing House of the Workers' and Peasants' Red Army, 1933. (Russian)


**STEKLOV INSTITUTE**

**MOSCOW, U.S.S.R.**
Section 1

Mathematical Logic and the Foundations of Mathematics