ON THE TRUNCATION ERROR WHICH ARISES FROM THE USE OF FINITE DIFFERENCES IN THE LAPLACIAN OPERATOR [1]

By Yoshimitsu Ogura

Johns Hopkins University[2]

(Original manuscript received 7 August 1957; revised manuscript received 19 June 1958)

ABSTRACT

The modification of the power spectrum for the Laplacian of a variable, associated with the use of finite differences instead of derivatives, is discussed for an isotropic scalar field. The results permit one to specify a finite difference scheme which reduces considerably the systematic truncation error in the Laplacian operator.

1. Introduction

In a recent report, Thompson (1955) has discussed the systematic error in computing the Laplacian of any variable by the usual method of finite differences. By systematic error is meant that part of the truncation error which does not depend upon the orientation of the finite difference grid. He also proposed in the report a simple scheme by which the systematic truncation errors in computing the Laplacian are considerably reduced.

On the other hand, the writer (1957) discussed the modification of the power spectrum for the Laplacian of a variable, which results from the use of finite differences, under the assumption that the variable is a steady, random function of space. It was also found that there is an error which tends to make the finite-difference equivalent of the Laplacian a systematic underestimate of the true Laplacian.

In this report, some attempts are made to estimate the magnitude of the systematic truncation error induced by finite-difference techniques in general, with the aim of finding a finite-difference scheme which yields a good approximation to the true Laplacian operator.

2. Systematic truncation errors as a function of wave number

In the finite-difference scheme proposed by Thompson (1955), the Laplacian of a variable ϕ at a grid point is computed essentially from values of ϕ at the reference point and 12 surrounding points. This scheme may be written in a more general form:

$$\nabla^2\phi = \frac{1}{a^2}[\alpha\phi_0+\beta(\phi_1+\phi_2+\phi_3+\phi_4) +\gamma(\phi_5+\phi_6+\phi_7+\phi_8)+\delta(\phi_9+\phi_{10}+\phi_{11}+\phi_{12})]. \quad (1)$$

This is our starting finite-difference form. In (1), a is the distance between adjacent points in a square grid, ϕ_0 is the value of ϕ at the point where the Laplacian is to be computed, and subscripts 1, 2, $\cdots$, 12 refer to the values of ϕ at grid points, as shown in fig. 1.

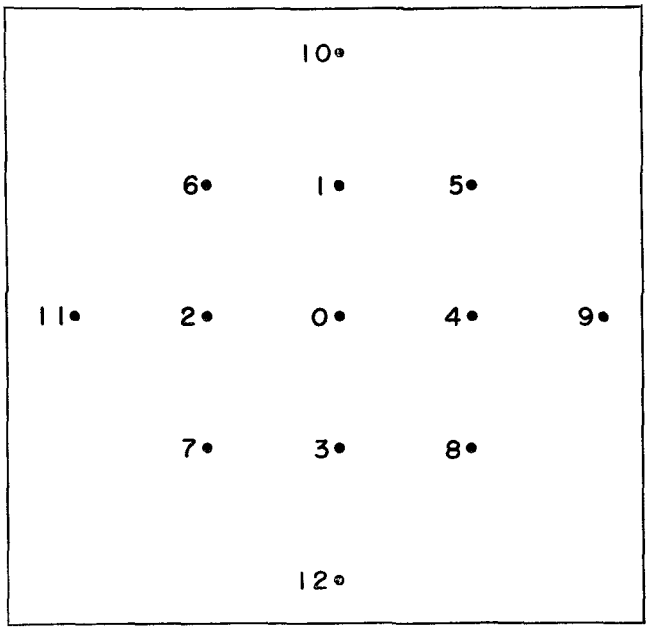

FIG. 1. Thirteen points in a square grid.

The constants α, β, γ, and δ are unspecified and free to be fixed. The symbol ∇^2 is used to distinguish the finite-difference equivalent of the Laplacian from the true Laplacian ∇^2. Our present concern is to estimate the systematic truncation errors induced by the use of (1) instead of derivatives and then to fix the values of parameters α, β, γ, and δ in such a way as to make the systematic error introduced by the use of finite-difference formula as small as possible.

First of all, the requirement for a zero value of Laplacian when ϕ is constant all over the domain

[1] The research reported in this paper was carried out at the University of Chicago under Contract CWB 9016 with the United States Weather Bureau.

[2] Present address: Massachusetts Institute of Technology.

will be met in (1) if

$$\alpha = -4(\beta + \gamma + \delta). \tag{2}$$

Consequently, (1) becomes

$$\nabla^2\phi = \frac{1}{a^2}[\beta(\phi_1 + \phi_2 + \phi_3 + \phi_4 - 4\phi_0) + \gamma(\phi_5 + \phi_6 + \phi_7 + \phi_8 - 4\phi_0) + \delta(\phi_9 + \phi_{10} + \phi_{11} + \phi_{12} - 4\phi_0)]. \tag{3}$$

For simplicity, we shall assume hereafter that ϕ is a steady, random function of space and moreover that the ϕ-field is isotropic in the statistical sense: the average value of any function of ϕ or of the derivatives of ϕ, defined relative to a particular set of axes, is unchanged by rotation and reflection of the axes. Then, following the procedure outlined in the previous paper (Ogura, 1957), we get the relation

$$\overline{(\nabla^2\phi)^2} = 2\pi \int_0^\infty k^5 F(k) dk, \tag{4}$$

where k is the magnitude of wave number vector and $F(k)$ is the power spectrum for the ϕ-field. The ensemble average of the variable is denoted by placing a bar over the unaveraged quantity.

Now, for the approximate value of the Laplacian (3), we have

$$\overline{(\nabla^2\phi)^2} = 2\pi \int_0^\infty k^5 F(k)\psi(ka) dk, \tag{5}$$

where

$$\psi(ka) = \frac{32}{\pi(ka)^4}\int_0^{\pi/2}\Big[-\beta\Big\{\sin^2\Big(\frac{ka}{2}\cos\theta\Big) + \sin^2\Big(\frac{ka}{2}\sin\theta\Big)\Big\} + \gamma\{\cos(ka\cos\theta) \times \cos(ka\sin\theta) - 1\} - \delta\{\sin^2(ka\cos\theta) + \sin^2(ka\sin\theta)\}\Big]^2 d\theta. \tag{6}$$

By comparing (4) and (5), we see that $\nabla^2\phi$ and $\nabla^2\phi$ differ by a factor of $\psi(ka)$, which is a function of not only wave number but also of three parameters β, γ, and δ. In other words, the response function $\psi(ka)$ gives a measure of systematic truncation errors, as a function of wave number, due to the use of finite differences.

Our task is to fix the values of three parameters β, γ, and δ in such a way that the truncation error is minimized. This can be achieved in various ways.

One method is to fix the values of β, γ, and δ by the condition that $\psi(ka)$ is unity for three specially chosen wave numbers, say k_a, k_b, and k_c. This procedure would be reasonable when the power spectrum $F(k)$ has three predominant peaks at k_a, k_b, and k_c, because the finite-difference form would then reflect quite well the main parts of the power spectrum $F(k)$.

A second method is to use a variational principle—*i.e.*, to determine β, γ, and δ by letting the value of $\int_{k_a}^{k_b}\{1 - \psi(ka)\}^2 dk$ be minimum, where k_a and k_b are certain wave numbers.

A third method—the one to be explored here—is to make values of $\psi(ka)$ close to unity for long waves. The reason for the choice of this method is that relatively short waves, more than long waves, are subject to instrumental and round-off errors and also to errors due to interpolation processes in reading values of ϕ at grid points on wealther maps. For this reason, it is desirable that disturbances of very small wave number (large scale) are unaffected by finite-difference processes, while disturbances of very large wave number (small scale) are obliterated. In other words, the desired property for $\psi(ka)$ would be that $\psi(ka)$ is close to unity for small wave numbers and then decreases rapidly to zero at some wave number near the upper limit of the resolving power of our system for meteorological measurement.

Let the Taylor series of $\psi(ka)$ with respect to ka about the origin be

$$\psi(ka) = \psi(0) + \frac{(ka)^2}{2!}\psi''(0) + \frac{(ka)^4}{4!}\psi^{(\mathrm{iv})}(0) + \frac{(ka)^6}{6!}\psi^{(\mathrm{vi})}(0) + \cdots.$$

Then the requirement that $\psi(ka)$ is close to unity for small wave number will be met if the following three equations are satisfied:

$$\left.\begin{aligned}\psi(0) &= 1,\\ \psi''(0) &= 0,\\ \psi^{(\mathrm{iv})}(0) &= 0.\end{aligned}\right\} \tag{7}$$

From (6), the preceding equations become

$$\psi(0) = (\beta + 2\gamma + 4\delta)^2 = 1,$$

$$\psi''(0) = -\tfrac{1}{4}(\beta + 2\gamma + 4\delta)(\beta + 4\delta + 16\delta) = 0, \tag{8}$$

or

$$\beta + 4\gamma + 16\delta = 0, \tag{9}$$

$$\psi^{(\mathrm{iv})}(0) = \frac{1}{192}[(\beta + 2\gamma + 16\delta) \times (19\beta + 98\gamma + 304\delta) + 108\gamma^2 + 16(\beta + 2\gamma + 4\delta)(\beta + 8\gamma + 64\delta)] = 0. \tag{10}$$

There are four sets of roots of equations (8), (9), (10): Taking $\beta + 2\gamma + 4\delta = \pm 1$ for (10), we get

$$\gamma^2 \mp \gamma - 1 = 0,$$

and

$$\alpha = 2\gamma \mp 5, \quad \beta = -\tfrac{4}{3}\gamma \pm \tfrac{4}{3}, \quad \delta = -\tfrac{1}{6}\gamma \mp \tfrac{1}{12}.$$

Using the upper sign, we have eare two sets:

	α	β	γ	δ
(i)	−6.2360	2.1573	−0.6180	0.0197
(ii)	−1.7640	−0.8240	1.6180	−0.3530

Using the lower sign, we get two sets which are the negatives of (i) and (ii):

(iii): the negative of (i).

(iv): the negative of (ii).

As far as the value of $\psi(ka)$ is concerned, there is no special reason for the choice of a particular set out of four sets of roots noted above: Any one of them gives us the same value for $\psi(ka)$ and $1-\psi(ka)$ starts from the sixth-power term of ka for small values of ka, so that we have very small systematic truncation error for small k. But we may infer, from the following simple consideration, that the sets (iii) and (iv) are not adequate for our purpose: Let ϕ be given, for simplicity, by

$$\phi = A \sin kx.$$

Then we find

$$\nabla^2\phi = -Ak^2 \sin kx, \tag{11}$$

and

$$\nabla^2\phi = -Ak^2 \sin kx \left\{\frac{\sin\frac{ka}{2}}{ka/2}\right\}^2 \times\left((\beta + 2\gamma + 4\delta) - 4\delta \sin^2\frac{ka}{2}\right). \tag{12}$$

For sufficiently small ka, the last term in the bracket may be neglected and we see then that we should take $\beta + 2\gamma + 4\delta = +1$ for (8); otherwise, the phase of the deformed elementary wave (12) would be different from that of the original one (11) by the amount of half wave-length.

For practical purposes, one may reduce the time necessary for numerical computation by selecting values of $\alpha, \beta, \gamma, \delta$ which differ only slightly from those given in the set (i). Since δ in (i) is very small, we will choose the condition $\delta = 0$ instead of (10). Then (8) and (9) give

$$\text{(v)} \quad \alpha = -6, \quad \beta = 2, \quad \gamma = -\tfrac{1}{2}. \tag{13}$$

Of course, the systematic error is a little larger when values of the set (v) are employed than when values of the set (i) are employed. However, (i) requires 13 points for the computation of each Laplacian, while (v) requires only 9 points; further, (v) permits the Laplacian to be computed at points adjacent to the boundaries of the domain under consideration. Another advantage of (v) compared with (i) is that the former is simpler for use with a binary machine.

It is interesting to note here that the values (v) are exactly the same as those proposed by Knighting (1955) from a different point of view.

3. Examples of the magnitude of the truncation error

We shall evaluate in this section the magnitude of the systematic truncation errors for various finite-difference schemes.

I. The standard scheme for computing the Laplacian of ϕ by finite differences is

$$\nabla^2\phi = \frac{1}{a^2}[\phi_1 + \phi_2 + \phi_3 + \phi_4 - 4\phi_0]. \tag{14}$$

The response function in this case is

$$\psi_1(ka) = \frac{2}{\pi\left(\frac{ka}{2}\right)^4}\int_0^{\pi/2}\left\{\sin^2\left(\frac{ka}{2}\cos\theta\right) + \sin^2\left(\frac{ka}{2}\sin\theta\right)\right\}^2 d\theta.$$

This is the response function discussed in the previous paper (Ogura, 1957). When ka is small, it becomes

$$\psi_1(ka) = 1 - \tfrac{1}{8}(ka)^2 + \cdots.$$

II. The method of inverse averaging, proposed by Thompson (1955), consists of three steps:

(i) computing $\tilde{\phi}$ using the scheme:

$$\tilde{\phi} = \tfrac{1}{16}[20\phi_0 - (\phi_1 + \phi_2 + \phi_3 + \phi_4)],$$

(ii) computing $\nabla^2\tilde{\phi}$ by the standard scheme for a number of grid orientations, and

(iii) averaging $\nabla^2\tilde{\phi}$ over the various grid orientations.

Let us consider, for simplicity, $\nabla^2\tilde{\phi}$ computed for only one grid orientation. Then the procedure mentioned above is equivalent to computing the Laplacian with the following scheme:

$$\nabla^2\phi = \frac{1}{16a^2}[24(\phi_1 + \phi_2 + \phi_3 + \phi_4 - 4\phi_0) - 2(\phi_5 + \phi_6 + \phi_7 + \phi_8 - 4\phi_0) - (\phi_9 + \phi_{10} + \phi_{11} + \phi_{12} - 4\phi_0)]. \tag{15}$$

The response function for this scheme is

$$\psi_2(ka) = \frac{2}{\pi\left(\frac{ka}{2}\right)^4}\int_0^{\pi/2}\left[1 + \tfrac{1}{4}\left\{\sin^2\left(\frac{ka}{2}\cos\theta\right) + \sin^2\left(\frac{ka}{2}\sin\theta\right)\right\}\right]^2\left[\sin^2\left(\frac{ka}{2}\cos\theta\right) + \sin^2\left(\frac{ka}{2}\sin\theta\right)\right]^2 d\theta,$$

and it will be

$$\psi_2(ka) = 1 - \left(\frac{55}{16}\right)\frac{1}{288}(ka)^4 + \cdots,$$

when ka is sufficiently small.

III. The 9-points scheme, proposed in section 2, is

$$\nabla^2\phi = \frac{1}{a^2}[2(\phi_1 + \phi_2 + \phi_3 + \phi_4 - 4\phi_0) - \tfrac{1}{2}(\phi_5 + \phi_6 + \phi_7 + \phi_8 - 4\phi_0)]. \quad (16)$$

The response function is

$$\psi_3(ka) = \frac{2}{\pi\left(\frac{ka}{2}\right)^4}\int_0^{\pi/2}\left[\sin^2\left(\frac{ka}{2}\cos\theta\right) + \sin^2\left(\frac{ka}{2}\sin\theta\right) + 2\sin^2\left(\frac{ka}{2}\cos\theta\right)\times\sin^2\left(\frac{ka}{2}\sin\theta\right)\right]^2 d\theta,$$

and

$$\psi_3(ka) = 1 - \frac{1}{288}(ka)^4 + \cdots,$$

for small values of ka.

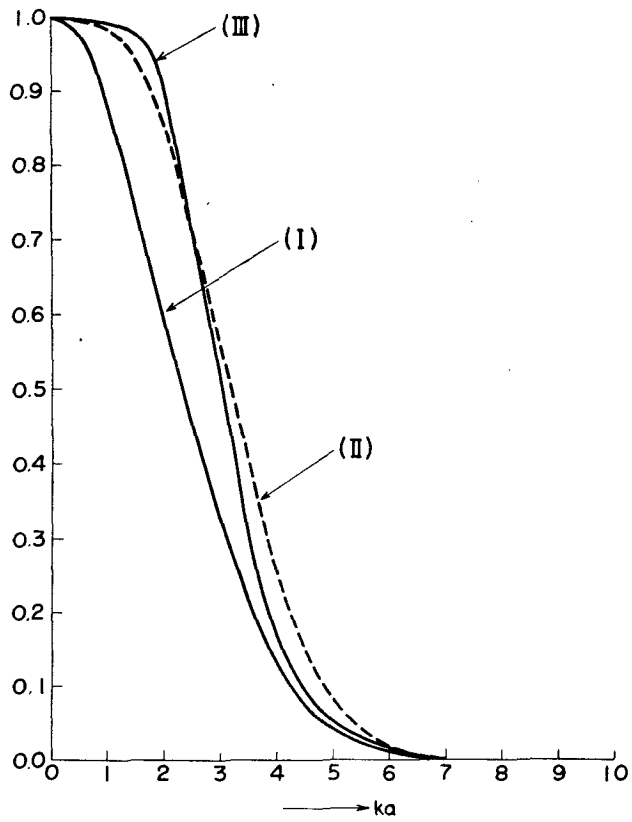

FIG. 2. The response function against wave number for (I) standard operator, (II) inverse averaging scheme, and (III) 9-point scheme. k: wave number. a: grid size.

Table 1 shows some examples of values of the response function, as a function of wave length $L(= 2\pi/k)$, for the three schemes mentioned above.

TABLE 1. Values of the response function.

L/a	10	8	6	4
Standard scheme	0.954	0.927	0.871	0.722
Inverse averaging	0.998	0.996	0.986	0.936
9-point scheme	1.000	0.999	0.994	0.971

Fig. 2 presents the response function as a function of ka. From the table and figure, we can see the systematic truncation error is considerably reduced for disturbances of wave length equal to or greater than four grid intervals.

Fig. 3 presents numerical examples of the deformed spectral distribution of vorticity, computed from (14), (15), and (16). The original spectrum of vorticity is assumed as

$$G(\xi) \equiv \xi^5 F(\xi) \approx \xi^3[1 + \xi^2]^{-\frac{3}{4}},$$

where $\xi = k/k_0$ and k_0 is a characteristic wave number for the turbulent flow under consideration. This functional form of vorticity spectrum is employed here because some investigations show that this form seems to represent fairly well the actual spectrum for large-scale atmospheric turbulence (for example, Ogura and Miyakoda, 1954; and Ogura, 1957). In fig. 3, the spectrum of vorticity is measured in an arbitrary unit and a is taken as 300 km. The characteristic length for large-scale atmospheric turbulence is assumed to be 9000 km as before (Ogura, 1957).

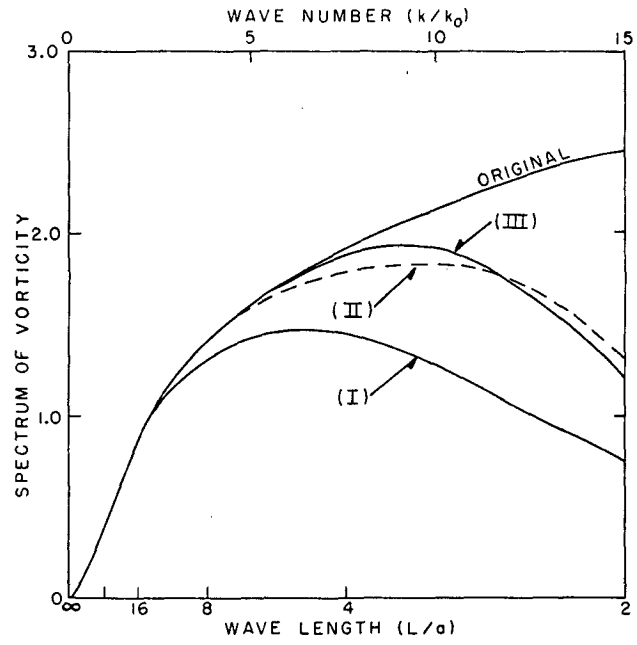

FIG. 3. Spectral distribution of vorticity computed by (I) standard operator, (II) inverse averaging scheme, and (III) 9-point scheme. a: size of finite differences. k_0: characteristic wave number of turbulence.

4. Random errors

So far our attention has been restricted to the systematic truncation error which results from the use of finite differences instead of derivatives. In other words, our analysis was carried out under the assumption that true values of ϕ are given at all grid points. This assumption can not be valid in actual cases, because values of ϕ at grid points are more or less subject to random errors associated with various sources, such as instrumental and round-off errors and errors due to the interpolation process in reading values of ϕ at grid points on weather charts. Then any quantity computed by making use of these values is also subject to unsystematic errors. In this section, we shall discuss quite briefly how the errors involved in ϕ themselves are reflected to the computed $\nabla^2\phi$.

Let a quantity f be given in general as a linear combination of $\phi_0, \phi_1, \cdots, \phi_n$:

$$f = \sum_{i=0}^{n} c_i\phi_i, \tag{17}$$

and suppose that each value of ϕ is expressed as

$$\phi_i = {}_t\phi_i + \phi_i',$$

where ${}_t\phi_i$ denotes the true value of ϕ at the i-point and ϕ_i' indicates the nonsystematic error at that point. Then (17) becomes

$$f = f_t + f', \tag{18}$$

where

$$f_t = \sum_{i=0}^{n} c_i{}_t\phi_i, \quad f' = \sum_{i=0}^{n} c_i\phi_i'.$$

Apparently the second term in the right-hand side of (18) represents the error of f. To estimate the magnitude of this error, we shall express the statistical frequency of f' in terms of the frequency function for ϕ_i'. If we can assume, for simplicity, that ϕ_i' are independent of each other—*i.e.*, that there are no correlations between ϕ_i' and ϕ_j' when $i \neq j$, then the task can be achieved very easily with the aid of random-walk theory. The result is

$$p(f') = \frac{2}{\pi}\int_0^\infty \cos f't \prod_{i=0}^{n} I_i(c_it)dt,$$

where

$$I(c_it) = \int_{-\infty}^{\infty} p_i(\phi_i') \cos \phi'_i c_i t d\phi_i',$$

and $p(f')$ and $p(\phi_i')$ represent the frequency functions for f' and ϕ_i', respectively.

We shall consider here the simple case in which $p(\phi_i')$ is given by

$$p_i(\phi_i') = \frac{2}{\sqrt{\pi}\sigma}\exp(-\phi_i'^2/\sigma^2),$$

that is a normal distribution with standard deviation σ which is assumed common for all $p_i(\phi_i')$. Then $p(f')$ is

$$p(f') = \frac{2}{\sqrt{\pi}\sigma(\sum_{i=0}^{n} c_i^2)^{\frac{1}{2}}}\exp\{-f'^2/\sigma^2(\sum_{i=0}^{n} c_i^2)\}. \tag{19}$$

In other words, the frequency distribution of f' is also normal, with standard deviation $\sigma(\sum c_i^2)^{\frac{1}{2}}$. For the three schemes considered in section 3 (equations 14, 15, and 16), the standard deviations of $\nabla^2\phi$ are respectively (I) 4.47 σ/a^2, (II) 6.05 σ/a^2, (III) 7.28 σ/a^2. Consequently we see that, although the systematic truncation error will be reduced considerably by making use of more grid points than five in the computation, we have at the same time more chances to have large non-systematic errors than we do with the standard method.

One method to reduce non-systematic errors—as was suggested by Thompson (1955)—may be to compute $\nabla^2\phi$ for a number of grid orientations and average the $\nabla^2\phi$ thus computed. For example, we shall take the following form:

$$\nabla^2\phi = \tfrac{1}{2}\left[\frac{1}{a^2}(\phi_1 + \phi_2 + \phi_3 + \phi_4 - 4\phi_0) + \frac{1}{(\sqrt{2}a)^2}(\phi_5 + \phi_6 + \phi_7 + \phi_8 - 4\phi_0)\right]. \tag{20}$$

The second term in the right-hand side of the equation represents the finite-difference equivalent of the Laplacian with diagonal grid orientation. For this scheme, the standard deviation of $\nabla^2\phi$ is reduced to 3.20 σ/a^2, in contrast to 4.47 σ/a^2 for the standard scheme.

Strictly speaking, it is to be noted that the mean square value of the first term in the right side of (20) is not equal in general to that of the second term because the grid sizes for these two terms are different (Ogura, 1957). In this sense, there remains some doubt as to the logical bases for taking the average such as (20).

Of course, equation (20) can be regarded as a finite-difference approximation of the Laplacian. Following the procedure mentioned in section 2, we find that the response function for this scheme is

$$\psi_4(ka) = \frac{2}{\pi\left(\frac{ka}{2}\right)^4}\int_0^{\pi/2}\left[\sin^2\left(\frac{ka}{2}\cos\theta\right) + \sin^2\left(\frac{ka}{2}\sin\theta\right) - \sin^2\left(\frac{ka}{2}\cos\theta\right) \times \sin^2\left(\frac{ka}{2}\sin\theta\right)\right]^2 d\theta, \tag{21}$$

and

$$\psi_4(ka) = 1 - \frac{9}{48}(ka)^2 + \cdots,$$

when ka is sufficiently small.

The following table shows some numerical examples of $\psi_4(ka)$ as a function of wave length:

L/a	10	8	6	4
$\psi_4(ka)$	0.934	0.888	0.814	0.624

By comparing this table and table 1, we see the systematic truncation error for scheme (20) is larger than that for the standard scheme, although nine points are used in the former scheme.

The analysis carried out in this section is based upon the assumption that the distribution of errors involved in ϕ themselves is random. The analysis can be extended easily to include more general cases—that is, cases with non-zero correlation between errors at different grid points. From the discussion mentioned above, however, we see that the non-systematic error tends to be large when the systematic error is reduced. Only when the random errors are small, or caused to be small by some suitable smoothing process, can the schemes discussed in section 2 give accurate value of the Laplacian.

Acknowledgment.—The work reported in this paper was initiated and completed during a visit of the writer to the University of Chicago. He is greatly indebted to Professor George W. Platzman for his helpful discussions on this work. Thanks are also due to Dr. Akira Kasahara for his valuable suggestions.

REFERENCES

Knighting, E., 1955: *The reduction of truncation error in symmetrical operators.* Tech. Memo. No. 3, Joint Num. Wea. Pred. Unit, 5 pp.

Ogura, Y., 1957: Spectrum modification due to the use of finite differences. *J. Meteor.*, **14**, 77–80.

——, and K. Miyakoda, 1954: Note on the pressure fluctuations in isotropic turbulence. *J. meteor. Soc. Jap.*, **32**, 160–166.

Thompson, P. D., 1955: *Reduction of trucation errors in the computation of geostrophic vorticity, the Laplacian operator and its inverse.* Tech. Memo. No. 2, Joint Num. Wea. Pred. Unit, 9 pp.