We note that the results mentioned above can be extended without any essential difficulties to superpositions of the form

$$\sum_{i=1}^{N} p_i(x_1, \ldots, x_n) f_i(q_i(x_1, \ldots, x_n)),$$

where $\{p_i\}$ are preassigned continuous functions, $\{q_i\}$ are preassigned smooth functions and $\{f_i\}$ are arbitrary continuous functions of one variable. But as it turns out this does not apply to superpositions of the form

$$\sum_{i=1}^{N} p_i(x_1, \ldots, x_n) f_i(q_{1,i}(x_1, \ldots, x_n), \ldots, q_{k,i}(x_1, \ldots, x_n)),$$

where $\{p_i\}$ are fixed continuous functions of n variables; and $\{q_{1i}\}, \ldots, \{q_{ki}\}$ are fixed smooth functions of n variables ($k < n$). Fridman answered that question only for $n = 3, 4$, $k = 2$ and $\{p_i\} \equiv 1$.

Also it is not known to what extent the problem of superpositions of smooth functions can be reduced to that of linear superpositions. "Such a reduction is proved only in the case of the so called stable" superpositions [10]. It turns out that not every analytic function of n variables can be represented by means of superpositions of smooth functions of a smaller number of variables it is assumed that the scheme is stable, i.e. for a small perturbation of a function represented the perturbations of the functions composing the superposition are comparatively small.

Chapter 2. — Superpositions of smooth functions

In this chapter we prove the existence of smooth functions of n variables ($n \geqslant 2$), not representable by superpositions of smooth functions of a smaller number of variables.

§ 1. *The notion of entropy*

We will denote by $C(\mathscr{I})$ the space of all functions defined on a set $\mathscr{I}$ and continuous on $\mathscr{I}$ (the norm is the maximum of the absolute value of the function). We fix a compact $F \subset C(\mathscr{I})$ and a positive number ε. A set $F^* \subset C(\mathscr{I})$ is called an ε-net of F if for any $f \in F$ there exists $f^* \subset F^*$

such that $\| f - f^* \| \leqslant \varepsilon$. We denote by $N_\varepsilon(F)$ the number of elements of a minimal ε-net of F. The number $H_\varepsilon(F) = \log_2 N_\varepsilon(F)$ is called the ε-entropy of the set F.

The notion of entropy arises in a natural way in connection with various problems of analysis. We consider an example.

Let f be a function. It is known only that f belongs to a compact F. For example a smoothness condition of f and estimates of derivatives are given. We consider the problem of tabulating the function f. The first part of the problem is to write down in a table some number (parameters of f). For example, the values of f at certain points or the Taylor coefficients of f can be taken as such parameters. The second part of the problem is to present a decoding algorithm universal for all $f \in F$ which allows f to be calculated at any point with the accuracy ε.

The complexity of a table is usually characterized by two factors—its volume (the total number of binary digits required to write down all the parameters of the table) and the complexity of the decoding algorithm. It is easy to see that the volume of the most economical table presenting f with the accuracy ε equals $H_\varepsilon(F)$. Moreover it is possible to characterize the decoding algorithm too in terms of the entropy [21], [22], [24], [25].

It will be shown in paragraphs 2 and 3 that the number of ε-distant smooth functions depends in an essential way on the number of variables. This enables us to construct smooth functions of n variables not representable by smooth functions of a smaller number of variables.

We present here estimates of the entropy for a few concrete classes.

1. Let F_S^n be the class of all real valued functions, defined on a cube $\mathscr{I}: \{ 0 \leqslant x_i \leqslant 1, \mathrm{i} = 1, \ldots, n \}$ whose partial derivatives of order up to S are bounded in modulas by a constant C. Then

$$c' \left(\frac{1}{\varepsilon}\right)^{n/s} \leqslant H_\varepsilon(F_s^n) \leqslant c'' \left(\frac{1}{\varepsilon}\right)^{n/s},$$

where $C' > 0$, $C'' > 0$ are independent of ε.

2. Let $F^c_{\rho_1, \rho_2, \ldots, \rho_n}$ be the space of functions analytic on the n-dimensional cube $\{ -1 \leqslant x_k \leqslant 1 \}$ $(k = 1, 2, \ldots, n)$ having analytic continuations in the region $E_\rho = E_{\rho_1} \times E_{\rho_2} \times \ldots \times E_{\rho_n}$ which are bounded in modulus in this region by the constant $C > 0$, where E_{ρ_k} is the region of the complex plane $z_k = x_k + iy_k$ bounded by the ellipse with semi-major axis ρ_k and with foci at the points -1, 1 of the real axis $(k = 1, 2, \ldots, n)$. Then

$$H_\varepsilon(F^c_{\rho_1,\rho_2,\dots,\rho_n}) - \frac{1}{(n+1)!}\prod_{k=1}^{n}\frac{1}{\log\ \rho_k}\left(\log\frac{c}{\varepsilon}\right)^{n+1} + O\left[\left(\log\frac{c}{\varepsilon}\right)^n \log\log\frac{c}{\varepsilon}\right].$$

3. Let $F^n_{s,c}$ be the class of real valued functions on the cube $\{-1 \leqslant x_k \leqslant 1\}$ $(k=1, \dots, n)$, bounded in modulus on that cube by the constant s_k and such that their analytic extensions are entire functions of order s_k with respect to $z_k = x_k + iy_k$ $(k=1, \dots, n)$. Then

$$H_\varepsilon(F^n_{s,c}) = \frac{1}{(n+1)!}\prod_{k=1}^{n} s_k \left(\log\frac{c}{\varepsilon}\right)^{n+1}\left(\log\ \log\frac{c}{\varepsilon}\right)^{-n} + $$
$$= O\left[\left(\log\frac{c}{\varepsilon}\right)^{n+1}\left(\log\ \log\frac{c}{\varepsilon}\right)^{-n-1}\right].$$

These estimates and other results connected with estimates of entropy and applications are to be found for example in [49]-[53].

§ 2. *The entropy of the space of smooth functions*

Here we give an estimate of the entropy of the class of S times differentiable functions of n variables. The lower estimate was obtained in [4], the upper one—in [23].

We fix integers $n \geqslant 1$ and $p \geqslant 0$ and numbers $0 \leqslant \alpha \leqslant 1$, $L > 0$, $C > 0$, $\rho > 0$. We will denote by $\mathcal{I}$ the cube $0 \leqslant x_i \leqslant \rho$ $(i=1, \dots, n)$ and by $F = F^{\rho, n}_{S, L, c}$ $(S = p + \alpha)$ the set of all real valued functions defined on $\mathcal{I}$ such that their partial derivatives of order p satisfy the condition Lip α with the constant L and

$$\left|\frac{\partial^{k_1+\dots+k_n} f(0)}{\partial^{k_1} x_1 \dots \partial^{k_n} x_n}\right| \leqslant c \quad \left(\sum_{i=1}^{n} k_i \leqslant p\right)$$

We say that the function $g(x)$ satisfies the condition Lip α with the constant L if for any x' and x''

$$|g(x') - g(x'')| \leqslant L(r(x', x''))^\alpha,$$

where $r(x', x'')$ is the distance between x' and x''.

THEOREM 2.2.1. *If* $\varepsilon > 0$ *is sufficiently small then*

$$A\rho^n\left(\frac{L}{\varepsilon}\right)^{n/s} \leqslant H_\varepsilon(F) \leqslant B\rho^n\left(\frac{L}{\varepsilon}\right)^{n/s},$$

where A *and* B *are positive constants depending only on* s *and* n.