A simple finite-difference grid with non-constant intervals

By HILDING SUNDQVIST, *International Meterological Institute in Stockholm*[1] and GEORGE VERONIS, *Dept. of Geology and Geophysics, Yale University*

(Manuscript received August 5, 1969)

ABSTRACT

A finite difference network with non-uniform grid intervals, h_i, such that $h_i = h_{i-1} + O(h_{i-1}^2)$ is presented. The use of the grid network for two specific choices of $h_i - h_{i-1}$ for a linear boundary layer problem shows that it has distinct advantages in computational efficiency and in accuracy of representation of the analytical solution.

1. Introduction

In the solution of partial differential equations by numerical methods one normally chooses finite-difference grid intervals so that the solution to the finite-difference system agrees as closely as possible with the continuous system which it represents. The choice of the number of points in the finite-difference network is often at least partly determined by practical considerations (electronic computer time etc.). Certainly the most commonly used finite-difference network is one with uniform grid intervals in a given direction. Spatial derivatives are normally represented by centered differences because the latter are accurate representations of the former to $O(h^2)$ where h is the uniform grid interval in the coordinate direction of differentiation.

A simple finite-difference grid with non-constant intervals can be constructed which gives the same accuracy as the uniform grid when derivatives are represented by centered differences. To show this suppose that the range $(0, L)$ of the independent coordinate x is divided into N intervals of non-constant length as shown in Fig. 1. We number the grid points i, $(1 \leq i \leq N+1)$ and represent the value of, x, at point i by x_i so that $x_1 = 0$ and $x_{N+1} = L$. The grid interval between x_i and x_{i+1} is represented by h_i.

Now consider an analytic function, $f(x)$, in the range $0 \leq x \leq L$, with $f_i = f(x_i)$. If we want to express the first and second derivatives of f by their finite difference approximations, we expand f_{i+1} and f_{i-1} in Taylor series about the point, x_i, to derive

$$f_{i+1} = f_i + h_i f_i' + \frac{h_i^2}{2} f_i'' + \frac{h_i^3}{6} f_i''' + \frac{h_i^4}{24} f_i^{IV} + \cdots \quad (1.1)$$

$$f_{i-1} = f_i - h_{i-1} f_i' + \frac{h_{i-1}^2}{2} f_i'' - \frac{h_{i-1}^3}{6} f_i''' + \frac{h_{i-1}^4}{24} f_i^{IV} - \cdots \quad (1.2)$$

where the primed f_i represent derivatives of f evaluated at x_i. We consider f_i' and f_i'' as the two unknowns given by (1.1) and (1.2). The higher order derivatives are also unknown, of course, but as is customary to this order, we are prepared to accept the error involved in neglecting the higher order derivatives. To solve for f_i' we take $h_{i-1}^2 \times (1.1) - h_i^2 \times (1.2)$ to get

$$f_i' = \frac{f_{i+1} - \left(\frac{h_i}{h_{i-1}}\right)^2 f_{i-1} - \left[1 - \left(\frac{h_i}{h_{i-1}}\right)^2\right] f_i}{h_i\left(1 + \frac{h_i}{h_{i-1}}\right)} - \frac{h_i h_{i-1}}{6} f_i''' + \cdots \quad (1.3)$$

To solve for f_i'' we take $h_{i-1} x (1.1) + h_i x (1.2)$ to get

$$f_i'' = \frac{2\left[f_{i+1} + \frac{h_i}{h_{i-1}} f_{i-1} - \left(1 + \frac{h_i}{h_{i-1}}\right) f_i\right]}{h_i h_{i-1}\left(1 + \frac{h_i}{h_{i-1}}\right)} - \frac{h_i - h_{i-1}}{3} f_i''' + \cdots \quad (1.4)$$

[1] Contr. no. 215.

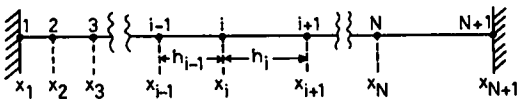

Fig. 1. The finite difference scheme with non-uniform intervals $h_i = x_{i+1} - x_i$.

The first term in expression (1.3) gives f_i' correct to $O(h_i h_{i-1})$, i.e., to second order in the space increments. The first term in expression (1.4) gives f_i'' correct to $O(h_i - h_{i-1})$, i.e. to first order in the difference of the space increments. In the special case of uniform grid intervals, $h_{i-1} \equiv h_i \equiv h$, we note that the last term in (1.4) vanishes identically and (1.3) and (1.4) reduce to

$$f_i' = \frac{f_{i+1} - f_{i-1}}{2h} + O(h^2) \tag{1.5}$$

$$f_i'' = \frac{f_{i+1} - 2f_i + f_{i-1}}{h^2} + O(h^2) \tag{1.6}$$

Hence, for this case both the first and second derivatives are correct to $O(h^2)$ when they are expressed in terms of the values f_{i-1}, f_i and f_{i+1}. The fortuitous cancellation of the first-order error in (1.4) plus ease of manipulation of the expressions to evaluate the finite differences are the main reasons for the great popularity of the uniform grid network.

When $h_{i-1} \neq h_i$, the error involved in neglecting the last term in (1.4) is $O(h_i - h_{i-1})$, i.e., it contains first-order terms in the increments. However, when $h_{i-1} - h_i = O(h_{i-1}^2)$ the error is reduced to second order. Hence, by this choice the f_i''' term contributes an error of the same order as the f^{IV} term and the latter is what is normally considered "acceptable" when a uniform grid is used.

Without additional advantages in using a non-uniform grid, there is no reason to discard the uniform grid which involves simpler manipulations of the finite difference expressions. However, the advantages of the non-uniform grid are easily demonstrated for the numerical analysis of boundary layer problems. In the next section we show how a simple choice of the h_i leads to a grid network which has smaller increments in the vicinity of the boundary $x = 0$ and a coarser network as x increases. In cases where the function f varies sharply near $x = 0$ and is smoother for larger x, just such a network is desirable. In section 3 the simple choice of the h_i is used to solve a particular boundary layer problem and the advantages in representation and computational efficiency are discussed. In section 4 we present a second choice for h_i which is even more accurate.

2. A specific choice of h_i

We choose

$$h_1 = h,\ h_2 = h_1\left(1 + \alpha\frac{h_1}{L}\right),\ \cdots h_i = h_{i-1}\left(1 + \alpha\frac{h_{i-1}}{L}\right),\ \cdots \tag{2.1}$$

where α is a constant of $O(1)$. We note that the coefficient of f_i''' in (1.4) is now $-\frac{1}{3}\alpha h_{i-1}^3$ and thus both f_i' and f_i'' are represented to an accuracy of $O(h_{i-1}^2)$. The case $\alpha = 0$ corresponds to a uniform grid.

In order to get a picture of the effect of the geometrical distortion of the uneven grid on the forms of the derivatives we rewrite expressions (1.3) and (1.4). First, consider the definitions

$$f_{i-\frac{1}{2}}' \equiv \frac{f_i - f_{i-1}}{h_{i-1}}, \quad f_{i+\frac{1}{2}}' \equiv \frac{f_{i+1} - f_i}{h_i} \tag{2.2}$$

These correspond to the finite difference forms of f' evaluated at $i - \frac{1}{2}$ and $i + \frac{1}{2}$ as if we were working with a uniform grid interval; i.e., within each interval they are evaluated as f_i' is in (1.5). Now it is a straight-forward matter to show that the expression for f_i' without the error terms can be written as

$$f_i' = \frac{h_{i-1}}{h_i + h_{i-1}} f_{i+\frac{1}{2}}' + \frac{h_i}{h_i + h_{i-1}} f_{i-\frac{1}{2}}'$$

$$= \frac{h_{i-1}}{h_i + h_{i-1}}\left[f_{i+\frac{1}{2}}' + f_{i-\frac{1}{2}}' + \frac{\alpha h_{i-1}}{L} f_{i-\frac{1}{2}}'\right] \tag{2.3}$$

The first equality in (2.3) shows that f_i' in (1.3) is expressed as a weighted average of $f_{i-\frac{1}{2}}'$ and $f_{i+\frac{1}{2}}'$. The coefficients of $f_{i-\frac{1}{2}}'$ and $f_{i+\frac{1}{2}}'$ reflect the relative distances of the point i to the points $i - \frac{1}{2}$ and $i + \frac{1}{2}$, with the greater weighting factor associated with the closer point. The second equality in (2.3) indicates that with the choice (2.1) for h_i, f_i' is proportional to the mean of $f_{i-\frac{1}{2}}'$ and $f_{i+\frac{1}{2}}'$ over the interval $(h_i + h_{i-1})$ plus a correction factor of $O(h_{i-1})$ which gives somewhat greater weight to the term $f_{i-\frac{1}{2}}'$. From Fig. 2 we can also appreciate graphically that an

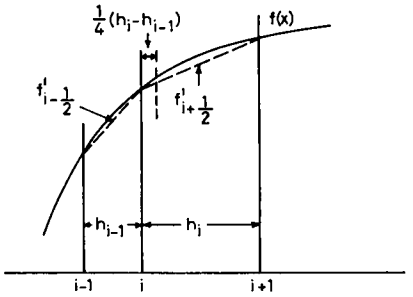

Fig. 2. A simple average of $f'_{i-\frac{1}{2}}$ and $f'_{i+\frac{1}{2}}$ would give f' at the point $x'_i = x_i + \frac{1}{4}(h_i - h_{i-1})$; however the desired accuracy in f'_i requires a correction term that gives more weight to $f'_{i-\frac{1}{2}}$ which is closer to the point i than is $f'_{i+\frac{1}{2}}$. The difference $f'_{i+\frac{1}{2}} - f'_{i-\frac{1}{2}}$ is proportional to f'' at x'_i and that approximates f''_i directly to $O(h^2_{i-1})$.

evaluation of f'_i in terms of $f'_{i-\frac{1}{2}}$ and $f'_{i+\frac{1}{2}}$ must reflect the stronger influence of the former because f'_i has a steeper slope than is given by a simple average of $f'_{i-\frac{1}{2}}$ and $f'_{i+\frac{1}{2}}$ over the interval $(i-\frac{1}{2},\ i+\frac{1}{2})$. Finally, we note that taking a simple average of $f'_{i-\frac{1}{2}}$ and $f'_{i+\frac{1}{2}}$ to evaluat ef'_i corresponds to treating the point i as if it were halfway between $i-\frac{1}{2}$ and $i+\frac{1}{2}$. From Fig. 2 it is evident that this deacreases both $f'_{i-\frac{1}{2}}$ and $f'_{i+\frac{1}{2}}$ and hence the mean. The weighting factor corrects for this decrease.

The second derivative in (1.4) can be rewritten without the error terms as

$$f'' = \frac{f'_{i+\frac{1}{2}} - f'_{i-\frac{1}{2}}}{\frac{1}{2}(h_i + h_{i-1})} \tag{2.4}$$

Note again that taking the simple difference of $f'_{i-\frac{1}{2}}$ and $f'_{i+\frac{1}{2}}$ to evaluate f''_i corresponds to treating the point i as if it were halfway between $i-\frac{1}{2}$ and $i+\frac{1}{2}$. Even though both $f'_{i-\frac{1}{2}}$ and $f'_{i+\frac{1}{2}}$ decrease by this procedure, their difference tends to cancel out the decrease. By our choice of h_i the net error that is made is $O(h^2_{i-1})$ and is therefore acceptable.

It may seem that little is to be gained by the small change in the grid length given by the choice (2.1). However, when many grid intervals are taken, the difference between h_N and h_1 can be quite large. In cases where the function, f, changes rapidly near $x=0$ and less rapidly near $x=L$, a substantial saving in computation can be gained by such a choice. In the next section we use the present procedure to analyze a

Table 1. *Values of h_i and x_i for 20 intervals in the range $0 \leqslant x \leqslant \pi$ using (2.1) with $\alpha = 2$*

i	h_i	x_i	i	h_i	x_i	i	h_i	x_i
1	.0731	0	8	.1057	.5962	15	.1870	1.530
2	.0765	.0731	9	.1129	.7019	16	.2092	1.717
3	.0802	.1496	10	.1210	.8148	17	.2371	1.926
4	.0843	.2298	11	.1303	.9357	18	.2729	2.163
5	.0888	.3141	12	.1411	1.066	19	.3203	2.436
6	.0938	.4029	13	.1538	1.207	20	.3856	2.756
7	.0994	.4967	14	.1688	1.361	21	—	3.142

boundary-layer problem which was put forth by Stommel (1948) as an idealized model of wind-driven ocean circulation. In that problem the function, ψ, varies sharply near $x=0$ and then changes slowly over the major part of the range $0 \leqslant x \leqslant \pi$. If a grid interval no larger than 0.1 is needed near $x=0$, at least 31 intervals are required for a uniform grid. Table 1 shows the intervals and values of x which are given by 20 intervals using (2.3) with $\alpha = 2$. Note that near $x=0$ the grid interval is less than 0.1, as desired, but near $x=\pi$ where a coarser grid is admissable the interval is 0.39.

3. Application to a boundary-layer problem

As an example of the use of the non-uniform grid interval we apply it to the theoretical model of wind-driven ocean circulation proposed by Stommel (1948). The equation for the model can be reduced to the boundary value problem given by

$$\varepsilon(\psi'' - \psi) + \psi' = -\sin x, \quad \varepsilon \ll 1, \quad \psi' = \frac{d\psi}{dx} \tag{3.1}$$

$$\psi = 0 \quad \text{at} \quad x = 0, \pi \tag{3.2}$$

The analytic solution is

$$\psi = \frac{1}{1+4\varepsilon^2}\left\{2\varepsilon \sin x + \cos x + \frac{1}{e^{\pi D_1} - e^{\pi D_2}} \times [(1+e^{\pi D_2})e^{D_1 x} - (1+e^{\pi D_1})e^{D_2 x}]\right\} \tag{3.3}$$

where

$$D_1 = \frac{-1-\sqrt{1+4\varepsilon^2}}{2\varepsilon}, \quad D_2 = \frac{-1+\sqrt{1+4\varepsilon^2}}{2\varepsilon}.$$

A plot of ψ vs. x is given in Fig. 3 for $\varepsilon = 0.05$.

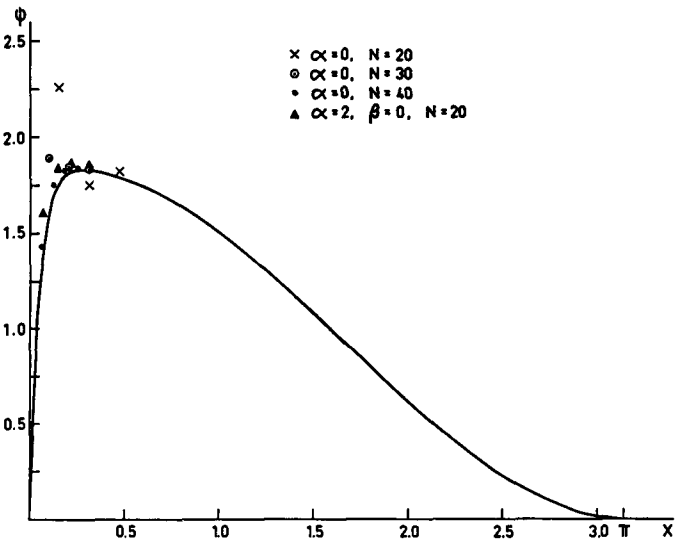

Fig. 3. The exact solution, ψ, of the problem posed by (3.1) and (3.2) is shown as the solid curve. The approximate solution obtained with $N = 20$, 30 and 40 for $\alpha = 0$, and with $N = 20$ for $\alpha = 2$ and $\beta = 0$ are shown by the designated points.

A numerical solution of the problem posed by (3.1) and (3.2) can be obtained by solving the finite difference analog of (3.1) which is

$$A_i \psi_{i+1} - B_i \psi_i + C_i \psi_{i-1} = -\frac{\sin x_i}{\varepsilon} \qquad (3.4)$$

where

$$A_i = \frac{2\varepsilon + h_{i-1}}{\varepsilon h_i (h_i + h_{i-1})} \qquad (3.4\,a)$$

$$B_i = 1 + \frac{2\varepsilon + h_{i-1} - h_i}{\varepsilon h_i h_{i-1}} \qquad (3.4\,b)$$

$$C_i = \frac{2\varepsilon - h_i}{\varepsilon h_{i-1}(h_i + h_{i-1})} \qquad (3.4\,c)$$

and

$$h_i = h_{i-1}\left(1 + \alpha \frac{h_{i-1}}{\pi}\right).$$

The solution can be derived by Gauss' elimination procedure as described in Richtmeyer and Morton (1967).[1]

The solution obtained with $\alpha = 0$ (i.e. a uniform grid) and 20 grid intervals ($N = 20$) is shown by the crosses in Fig. 3. Twenty intervals are obviously inadequate because the boundary-layer thickness is too small to be resolved by such a coarse grid. This inadequacy is reflected by the large distortions and the oscillatory behaviour near $x = 0$. For $x \geq 0.5$ the numerical solution approximates the analytic solution with an error of less than 1 %. Thirty grid intervals yield the solution denoted by the open circles in Fig. 3. Hence, the solution is an improvement over that with 20 intervals but there is still a distortion of similar character near $x = 0$. With 40 grid intervals the solution is much smoother (see black circles in Fig. 3) but considerable more than 100 points are required to approximate the analytical solution near $x = 0$ with an error of less than 1 %.

The solution obtained for $\alpha = 2$ and $N = 20$ is indicated by the black triangles in Fig. 3. The oscillatory behavior resulting from the use of the uniform grid of 20 points is no longer present. This improvement is due to the denser grid network near $x = 0$ when $\alpha = 2$.

As we observed earlier, the network with $\alpha > 0$ has larger grid intervals intervals near $x = \pi$. Since the error due to neglected terms is $O(h_{i-1}^2)$ this means that larger errors are

[1] The coefficient matrix is positive definite if $B_i > |A_i| + |C_i|$. The latter is a sufficient condition for the error of the results to be the same order as the truncation error of the finite-difference forms. In our calculations the condition is not satisfied for $N < 32$ and for these cases the error of the results can be considerably larger than the truncation error of the formula. However, this point is irrelevant to the comparisons which we make between the constant and non-constant grid calculations.

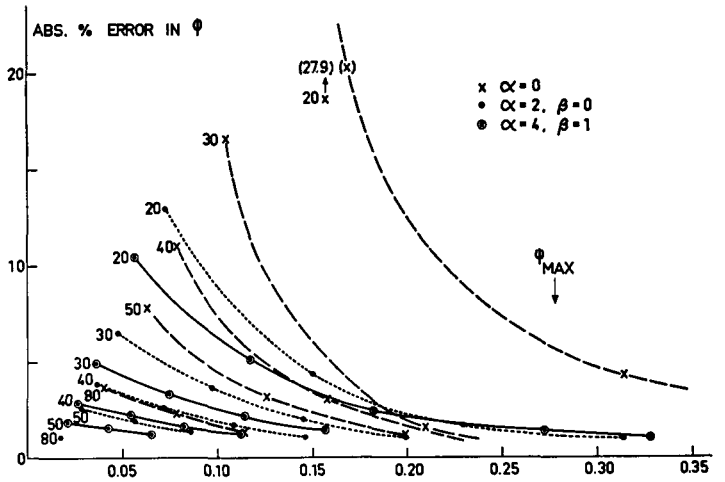

Fig. 4. Absolute value of percentage error in ψ, relative to the exact solution in the range $0 \leqslant x \leqslant 0.35$, and for various values of N and for three pairs of values of α and β. The curves connecting the grid points are drawn in as visual aids but have no meaning in the intervals between grid points. The associated value of N is written at the left extremity of each curve.

encountered near $x = \pi$ when $\alpha > 0$. E.g., with $\alpha = 2$, $N = 20$ the error at the last interior grid point is 3.4 % compared with 0.6 % when $\alpha = 0$. Thus if too large a value of α is used, the error near $x = \pi$ may become troublesome.[1] On the other hand, in order to take advantage of the non-uniform grid intervals the value of α cannot be chosen too small. For the problem that we have treated the value of $\alpha = 2$ seems to be optimum when $N = 20$.

As N is increased, the optimum value of α increases also. The reason for this is that increasing N tends to cut down the large interval near $x = \pi$ so that the associated error is decreased. Hence, increasing α provides the advantage of increasing the grid density near $x = 0$ with a tolerable error near $x = \pi$ when N is large. With 80 grid intervals the optimum choice is $\alpha = 3$. For this case all errors are considerably less than 1 %. With $\alpha = 2$ and $N = 50$ the error in ψ near $x = 0$ is smaller than that obtained with $\alpha = 0$ and $N = 80$. The accuracy obtained with $\alpha = 2$ and $\alpha = 0$ for different values of N is summarized in Fig. 4 where we plot the percentage error in ψ (only those errors > 1 % are shown) vs. position for $0 \leqslant x \leqslant 0.35$.

[1] For boundary layer problems the function often becomes smoother with increasing distance from the boundary so that the distortion in curvature at large distances may not be serious. For the present problem the larger errors near $x = \pi$ where ψ and ψ' are small are not as serious as large errors near $x = 0$ where ψ and ψ' are large.

4. A second expression for h_i

Because of the larger errors which occur near $x = \pi$ with $\alpha > 0$ we have modified the expression for h_i by taking

$$h_i = h_{i-1}\left[1 + \frac{\alpha}{\pi}\left(\frac{\pi - x_i}{\pi}\right)^{\beta} h_{i-1}\right] \qquad (4.1)$$

where α and β are constants. With $\beta = 0$ (4.1) reduces to (2.1). With $\beta > 0$ the form (4.1) tends to cut down the increase of h_i as i increases. E.g., for $N = 20$ with $\alpha = 2$, $\beta = 1$ the last grid interval is 0.225 as compared to 0.386 with $\alpha = 2$, $\beta = 0$. This means that the error near $x = \pi$ can be reduced considerably. It also means that the grid density near $x = 0$ is reduced but we can use larger values of α to increase the grid density near $x = 0$. The consequence of increasing α is that larger errors are obtained over the entire range. Hence, one must chose α and β so that the errors are minimized in some sense.

Numerical tests with the problem of section 3 were made for $N = 20$ with $\alpha = 1, 2, 3, 4$ and $\beta = 1, 2, 3, 4$. The best pair of values are $\alpha = 4$, $\beta = 1$. The error near $x = \pi$ for this choice is only 2.5 %.

Fig. 4 contains a comparision of the errors in ψ obtained by use of (α, β) given by (0,0), (2,0) and (4,1) in the range $0 \leqslant x \leqslant 0.35$.

The case $\alpha = 0$ is obviously the least accurate of the three near $x = 0$. For $N = 20$ the spurious oscillatory behavior shown by the crosses in Fig. 3 appears here as a large positive error (27.9%) at $x = 0.157$ and a negative error[1] (−4.2%) at $x = 0.314$. As N is increased the error decreases but even for $N = 80$ the value of ψ at the first three grid points contains an error greater than 1%.

Furthermore, the first derivative of ψ is very badly represented by $\alpha = 0$, $N = 20$. Much too large a value for ψ' occurs at $x = 0$ and a spurious large negative value occurs at the third grid point. Since ψ' corresponds to the velocity for this problem a serious error in interpretation would result from the use of $\alpha = 0$, $N = 20$. This error is effectively eliminated when $\alpha \geqslant 2$.

With $\alpha = 2$, $\beta = 0$ the results are significantly inproved. The error for $N = 20$ is less than half that obtained in the previous case. The improvement as N is increased is even greater. With $N = 80$ the error in ψ at the first grid point is 1% and is considerably less than 1% for all the remaining grid points.

Use of $\alpha = 4$, $\beta = 1$ improves the results even more. When $N = 50$ errors between 1% and 2% occur only for $x < 0.07$ and are elsewhere less than 1%. The vast improvement over the results obtained with the uniform grid with $N = 50$ is obvious. Even with only 30 grid points the solution is as good as that obtained with $\alpha = 0$ and $N = 60$ near $x = 0$ though there is now an error of 1.3% at $x = 2.9$. The results for this case are also significantly better than those obtained for $\alpha = 2$, $\beta = 0$.

A single calculation for $N = 50$ with $\alpha = 5$, $\beta = 1$ shows that the results are even better than for $\alpha = 4$, $\beta = 1$. The reason for this is obvious. With $\beta = 1$ the very large h_i near $x = \pi$ which result from use of large α are reduced, hence larger α can be used to advantage.

Concluding remarks

The non-uniform grid that we have used has distinct advantages for boundary layer problems. There are, of course, many different ways to choose the h_i and we have presented but two of them. Other choices may be more suitable for specific problems (even for this one) but we have made no attempt at finding an optimum general choice. Furthermore, there are other properties such as numerical stability considerations when such a grid is used in a prognostic equation, which should be understood. It is evident that more complete investigation of the consequences of chosing $h_i - h_{i-1} = O(h_{i-1}^2)$ would be very desirable.

Acknowledgement

The research was supported in part by NSF Grant-1416.

REFERENCES

Richtmyer, R. D. & Morton, K. W. 1967. *Difference methods for initial-value problems*, 198–201. Interscience Publ., New York.

Stommel, H. 1948. The westward intensification of winddriven ocean currents. *Trans. Am. Geoph. Union 29*, 202–206.

[1] The absolute value is plotted in Fig. 4.

ПРОСТАЯ КОНЕЧНО-РАЗНОСТНАЯ СЕТКА С НЕПОСТОЯННЫМИ ИНТЕРВАЛАМИ

Описана конечно-разностная схема с неоднородными интервалами h_i такими, что $h_i = h_{i-1} + O(h_{i-1}^2)$. Использование этой сетки для двух специальных случаев выбора $h_i - h_{i-1}$ для линейной задачи о пограничном слое показывает, что она имеет отчетливые преимущества в эффективности вычислений и в точности представления аналитического решения задачи.