On the Truncation Error in the Solution of Laplace's Equation by Finite Differences [1]

Wolfgang Wasow

The difference between the solution of the Laplace differential equation and the Laplace difference equation, defined in the same rectangle and assuming the same boundary values, is estimated under the assumption that the boundary function possesses a bounded third derivative. The bound obtained is of the order of magnitude of the square of the mesh length.

1. Introduction

Probably the most widely used approach for the numerical solution of differential equations for which simple analytic solutions are not available consists in replacing the differential equation by a finite difference equation. The difference between the exact solutions of these two problems is usually called—somewhat inappropriately—the *truncation error.*

It is highly desirable, for numerical calculations, to have adequate estimates of this truncation error. There exist in the literature numerous investigations of this problem. We shall be primarily concerned with the case of the first boundary value problem for elliptic partial linear differential equations. Some references to the literature for this case are listed at the end of this paper. All these investigations, as far as they are mathematically complete, suffer from the defect that the estimates given depend on bounds for the derivatives of the unknown solution of the differential equation itself. Such bounds can in general not be found without effectively solving the differential equation. Thus the value of these estimates consists essentially in giving information about the order of magnitude of the truncation error in terms of the mesh length used, and in helping to form reasonable guesses as to the size of the error.

It turns out that, with the difference expressions used in most numerical work, and for a square net of mesh length h the truncation error is of the order $O(h^2)$, provided the boundary values and the equations of the boundary curve C possess *bounded fourth derivatives*, and provided the boundary values of the difference equation problem are suitably adjusted (see, e. g. [3] [2]). This estimate is then valid in the whole domain B.

The bound for the truncation error given by P. C. Rosenbloom in [6] is somewhat different. It depends only on the modulus of continuity in $B+C$ of the unknown harmonic function, but it is a pointwise estimate that deteriorates indefinitely at points near the boundary.

Thus there remain a number of unsolved problems of theoretical as well as computational interest: The ideal would be a pointwise estimate of the truncation error, valid if the boundary curve and the boundary values are piecewise analytic. This error estimate

should have the highest possible order in h, and, at the same time, it should not be unduly wasteful. It should, of course, depend only on quantities that can be easily computed from the data.

From the *theoretical* viewpoint the order of magnitude of the error is the most interesting quantity. In particular, one would like to know the exact way in which a weakening of the smoothness required of the data affects the global and local order of the error.

For *computational* purposes an estimate will be most useful if it yields small bounds for the error at moderate sizes of h. Thus $10h$ is a better bound in practice than $10,000h^2$.

The contents of this note are intended as a first exploratory step toward the examination of these questions. The error estimate for Laplace's equation in a rectangle given below in formulas (18) and (19) is of order $O(h^2)$, with coefficients of moderate size, although it depends on bounds for the *third* derivatives of the boundary values only, and in spite of the fact that a rectangle is a domain with corners. It is hoped that this special result may help to stimulate more general investigations.

2. Truncation Error for Dirichlet's Problem in a Rectangle

Let B be the rectangle with vertices $(0,0)$, $(1,0)$, $(1,b)$, $(0,b)$ where b is a rational number. The prescribed boundary function f is assumed to be continuous on the boundary C of B and to possess bounded *third* derivatives on each closed side of C.

We denote the Laplacian, as usual, by Δu and write $\Delta_h v$ for the expression

$$\frac{1}{h^2}[v(x+h,y)+v(x-h,y)+v(x,y+h)$$
$$+v(x,y-h)-4v(x,y)].$$

For any value of h such that $1/h$ and b/h are integers we denote by $u_h(x,y)$ the function, defined at all netpoints (nh, mh) in $B+C$, where n,m are integers, for which

$$\Delta_h u_h = 0, \quad \text{in } B$$

and

$$u_h = f, \quad \text{on } C.$$

If u designates the solution of the corresponding Dirichlet problem for the operator Δ, then we are concerned with the truncation error $u-u_h$.

It is desirable to deal with boundary values that vanish at the four vertices. To this end we introduce the harmonic polynominal $Q(x,y)$, defined by $b \cdot Q(x,y) = A_{00}(x-1)(y-b) - A_{10}x(y-b) + A_{10}xy - A_{01}(x-1)y$, where A_{00}, A_{10}, etc. are the values of f at the vertices. This function solves both the difference and the differential equation. If we replace the boundary values f by

$$f^* = f - Q,$$

we obtain therefore a new problem for which the truncation error is the same as for the original problem, since the truncation error corresponding to the boundary problem determined by the values of $Q(x,y)$ on C is zero. It is clear that f^* vanishes at the vertices. Also the second and higher derivatives of f^* are the same as those of f.

The truncation error can be considered the sum of four terms, each corresponding to boundary values that are zero except on one of the four sides of the rectangle. Therefore, we assume temporarily that f^* is zero except on the side $y=0$. There $f^* = f^*(x)$ is a function of x alone.

The solutions of $\Delta u = 0$ and $\Delta_h u_h = 0$ for these boundary values can be written, respectively,

$$u(x,y) = \sum_{n=1}^{\infty} c_n g(y, n\pi) \sin n\pi x, \qquad (1)$$

$$u_h(x,y) = \sum_{n=1}^{1/h} \gamma_n g(y, \beta_n/h) \sin n\pi x, \qquad (2)$$

where

$$c_n = 2\int_0^1 f^*(t) \sin n\pi t\, dt \qquad (3)$$

$$\gamma_n = 2h\sum_{r=1}^{1/h} f^*(rh) \sin n\pi r h \qquad (4)$$

$$g(y,z) = \sinh (b-y)z/\sinh bz, \qquad (5)$$

and β_n is the solution of

$$\sinh \frac{\beta_n}{2} = \sin \frac{nh\pi}{2}. \qquad (6)$$

These well-known formulas can be readily verified by substitution.

We next collect a few facts needed for the comparison of u and u_h.

To find an estimate of the Fourier coefficients c_n we integrate the right member of formula (3) by parts three times, obtaining the formula

$$c_n = \frac{2}{(n\pi)^3}[(-1)^{n-1}f''(1) - f''(0)]$$

$$+ \frac{2}{(n\pi)^3}\int_0^1 f'''(t) \cos n\pi t\, dt. \qquad (7)$$

This implies

$$|c_n| \leq Kn^{-3}, \qquad (7)$$

where

$$K = \frac{2}{\pi^3}[|f''(1)| + |f''(0)|] + \frac{4}{\pi^3} \max |f'''(x)|. \qquad (8)$$

The Fourier coefficients c_n and the "interpolation coefficients", γ_n, are related by the equation

$$\gamma_n = c_n + \sum_{k=1}^{\infty}(c_{2k/h+n} - c_{2k/h-n}), \quad (n=1,2,\ldots,1/h).$$

This relation can readily be proved by expressing $f^*(rh)$ in (4) by its Fourier series. (See also the references in [7].) Combining (7) with the last formula we find, for $n \leq 1/h$,

$$|c_n - \gamma_n| \leq K\sum_{k=1}^{\infty}\{(2k/h+n)^{-3} + (2k/h-n)^{-3}\}$$

$$\leq K\frac{h^3}{8}\left\{\left(1+\frac{nh}{2}\right)^{-3} + \left(1-\frac{nh}{2}\right)^{-3}\right.$$

$$\left. + \int_1^{\infty}\left[\left(x+\frac{nh}{2}\right)^{-3} + \left(x-\frac{nh}{2}\right)^{-3}\right]dx\right\}.$$

Hence,

$$|c_n - \gamma_n| \leq 1.7Kh^3, \quad \text{for } n \leq 1/h. \qquad (9)$$

Next we need an estimate of the difference $g(y, n\pi) - g(y, \beta_n/h)$. To this end we express $\beta_n/2$, as given by (6), in terms of $nh\pi/2$ by means of Taylor's formula. A straightforward calculation shows that

$$\frac{\beta_n}{2} = \frac{nh\pi}{2} + \left(\frac{nh\pi}{2}\right)^3 R$$

where R is the value of

$$-\frac{1}{3}\cos \sigma (1 - 2\sin^2 \sigma)(1 + \sin^2 \sigma)^{-5/2}$$

at a certain mean value of σ in the interval

$$0 < \sigma < nh\pi/2.$$

Hence

$$|\beta_n - nh\pi| \leq (nh\pi)^3/12. \qquad (10)$$

In order to estimate the difference $g(y, n\pi) - g(y, \beta_n/h)$ with the help of the theorem of the mean we have to estimate the right member of the formula

$$\frac{d}{dz}g(y,z) = \frac{1}{2}\sinh^{-2} bz$$

$$\times \{(2b-y)\sinh yz - y\sinh(2b-y)z\}. \qquad (11)$$

We observe that $\sinh t/t$ is an increasing function of t, for $t > 0$. Therefore,

$$0 \leq \frac{d}{dz}g(y,z) \leq \frac{y}{2}\sinh^{-2} bz \sinh(2b-y)z,$$

$$\text{for } z \geq 0, y \geq 0. \qquad (12)$$

To this expression we shall apply the inequalities

$$\sinh kt \le e^{(k-1)t}\sinh t, \quad \text{for } 0 \le k \le 1, \qquad (13)$$

and

$$\sinh t \ge \frac{1}{2}(1-e^{-2t_1})e^t, \quad t \ge t_1 > 0. \qquad (14)$$

The second of these inequalities is obvious. The first follows from

$$\sinh kt/\sinh t = e^{(k-1)t}(1-e^{-2kt})/(1-e^{-2t}).$$

Also, we shall need the inequalities

$$\frac{4}{3}n \le \beta_n/h \le \pi n, \quad \text{for } 1 \le n \le 1/h. \qquad (15)$$

The second of these is a consequence of (6); the first can be proved as follows:

$$\sinh\frac{\beta_n}{2} = \frac{nh\pi}{2}\left(\sin\frac{nh\pi}{2}\bigg/\frac{nh\pi}{2}\right) \ge \frac{nh\pi}{2}\cdot\frac{2}{\pi} = nh,$$

and

$$\sinh\frac{\beta_n}{2} \le \frac{nh\pi}{2} < \frac{\pi}{2}.$$

Since $\sinh x/x$ is an increasing function of x for $x = 0$,

$$\sinh\frac{\beta_n}{2}\bigg/\frac{\beta_n}{2} \le \frac{\pi}{2}\bigg/\operatorname{arsinh}\frac{\pi}{2} \le \frac{3}{2}.$$

Hence

$$\frac{\beta_n}{2} \ge \frac{2}{3}\sinh\frac{\beta_n}{2} \ge \frac{2}{3}nh,$$

which completes the proof of (15). Applying (13), (14) and (15) to (12) we obtain

$$\left|\frac{d}{dz}g(y,z)\right| \le (1-e^{-8b/3})^{-1}ye^{-z}, \text{ for }\begin{cases}1 \le n \le 1/h \\ \beta_n/h < z < n\pi \\ 0 \le y \le b\end{cases}$$

and therefore, using (10),

$$\left|g\left(y,\frac{\beta_n}{h}\right) - g(y,n\pi)\right| \le (1-e^{-8b/3})^{-1}\frac{\pi^3}{12}ye^{-4n\pi/3}n^3h^3,$$
$$\text{for } 1 \le n \le 1/h. \quad (16)$$

We conclude these preparatory remarks with the inequality

$$0 \le g(y,z) \le 1, \quad \text{for } 0 \le y \le b, \qquad (17)$$

which is an immediate consequence of the definition of $g(y,z)$.

We are now ready to estimate $|u-u_h|$. To this end we write, using (1) and (2),

$$|u-u_h| \le R_1 + R_2 + R_3,$$

$$R_1 = \sum_{n=1}^{1/h}|c_n - \gamma_n|g(y,n\pi)$$

$$R_2 = \sum_{n=1}^{1/h}|\gamma_n|\cdot|g(y,n\pi) - g(y,\beta_n/h)|$$

$$R_3 = \sum_{n=1+1/h}^{\infty}|c_n|g(y,n\pi).$$

From (9) and (17) we conclude that

$$R_1 \le 1.7Kh^2$$

and from (7) it follows that

$$R_3 \le K\sum_{n=1+1/h}^{\infty}n^{-3} = Kh^2/2.$$

In order to estimate R_2, we note that, in view of (7) and (9),

$$|\gamma_n| \le K(n^{-3} + 1.7h^3).$$

Using also (16) we obtain, by summation,

$$R_2 \le (1-e^{-8b/3})^{-1}K(2.7)\frac{\pi^3}{12}y(e^{4y/3}-1)^{-1}h^2.$$

Since

$$y/(e^{4y/3}-1) \le \frac{3}{4}, \quad \text{for } y \ge 0,$$

this yields

$$R_2 \le 5.24(1-e^{-8b/3})^{-1}Kh^2$$

and, hence, adding these inequalities,

$$|u-u_h| \le [2.3 + 5.24(1-e^{-8b/3})^{-1}]Kh^2.$$

If we denote by M_2 the numerical maximum of the second derivative of the given boundary function at the vertices of the rectangle, and by M_3 the numerical maximum, anywhere on C, of the third derivative of the boundary function, the last inequality can be replaced, in view of (8), by

$$|u-u_h| \le [.297 + .676(1-e^{-8b/3})^{-1}][M_2 + .319M_3]h^2.$$

Finally, if the sides of the rectangle are A, B instead of $1,b$ a homothetic transformation yields the formula

$$|u-u_h| \le [.297 + .676(1-e^{-8B/3A})^{-1}][M_2 + .319AM_3]h^2.$$

We recall that this formula is valid only for boundary values that vanish except on the side $y=0$. For the full truncation error three analogous expressions have to be added to the right member of the last inequality, leading—with some rounding off to shorten the expression—to

$$|u-u_h| \le [.6 + 1.4(1-e^{-8B/3A})^{-1}][M_2 + .32AM_3]h^2$$
$$+ [.6 + 1.4(1-e^{-8A/3B})^{-1}][M_2 + .32BM_3]h^2. \qquad (18)$$

For the square of side A this reduces to

$$|u-u_h| \leq (4.3 M_2 + 1.4 A M_3) h^2. \qquad (19)$$

Added in proof: The function

$$u_h^*(x,y) = \sum_{n=1}^{\infty} c_n g(y, \beta_n/h) \sin n\pi x$$

coincides with $u_h(x,y)$ at all net points. Its use instead of $u_h(x,y)$ simplifies the calculations somewhat and leads to better estimates for the truncation error, namely

$$|u-u_h^*| \leq [.13 + .5(1-e^{-\pi B/3A})^{-1}][M_3 + 0.32 A M_3] h^2$$

$$+ [.13 + .5(1-e^{-\pi B/3A})^{-1}][M_2 + 0.32 B M_3] h^2$$

and

$$|u-u_h^*| \leq (1.4 M_2 + 0.43 A M_3) h^2$$

instead of (18) and (19).

3. References

[1] L. F. Richardson, The approximate arithmetic solution by finite differences of physical problems involving differential equations, etc., Phil. Trans. [A] **210**, 307 to 357 (1911).
[2] S. Gerschgorin, Fehlerabschätzung für das Differenzenverfahren zur Lösung partieller Differentialgleichungen, Z. Angew. Math. Mech. [4] **10**, 373 to 382 (1930).
[3] L. Collatz, Bemerkungen zur Fehlerabschätzung für das Differenzenverfahren bei partiellen Differentialgleichungen, Z. Angew. Math. und Mech. **12**, 56 to 57 (1933).
[4] L. Collatz, Das Differenzenzenverfahren mit höherer Approximation für lineare Differentialgleichungen, Schriften des Mathematischen Seminars, Univ. Berlin **8**, 1 to 34 (1935).
[5] M. Inoue, Sur les fonctions de noeud et leurs applications à l'integration numérique des équations aux dérivées partielles, Mem. Faculty Science, Kyusyu Univ. [A], IV, No. 2, 107 to 178 (1949).
[6] P. C. Rosenbloom, The difference equation method for solving the Dirichlet problem, To appear in Proc. of a Symposium on the Construction and Application of Conformal Maps, NBS Applied Mathematics Series.
[7] H. Burkhardt, Trigonometrische Interpolation, Enzyklopädie der mathematischen Wissenschaften [A] II, 9A, p. 651 (1904).

Los Angeles, July 26, 1951.

○