U. S. Department of Commerce National Bureau of Standards

RESEARCH PAPER RP1235

Part of Journal of Research of the National Bureau of Standards, Volume 23, August 1939

APPROXIMATION TO A FUNCTION OF ONE VARIABLE FROM A SET OF ITS MEAN VALUES

By Martin Greenspan

ABSTRACT

Physical data intended to represent the variation of a function of a single variable may be actually mean values of the function over intervals of the argument. Formulas for approximating to the values of the function from such data and examples of their use are presented in this paper. These formulas have been applied to problems of strain distribution in the Engineering Mechanics Section of the National Bureau of Standards.

CONTENTS

I. INTRODUCTION

Methods of measurement to determine the variation of a function with its argument are frequently such that only the mean values of the function over intervals of the argument are obtained. For example, a strain gage measures the total extension over the gage length and the mean strain is computed by dividing the extension by the gage length. If the variation of strain along the gage length is not nearly linear, the measured value may differ considerably from the strain at the middle of the gage length. This difference may be reduced by the use of a shorter gage length but as a result the sensitivity of the gage would be decreased and the gage would be more difficult to construct.

Another example is the measurement of heat capacity; the heat capacity, dQ/dt, of a body at temperature t is usually approximated by $\Delta Q/\Delta t$, where ΔQ is the quantity of heat required to change the temperature of the body by an amount Δt, where the interval Δt includes t.

A method of correcting such data was obtained by Strutt [1].[1] Strutt's formula is equivalent to the first correction-term of eq 12 in this paper. It will be seen that this is inadequate in many

[1] Figures in brackets indicate the literature references at the end of this paper.

cases. An analogous method for cases where the function is defined for discrete values of the argument only is King's formula for quinquennial sums [2].

Runge [3] has proposed a method of correcting observed spectral-energy distributions which has been applied by Stang [4], and by Runge to some data of Paschen. In this case the observed intensity is a double integral of the desired intensity, the error arising from the finite width of both the collimating slit and the bolometer strip. Runge's formula is not applicable to the problem considered here.

The statistical data which specify frequency distributions are often obtained in an approximate form analogous to the cases of strain and heat capacity. Thus, in a table of heights of individuals, the entry corresponding to 70 in. might be the number of individuals whose heights lie between 69½ and 70½ in. A method of computing the "true" moments of the distribution from the "rough" moments obtained from the data has been given by Sheppard [5]. Sheppard's work is the closest approach to a solution of the present problem found in the literature, but the method is in general not directly applicable.

II. DERIVATION OF THE FORMULAS

1. CENTRAL-DIFFERENCE FORMULA

Suppose that it is desired to determine the values of $f(x)$ corresponding to successive values of x differing by w, and that the method of measurement is such that the result, $F(x)$, is the average value of $f(x)$ over the interval g from $x-g/2$ to $x+g/2$. Then

$$F(x)=\frac{1}{kw}\int_{x-\frac{kw}{2}}^{x+\frac{kw}{2}} f(x)dx, \tag{1}$$

where $k=g/w$. This may be written symbolically as

$$F(x)=\frac{1}{kw}\delta_k\frac{f(x)}{D}, \tag{2}$$

where $D=d/dx$ and δ_k is defined from

$$\delta_k\psi(x)=\psi\left(x+\frac{kw}{2}\right)-\psi\left(x-\frac{kw}{2}\right). \tag{3}$$

Eq 1 may be written

$$f(x)=kw\frac{D}{\delta_k}F(x). \tag{4}$$

It is desired to calculate $f(x)$ in terms of $F(x)$ and its successive central differences, $\delta^n F(x)$, defined by repeated application of

$$\delta^n\psi(x)=\delta[\delta^{n-1}\psi(x)]$$

to eq 3 with $k=1$. The operators δ_k and D in eq 4 are thus to be replaced by equivalent expressions in terms of the operator δ. These expressions are obtained by expanding the right-hand side of eq 3 in Taylor's series about x, thus:

$$\delta_k\psi(x)=e^{\frac{kw}{2}D}\,\psi(x)-e^{-\frac{kw}{2}D}\,\psi(x),$$

or

$$\delta_k=2\sinh\frac{kw}{2}D, \tag{5}$$

from which, with $k=1$,

$$D=\frac{2}{w}\sinh^{-1}\frac{\delta}{2}. \tag{6}$$

The substitution of D from eq 6 into eq 5 gives

$$\delta_k=2\sinh\left(k\sinh^{-1}\frac{\delta}{2}\right). \tag{7}$$

Replacement of D and δ_k in eq 4 by the values given in eq 6 and 7 gives

$$f(x)=\frac{k\sinh^{-1}\frac{\delta}{2}}{\sinh\left(k\sinh^{-1}\frac{\delta}{2}\right)}F(x). \tag{8}$$

This is in the form $f(x)=p\operatorname{csch}p\,F(x)$, where $p=k\sinh^{-1}(\delta/2)$, and by expansion of $p\operatorname{csch}p$ in Maclaurin's series may be written

$$f(x)=\left[1+2\sum_{n=1}^{\infty}\frac{(-1)^n(2^{2n-1}-1)}{(2n)!}B_{2n-1}\left(k\sinh^{-1}\frac{\delta}{2}\right)^{2n}\right]F(x), \tag{9}$$

where B_{2n-1} are the successive Bernoulli's numbers.[2] Expansion of $\sinh^{-1}(\delta/2)$ in Maclaurin's series in eq 9 gives the double series

$$f(x)=\left\{1+2\sum_{n=1}^{\infty}(-1)^n\frac{(2^{2n-1}-1)}{(2n)!}B_{2n-1}\right.$$
$$\left.\left[\frac{k\delta}{2}\left(1+\sum_{m=1}^{\infty}(-1)^m\frac{(2m)!}{(2^m m!)^2(2m+1)}(\delta/2)^{2m}\right)\right]^{2n}\right\}F(x), \tag{10}$$

or

$$f(x)=\left[1-\frac{k^2}{24}\delta^2+\frac{k^2(20+7k^2)}{5760}\delta^4-\frac{k^2(448+196k^2+31k^4)}{967{,}680}\delta^6+\ldots\right]F(x). \tag{11}$$

When $k=1$, which is the case when the intervals g are contiguous, eq 11 reduces to

$$f(x)=\left(1-\frac{1}{24}\delta^2+\frac{3}{640}\delta^4-\frac{5}{7168}\delta^6+\ldots\right)F(x). \tag{12}$$

The numerical values of the coefficients of the first three correction terms in eq 11 are given for four values of k, in table 1.

[2] What are here denoted by B_{2n-1} are sometimes denoted by B_n.

TABLE 1.—*Values of* C_2, C_4, *and* C_6 *in* $f(x)=(1+C_2\delta^2+C_4\delta^4+C_6\delta^6+\ldots)\,F(x)$

k	$-C_2$	C_4	$-C_6$	k	$-C_2$	C_4	$-C_6$
1	0.0417	0.00469	0.000714	3	0.375	0.130	0.0439
2	.167	.0333	.00714	4	.667	.367	.191

2. DESCENDING-DIFFERENCE FORMULA

Equations 11 and 12 cannot be used to calculate $f(x)$ at $x=b$, where b is near a boundary of the interval of x for which $f(x)$ is defined, or where values of x on only one side of b are accessible to the method of measurement, because the necessary central differences are not available. For such cases an ascending- or descending-difference formula may be used. It is convenient to consider the measured value $F(x)$ to correspond to the initial point of the interval $g=kw$, so that

$$F(x)=\frac{1}{kw}\int_{z}^{z+kw} f(x)dx, \tag{13}$$

or, symbolically,

$$F(x)=\frac{1}{kw}\Delta_k\frac{f(x)}{D}, \tag{14}$$

where $D = d/dx$ and Δ_k, the descending difference, is defined from

$$\Delta_k\psi(x)=\psi(x+kw)-\psi(x). \tag{15}$$

Equation 14 may be written

$$f(x)=kw\,\frac{D}{\Delta_k}F(x)\cdot \tag{16}$$

It is desired to calculate $f(x)$ in terms of $F(x)$ and its successive descending differences $\Delta^n F(x)$, defined by repeated application of

$$\Delta^n\psi(x)=\Delta[\Delta^{n-1}\psi(x)]$$

to eq 15 with $k=1$. The operators Δ_k and D in eq 16 are thus to be replaced by equivalent expressions in terms of the operator Δ. Equation 15 may be written

$$(1+\Delta_k)\psi(x)=\psi(x+kw)=e^{kwD}\psi(x),$$

or

$$1+\Delta_k=e^{kwD}, \tag{17}$$

and

$$1+\Delta=e^{wD}. \tag{18}$$

Equation 17 and 18 give

$$wD=\log\,(1+\Delta), \tag{19}$$

and

$$\Delta_k=(1+\Delta)^k-1. \tag{20}$$

Replacement of D and Δ_k in eq 16 by the values given in eq 19 and 20 gives

$$f(x)=\frac{\log\,(1+\Delta)^k}{(1+\Delta)^k-1}F(x). \tag{21}$$

This is in the form $f(x)=\dfrac{p}{e^p-1}F(x)$, where $p=k\,\log\,(1+\Delta)$, and by expansion of $p/(e^p-1)$ in Maclaurin's series may be written

$$f(x)=\left\{1-\frac{k}{2}\log\,(1+\Delta)+\sum_{n=1}^{\infty}\frac{(-1)^{n+1}B_{2n-1}}{(2n)!}[k\,\log\,(1+\Delta)]^{2n}\right\}F(x), \tag{22}$$

where B_{2n-1} are the successive Bernoulli's numbers.[3] Expansion of log $(1+\Delta)$ in Taylor's series in eq 22 gives the double series

$$f(x)=\left\{1-\frac{k}{2}\sum_{m=1}^{\infty}(-1)^{m+1}\frac{\Delta^m}{m}+\sum_{n=1}^{\infty}\frac{(-1)^{n+1}B_{2n-1}}{(2n)!}k^{2n}\left[\sum_{m=1}^{\infty}(-1)^{m+1}\frac{\Delta^m}{m}\right]^{2n}\right\}F(x),$$

or (23)

$$f(x)=\left[1-\frac{k}{2}\Delta+\frac{k}{4}\left(1+\frac{k}{3}\right)\Delta^2-\frac{k}{6}\left(1+\frac{k}{2}\right)\Delta^3+\frac{k}{8}\left(1+\frac{11k}{18}-\frac{k^3}{90}\right)\Delta^4\right.$$
$$\left.-\frac{k}{2}\left(\frac{1}{5}+\frac{5k}{36}-\frac{k^3}{180}\right)\Delta^5+\frac{k}{12}\left(1+\frac{137k}{180}-\frac{17k^3}{360}+\frac{k^5}{2520}\right)\Delta^6+\ldots\right]F(x). \tag{24}$$

When $k=1$, which is the case when the intervals g are contiguous, eq 24 reduces to

$$f(x)=\left(1-\frac{1}{2}\Delta+\frac{1}{3}\Delta^2-\frac{1}{4}\Delta^3+\frac{1}{5}\Delta^4-\frac{1}{6}\Delta^5+\frac{1}{7}\Delta^6-\ldots+\ldots\right)F(x). \tag{25}$$

The numerical values of the coefficients of the first six correction terms in eq 24 are given, for four values of k, in table 2.

TABLE 2.—*Values of* C_1, C_2, *etc. in* $f(x)=(1+C_1\Delta+C_2\Delta^2+C_3\Delta^3+\ldots)F(x)$

k	$-C_1$	C_2	$-C_3$	C_4	$-C_5$	C_6
1	0.500	0.333	0.250	0.200	0.167	0.143
2	1.000	.834	.667	.533	.433	.360
3	1.500	1.500	1.250	.950	.700	.526
4	2.000	2.333	2.000	1.367	.800	.476

III. ACCURACY AND LIMITATIONS

The derivations of eq 11 and 24 involve the expansion of operators in terms of ascending powers (orders) of the symbols δ and Δ, respectively. This procedure is strictly valid if the function $F(x)$ upon which the expanded expression operates is a polynomial, because δ and Δ obey the commutative, associative, and distributive laws; and because, since the operation δ or Δ on a polynomial of finite degree n reduces its degree by one, all differences of $F(x)$ of order higher than the nth vanish. The series therefore terminates, and is exact.

However, if $F(x)$ is not a polynomial, the formulas 11 and 24 can be applied only on the assumption that $F(x)$ may be sufficiently well represented by a polynomial of degree n. The error involved in such a procedure might be estimated by considering the remainder after

[3] See footnote 2.

the nth difference term of the series, but this is not possible when, as is usually the case, the characteristics of $F(x)$ are unknown. The justification for the use of formulas 11 and 24 therefore rests on the fact that distributions encountered in practice are frequently such that they may be approximated to by polynomials of low degree.

It is obvious that, so far as the data $F(x)$ are concerned, the function $f(x)$ is undetermined to the extent of an additive arbitrary periodic function of period g the average value of which over any interval g is zero. This indetermination can be reduced only by reducing the interval g over which $f(x)$ is averaged in the measurement. Of course, a lower limit to g is set by other considerations.

The value for the interval w between observations may ordinarily be the same as that which would for various reasons have been used if the corrections were not to be applied. In cases where there are not enough data to provide differences of sufficiently high order, w can be decreased. The highest order of differences which can be advantageously used is determined by the accuracy of the data. The successive differences are increasingly affected by the errors of measurement, and the differences of a certain order, and all succeeding differences, will consist mostly of accumulated error. It may be desirable to graduate the data $F(x)$ before applying eq 11 or 24.

IV. EXAMPLES

It was thought desirable to include the following two examples which indicate roughly the order of accuracy to be expected from the use of eq 11 and 24. For these examples hypothetical data were obtained by computing mean values of strain over intervals of length g from theoretically known strain distributions. These mean values represent data that would be obtained by a perfect strain gage of gage length g.

1. CENTRAL-DIFFERENCE FORMULA

Consider (fig. 1) a very long isotropic elastic strip of breadth $2b$ containing a centrally located circular hole of diameter b. The strip is in a state of generalized plane stress such that the stress at cross sections remote from the hole is of magnitude $\widehat{xx} \rightarrow T$, uniform across the section and normal to it. The boundaries of the strip are free. The curve of figure 1 shows according to Howland [6] the variation of stress along an edge of the strip near the hole. Suppose that in a certain specimen the stresses along an edge were actually as given by the curve of figure 1 and it were desired to determine them by means of a strain gage. (Since the edge of the strip is a free boundary, the stress is proportional to the strain.) The data that would be obtained by the use of a strain gage of gage length $(2/3)b$ and the results of applying eq 11 are shown in figure 1. It is evident that the corrected values conform much more closely than the data to the original distribution.

2. DESCENDING-DIFFERENCE FORMULA

Consider (fig. 2) an isotropic elastic flat circular plate of radius a and thickness $h \ll a$, clamped around the edge and subjected to a hydrostatic pressure $p=18.8E(h^4/a^4)$ on one face. (E is Young's modulus of elasticity and Poisson's ratio is taken as 1/4.) The curve of figure 2 shows according to Nádai [7] the variation along a radius of

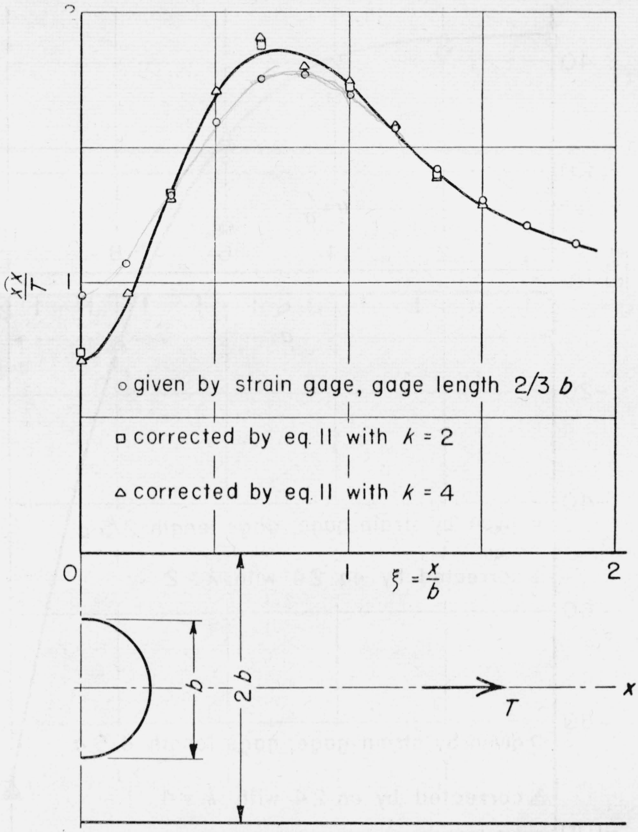

FIGURE 1.—*Problem of a tension member containing a hole.*

$\epsilon_r \Big/ \dfrac{h^2}{12a^2}$, where ϵ_r is the normal strain in the unloaded face in the radial direction. Suppose that in a certain specimen the radial strains were actually as given by the curve of figure 2 and it were desired to determine them by means of a strain gage. If the value of the strain at the clamped edge were desired, eq 24 would be used with the transformation $x=1-u$. The data that would be obtained by the use of strain gages of gage lengths $(2/5)a$ and $(4/5)a$ and the results of applying eq 24 are shown in figure 2.

The corrected values conform much more closely than the data to the original distribution. Furthermore, the value of $\epsilon_r \Big/ \dfrac{h^2}{12a^2}$ at the clamped edge, which is inaccessible to the strain gage, is accurately obtained.

3. NOTATION

Various notations for differences are in use. Tables 3 and 4 show how the differences used in this paper are computed. The tables are arranged so that the difference corresponding to any entry, $F(x)$, is on a horizontal line through that entry.

The interval between successive values of the argument x is constant.

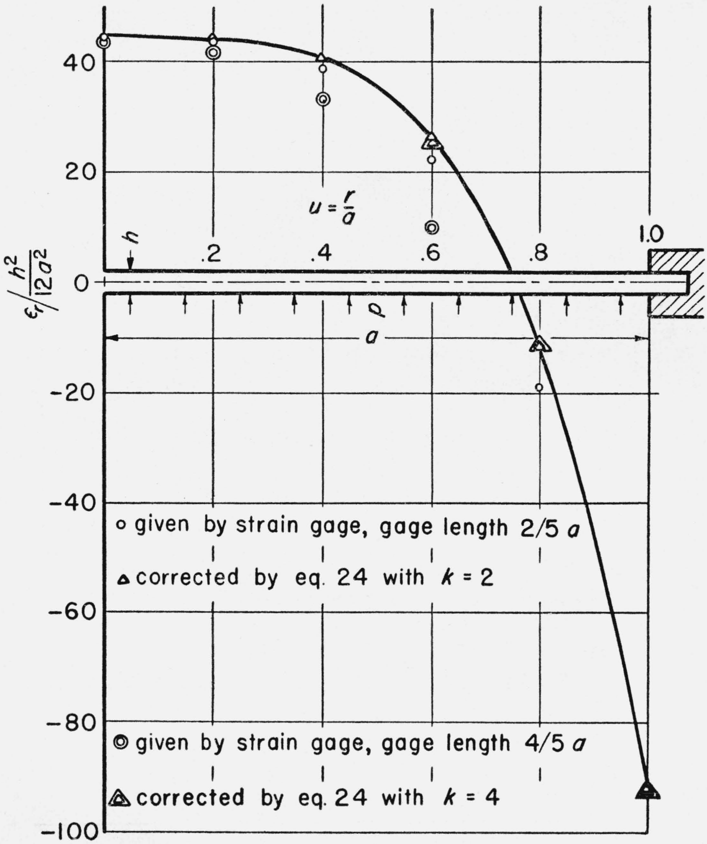

FIGURE 2.—*Problem of a circular flat plate subjected to hydrostatic pressure.*

TABLE 3.—*Central differences*

$F(x)$	$\delta F(x)$	$\delta^2 F(x)$	$\delta^3 F(x)$	$\delta^4 F(x)$
$F(x_0)$				
	$\delta F(x_{1/2})=F(x_1)-F(x_0)$			
$F(x_1)$		$\delta^2 F(x_1)=\delta F(x_{3/2})-\delta F(x_{1/2})$		
	$\delta F(x_{3/2})=F(x_2)-F(x_1)$		$\delta^3 F(x_{3/2})=\delta^2 F(x_2)-\delta^2 F(x_1)$	
$F(x_2)$		$\delta^2 F(x_2)=\delta F(x_{5/2})-\delta F(x_{3/2})$		$\delta^4 F(x_2)=\delta^3 F(x_{5/2})-\delta^3 F(x_{3/2})$
	$\delta F(x_{5/2})=F(x_3)-F(x_2)$		$\delta^3 F(x_{5/2})=\delta^2 F(x_3)-\delta^2 F(x_2)$	
$F(x_3)$		$\delta^2 F(x_3)=\delta F(x_{7/2})-\delta F(x_{5/2})$		$\delta^4 F(x_3)=\delta^3 F(x_{7/2})-\delta^3 F(x_{5/2})$
	$\delta F(x_{7/2})=F(x_4)-F(x_3)$		$\delta^3 F(x_{7/2})=\delta^2 F(x_4)-\delta^2 F(x_3)$	
$F(x_4)$		$\delta^2 F(x_4)=\delta F(x_{9/2})-\delta F(x_{7/2})$		$\delta^4 F(x_4)=\delta^3 F(x_{9/2})-\delta^3 F(x_{7/2})$
	$\delta F(x_{9/2})=F(x_5)-F(x_4)$		$\delta^3 F(x_{9/2})=\delta^2 F(x_5)-\delta^2 F(x_4)$	
$F(x_5)$		$\delta^2 F(x_5)=\delta F(x_{11/2})-\delta F(x_{9/2})$		$\delta^4 F(x_5)=\delta^3 F(x_{11/2})-\delta^3 F(x_{9/2})$
	$\delta F(x_{11/2})=F(x_6)-F(x_5)$		$\delta^3 F(x_{11/2})=\delta^2 F(x_6)-\delta^2 F(x_5)$	
$F(x_6)$		$\delta^2 F(x_6)=\delta F(x_{13/2})-\delta F(x_{11/2})$		
	$\delta F(x_{13/2})=F(x_7)-F(x_6)$			
$F(x_7)$				

TABLE 4.—*Descending differences*

$F(x)$	$\Delta F(x)$	$\Delta^2 F(x)$	$\Delta^3 F(x)$	$\Delta^4 F(x)$
$F(x_0)$	$\Delta F(x_0)=F(x_1)-F(x_0)$	$\Delta^2 F(x_0)=\Delta F(x_1)-\Delta F(x_0)$	$\Delta^3 F(x_0)=\Delta^2 F(x_1)-\Delta^2 F(x_0)$	$\Delta^4 F(x_0)=\Delta^3 F(x_1)-\Delta^3 F(x_0)$
$F(x_1)$	$\Delta F(x_1)=F(x_2)-F(x_1)$	$\Delta^2 F(x_1)=\Delta F(x_2)-\Delta F(x_1)$	$\Delta^3 F(x_1)=\Delta^2 F(x_2)-\Delta^2 F(x_1)$	$\Delta^4 F(x_1)=\Delta^3 F(x_2)-\Delta^3 F(x_1)$
$F(x_2)$	$\Delta F(x_2)=F(x_3)-F(x_2)$	$\Delta^2 F(x_2)=\Delta F(x_3)-\Delta F(x_2)$	$\Delta^3 F(x_2)=\Delta^2 F(x_3)-\Delta^2 F(x_2)$	$\Delta^4 F(x_2)=\Delta^3 F(x_3)-\Delta^3 F(x_2)$
$F(x_3)$	$\Delta F(x_3)=F(x_4)-F(x_3)$	$\Delta^2 F(x_3)=\Delta F(x_4)-\Delta F(x_3)$	$\Delta^3 F(x_3)=\Delta^2 F(x_4)-\Delta^2 F(x_3)$	$\Delta^4 F(x_3)=\Delta^3 F(x_4)-\Delta^3 F(x_3)$
$F(x_4)$	$\Delta F(x_4)=F(x_5)-F(x_4)$	$\Delta^2 F(x_4)=\Delta F(x_5)-\Delta F(x_4)$	$\Delta^3 F(x_4)=\Delta^2 F(x_5)-\Delta^2 F(x_4)$	
$F(x_5)$	$\Delta F(x_5)=F(x_6)-F(x_5)$	$\Delta^2 F(x_5)=\Delta F(x_6)-\Delta F(x_5)$		
$F(x_6)$	$\Delta F(x_6)=F(x_7)-F(x_6)$			
$F(x_7)$				

V. REFERENCES

[1] J. W. Strutt, *On a correction sometimes required in curves professing to represent the connection between two physical magnitudes.* Phil. Mag. [4] **42,** 441 (1871).

[2] G. King, *On a new method of constructing and graduating mortality and other tables.* J. Inst. Actuaries **43,** 109 (1909).

[3] C. Runge, *Über die Differentiation empirischer Functionen.* Z. Math. Phys. **42,** 205 (1897).

[4] A. H. Stang, *The infra-red absorption spectrum of naphthalene and some of its mono-derivatives in solution.* Phys. Rev. [N. S.] **9,** 542 (1917).

[5] W. F. Sheppard, *On the calculation of the most probable values of frequency-constants, for data arranged according to equidistant divisions of a scale.* Proc. London Math. Soc. **29,** 353 (1898).

[6] R. C. J. Howland, *On the stresses in the neighborhood of a circular hole in a strip under tension.* Phil. Trans. Roy. Soc. [A] **229,** 49 (1930).

[7] A. Nadai, Die elastischen Platten, p. 291 (Julius Springer, Berlin, 1925).

WASHINGTON, May 23, 1939.