J. SIAM Numer. Anal.
Ser. B, Vol. 1
Printed in U.S.A., 1964

A PRECISION APPROXIMATION OF THE GAMMA FUNCTION*

C. LANCZOS†

The gamma function is one of the most interesting transcendentals of higher mathematics. It is an analytic function of the complex variable z which in any finite domain has no singularities other than simple poles, situated at the points $x = 0, -1, -2, -3, \cdots$. For integer values the gamma function coincides with the ordinary factorials, according to the relation

$$\Gamma(n + 1) = n! \tag{1}$$

The normalization of the gamma function to $\Gamma(n + 1)$ instead of $\Gamma(n)$ is due to Legendre and void of any rationality. This unfortunate circumstance compels us to utilize the notation $z!$ instead of $\Gamma(z + 1)$, although this notation is obviously highly unsatisfactory from the operational point of view.

Euler gave the well-known integral representation of the factorial function:

$$z! = \Gamma(z + 1) = \int_0^\infty t^z e^{-t}\, dt, \tag{2}$$

which defines $z!$ as an analytical function of z for all z whose real part is greater than -1. We will restrict ourselves to the right complex half plane by the condition

$$\operatorname{Re} z \geqq 0 \tag{3}$$

because the reflexion theorem

$$(-z)!\, z! = \frac{\pi z}{\sin \pi z} \tag{4}$$

reduces the definition of $(-z)!$ to the definition of $z!$. Our following discussions will be based on Euler's integral, which we will transform into a form that is particularly well suited for approximation purposes.

Replacing t by αt we obtain

$$z! = \alpha^{z+1} \int_0^\infty t^z e^{-\alpha t}\, dt. \tag{5}$$

Since z is a constant with respect to the process of integration, it is permissible to put

$$\alpha = 1 + \rho z, \tag{6}$$

* Received by the editors November 4, 1963.

† School of Theoretical Physics, Dublin Institute for Advanced Studies, 64 Merrion Square, Dublin, Ireland.

where we will consider ρ as some positive constant. This yields

$$(7) \qquad z! = (1 + \rho z)^{z+1} \int_0^\infty (te^{-\rho t})^z e^{-t}\, dt.$$

Furthermore, we put the constant factor $(e\rho)^{-z}$ in front and its reciprocal behind the integral sign, thus changing (7) to

$$(8) \qquad z! = (1 + \rho z)^{z+1} (e\rho)^{-z} \int_0^\infty (\rho t e^{1-\rho t})^z e^{-t}\, dt,$$

and now we put

$$e^{1-\rho t} = v.$$

With this change of variable our definition of $z!$ becomes

$$(9) \qquad \begin{aligned} z! &= (1 + \rho z)^{z+1} \rho^{-(z+1)} e^{-z} \int_0^e [v(1 - \log v)]^z \left(\frac{v}{e}\right)^{1/\rho} \frac{dv}{v} \\ &= \left(\frac{1}{\rho} + z\right)^{z+1} e^{-z-1/\rho} \int_0^e [v(1 - \log v)]^z v^{1/\rho - 1}\, dv. \end{aligned}$$

Finally we replace ρ by a new constant γ, defined by

$$1/\rho = \gamma + 1.$$

Hence the final form of the definite integral, on which we want to base the investigation of $z!$, becomes

$$(10) \qquad z! = (z + \gamma + 1)^{z+1} e^{-(z+\gamma+1)} \int_0^e [v(1 - \log v)]^z v^\gamma\, dv.$$

The constant γ, which we have at our disposal, will turn out later to be of great value for the precision of our approximation. For the time being we carry it along as a given positive (or zero) constant.

Replacing z by $z - \frac{1}{2}$ we can equally put

$$(11) \qquad (z - \tfrac{1}{2})! = (z + \gamma + \tfrac{1}{2})^{z+\frac{1}{2}} e^{-(z+\gamma+\frac{1}{2})} \int_0^e [v(1 - \log v)]^{z-\frac{1}{2}} v^\gamma\, dv.$$

Our aim will be to study more closely the integral transform

$$(12) \qquad F(z) = \int_0^e [v(1 - \log v)]^{z-\frac{1}{2}} v^\gamma\, dv.$$

The function $v(1 - \log v)$ has the following course. It starts at $v = 0$ with zero value and returns to zero at $v = e$. It reaches its maximum value 1 at $v = 1$. (See Fig. 1.)

We will now transform the integration variable v into a new variable x

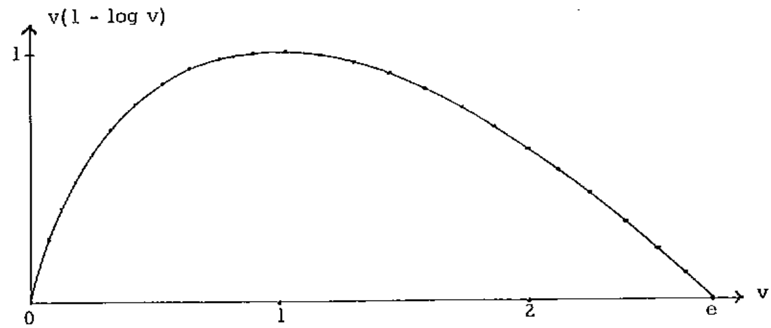


FIGURE 1

by the implicit transcendental equation

$$v(1 - \log v) = 1 - x^2, \tag{13}$$

with the boundary conditions:

$$v = 0 \quad \text{corresponds to} \quad x = -1,$$

$$v = 1 \quad \text{corresponds to} \quad x = 0,$$

$$v = e \quad \text{corresponds to} \quad x = 1.$$

Then $F(z)$ appears in the form

$$F(z) = \int_{-1}^{+1} (1 - x^2)^{z-\frac{1}{2}} v^\gamma \frac{dv}{dx}\, dx, \tag{14}$$

and if we change x to the angle variable θ by putting

$$x = \sin\theta, \tag{15}$$

we obtain the integral transform

$$F(z) = \int_{-\pi/2}^{+\pi/2} \cos^{2z}\theta\, f(\theta)\, d\theta, \tag{16}$$

where we have put

$$f(\theta) = v^\gamma \frac{dv}{dx} = v^\gamma \frac{dv}{\cos\theta\, d\theta}. \tag{17}$$

Now the implicit equation (13) can be changed to the differential equation

$$v' \log v = 2x,$$

or, substituting $\log v$ from (13), we obtain for $v(x)$ the nonlinear differential equation of first order

$$\tfrac{1}{2}(v^2)' - (1 - x^2)v' - 2xv = 0, \tag{18}$$

with the boundary condition

$$v(0) = 1.$$

The study of this differential equation shows that $v(x)$ is a function of the complex variable x which is analytical inside the unit circle $|x| < 1$, the only singular point occurring at $x = -1$, where $v(x)$ has a logarithmic singularity. Although $v(-1)$ goes to zero, and even

$$v'(x) = \frac{2x}{\log v}$$

goes to zero, yet $v''(-1)$ goes to infinity.

We thus see that $v(x)$ can be expanded in a Taylor series around $x = 0$ which converges uniformly for all $|x| < 1$ but the convergence becomes infinitely slow in the vicinity of the point $x = -1$. By putting the formal expansion

$$v(x) = 1 + a_1x + a_2x^2 + \cdots,$$

in the differential equation (18), we obtain recurrence relations for the a_k which can be solved in succession, obtaining

$$a_1 = \sqrt{2}, \qquad a_3 = -\frac{\sqrt{2}}{36}, \qquad a_5 = -\frac{23}{27}\frac{\sqrt{2}}{160}, \cdots,$$

$$a_2 = \frac{1}{3}, \qquad a_4 = \frac{2}{135}, \qquad a_6 = \frac{38}{8505}, \cdots,$$

and thus

$$\frac{dv}{dx} = a_1 + 3a_3\,x^2 + 5a_5\,x^4 + \cdots + 2a_2\,x + 4a_4\,x^3 + \cdots.$$

If for the time being we put $\gamma = 0$ and substitute this expansion in (16) for $f(\theta)$, all the odd powers vanish on account of the integration and we obtain

$$F(z) = \sqrt{2}\int_{-\pi/2}^{+\pi/2} \cos^{2z}\theta\left(1 - \frac{3}{36}\sin^2\theta - \frac{23}{27}\frac{5}{160}\sin^4\theta - \frac{11237}{5443200}\,7\sin^6\theta - \cdots\right)d\theta. \tag{19}$$

Hence we need an infinity of definite integrals which are, however, available in closed form:

$$\int_{-\pi/2}^{+\pi/2} \cos^{2z}\theta\,\sin^{2k}\theta\,d\theta = \frac{(z - \frac{1}{2})!\,(k - \frac{1}{2})!}{(z + k)!}. \tag{20}$$

Hence

$$F(z) = \sqrt{2\pi}\,\frac{(z-\frac{1}{2})!}{z!}\left[1 - \frac{1}{24}\frac{1}{z+1} - \frac{23}{1152}\frac{1}{(z+1)(z+2)} - \frac{11237}{414720}\frac{1}{(z+1)(z+2)(z+3)} - \cdots\right],$$

and substituting in (11) we obtain the following convergent expansion for $z!$:

$$z! = \sqrt{2\pi}(z+\tfrac{1}{2})^{z+\frac{1}{2}}e^{-(z+\frac{1}{2})}\left[1 - \frac{1}{24}\frac{1}{z+1} - \frac{23}{1152}\frac{1}{(z+1)(z+2)} - \frac{11237}{414720}\frac{1}{(z+1)(z+2)(z+3)} - \cdots\right]. \tag{21}$$

It is of interest to compare this formula with the celebrated formula of Stirling:

$$z! = \sqrt{2\pi}z^{z+\frac{1}{2}}e^{-z}\left[1 + \frac{1}{12z} + \frac{1}{288z^2} + \frac{139}{51840}\frac{1}{z^3}\cdots\right].$$

The latter formula is divergent for all values of z, but can be used in the sense of an asymptotic expansion by terminating the series at the proper point. On the other hand, the expansion (21) converges for all values of z in the right half plane, but the convergence is too slow to be of practical value.

Our next aim will be to discover some means by which the convergence of our expansion can be improved. This can be done on the basis of the following consideration. The Taylor series is an extrapolating series and thus naturally of slow convergence. We can hope for much better results if we operate with an orthogonal set of functions which within its domain of orthogonality interpolates rather than extrapolates the given function. We have considered (16) as an integral transform of the function $f(\theta)$ to $F(z)$. In the integrand of (16) $f(\theta)$ is multiplied by an even function of θ. If we write $f(\theta)$ as a sum of its even and odd part:

$$f(\theta) = \tfrac{1}{2}[f(\theta) + f(-\theta)] + \tfrac{1}{2}[f(\theta) - f(-\theta)]$$

we notice that the contribution of the second part to the definite integral (16) vanishes, since the integration is taken between equal $\pm$ limits. Hence $f(\theta)$ can be replaced (for $\gamma = 0$) by

$$\tfrac{1}{2}[f(\theta) + f(-\theta)] = \frac{1}{2\cos\theta}\left[\frac{dv}{d\theta}(\theta) + \frac{dv}{d\theta}(-\theta)\right] = a_1 + 3a_3\sin^2\theta + 5a_5\sin^4\theta + \cdots \tag{22}$$

With this replacement, $f(\theta)$ is an *even* and *periodic* function of θ, of period π. Such a function can be expanded into a convergent Fourier series, of the form

$$(23) \qquad f(\theta) = \tfrac{1}{2}c_0 + c_1\cos 2\theta + c_2\cos 4\theta + \cdots ,$$

with

$$(24) \qquad c_k = \frac{2}{\pi}\int_{-\pi/2}^{+\pi/2} f(\theta)\cos 2k\theta \, d\theta.$$

We might think that it will be difficult to evaluate these integrals, in view of the fact that $f(\theta)$ is a complicated function. In fact, they are explicitly available, with the help of the integral tranform $F(z)$. For this purpose we express $\cos 2k\theta$ in terms of $\cos^{2\alpha}\theta$, making use of the coefficients of the Chebyshev polynomials:

$$T_{2k}(\xi) = \sum_{\alpha=0}^{k} C_{2\alpha}^{2k}\,\xi^{2\alpha},$$

$$\cos 2k\theta = \sum_{\alpha=0}^{k} C_{2\alpha}^{2k}\cos^{2\alpha}\theta.$$

Hence

$$(25) \qquad c_k = \frac{2}{\pi}\int_{-\pi/2}^{+\pi/2} f(\theta)\sum_{\alpha=0}^{k} C_{2\alpha}^{2k}\cos^{2\alpha}\theta \, d\theta = \frac{2}{\pi}\sum_{\alpha=0}^{k} C_{2\alpha}^{2k} F(\alpha).$$

We see that by *weighting the function values* $F(0), F(1), \cdots, F(k)$ *by the coefficients of the Chebyshev polynomials, we obtain the numerical values of the Fourier coefficients* c_k *in explicit form.*

Having obtained the c_k we now substitute the expansion (23) in (16), obtaining

$$(26) \qquad F(z) = \int_{-\pi/2}^{+\pi/2}\cos^{2z}\theta\left(\tfrac{1}{2}c_0 + c_1\cos 2\theta + c_2\cos 4\theta + \cdots\right) d\theta.$$

The definite integrals demanded by this formula are once more available in closed form:

$$\int_{-\pi/2}^{+\pi/2}\cos^{2z}\theta\cos 2k\theta \, d\theta = \sqrt{\pi}\,\frac{(z-\frac{1}{2})!}{(z+k)!}\,\frac{z!}{(z-k)!},$$

which yields

$$F(z) = \sqrt{\pi}\,\frac{(z-\frac{1}{2})!}{z!}\left[\frac{1}{2}c_0 + c_1\frac{z}{z+1} + c_2\frac{z(z-1)}{(z+1)(z+2)} + \cdots\right].$$

Finally it will be advantageous to take the numerical factor $\sqrt{2}$ in front and write our formula in the form

$$F(z) = \sqrt{2\pi}\,\frac{(z-\frac{1}{2})!}{z!}\left[\frac{1}{2}c_0' + c_1'\frac{z}{z+1} + c_2'\frac{z(z-1)}{(z+1)(z+2)} + \cdots\right], \tag{27}$$

with

$$c_k' = \frac{\sqrt{2}}{\pi}\sum_{\alpha=0}^{k} C_{2\alpha}^{2k}\,F(\alpha), \tag{28}$$

where

$$F(\alpha) = (\alpha - \tfrac{1}{2})!\,(\alpha+\tfrac{1}{2})^{-(\alpha+\frac{1}{2})}\,e^{\alpha+\frac{1}{2}}, \tag{29}$$

and

$$z! = \sqrt{2\pi}(z+\tfrac{1}{2})^{z+\frac{1}{2}}e^{-(z+\frac{1}{2})}\left[\frac{1}{2}c_0' + c_1'\frac{z}{z+1} + \cdots\right]. \tag{30}$$

This expansion is not yet sufficient for an effective approximation of the factorial function, although it is of interest that the two-term formula

$$z! = \sqrt{2\pi}(z+\tfrac{1}{2})^{z+\frac{1}{2}}e^{-(z+\frac{1}{2})}\left(1.0163 - \frac{0.0861}{z+1} + \epsilon\right), \quad |\epsilon| < 0.02, \tag{31}$$

is already accurate to a relative error not exceeding 2% at any point of the right complex half plane.

The reason $v(\theta)$ has a slowly convergent Fourier series is that the function dv/dx, although going to zero at $x = -1$, has an infinitely large tangent at that point. Hence there is a relatively mild "kink" at the end of the range which is of small extension. We can imagine that by taking the square, or cube, or some still higher power of this kink, the influence of the singularity will be greatly reduced and thus the convergence of the Fourier series greatly increased. This now can be achieved by making use of the constant γ which is equivalent to replacing v by $v^{\gamma+1}/(\gamma+1)$, as the formula (17) shows. For $\gamma = 0$ the accuracy of our approximation does not improve much by taking in more terms of the expansion, on account of the slow convergence of the coefficients c_k'. But let us now choose $\gamma = 1$. Here the successive coefficients diminish much more rapidly and when we reach the point beyond which the convergence becomes slow, the coefficients are already very small. Generally, the higher γ becomes, the smaller will be the value of the coefficients at which the convergence begins to slow down. At the same time, however, we have to wait longer, before the asymptotic stage is reached. Hence a large value of γ is advocated if very high accuracy is demanded, but then the required number of terms will also be larger.

We now have

(32) $$z! = \sqrt{2\pi}(z + \gamma + \tfrac{1}{2})^{z+\frac{1}{2}} e^{-(z+\gamma+\frac{1}{2})} A_\gamma(z),$$

where

(33) $$A_\gamma(z) = \tfrac{1}{2}\rho_0 + \rho_1 \frac{z}{z+1} + \rho_2 \frac{z(z-1)}{(z+1)(z+2)} + \cdots,$$

with

(34) $$\rho_k = \sum_{\alpha=0}^{k} C_{2\alpha}^{2k} F(\alpha),$$

where

(35) $$F(\alpha) = \frac{\sqrt{2}}{\pi} (\alpha - \tfrac{1}{2})! \, (\alpha + \gamma + \tfrac{1}{2})^{-(\alpha+\frac{1}{2})} e^{\alpha+\gamma+\frac{1}{2}}.$$

Table 1 gives a good idea of the manner in which the ρ_k coefficients decrease with increasing values of γ.

TABLE 1

	$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$	$\gamma = 3$
$\frac{1}{2}\rho_0$	1.459843	2.0844142	3.07380467	7.06165881
ρ_1	−0.460642	−1.0846349	−2.11237574	−6.59935794
ρ_2	0.001054	0.0001207	0.03862116	0.53965224
ρ_3	−0.000338	0.0001145	−0.00005100	−0.00195197
ρ_4	0.000118	−0.0000171	0.00000048	−0.00000133
ρ_5	−0.000051	0.0000018	0.00000067	0.00000022

A good check on the accuracy of the truncated Fourier series is provided by evaluating the approximate value of $f_\gamma(\theta)$ at $\theta = -\pi/2$, that is by forming the alternate sum

$$\tfrac{1}{2}\rho_0 - \rho_1 + \rho_2 - \rho_3 + \cdots = f_\gamma\left(-\frac{\pi}{2}\right) = \frac{e^\gamma}{\sqrt{2}}.$$

We know from the theory of the Fourier series that the maximum local error (after reaching the asymptotic stage) can be expected near to the point of singularity. Let this error be η. Then a simple estimation shows that the influence of this error on the integral transform (16) (for values of z which stay within the right complex half plane), cannot be greater than $(\pi/2)\eta$. Thus we can give a definite error bound for the approximation obtained.

Finally, it is convenient to resolve the rational functions which appear in (34), into their constituent partial fractions:

$$\frac{z}{z+1} = 1 - \frac{1}{z+1}, \qquad \frac{z(z-1)}{(z+1)(z+2)} = 1 + \frac{2}{z+1} - \frac{6}{z+3},$$

$$\frac{z(z-1)(z-2)}{(z+1)(z+2)(z+3)} = 1 - \frac{3}{z+1} + \frac{24}{z+2} - \frac{30}{z+3},$$

$$\frac{z(z-1)(z-2)(z-3)}{(z+1)(z+2)(z+3)(z+4)} = 1 + \frac{4}{z+1} - \frac{60}{z+2} + \frac{180}{z+3} - \frac{140}{z+4}.$$

The final results can be tabulated as follows:

$$z! = (z+\gamma+\tfrac{1}{2})^{z+\frac{1}{2}} e^{-(z+\gamma+\frac{1}{2})} \sqrt{2\pi}[A_\gamma(z) + \epsilon].$$

$\underline{\gamma = 1}$

$$A_1(z) = 0.9992 + \frac{0.46064}{z+1}, \qquad |\epsilon| < 0.001,$$

$\underline{\gamma = 1.5}$

$$A_{1.5}(z) = 0.999779 + \frac{1.084635}{z+1}, \qquad |\epsilon| < 0.00024,$$

$\underline{\gamma = 2}$

$$A_2(z) = 1.0000509 + \frac{2.18961806}{z+1} - \frac{0.23172696}{z+2}, \qquad |\epsilon| < 5.1\cdot 10^{-5},$$

$$A_2(z) = 0.99999909 + \frac{2.18977107}{z+1} - \frac{0.23295108}{z+2} + \frac{0.00153015}{z+3}, \qquad |\epsilon| < 1.5\cdot 10^{-6},$$

$\underline{\gamma = 3}$

$$A_3(z) = 1.00000114 + \frac{7.68451833}{z+1} - \frac{3.28476072}{z+2} + \frac{0.05855910}{z+3}, \qquad |\epsilon| < 1.4\cdot 10^{-6},$$

$\underline{\gamma = 4}$

$$A_4(z) = 0.9999999469 + \frac{24.7158060592}{z+1} - \frac{19.2112843044}{z+2} + \frac{2.4635062800}{z+3} - \frac{0.0096933620}{z+4}, \qquad |\epsilon| < 5\cdot 10^{-8},$$

$\underline{\gamma = 5}$

$$A_5(z) = 1.000000000178 + \frac{76.180091729406}{z+1} - \frac{86.505320327112}{z+2} + \frac{24.014098222230}{z+3} - \frac{1.231739516140}{z+4} + \frac{0.001208580030}{z+5} - \frac{0.000005363820}{z+6}, \qquad |\epsilon| < 2\cdot 10^{-10}.$$

Particularly remarkable is the approximation of only two terms ($\gamma = 1.5$):

$$z! = (z+2)^{z+\frac{1}{2}} e^{-(z+2)} \sqrt{2\pi} \left(0.999779 + \frac{1.084635}{z+1}\right),$$

which is correct up to a relative error of $2.4\cdot 10^{-4}$ everywhere in the right complex half plane.

The error of the truncated series would rapidly increase if z moved over to the negative half plane. It is of interest to observe, however, that the convergence of the infinite expansion extends even to the negative realm and is in fact limited by the straight line

$$\operatorname{Re} z = -(\gamma + \tfrac{1}{2}).$$

The larger γ becomes, the more the domain shrinks in which the series diverges. If γ grows to infinity, we obtain a representation of the factorial function which holds *everywhere* in the complex plane. In this case we are able to give the coefficients of the series (34) in explicit form, due to the extreme nature of the function v^γ. We thus obtain the following limit relation:

$$z! = 2 \lim_{\gamma\to\infty} \gamma^z \left[\frac{1}{2} - e^{-1/\gamma} \frac{z}{z+1} + e^{-4/\gamma} \frac{z(z-1)}{(z+1)(z+2)} - \cdots\right]$$
$$= 2 \lim_{\gamma\to\infty} \gamma^z \sum_{k=0}^{\infty}{}' \, (-1)^k e^{-k^2/\gamma} \frac{\binom{z}{k}}{\binom{z+k}{k}}$$

valid for all values of z. The proper interpretation of this peculiar limit law and its possible relation to the same series but with finite values of γ, requires further investigation.

Acknowledgments. The basic idea of this approximation method goes back to the year 1952, when the author was associated with the Institute for Numerical Analysis, University of California, Los Angeles. The final

formulation was accomplished in the winter of 1959–60, when by the kind invitation of Professor Rudolph E. Langer the author had the privilege of enjoying the excellent research opportunities provided by the Mathematics Research Center, U.S. Army, University of Wisconsin, Madison, Wisconsin.

REFERENCES

[1] HAROLD T. DAVIS, *Tables of the Higher Mathematical Functions*, vol. 1, The Principia Press, Bloomington, Indiana, 1933.

[2] E. T. WHITTAKER, AND G. N. WATSON, *A Course of Modern Analysis*, Chapter 12, 4th ed., Cambridge University Press, 1935.

J. SIAM Numer. Anal.
Ser. B, Vol. 1
Printed in U.S.A., 1964

ENTROPY AND ITS APPLICATIONS*

G. G. LORENTZ†

1. Metric entropy. Examples. I will explain here the notion of metric entropy, and some of its applications in analysis.

For a compact metric space A, the *metric entropy* was devised by Kolmogorov to describe the massivity, the thickness of the set A. It is a function of a positive argument $\epsilon > 0$, $H_\epsilon(A)$, and its mode of convergence to infinity for $\epsilon \to 0$ describes the set A. A companion invariant of A is the *capacity* $C_\epsilon(A)$ of A.

Definitions. Subsets $U_1, \cdots, U_n$ of A are an ϵ-*cover* of A if the diameter of each U_i is $\leqq 2\epsilon$ and if $A \subset \bigcup_{i=1}^n U_i$. The number n depends on the cover, but

$$\min n = N(\epsilon)$$

is an invariant of the set A;

$$H_\epsilon(A) = \log N_\epsilon(A) \tag{1}$$

is the entropy of A. We are interested, of course, in the asymptotic behavior of $H_\epsilon(A)$ for $\epsilon \to 0$.

To define capacity, we say that the points $y_1, \cdots, y_m$ of A are ϵ-distinguishable if $\rho(y_i, y_j) > \epsilon$, $i \neq j$. Then

$$C_\epsilon(A) = \log M(\epsilon) \tag{2}$$

where

$$M(\epsilon) = \max m$$

is the capacity of A.

One should bear in mind that H_ϵ and C_ϵ are only defined in the set A and its metric are given. An inequality which is easy to establish but is quite important is the following:

$$C_{2\epsilon}(A) \leqq H_\epsilon(A) \leqq C_\epsilon(A). \tag{3}$$

This is used to obtain asymptotic estimates of H_ϵ and C_ϵ; an exact determination is rarely possible. For more details, see [1, 2].

We give some examples.

Example 1. If A is an s-dimensional compact set with interior points, with its euclidean metric, then

$$H_\epsilon(A) \sim s \log \frac{1}{\epsilon}, \quad \text{and even} \quad H_\epsilon(A) = s \log \frac{1}{\epsilon} + O(1). \tag{4}$$

* Received by the editors November 4, 1963. This review has been prepared with support (in part) of the grant AF-AFOSR-424-63 of the Office of Scientific Research of the U. S. Air Force.

† Department of Mathematics, Syracuse University, Syracuse, New York.

Example 2. Let $A_r(C)$, $r > 1$, be the set of all functions analytic in $|z| < r$, continuous in $|z| \leqq r$, which satisfy $|f(z)| \leqq C$ in $|z| \leqq r$. We take as distance the uniform norm on $|z| \leqq 1$: $\|f\| = \max_{|z|\leqq 1} |f(z)|$. Then

$$(5) \qquad H_\epsilon(A) \sim \frac{1}{\log r}\left(\log\frac{1}{\epsilon}\right)^2.$$

The first factor describes the geometric situation, the exponent 2 the number of complex variables. It has to be replaced by $s + 1$ for analytic functions of s variables. Relation (5) is obtained by means of the parametrization $f = \Sigma a_n z^n$ and the determination of the amount of freedom for the a_n.

Example 3. Let $\Lambda^s_{p+\alpha}$, $p = 0, 1, \cdots, 0 < \alpha \leqq 1$, $s = 1, 2, \cdots$, be the set of functions $f(x) = f(x_1, \cdots, x_s)$ of s variables on the unit cube such that all partial derivatives $D^r f$, $0 \leqq r \leqq p$ exist, satisfy $|D^r f| \leqq C_r$, and also the Lipschitz condition $|D^p f(x) - D^p f(x')| \leqq M\,|x - x'|^\alpha$. We give $\Lambda^s_{p+\alpha}$ the uniform norm. Then

$$(6) \qquad H_\epsilon(\Lambda^s_{p+\alpha}) \approx \epsilon^{-s/(p+\alpha)}.$$

This formula confirms the idea that "large" sets A have large entropy. The larger $p + \alpha$, the smaller, simpler is the set $A = \Lambda^s_{p+\alpha}$; the larger s, the larger, more complicated is A. The means of proof is here the finite Taylor formula for f, which allows us to pass from one point x of the cube Q to another x', at a small distance from x.

Example 4. In the space l^2, with the usual norm, let E be the ellipsoid

$$E: \qquad \sum_1^\infty d_n^{-2} x_n^2 \leqq 1.$$

For this set E, Kolmogorov proves

$$(7) \qquad \int_0^{1/2\epsilon} \frac{m(t)}{t}\,dt \leqq H_\epsilon(E) \leqq C\int_0^{2/\epsilon} \frac{m(t)}{t}\,dt,$$

where $m(t)$ is defined by

$$m(t) = \max\left[n : d_n \geqq \frac{1}{t}\right].$$

This is obtained in the following way: E is approximated by a finite-dimensional ellipsoid E_s. If a covering $\bigcup_{i=1}^n U_i$ of E_s is given, we can write an obvious inequality for the volumes of E_s and the U_i. This allows us to estimate n. A similar consideration is applied to a set of ϵ-distinguishable points of E_s.

Mityagin [3] has applied inequalities of type (7) in spaces l^p to characterize and classify the nuclear spaces.

2. Applications of entropy. In 1960, a Russian book of Vitushkin appeared, called "Estimation of the Complexity of the Tabulation Problem." It was later translated into English as "Theory of Transmission and Processing of Information." The main subject of the book is nonlinear approximation. To explain what this means, let us consider an example.

One of the "direct" theorems of approximation is the following: If $F = \text{Lip } \alpha$, i.e., if F is the family of 2π-periodic functions f which satisfy $|f(x) - f(x')| \leqq |x - x'|^\alpha$, then for some $M > 0$ the following is true: for each $f \in \text{Lip } \alpha$ there is a trigonometric polynomial T_n of degree n so that $|f(x) - T_n(x)| \leqq M/n^\alpha$. If M is replaced by a small number $m > 0$, this is no longer true. This means the following: if for each $f \in \text{Lip } \alpha$, one has $|f(x) - T_n(x)| < \epsilon$ for properly chosen T_n, then $\epsilon > m/n^\alpha$. Solving for n, we obtain

$$n > \text{Const } \epsilon^{-1/\alpha} \geqq \text{Const } H_\epsilon(F) \tag{1}$$

(see Example 3). Vitushkin's algorithms of approximation are not linear combinations

$$P_n(x) = a_1\phi_1(x) + \cdots + a_n\phi_n(x),$$

not even generalized rational functions

$$P(x) = \frac{a_1\,\phi_1(x) + \cdots + a_n\,\phi_n(x)}{a_1\,\psi_1(x) + \cdots + a_n\,\psi_n(x)},$$

they are more generally of the form

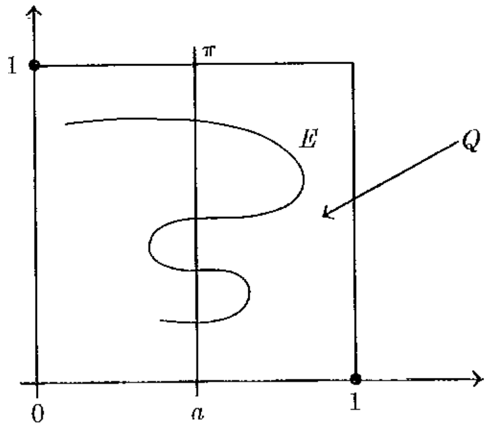
$$P(x) = \frac{Q_k(a_1, \cdots, a_n; x)}{R_k(a_1, \cdots, a_n; x)}, \qquad x \in Q; \tag{2}$$

where x is a point in an s-dimensional cube Q, and Q_k, R_k are polynomials of degree k in the constants $a_1, \cdots, a_n$, whose coefficients are *given* functions of x. If a family F of functions on Q is given, and for each $f \in F$ there is a P of the form (2) with $|f(x) - P(x)| < \epsilon$, $x \in Q$, then P is called an *ϵ-algorithm*. Vitushkin asks the following question. It is known that F has an ϵ-algorithm (2). What can be said about F, ϵ, k, n? The answer is roughly

$$n \log(k + 1) \geqq CH_{c_1\epsilon}(F) \tag{3}$$

(for some families F.)

Before formulating the results more precisely, I will describe the means involved in Vitushkin's proof. The first two of these refer to an s-dimensional space R_s. In order to define the *variation* of a point set E in an s-dimensional cube Q assume first that E is a curve in the square $Q(s = 2)$. For each a, let $n(a)$ be the number of intersections of $\pi: x = a$ with E, and



$n'(b)$ the same number for $\pi : y = b$. Then

$$v_1(E) = \frac{1}{2}\int_0^1 n(a)\,da + \frac{1}{2}\int_0^1 n'(b)\,db$$

is the one-dimensional variation $v_1(E)$ of E.

In the general case, for the variation $v_r(E)$ of dimension r, one replaces $n(a)$ with the *number of components* of $E \cap \pi$ contained entirely inside π, where π is an r-dimensional square, parallel to one of the sides of Q. The sum of variations of all dimensions is the variation $v(E)$:

$$v(E) = v_0(E) + v_1(E) + \cdots + v_s(E).$$

Theorems are proved about variation; the use of the entropy is essential here. For instance: if E in R_s is a surface of fairly low dimension $m < s$ and if $v(E)$ is not very large, then $Q - E$ contains a cube with a fairly large side $\epsilon > 0$.

The other tool is some theorems of Oleinik, which estimate the number of bounded connected components of the algebraic surface

$$Q_k(y_1, \cdots, y_n) = \text{Const};$$

Q_k is a polynomial of degree k in n variables. This number is of the order of k^n.

Finally it is necessary to connect these results concerning finite-dimensional spaces with the properties of the infinite-dimensional family of functions $F = \{f(x)\}$, $x \in Q$. This is achieved by introducing points $x_1, \cdots, x_s$ in Q and by replacing the functions $f(x)$, $P(x)$ by the finite sets $f(x_j)$, $P(x_j)$, $j = 1, \cdots, s$.

Here is the weakest part of Vitushkin's work, because in order to introduce the points x_j in a satisfactory fashion, Vitushkin is forced to assume hypotheses, two different ones for continuous and for analytic functions, which arbitrarily postulate a connection between the points x_j and the entropy $H_\epsilon(F)$ of F. The justification of the hypotheses is only that they happen to work for important classes F.

For instance for the classes of type Λ (example 3) the hypothesis is as follows:

For each $\epsilon > 0$, one can find $s(\epsilon)$ points x_j such that:

$$\begin{cases} \text{(a) } s(\epsilon) \text{ is large: } s(\epsilon) \geqq CH_{c_1\epsilon}(F); \\ \text{(b) } s(\epsilon) \text{ is not very large: the values } f(x_j), j = 1, \cdots, s, f \in F \text{ can be} \\ \quad \text{prescribed arbitrarily in the range } |f(x_j)| \leqq \epsilon. \end{cases}$$

This leads to (3). For analytic families one gets somewhat better inequalities of the type

$$n \log (k + 1) \geqq [1 - o(1)]H_\epsilon(F). \tag{4}$$

The hypothesis which leads to the inequalities of type (4) is harder to describe. Its main content is the following.

For a selection of points $x_1, \cdots, x_s$ in Q we denote by F' the set of all points of the s-dimensional space R_s of the form $\{f(x_j)\}_{j=1}^s$, $f \in F$. The norm in R_s is taken to be the maximum of the absolute values of the coordinates. In this norm, F' has the entropy $H\delta(F')$. We assume:

There is a function $\delta(\epsilon) > \epsilon$, defined for each $\epsilon > 0$ and an integer $s(\epsilon, B)$, defined for $B > 0$, $\epsilon > 0$, with the properties: $\delta(\epsilon)$ decreases to zero for $\epsilon \to 0$;

If $B > 0$ is given and $\epsilon > 0$ is sufficiently small, one can find $s = s(\epsilon, B)$ points $x_j \in Q$ for which

$$\begin{cases} \text{(a) } s(\epsilon, B)/H_{\delta(\epsilon)}(F) \to 0 \quad \text{for } \epsilon \to 0; \\ \text{(b) } H_{B\epsilon}(F') \geqq H_{\delta(\epsilon)}(F). \end{cases}$$

Vitushkin's work is of great originality and depth. We have omitted details in some of the above statements and conditions. In particular, Vitushkin's definition of an ϵ-algorithm is based on a formula more general than (2), so that for example maxima and minima of finitely many functions P are also allowed.

We discuss another application. Can one represent "bad," complicated functions by means of superpositions of simple, "good" functions? Kolmogorov's representation theorem for continuous functions of two variables establishes a case when this is possible. Kolmogorov's formula is as follows:

$$f(x, y) = \sum_{q=1}^{5} g(\phi_q(x) + \psi_q(y)). \tag{5}$$

Here ϕ_q, ψ_q are fixed continuous functions, and g is a continuous function depending on f. (In his dissertation [5], D. A. Sprecher announces some improvements on Kolmogorov's theorem; in particular, he shows that in the formula (5) one can replace the functions $\psi_q(y)$ by $\lambda\phi_q(y)$, $q = 1, \cdots, 5$, λ being some fixed constant.) We agree here to consider arbitrary continuous functions of two variables to be complicated, sums and continuous functions of one variable to be simple.

There are obvious negative examples. Consider functions $f(x, y)$ which are superpositions of other functions, for example

$$f(x, y) = g(x, h(y, k(x)), l(x, y)). \tag{6}$$

One cannot represent all p-times continuously differentiable functions f by superpositions of functions which are $q > p$ times continuously differentiable. Vitushkin and Kolmogorov [1] give a first nontrivial negative case. By a method which combines entropy and category, they show that in the above statement one can replace p by p/s, where s denotes the number of variables.

It is interesting to compare the entropy of a set A with the properties of A which are important in the theory of approximation.

Assume that A is a compact subset of a Banach space X, and that $\Phi = \{\phi_1, \cdots, \phi_n, \cdots\}$ is a sequence of elements which span X. Then

$$E_n^\Phi(f) = \min_{a_i} \| f - \sum_{i=1}^n a_i\phi_i \| \tag{7}$$

is the degree of approximation of a function f, and

$$d_n = E_n^\Phi(A) = \sup_{f \in A} E_n^\Phi(f) \tag{8}$$

is the degree of approximation of A. How do d_n and $H_\epsilon(A)$ compare? The following theorem gives an answer to this when $X = L^2$. Other spaces also could be treated in this way.

THEOREM. *Let Φ and a sequence $d_n \searrow 0$ be given, and let A be the set of all $f \in L^2$ with*

$$E_n^\Phi(f) \leqq d_n, \qquad n = 0, 1, 2, \cdots. \tag{9}$$

Then

$$\int_0^{1/2\epsilon} \frac{m(t)}{t}\, dt \leqq H_\epsilon(A) \leqq C \int_0^{4/\epsilon} \frac{m(t)}{t}\, dt, \tag{10}$$

where $m(t)$ is defined by

$$m(t) = \max \left[n : d_n \geqq \frac{1}{t} \right]. \tag{11}$$

As a corollary we obtain that the entropy of the set Lip $(2, \alpha)$ in the L^2-metric is of the order $\epsilon^{-1/\alpha}$.

Another corollary is the following. The system Φ and the set of the $d_n \searrow 0$ describe the set A of the theorem. Could it happen that the set A would be better described by another sequence $\Phi' = \{\phi'_n\}$? Let $d_n'/d_n \to 0$ and let A' be the set which corresponds to Φ' and d_n'. Could it be that $A \subset A'$? The answer is very emphatically no: one has

$$\lim_{\epsilon \to 0} \frac{H_\epsilon(A \cap A')}{H_\epsilon(A)} = 0. \tag{12}$$

Concluding, we shall mention some more theoretic aspects of entropy, which concern the classification of linear topological spaces. If X is a linear topological space, if $A \subset X$ is compact, and if U is a neighborhood of the origin in X, then A can be covered by finitely many translations of the set ϵU: $A \subset \bigcup_{i=1}^{i=n} (x_1 + \epsilon U)$. We put $N_\epsilon(A; U) = \min m$ (compare the function $N_\epsilon(A)$, defined in §1). By means of $N_\epsilon(A; U)$ as a function of ϵ, Gel'fand and Mityagin have characterized the *nuclear spaces* of Grothendieck among all linear topological spaces of type (F). The necessary and sufficient condition is here $\log \log N_\epsilon(A; U)/\log \epsilon \to 0$. Kolmogorov, using the same $N_\epsilon(A; U)$, has defined the *approximate dimension* of a space X. This notion allowed him to show that the linear topological spaces of analytic functions of s complex variables, defined in a given bounded domain, are topologically different for different s. Compare the paper Mityagin [3].

REFERENCES

[1] A. N. Kolmogorov and V. N. Tikhomirov, *ε-entropy and ε-capacity in function spaces*, Uspehi Mat. Nauk 14, no. 2 (86), (1959), pp. 3–86. Translated: American Mathematical Society Translations, Series 2, vol. 17, 1961, pp. 277–364.

[2] G. G. Lorentz, *Metric entropy, widths, and superpositions of functions*, Amer. Math. Monthly, 69 (1962), pp. 469–485.

[3] B. S. Mityagin, *Approximative dimension and bases in nuclear spaces*, Uspehi Mat. Nauk 16, no. 4 (100) (1961), pp. 63–132. Translated: Russian Math. Surveys, vol. 16, No. 4, 1961, pp. 59–127.

[4] A. G. Vitushkin, *Estimation of complexity of the tabulation problem*, Gosizdat, Moscow 1959. Translation: Pergamon Press, New York, 1961.

[5] D. A. Sprecher, dissertation, University of Maryland, 1963.