MATHEMATICS OF COMPUTATION, VOLUME 28, NUMBER 126, APRIL 1974, PAGES 369–391

High-Order Finite-Difference Methods for Poisson's Equation

By H. J. van Linde

Abstract. Finite-difference approximations to the three boundary value problems for Poisson's equation are given with discretization errors of $O(h^3)$ for the mixed boundary value problem, $O(h^3|\ln h|)$ for the Neumann problem and $O(h^4)$ for the Dirichlet problem, respectively. These error bounds are an improvement upon similar results obtained by Bramble and Hubbard; moreover, all resulting coefficient matrices are of positive type.

I. Introduction. In this paper, we shall consider the solution by finite-difference methods of the three boundary value problems for Poisson's equation

$$-\Delta u = f \quad \text{in } R, \tag{1.1}$$

R a bounded connected open set in the (x, y) plane with boundary C. The symbol Δ denotes the Laplace operator

$$\Delta \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2.$$

The Dirichlet problem for this equation is

$$\begin{aligned} -\Delta u &= f \quad \text{in } R, \\ u &= g \quad \text{on } C. \end{aligned} \tag{1.2}$$

It is well known that a unique solution exists under very general assumptions on R and the known functions f and g.

The Neumann problem is

$$\begin{aligned} -\Delta u &= f \quad \text{in } R, \\ \partial u/\partial n &= g \quad \text{on } C, \end{aligned} \tag{1.3}$$

$\partial/\partial n$ denoting differentiation with respect to the outward-directed normal on C. From Green's first identity, it follows that f and g must satisfy the relation

$$\int_R f\, d\sigma + \int_C g\, ds = 0. \tag{1.4}$$

Again, under general assumptions, a solution, unique except for an additive constant, exists.

Finally, the third (or Robin) boundary value problem can be formulated as

Received February 20, 1973.

AMS (MOS) subject classifications (1970). Primary 35J05, 65N05, 65N10; Secondary 39A05, 41A25.

$$\begin{aligned} -\Delta u &= f \quad \text{on } R, \\ \partial u/\partial n + \alpha u &= g \quad \text{on } C_1, \qquad (1.5) \\ u &= g_1 \quad \text{on } C_2. \end{aligned}$$

It is assumed here that the boundary C consists of the two parts C_1 and C_2. We require that the function α be piecewise continuous on C_1 with a finite number of discontinuities and twice piecewise differentiable.

Further, at all points of continuity, either $\alpha = 0$ (the set $C_1^{(1)}$) or $\alpha \geqq \alpha_m > 0$, where α_m is a constant (the set $C_1^{(2)}$). We need only consider the cases where either C_2 or $C_1^{(2)}$ contains a nonempty subset of C, since, otherwise, we again have the Neumann problem; these cases provide a unique solution under general assumptions on R and f, g and g_1.

The most accurate finite-difference schemes to date for Poisson's equation have been devised by Bramble and Hubbard. Covering the region R by a square net with mesh width h, they formulated finite-difference analogues with an error estimate of $O(h^4)$ for the Dirichlet problem [1], $O(h^2 \,|\ln h|)$ for the Neumann problem [2], and $O(h^2)$ for the Robin problem [3].

In this work, which is based upon the author's thesis [4], we shall propose a finite-difference analogue for the third boundary value problem with an error estimate of $O(h^3)$ and one for the Neumann problem that converges as $O(h^3 \,|\ln h|)$. An $O(h^4)$ approximation for the Dirichlet problem will be given, which is of positive type, with an error bound which is never worse than the one proposed by Bramble and Hubbard [1].

We shall cover the region R under consideration by a square net with mesh width h and we shall call the crossings of the net lines mesh points. We introduce a point set R_h, consisting of all those mesh points of R whose eight nearest neighbors are also in R.

The intersection points of the net with the boundary C of R make up a set C_h, subdivided for the third problem in C_{1h} and C_{2h}. Together, the mesh points of R which are not in R_h form a set $C_h{}^*$. This set may be divided into two sets $C_{1h}{}^*$ and $C_{2h}{}^*$ for the Robin problem. The exact way in which this is done will be considered later.

We have to define a suitable finite-difference approximation Δ_h to the Laplace operator Δ in R_h and $C_h{}^*$ and, in the case of the Neumann and Robin problems, an analogue δ_n for the operator $\partial/\partial n$ on C_h. From the work of Bramble and Hubbard, it can be inferred that, in order that the above proposed error estimates be attained, we need an approximation Δ_h to Δ with a truncation error of $O(h^4)$ in R_h, $O(h^2)$ in $C_{1h}{}^*$ and $O(h)$ in $C_{2h}{}^*$. We shall also have to find an approximation δ_n for $\partial/\partial n$ on C_h with a truncation error of $O(h^3)$.

In [3], Bramble and Hubbard gave an approximation to the operator $\partial/\partial n$ with a truncation error of $O(h^2)$. In Section II, we shall show that an easier proof of their results can be given which also makes the results valid under less severe restrictions. Moreover, this different approach makes it possible to construct an analogue to $\partial/\partial n$ which can be shown to have a truncation error of $O(h^3)$, the proof of which under the original approach would have been prohibitive.

In Section III, a suitable approximation to the Laplace operator for the set $C_h{}^*$

will be derived with a truncation error of $O(h^2)$.

In R_h, we shall use the well-known nine-point approximation to Δ; if (x, y) is a point of R_h, then

$$\begin{aligned}\Delta_h V(x, y) = \frac{1}{6h^2} \{4[V(x, y+h) + V(x, y-h) + V(x+h, y) + V(x-h, y)]\\ + V(x+h, y+h) + V(x+h, y-h) + V(x-h, y+h)\\ + V(x-h, y-h) - 20V(x, y)\}.\end{aligned} \tag{1.6}$$

For $u \in C^{(7)}(\bar{R})$, the inequality

$$\left|\Delta_h u(P) - \Delta u(P) - \frac{h^2}{12}\Delta\Delta u(P)\right| \leqq \frac{1}{30} M_6 h^4 + O(h^5) \tag{1.7}$$

holds for $P \in R_h$, using the notation

$$M_j = \sup_{P\in R} \{|\partial^j u(P)/\partial x^i \partial y^{j-i}| \mid i = 0, 1, \cdots, j\}. \tag{1.8}$$

A remark may be made on the fact that (1.7) does not hold for $u \in C^{(6)}(\bar{R})$; the truncation error in that case is still of $O(h^4)$, but the upper bound is greater than the one given in (1.7).

We shall also need the inequality

$$\left|\Delta_h u(P) - \Delta u(P) - \frac{h^2}{12}\Delta\Delta u(P)\right| \leqq \tfrac{1}{5} M_5 h^3 \tag{1.9}$$

which holds for $P \in R_h$, if $u \in C^{(5)}(\bar{R})$.

In $C_{2h}{}^*$, we shall use the operator introduced by Shortley and Weller [5] for points like (x, y) in Fig. 1.1.

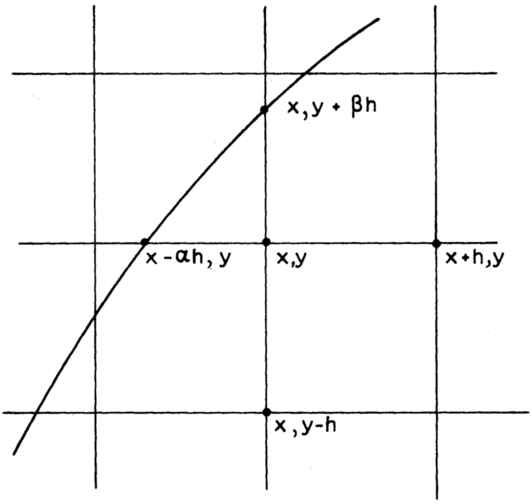

FIGURE 1.1. *Situation of Mesh Points Near the Boundary*

We then approximate Δ by

$$\begin{aligned}\Delta_h V(x, y) = 2h^{-2}\Big\{&\frac{1}{\alpha(1+\alpha)} V(x-\alpha h, y) + \frac{1}{1+\alpha} V(x+h, y)\\ &+ \frac{1}{\beta(1+\beta)} V(x, y+\beta h) + \frac{1}{1+\beta} V(x, y-h)\\ &- \left(\frac{1}{\alpha}+\frac{1}{\beta}\right) V(x, y)\Big\},\end{aligned} \tag{1.10}$$

α and β may equal 1; if $\alpha = \beta = 1$, the operator (1.10) becomes the usual five-point difference analogue for the Laplace operator. Of course, the orientation may be different from the one given in Fig. 1.1. Appropriate changes in (1.10) should then be made.

For $u \in C^{(3)}(\bar{R})$, we have, in a point P of $C_h{}^*$ (or $C_{1h}{}^*$ or $C_{2h}{}^*$),

$$|\Delta_h u(P) - \Delta u(P)| \leqq \tfrac{2}{3} M_3 h. \tag{1.11}$$

These operators, and the ones derived in Sections II and III, shall be used in Sections IV–VI to derive error estimates for the Robin, Neumann and Dirichlet problems, respectively.

We shall approximate the solution of the boundary value problems (1.2), (1.3) and (1.5) by finite-difference methods; that is, we shall solve a set of n simultaneous linear equations in n unknowns. The operators by which the various differential operators are approximated will be chosen in such a way that the coefficient matrix A of the resulting system of linear equations will have a very useful property, both for estimating the discretization error and for actually solving the systems, the matrix being of positive type [6]. Generally, in problems of this type, one has to attend to three things; first one has to establish the convergence of the proposed method, then one has to show that the resulting system of linear equations, which will have a sparse coefficient matrix, can be solved by iterative methods, and, finally, the stability of the method has to be investigated.

The advantage of methods which lead to coefficient matrices which are of positive type is that only one problem is left to deal with. Once one has proven the convergence of the method, in which proof the fact that the matrix is of positive type plays a crucial role, it can be concluded at once from the Stein-Rosenberg theorem [7] that the Jacobi and Gauss-Seidel methods are both convergent, and it can also be seen that the stability is guaranteed.

II. An $O(h^3)$ Finite-Difference Operator for $\partial/\partial n$. Bramble and Hubbard [3] are the first to have given an $O(h^2)$ approximation to the third boundary value problem, using an $O(h^2)$ approximation to $\partial/\partial n$. Before this, a convergence proof had only been given once, for an $O(h)$ approximation, in a paper by Batschelet [8]. We shall now first inspect the operator of Bramble and Hubbard and give a different derivation of their results, which can then be extended to yield an $O(h^3)$ operator.

The basis on which Bramble and Hubbard's proof rests, which also determines the extent to which their results are valid, is the question if and under what circumstances the system

$$\sum_{i=1}^{3} a_i y_i = 1,$$

$$\sum_{i=1}^{3} a_i x_i[1 + y_i(\alpha(P) + K(P))] = 0, \tag{2.1}$$

$$\sum_{i=1}^{3} a_i[x_i^2 - y_i^2] = 0,$$

has a nonnegative solution a_i, under certain assumptions for α and K, and with x_i, y_i satisfying

$$\begin{aligned} 4\epsilon &> x_1 > y_1 + \epsilon > 2\epsilon, \\ 4\epsilon &> -x_2 > y_2 + \epsilon > 2\epsilon, \\ 6\epsilon &\geqq y_3 > |x_3| + 5\epsilon. \end{aligned} \tag{2.2}$$

α is the constant mentioned in (1.5), K is the signed curvature of the boundary in the point under consideration (see [3]), and ϵ is a given positive constant, dependent on h, later chosen to be $3h/2$.

They show, by giving bounds for the determinants connected with a slightly modified system, that the a_i satisfy the inequality

$$a_i > h^{-1}\left[\frac{1 - (84\,|\alpha + K|_M)h}{96 + (756\,|\alpha + K|_M)h}\right] \tag{2.3}$$

where $|\alpha + K|_M = \max_{P \in C_1} |\alpha(P) + K(P)|$, which places a rather heavy restriction on h to make the a_i nonnegative.

We shall now give a new proof for the contention that the a_i satisfying (2.1) are nonnegative, provided h is chosen sufficiently small. It will turn out that we shall have to place hardly any restriction on h, apart from

$$h < 4/51\bar{K} \tag{2.4}$$

which should already have been imposed for other reasons (see [3]); $\bar{K}$ is the maximum positive curvature of C_1.

We call $\alpha(P) + K(P) = q$ and write (2.1) as

$$\begin{bmatrix} y_1 & y_2 & y_3 \\ x_1(1 + qy_1) & x_2(1 + qy_2) & x_3(1 + qy_3) \\ x_1^2 - y_1^2 & x_2^2 - y_2^2 & x_3^2 - y_3^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}. \tag{2.5}$$

From (2.2) it is clear that all $y_i > 0$. This implies that the first equation in (2.1) can always be satisfied with positive a_i by appropriate scaling, without losing the positivity of the a_i. The only remaining condition now is that the vector a (with a_1, a_2 and a_3 as its components) is perpendicular to the plane spanned by the vectors

$$\begin{bmatrix} x_1(1 + qy_1) \\ x_2(1 + qy_2) \\ x_3(1 + qy_3) \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x_1^2 - y_1^2 \\ x_2^2 - y_2^2 \\ x_3^2 - y_3^2 \end{bmatrix}.$$

A nonnegative vector a with this property can always be constructed if there is no vector in the above-mentioned plane with the same sign for all its components.

We may consider, without loss of generality, the vector with components

$$\lambda x_i(1 + qy_i) + (x_i^2 - y_i^2), \qquad i = 1, 2, 3. \tag{2.6}$$

We have to prove that this vector has, for any λ, two components of opposite sign. We shall discuss two cases, depending on the sign of q.

First, we consider the case $q \geqq 0$; this case occurs very often, for instance, for all convex regions R. Let the first two components of (2.6) have the same sign, otherwise our assertion has already been proved. It then follows directly from (2.2) that both are positive. Now suppose that the third component is also positive. This clearly implies $\lambda x_3 > 0$. Without loss of generality, we take both x_3 and $\lambda > 0$. From (2.2) then follows immediately that

$$\lambda(1 + qy_3) > 35\epsilon \tag{2.7}$$

has to be true and therefore also

$$\lambda(1 + qy_i) > 5\epsilon, \qquad i = 1, 2, \tag{2.8}$$

which leads to a contradiction, because under (2.8) the first and second components cannot both be positive. Therefore, the third component is negative, which is what we wanted to prove.

Now consider the case $q < 0$. We shall conduct the proof along similar lines as in the first case. Again, we only have to inspect the case where the first two components have the same sign. Clearly, we have to prevent $(1 + qy_i)$ from becoming zero for $i = 1, 2$, otherwise we will not be able to arrive at a contradiction, because the first and second components will then be positive, irrespective of the value of λ. We therefore take h so small that, with $\epsilon = o(1)$ as $h \to 0$,

$$1 + 3\epsilon q > 0. \tag{2.9}$$

Then, again, the first two components are positive, and we must have

$$|\lambda| \leqq 15\epsilon/4(1 + 3\epsilon q). \tag{2.10}$$

Assuming that the third component is also positive, we arrive at

$$|\lambda| > 35\epsilon \tag{2.11}$$

using the fact that $1 + qy_3 > -1$. The relations (2.10) and (2.11) lead to a contradiction if

$$1 + 3\epsilon q \geqq 3/28, \tag{2.12}$$

which does not violate our earlier condition (2.9). Taking $\epsilon = 3h/2$ as in [3], (2.12) yields

$$h \leqq \frac{25}{126} \cdot \frac{1}{|q|}. \tag{2.13}$$

We have therefore proved our assertion under this condition.

We now have shown that the system (2.1) always has a nonnegative solution, either under the earlier condition (2.4) alone, or under (2.4) and (2.13).

Both conditions say that h may not be too large compared to the radius of curvature, which is quite natural.

We shall now extend the above-mentioned method to construct a positive type $O(h^3)$ approximation to the operator $\partial/\partial n$. Throughout, we shall assume that a square net with mesh width h is placed over the region R.

We consider an arbitrary point 0 on the boundary C_1, where C_1 is sufficiently smooth, and introduce two coordinate systems with origin 0:

(a) a right-handed Cartesian coordinate system, with the x-axis tangent to C_1 at 0, and the positive y-axis along the inward-directed normal at 0,

(b) geodesic normal coordinates with s the arc-length along C_1 and n the outward-directed normal.

This situation is given in Fig. 2.1 with $\phi = 0$ in the point 0. For a sufficiently often differentiable function v, the following relations hold on C_1:

$$\begin{aligned} v_s &= v_x \cos\phi + v_y \sin\phi, \\ v_n &= v_x \sin\phi - v_y \cos\phi. \end{aligned} \tag{2.14}$$

Subscripts denote the indicated partial differentiation.

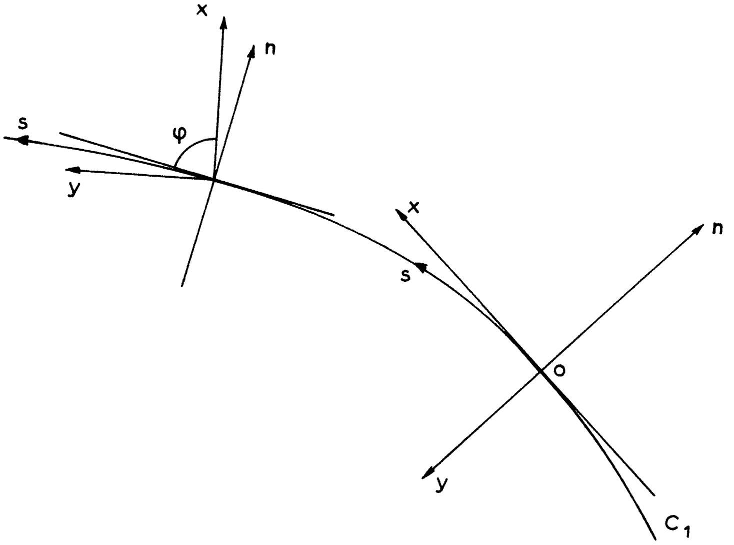

FIGURE 2.1. *Situation of the Coordinate Systems*

Taking $\phi = 0$ and differentiating further, we obtain the following relations between the various partial derivatives in the point 0:

$$\begin{gathered} v_x = v_s, \qquad v_y = -v_n, \qquad v_{xy} = -v_{ns} + Kv_s, \\ -v_{xxy} = (v_{ns} - Kv_s)_s + K(v_{yy} - v_{xx}). \end{gathered} \tag{2.15}$$

By using the same technique as in [3], now using four points instead of three, and taking Taylor expansions including the third-order terms, we obtain the relation

$$
\begin{aligned}
&\sum_{i=1}^{4} a_i\{v(P_i) - [1 + x_i y_i \alpha_s(0) + \tfrac{1}{6}(3x_i^2 y_i - y_i^3)\alpha_{ss}(0)]v(0)\} \\
&= -v_n(0) + \sum_{i=1}^{4} a_i\{[\tfrac{1}{2}y_i^2 - \tfrac{1}{6}(3x_i^2 y_i - y_i^3)K(0)]\Delta v(0) \\
&\qquad - x_i y_i (v_n + \alpha v)_s(0) - \tfrac{1}{6}(3x_i^2 y_i - y_i^3)(v_n + \alpha v)_{ss}(0) \\
&\qquad + \tfrac{1}{2}x_i y_i^2 (\Delta v)_x(0) + \tfrac{1}{6}y_i^3(\Delta v)_x(0)\} \\
&\quad + O(a_i h^4)
\end{aligned}
\tag{2.16}
$$

between the boundary function v and some of its derivatives in the boundary point 0 and the values of v in four interior points P_i.

We want to show that it is possible to choose the P_i so that $a_i \geqq 0$, for $i = 1, \cdots, 4$, because this will be useful in later applications. Apart from the fact that they must be chosen to satisfy this condition, the P_i must satisfy two further requirements: first, that they lie in R and, secondly, that they lie in the neighborhood of the point 0.

Instead of

$$
\begin{aligned}
\sum_{i=1}^{4} a_i\{y_i + \tfrac{1}{6}(3x_i^2 y_i - y_i^3)(\alpha + K)K\} &= 1, \\
\sum_{i=1}^{4} a_i\{x_i + x_i y_i(\alpha + K) + \tfrac{1}{6}(3x_i^2 y_i - y_i^3)(2\alpha_s + K_s)\} &= 0, \\
\sum_{i=1}^{4} a_i\{(x_i^2 - y_i^2) + \tfrac{1}{3}(3x_i^2 y_i - y_i^3)(\alpha + 3K)\} &= 0, \\
\sum_{i=1}^{4} a_i\{x_i^3 - 3x_i y_i^2\} &= 0
\end{aligned}
\tag{2.17}
$$

(where α, K and their derivatives are taken in the point 0) by which relations the a_i are defined, and which are used in [4] to derive (2.16), we consider the approximating system

$$
\begin{aligned}
\sum_{i=1}^{4} \bar{a}_i y_i &= 1, \\
\sum_{i=1}^{4} \bar{a}_i x_i &= 0, \\
\sum_{i=1}^{4} \bar{a}_i (x_i^2 - y_i^2) &= 0, \\
\sum_{i=1}^{4} \bar{a}_i (x_i^3 - 3x_i y_i^2) &= 0.
\end{aligned}
\tag{2.18}
$$

The solution $\bar{a}_i$ of (2.18) will be close to a_i, since the x_i and y_i are small. We write (2.18) in matrix form as

$$
\begin{bmatrix}
y_1 & y_2 & y_3 & y_4 \\
x_1 & x_2 & x_3 & x_4 \\
x_1^2 - y_1^2 & x_2^2 - y_2^2 & x_3^2 - y_3^2 & x_4^2 - y_4^2 \\
x_1^3 - 3x_1 y_1^2 & x_2^3 - 3x_2 y_2^2 & x_3^3 - 3x_3 y_3^2 & x_4^3 - 3x_4 y_4^2
\end{bmatrix}
\cdot
\begin{bmatrix} \bar{a}_1 \\ \bar{a}_2 \\ \bar{a}_3 \\ \bar{a}_4 \end{bmatrix}
=
\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.
\tag{2.19}
$$

We shall now make a definite choice for the $P_i = (x_i, y_i)$, so that

$$\begin{aligned} 3\epsilon &\geqq x_1 \geqq y_1 + \epsilon \geqq 2\epsilon, \\ 3\epsilon &\geqq -x_2 \geqq y_2 + \epsilon \geqq 2\epsilon, \\ 3\epsilon &\geqq y_3 \geqq x_3 + \epsilon \geqq 2\epsilon, \\ 3\epsilon &\geqq y_4 \geqq -x_4 + \epsilon \geqq 2\epsilon. \end{aligned} \tag{2.20}$$

Geometrically, this means that in Fig. 2.2, $P_1 \in$ I, $P_2 \in$ II, $P_3 \in$ III and $P_4 \in$ IV.

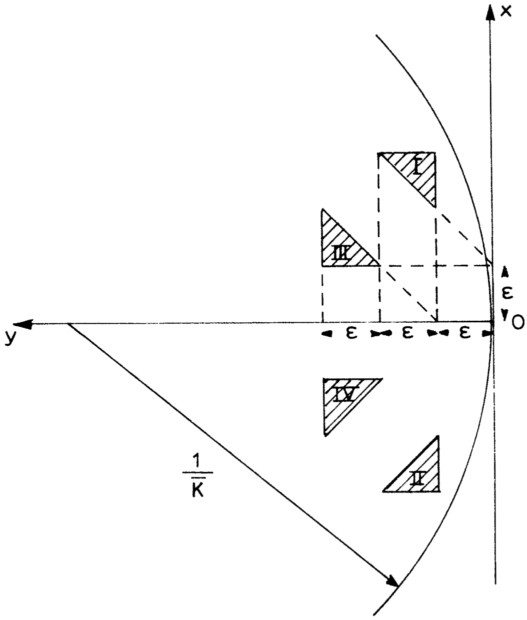

FIGURE 2.2. *Position of the Points P_i*

It is shown in [4], using the same method of proof as was used above, that (2.18) always yields a positive set of $\bar{a}_i$ when the P_i satisfy (2.20), which means that a nonnegative set of a_i can be found for (2.17).

It is possible to find an upper bound for h, below which (2.17) has such a nonnegative solution. We shall however refrain from doing so. The upper bound we were able to find is so small as to be of little practical value; while, on the other hand, it is clear that it is certainly not optimal. For practical purposes, the absence of a definite bound on h poses no problems, because it is always discovered, if h is taken too large, by the occurrence of negative coefficients. In the computation of the examples given in [4], we never met with difficulties, taking any h satisfying the other conditions we imposed upon it.

We now have for h sufficiently small $a_i \geqq 0, \quad i = 1, \cdots, 4$, and also

$$a_i < Mh^{-1}, \qquad i = 1, \cdots, 4, \tag{2.21}$$

with M a fixed positive constant. The last of these two inequalities follows from (2.17) and (2.20).

We now define an operator δ_n for points P on C_1 as follows:

$$(2.22) \qquad \delta_n V(P) = \sum_{i=1}^{4} a_i\{[1 + x_i y_i \alpha_s(P) + \tfrac{1}{6}(3x_i^2 y_i - y_i^3)\alpha_{ss}(P)]\, V(P) - V(P_i)\},$$

where the a_i are defined by (2.17) and the P_i satisfy (2.20).

Using (2.16), it is now clear that the function u in (1.5) satisfies

$$(2.23) \quad \Bigg| \delta_n u(P) + \alpha(P)u(P) - \Bigg\{ g(P) + \sum_{i=1}^{4} a_i[(\tfrac{1}{2}y_i^2 - \tfrac{1}{6}(3x_i^2 y_i - y_i^3)K(P))f(P) + x_i y_i g_s(P) + \tfrac{1}{6}(3x_i^2 y_i - y_i^3) g_{ss}(P) + \tfrac{1}{2}x_i y_i^2 f_s(P) - \tfrac{1}{6}y_i^3 f_n(P)] \Bigg\} \Bigg| \leqq k_1 h^3$$

with k_1 a positive constant.

We have therefore found in (2.22) an $O(h^3)$ approximation δ_n to the operator $\partial/\partial n$, which we shall use in later sections to derive some error estimates for the various boundary value problems.

III. An $O(h^2)$ Positive Type Finite-Difference Laplace Operator on the Set C_h*. In section I, we mentioned that, in the set C_h*, we need an approximation Δ_h to the Laplace operator Δ with a truncation error of $O(h^2)$. C_h* consists of the net points in R whose eight nearest neighbors are not all in R, roughly speaking, the net points near but not on the boundary. Bramble and Hubbard gave such an operator in [1]. We shall, however, not use it because its application results in a coefficient matrix which is not of positive type. To ensure that the resulting matrix has this useful property, it is necessary that in the formula

$$(3.1) \qquad \Delta_h V(Q) = \sum_i \lambda_i\{V(Q_i) - V(Q)\},$$

which may be considered as the general form of the approximation we have in mind, all λ_i are positive. This is not the case in the above-mentioned approximation in [1]. We have shown in [4] that derivation of an approximation with positive coefficients is possible and that, moreover, the use of this approximation for the Dirichlet problem results in an upper bound for the discretization error which is never larger than that of Bramble and Hubbard.

In Fig. 3.1(a)–(e) we give the five fundamentally different configurations we shall distinguish. Throughout, we assume the shaded region to be in R. We made no restriction to convex regions, although this might seem to be the case from the figures. The only assumption at this stage is that all α_i in the above configurations satisfy

$$(3.2) \qquad 0 < \alpha_i \leqq 1.$$

In [4], we formulated further restrictions on the α_i which are consistent with the restrictions imposed upon h in Section II, in the sense that they are always fulfilled.

Since we stipulated that h is not too large compared to the radius of curvature, we may exclude occurrence of a situation as in Fig. 3.2(a) while the configurations given in Fig. 3.2(b)–(d) may be considered as special cases of the one given in Fig. 3.1(b). In Fig. 3.2(c), (d) the quantity α_1 from Fig. 3.1(b) should be taken equal tȯ 1.

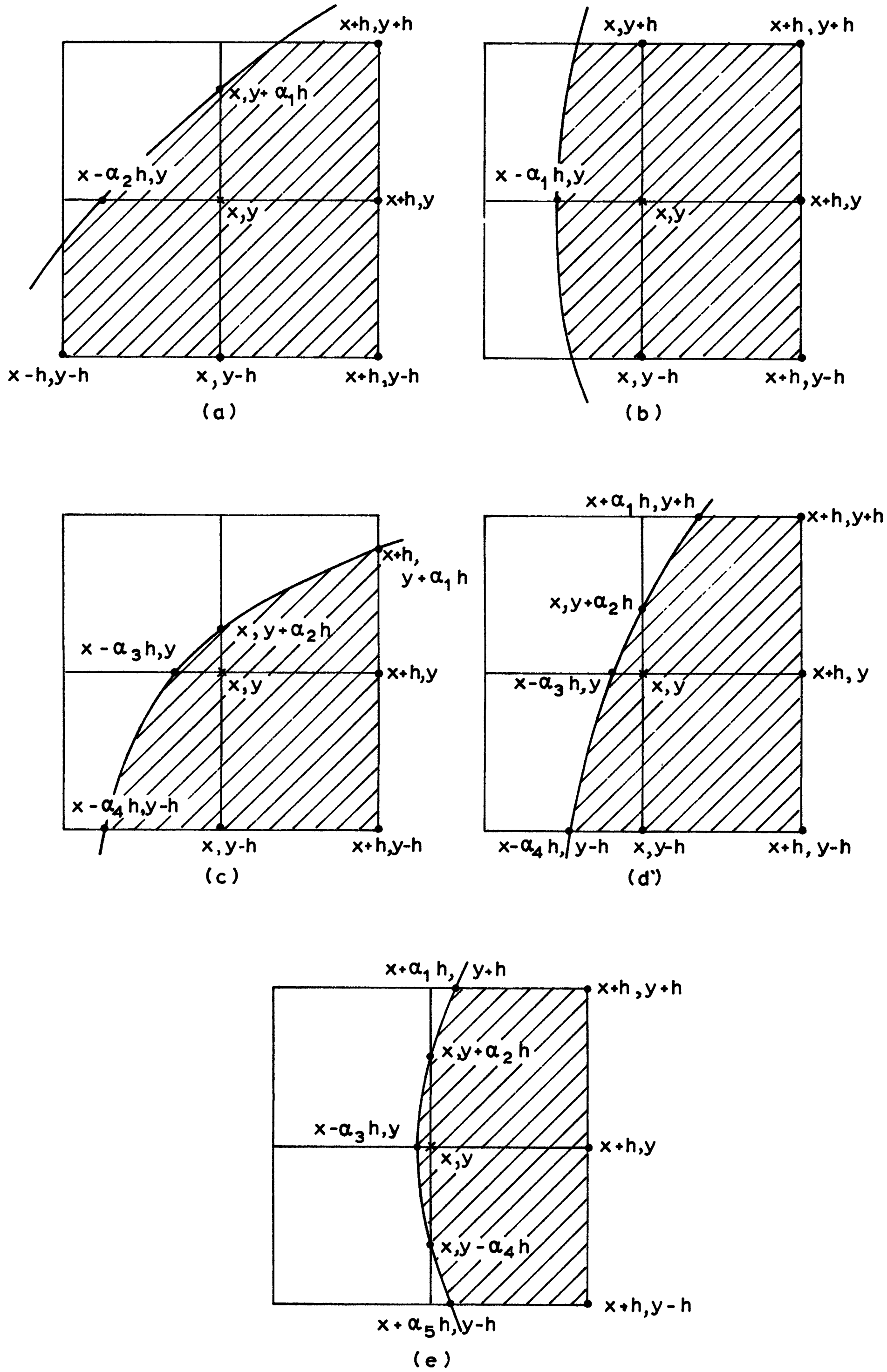

FIGURE 3.1. *Configurations Under Consideration*

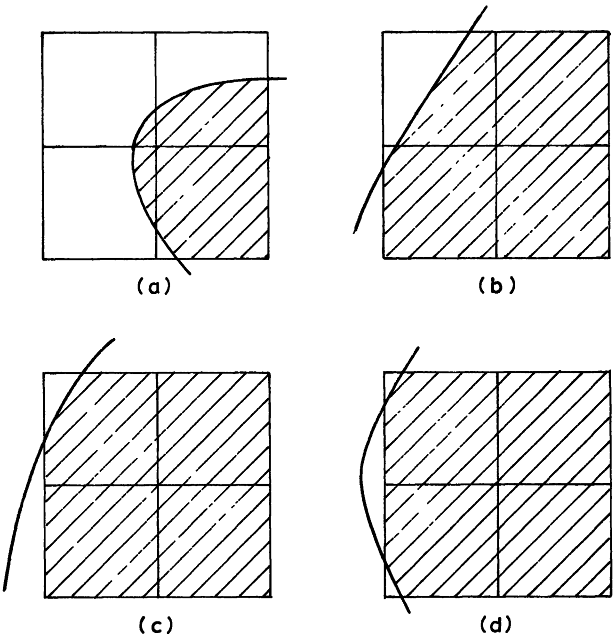

FIGURE 3.2. *Special Cases of Fig.* 3.1(b)

Apart from this, Fig. 3.1 gives all essentially different possibilities after appropriate rotation and reflection. For all five cases, we shall give a formula of type (3.1) with positive λ_i. We shall see that, to attain this goal, a further subdivision of these five cases will be necessary. For each subcase, we shall give an inequality of the type

$$|\Delta u(P) + F(\Delta_x, \Delta_y)u(P) - \Delta_h u(P)| \leqq cM_4h^2 + O(h^3) \tag{3.3}$$

with c a constant, the value of which shall be explicitly given, and M_4 defined by (1.8). F denotes an operator of the type $h\{\alpha\Delta_x + \beta\Delta_y\}$. We here assume $\Delta_x u = f_x$ and $\Delta_y u = f_y$ to be known in R.

The derivation of the formulae (3.1) has as its underlying principle that a Taylor series expansion is made for each of the $V(Q_i)$ relative to the point Q. These are then multiplied by appropriate constants λ_i and summed. The λ_i must be so chosen that after this summation all third and lower derivatives vanish or can be expressed in terms of the Laplace operator and its derivatives. This gives a set of simultaneous linear equations for the λ_i. The freedom that most of the thus obtained systems still leave in the choice of the λ_i is used to make them positive and keep them as simple as possible. The formulae (3.3) follow after some computation from (3.1) by summing over the various fourth derivatives, again multiplied by the λ_i, in the above-mentioned Taylor expansions.

Case I. The point configuration for this case is given in Fig. 3.1(a). We define Δ_h as

$$\begin{aligned}\Delta_h V(x, y) = \mu h^{-2}\{&\lambda_1 V(x+h, y) + \lambda_2 V(x, y-h) + \lambda_3 V(x+h, y+h)\\ &+ \lambda_4 V(x-h, y-h) + \lambda_5 V(x+h, y-h)\\ &+ \lambda_6 V(x, y+\alpha_1 h) + \lambda_7 V(x-\alpha_2 h, y) - \lambda_8 V(x, y)\}\end{aligned} \tag{3.4}$$

with

$$\lambda_1 = \frac{2(2+\alpha_1)(2+\alpha_2^2)}{(1+\alpha_2)}, \qquad \lambda_2 = \frac{2(2+\alpha_2)(2+\alpha_1^2)}{(1+\alpha_1)},$$

$$\lambda_3 = (2+\alpha_1)(1-\alpha_2), \qquad \lambda_4 = (2+\alpha_2)(1-\alpha_1),$$

$$\lambda_5 = 4 - \alpha_1 - \alpha_2 - 2\alpha_1\alpha_2,$$

$$\lambda_6 = \frac{6(2+\alpha_2)}{\alpha_1(1+\alpha_1)}, \qquad \lambda_7 = \frac{6(2+\alpha_1)}{\alpha_2(1+\alpha_2)},$$

$$\lambda_8 = \frac{6(\alpha_1^2+\alpha_2^2+2\alpha_1+2\alpha_2)}{\alpha_1\alpha_2}, \qquad \mu = \frac{1}{8+\alpha_1+\alpha_2-\alpha_1\alpha_2}.$$

For $\alpha_1 = \alpha_2 = 1$, Δ_h becomes the ordinary five-point Laplace operator. For $u \in C^{(5)}(\bar{R})$, we have

$$|\Delta u(P) - h\mu\lambda_3\Delta_x u(P) + h\mu\lambda_4\Delta_y u(P) - \Delta_h u(P)| \leqq \tfrac{5}{12} M_4 h^2 + O(h^3). \tag{3.5}$$

Case II. The point configuration for this case is given in Fig. 3.1(b). We define Δ_h as

$$\begin{aligned} \Delta_h V(x, y) = h^{-2}\{&\lambda_1 V(x+h, y) + \lambda_2 V(x, y-h) + \lambda_2 V(x, y+h) \\ &+ \lambda_3 V(x+h, y+h) + \lambda_3 V(x+h, y-h) \\ &+ \lambda_4 V(x-\alpha_1 h, y) - \lambda_5 V(x, y)\} \end{aligned} \tag{3.6}$$

with

$$\lambda_1 = \frac{2(2+\alpha_1^2)}{3(1+\alpha_1)}, \qquad \lambda_2 = \frac{2+\alpha_1}{3}, \qquad \lambda_3 = \frac{1-\alpha_1}{3},$$

$$\lambda_4 = \frac{2}{\alpha_1(1+\alpha_1)}, \qquad \lambda_5 = \frac{2(3+2\alpha_1+\alpha_1^2)}{3\alpha_1}.$$

For $u \in C^{(5)}(\bar{R})$, we have

$$|\Delta u(P) - h\lambda_3\Delta_x u(P) - \Delta_h u(P)| \leqq \tfrac{1}{3}M_4 h^2 + O(h^3) \tag{3.7a}$$

or, if $\alpha_1 = 1$, and we have the ordinary five-point operator

$$|\Delta u(P) - \Delta_h u(P)| \leqq \tfrac{1}{6}M_4 h^2 + O(h^3). \tag{3.7b}$$

Case III. The point configuration for this case is given in Fig. 3.1(c). We define Δ_h as

$$\begin{aligned} \Delta_h V(x, y) = \mu h^{-2}\{&\lambda_5 V(x, y-h) + \lambda_6 V(x+h, y) + \lambda_7 V(x+h, y-h) \\ &+ \lambda_1 V(x+h, y+\alpha_1 h) + \lambda_2 V(x, y+\alpha_2 h) \\ &+ \lambda_3 V(x-\alpha_3 h, y) + \lambda_4 V(x-\alpha_4 h, y-h) - \lambda_8 V(x, y)\} \end{aligned} \tag{3.8}$$

with

$$\lambda_1 = \frac{1}{\alpha_1(1+\alpha_1)}\{3 - 2\alpha_3 - \alpha_4 - \alpha_2\alpha_3 + \alpha_2\alpha_4\},$$

$$\lambda_2 = \frac{1}{\alpha_2(1+\alpha_2)}\{6 + 2\alpha_3 + \alpha_4 + \alpha_1\alpha_3 - \alpha_1\alpha_4\},$$

$$\lambda_3 = \frac{1}{\alpha_3(1 + \alpha_3)} \{6 + 2\alpha_2 + \alpha_1 + \alpha_2\alpha_4 - \alpha_1\alpha_4\},$$

$$\lambda_4 = \frac{1}{\alpha_4(1 + \alpha_4)} \{3 - 2\alpha_2 - \alpha_1 - \alpha_2\alpha_3 + \alpha_1\alpha_3\},$$

$$\lambda_5 = \alpha_2\lambda_2 - (1 + \alpha_4)\lambda_4, \qquad \lambda_6 = \alpha_3\lambda_3 - (1 + \alpha_1)\lambda_1,$$

$$\lambda_7 = \alpha_1\lambda_1 + \alpha_4\lambda_4, \qquad \lambda_8 = (1 + \alpha_2)\lambda_2 + (1 + \alpha_3)\lambda_3,$$

$$\mu = \frac{2}{9 + (\alpha_1 - \alpha_2)(\alpha_3 - \alpha_4)}.$$

It is shown in [4] that, as in the preceding and following cases, all λ_i and μ are nonnegative, which is not always obvious, under no further restrictions than the ones imposed upon h in Section II. For $u \in C^{(5)}(\bar{R})$, we have

$$\left|\Delta u(P) - h\mu \frac{\alpha_1(1 + \alpha_1)}{2} \lambda_1\Delta_x u(P) + h\mu \frac{\alpha_4(1 + \alpha_4)}{2} \lambda_4\Delta_y u(P) - \Delta_h u(P)\right| \tag{3.9}$$
$$< \tfrac{75}{128} M_4h^2 + O(h^3).$$

Case IV. The point configuration for this case is given in Fig. 3.1(d). We first formulate two further conditions:

$$1 - \alpha_2^2 - 3\alpha_4 \leqq 0 \tag{3.10}$$

and

$$3\alpha_1(1 - \alpha_1)(\alpha_2 - \alpha_3) + (1 - \alpha_2)(1 + \alpha_1)(2 - 2\alpha_3 - 5\alpha_1 + \alpha_1\alpha_3 + \alpha_1^2) \geqq 0. \tag{3.11}$$

The origin of these complicated conditions can be found in [4]. We now subdivide this case:

(a) (3.10) is valid.
(b) (3.11) is valid and (3.10) is not.
(c) Neither (3.10) nor (3.11) is fulfilled.
(a) We define Δ_h as

$$\begin{aligned}\Delta_h V(x, y) = \mu h^{-2}\{&\lambda_5 V(x + h, y) + \lambda_6 V(x, y - h) + \lambda_7 V(x + h, y - h)\\ &+ \lambda_0 V(x + h, y + h) + \lambda_2 V(x, y + \alpha_2 h)\\ &+ \lambda_3 V(x - \alpha_3 h, y) + \lambda_4 V(x - \alpha_4 h, y - h) - \lambda_8 V(x, y)\}\end{aligned} \tag{3.12}$$

with

$$\lambda_0 = 3 - 2\alpha_3 - \alpha_4 - \alpha_2\alpha_3 + \alpha_2\alpha_4, \qquad \lambda_2 = \frac{6(2 + \alpha_3)}{\alpha_2(1 + \alpha_2)},$$

$$\lambda_3 = \frac{-2(2 + \alpha_4)(1 - \alpha_2) + 18}{\alpha_3(1 + \alpha_3)}, \qquad \lambda_4 = \frac{2(1 - \alpha_2)(2 + \alpha_3)}{\alpha_4(1 + \alpha_4)},$$

$$\lambda_5 = \alpha_3\lambda_3 - 2\lambda_0, \qquad \lambda_6 = \alpha_2\lambda_2 - (1 + \alpha_4)\lambda_4,$$

$$\lambda_7 = \lambda_0 + \alpha_4\lambda_4, \qquad \lambda_8 = (1 + \alpha_2)\lambda_2 + (1 + \alpha_3)\lambda_3,$$

$$\mu = \frac{1}{9 + (1 - \alpha_2)(\alpha_3 - \alpha_4)}.$$

For $u \in C^{(5)}(\bar{R})$, we have

$$|\Delta u(P) - h\mu\lambda_0\Delta_x u(P) + h\mu(1-\alpha_2)(2+\alpha_3)\Delta_y u(P) - \Delta_h u(P)| < \frac{113}{216}M_4h^2 + O(h^3)\cdot \tag{3.13}$$

(b) We define Δ_h as

$$\begin{aligned}\Delta_h V(x,y) = \mu h^{-2}\{&\lambda_5 V(x+h,y) + \lambda_6 V(x,y-h) + \lambda_7 V(x+h,y-h)\\ &+ \lambda_0 V(x+h,y+h) + \lambda_1 V(x+\alpha_1 h, y+h)\\ &+ \lambda_2(x, y+\alpha_2 h) + \lambda_3 V(x-\alpha_3 h, y) - \lambda_8 V(x,y)\}\end{aligned} \tag{3.14}$$

with

$$\begin{aligned}\lambda_0 &= \alpha_3(1+\alpha_3)\alpha_2(1+\alpha_2)\\ &\quad\cdot\{3\alpha_1(1-\alpha_1)(\alpha_2-\alpha_3) + (1-\alpha_2)(1+\alpha_1)(2-2\alpha_3-5\alpha_1+\alpha_1\alpha_3+\alpha_1^2)\},\\ \lambda_1 &= 2\alpha_2(1-\alpha_2)\alpha_3(1+\alpha_3)(2+\alpha_3)(1+\alpha_2),\\ \lambda_2 &= 6\alpha_1(1-\alpha_1)\alpha_3(1+\alpha_3)(2+\alpha_3),\\ \lambda_3 &= 2\alpha_2(1+\alpha_2)(1-\alpha_1)\{(6-\alpha_1)(1+\alpha_1)(1-\alpha_2) + 3\alpha_1(2+\alpha_2)\},\\ \lambda_5 &= \alpha_3\lambda_3 - 2\lambda_0 - 2\alpha_1\lambda_1,\\ \lambda_6 &= (1-\alpha_1)\lambda_1 + \alpha_2\lambda_2,\\ \lambda_7 &= \lambda_0 + \alpha_1\lambda_1,\\ \lambda_8 &= 2(1-\alpha_1)\lambda_1 + (1+\alpha_2)\lambda_2 + (1+\alpha_3)\lambda_3,\\ \mu &= \frac{2}{2\lambda_0 + 2\lambda_1 + \alpha_2(1+\alpha_2)\lambda_2}.\end{aligned}$$

For $u \in C^{(5)}(\bar{R})$, we have

$$|\Delta u(P) - h\mu\lambda_7\Delta_x u(P) + \tfrac{1}{2}h\mu\alpha_1(1-\alpha_1)\lambda_1\Delta_y u(P) - \Delta_h u(P)| < \tfrac{2}{3}M_4h^2 + O(h^3). \tag{3.15}$$

(c) It is not certain whether this case can actually occur under the restriction we have already made upon h. Anyway, it is unlikely that we shall meet it in practical problems. For the sake of completeness, we shall show that a satisfying definition for Δ_h can be given in this case also. We shall call the operator of Case IV(a) $\Delta_h^{(a)}$ and that of Case IV(b) $\Delta_h^{(b)}$. We now define Δ_h as

$$\Delta_h V(P) = k_a\Delta_h^{(a)} V(P) + k_b\Delta_h^{(b)} V(P). \tag{3.16}$$

k_a and k_b are two constants, satisfying $k_a + k_b = 1$ and further so chosen that the coefficient λ_6 of $V(x, y-h)$ in (3.16) is zero; k_a and k_b are thus both positive. The coefficients λ_i in (3.16) are

$$\lambda_i = k_a\lambda_i^{(a)} + k_b\lambda_i^{(b)}, \qquad i = 0, \cdots, 7,$$

using an obvious notation. Clearly, a formula of type (3.3) can also be given, which is a linear combination of (3.13) and (3.15).

Case V. The point configuration for this case is given in Fig. 3.1(e). Its occurrence in practical computation can almost certainly be prevented by a suitable choice of h; for completeness sake it is included.

We define Δ_h as

$$\begin{aligned}\Delta_h V(x, y) = \mu h^{-2}\{&\lambda_6 V(x + h, y) + \lambda_7 V(x + h, y - h) + \lambda_8 V(x + h, y + h)\\ &+ \lambda_1 V(x+\alpha_1 h, y+h) + \lambda_2 V(x, y+\alpha_2 h) + \lambda_3 V(x-\alpha_3 h, y)\\ &+ \lambda_4 V(x, y - \alpha_4 h) + \lambda_5 V(x + \alpha_5 h, y - h) - \lambda_9 V(x, y)\}\end{aligned} \tag{3.17}$$

with

$$\lambda_1 = \frac{1 + 3\alpha_5 - \alpha_2^2}{1 - \alpha_1}, \qquad \lambda_2 = \frac{1 + 3\alpha_1 - \alpha_4^2}{\alpha_2},$$

$$\lambda_4 = \frac{1 + 3\alpha_5 - \alpha_2^2}{\alpha_4}, \qquad \lambda_5 = \frac{1 + 3\alpha_1 - \alpha_4^2}{\alpha_5},$$

$$\lambda_3 = \frac{(1 + 3\alpha_5 - \alpha_2^2)(3 + 2\alpha_1 + 3\alpha_4 - \alpha_1^2) + (1 + 3\alpha_1 - \alpha_4^2)(3 + 2\alpha_5 + 3\alpha_2 - \alpha_5^2)}{\alpha_3(1 + \alpha_3)(2 + \alpha_3)},$$

$$\lambda_7 = \tfrac{1}{2}\{\alpha_3(1 + \alpha_3)\lambda_3 - (1 - \alpha_1^2)\lambda_1 - \alpha_2^2\lambda_2 - \alpha_4^2\lambda_4 - (1 + 2\alpha_5 - \alpha_5^2)\lambda_5\},$$

$$\lambda_8 = \tfrac{1}{2}\{\alpha_3(1 + \alpha_3)\lambda_3 - (1 + 2\alpha_1 - \alpha_1^2)\lambda_1 - \alpha_2^2\lambda_2 - \alpha_4^2\lambda_4 - (1 - \alpha_5^2)\lambda_5\},$$

$$\lambda_6 = \alpha_3\lambda_3 - \alpha_1\lambda_1 - \alpha_5\lambda_5 - \lambda_7 - \lambda_8,$$

$$\lambda_9 = (1 - \alpha_1)\lambda_1 + \lambda_2 + (1 + \alpha_3)\lambda_3 + \lambda_4 + (1 - \alpha_5)\lambda_5,$$

$$\mu = \frac{2}{\alpha_3(1 + \alpha_3)\lambda_3 - \alpha_1(1 - \alpha_1)\lambda_1 - \alpha_5(1 - \alpha_5)\lambda_5}.$$

For $u \in C^{(5)}(\bar{R})$, we have

$$\begin{aligned}|\Delta u(P) &+ \tfrac{1}{2}h\mu(\lambda_7 + \lambda_8 + \alpha_1\lambda_1 + \alpha_5\lambda_5)\Delta_x u(P)\\ &+ \tfrac{1}{2}h\mu(\alpha_5(1 - \alpha_5) - \alpha_1(1 - \alpha_1))\Delta_y u(P) - \Delta_h u(P)|\\ &\leqq \tfrac{17}{24} M_4 h^2 + O(h^3).\end{aligned} \tag{3.18}$$

Now that we have given all the necessary formulae, a few general remarks about them must be made. First, it must be pointed out that the inequalities of type (3.3) which were given above do not, for the most part, contain best possible constants. We contented ourselves with relatively easily obtainable bounds which suffice for the later use we have in mind.

Secondly, there is the problem that most of the difference operators given above look rather complicated. It must therefore be emphasized that, in practical computation, the frequency of the use that is made of the various formulae is almost directly proportional to their simplicity. Let us take the unit circle as an example for R considering, for reasons of symmetry, only a quarter of its boundary. For $h = 1/10$, C_h^* then contains fourteen points: eleven of type II, two of type IV(a) and one of type I. For $h = 1/20$, these numbers are twenty-seven of type II, six of type IV(a) and three of type I. The intricate formulae of type III, IV(b), (c) and V are not used at all; the relatively simple type II occurs in three-quarters of the total number of cases.

Emphasis must also be put on the fact that, for our purposes, any approximation to Δ giving the desired truncation error will do as long as the resulting matrix is of positive type. The approximating operators given in this section and in Section II only serve as an illustration of the fact that such approximations can be given under quite general circumstances. In special cases, much easier methods leading to the desired results may be found.

IV. The Third Boundary Value Problem. We shall now approximate the third boundary value problem (1.5) by a finite-difference analogue, using the operators given in Sections II and III. Consider the approximation

$$\begin{aligned} -\Delta_h U(P) &= f^*(P), & P &\in R_h + C_{1h}^* + C_{2h}^*, \\ \delta_n U(P) + \alpha(P) U(P) &= g^*(P), & P &\in C_{1h}, \\ U(P) &= g_1(P), & P &\in C_{2h}, \end{aligned} \tag{4.1}$$

with Δ_h defined by (1.6) for $P \in R_h$, by (1.10) for $P \in C_{2h}{}^*$ and by the appropriate operator defined in Section III for $P \in C_{1h}{}^*$; δ_n is defined by (2.22). The sets R_h, C_{ih} and $C_{ih}{}^*$, $i = 1, 2$, are as in Section I. We already mentioned the division of $C_h{}^*$ into two sets $C_{1h}{}^*$ and $C_{2h}{}^*$, but have not yet discussed it in detail. Points lying in $C_h{}^*$ have a part of the boundary lying inside their nine-point molecule. A point $P \in C_h{}^*$ will be in $C_{ih}{}^*$, if this part of the boundary entirely belongs to C_i, for $i = 1, 2$. If the part of the boundary lying inside the nine-point molecule does not belong exclusively to C_1 or C_2, the corresponding point P is in $C_{2h}{}^*$ if C_2 cuts a main axis or if C_2 cuts the boundary of the molecule, while the main axes are entirely in R, and otherwise in $C_{1h}{}^*$.

The functions $f^*(P)$ and $g^*(P)$ in (4.1) are defined as

$$\begin{aligned} f^*(P) &= f(P) + \frac{h^2}{12}\Delta f(P), & P &\in R_h, \\ f^*(P) &= f(P) - hF(P), & P &\in C_{1h}^*, \\ f^*(P) &= f(P), & P &\in C_{2h}^*, \\ g^*(P) &= g(P) + F_1(P), & P &\in C_{1h}, \end{aligned} \tag{4.2}$$

with $hF(P) = F(\Delta_x, \Delta_y)u(P)$ (see (3.3)) and $F_1(P)$ a known function of f, g and their derivatives defined by considering

$$|\delta_n u(P) + \alpha(P)u(P) - g(P) - F_1(P)| \leqq k_1 h^3 \tag{4.3}$$

as an equivalent notation for (2.23).

The matrix of the system (4.1) is of positive type provided

$$\sum_{i=1}^{4} a_i[x_i y_i \alpha_s(P) + \tfrac{1}{6}(3x_i^2 y_i - y_i^3)\alpha_{ss}(P)] + \alpha(P) \geqq 0 \tag{4.4}$$

is true for all $P \in C_{1h}$. Since we stipulated that α is bounded away from zero, (4.4) can always be satisfied for h chosen sufficiently small. This matrix then possesses a nonnegative inverse. The general idea behind the following proof concerning the magnitude of the discretization error has been taken from Bramble and Hubbard

[3]; since we necessarily work with a different discrete Green's function, the entire detailed proof has to be given again for this case. As we have shown in [4], the use of a different and more complicated Green's function may severely complicate the proof of corresponding theorems. We now introduce the discrete Green's function $G_h(P, Q)$ for the region R under consideration, defined by

$$\begin{aligned} -\Delta_h G_h(P, Q) &= h^{-2}\delta(P, Q), & P &\in R_h + C_{1h}^* + C_{2h}^*, \\ \delta_n G_h(P, Q) + \alpha(Q) G_h(P, Q) &= h^{-1}\delta(P, Q), & P &\in C_{1h}, \\ G_h(P, Q) &= h^{-1}\delta(P, Q), & P &\in C_{2h}, \end{aligned} \tag{4.5}$$

with $Q \in R_h + C_{1h}{}^* + C_{2h}{}^* + C_{1h} + C_{2h}$. The symbol $\delta(P, Q)$ denotes the Kronecker delta. Here, and in the following sections, we assume the operators Δ_h and δ_n to be working on the first parameter. Clearly, $G_h(P, Q)$ is nonnegative, being the inverse of the coefficient matrix of (4.1), multiplied by a nonnegative diagonal matrix.

We now have, for any mesh-function V,

$$\begin{aligned} V(P) = h^2 &\sum_{Q \in R_h + C_{1h}{}^* + C_{2h}{}^*} G_h(P, Q)[-\Delta_h V(Q)] \\ &+ h \sum_{Q \in C_{1h}} G_h(P, Q)[\delta_n V(Q) + \alpha(Q) V(Q)] + h \sum_{Q \in C_{2h}} G_h(P, Q) V(Q) \end{aligned} \tag{4.6}$$

which follows from the fact that the coefficient matrix of (4.1) is nonsingular.

It must be pointed out that we used a definition slightly different from the one used by Bramble and Hubbard [3]. This difference consists of the inclusion of the factors h^{-1} in the second and third lines of (4.5). The reason for this is, that (4.6) is now more in agreement with the continuous representation of the solution of (1.5) by means of kernel functions

$$u(P) = \iint\limits_R G_1(P, Q) f(Q)\, d\sigma + \int_{C_1} G_2(P, Q) g(Q)\, ds + \int_{C_2} G_3(P, Q) g_1(Q)\, ds.$$

We first take $V(P) = 1$ in (4.6), which yields

$$h \sum_{Q \in C_{2h}} G_h(P, Q) \leqq 1. \tag{4.7}$$

We now suppose that a function $\phi \in C^{(3)}(\bar{R})$ exists satisfying

$$\begin{aligned} -\Delta\phi &\geqq 1 \quad \text{in } R, \\ \partial\phi/\partial n + \alpha\phi &\geqq 1 \quad \text{on } C_1. \end{aligned} \tag{4.8}$$

Then, for sufficiently small h,

$$\begin{aligned} -\Delta_h\phi(P) &\geqq \tfrac{1}{2}, & P &\in R_h + C_{1h}^* + C_{2h}^*, \\ \delta_n\phi(P) + \alpha(P)\phi(P) &\geqq \tfrac{1}{2}, & P &\in C_{1h}. \end{aligned}$$

If we now take $V(P) = \phi(P)$ in (4.6), we obtain

$$h^2 \sum_{Q \in R_h + C_{1h}{}^* + C_{2h}{}^*} G_h(P, Q) + h \sum_{Q \in C_{1h}} G_h(P, Q) \leqq 4\, |\phi|_M \tag{4.9}$$

with $|\phi|_M = \max_{P \in R_h + C_{1h}{}^* + C_{2h}{}^* + C_{1h} + C_{2h}} |\phi(P)|$.

We now introduce the sets $C_{1h}{}^{**}$ and $C_{2h}{}^{**}$, the subsets of $C_{1h}{}^*$ and $C_{2h}{}^*$ where

Δ_h is represented by the ordinary five-point formula. We now define a function $W(P)$ by $W(P) = 0$ on C_h, $W(P) = 1$ in R_h, in $C_{1h}{}^{**}$ and $C_{2h}{}^{**}$, and in those points of $(C_{1h}{}^{*} \cup C_{2h}{}^{*}) - (C_{1h}{}^{**} \cup C_{2h}{}^{**})$ which do not belong to a star, the centre of which is in $C_{1h}{}^{**} \cup C_{2h}{}^{**}$. In the points of $(C_{1h}{}^{*} \cup C_{2h}{}^{*}) - (C_{1h}{}^{**} \cup C_{2h}{}^{**})$ which belong to a star, the centre of which is in $C_{1h}{}^{**} \cup C_{2h}{}^{**}$, $W(P) = 7/8$. We can then show

$$-\Delta_h W(P) \geqq \tfrac{1}{4}h^{-2}, \qquad P \in C_{1h}^* + C_{2h}^*,$$
$$-\Delta_h W(P) \geqq 0, \qquad P \in R_h.$$

Taking $V(P) = W(P)$ in (4.6), we have

$$1 \geqq h^2 \sum_{Q \in C_{1h}{}^* + C_{2h}{}^*} G_h(P, Q)[-\Delta_h W(Q)] + h \sum_{Q \in C_{1h}} G_h(P, Q)[\delta_n W(Q) + \alpha(Q) W(Q)]$$

or

$$\text{(4.10)} \qquad \sum_{Q \in C_{1h}{}^* + C_{2h}{}^*} G_h(P, Q) \leqq 4 + 4 \max_{\bar{Q} \in C_{1h}} \left[\sum_{i=1}^{4} a_i(\bar{Q})\right] h \sum_{Q \in C_{1h}} G_h(P, Q).$$

We also have (see (2.17))

$$\begin{aligned} 1 &= \sum_{i=1}^{4} a_i\{y_i + \tfrac{1}{6}(3x_i^2 y_i - y_i^3)(\alpha + K)K\} \\ \text{(4.11)} \qquad &\geqq \left[\sum_{i=1}^{4} a_i\right] \min\{y_i + \tfrac{1}{6}(3x_i^2 y_i - y_i^3)(\alpha + K)K\} \\ &\geqq \left[\sum_{i=1}^{4} a_i\right]\left\{\frac{3h}{\sqrt{2}} + O(h^3)\right\} \geqq \left[\sum_{i=1}^{4} a_i\right]\frac{3h}{2} \end{aligned}$$

for sufficiently small h, for any $P \in C_{1h}$. We now have $\sum_{i=1}^{4} a_i \leqq \frac{2}{3}h^{-1}$ and this yields, with (4.9) and (4.10),

$$\text{(4.12)} \qquad h \sum_{Q \in C_{1h}{}^* + C_{2h}{}^*} G_h(P, Q) \leqq 4(h + \tfrac{8}{3}\, |\phi|_M)\cdot$$

We shall derive a sharper bound for $Q \in C_{2h}{}^*$. Take $W(P) = 0$ on C_h, $W(P) = 1$ everywhere in $R_h + C_{1h}{}^* + C_{2h}{}^{**}$ and in those points of $C_{2h}{}^* - C_{2h}{}^{**}$ which do not belong to a five-point star, the centre of which is in $C_{2h}{}^{**}$. In the points of $C_{2h}{}^* - C_{2h}{}^{**}$ which belong to a five-point star, the centre of which is in $C_{2h}{}^{**}$, $W(P) = 7/8$. We then have

$$-\Delta_h W(P) \geqq \tfrac{1}{4}h^{-2}, \qquad P \in C_{2h}^*,$$
$$-\Delta_h W(P) \geqq 0, \qquad P \in R_h + C_{1h}^*.$$

$V(P) = W(P)$ in (4.6) then yields

$$\text{(4.13)} \qquad \sum_{Q \in C_{2h}{}^*} G_h(P, Q) \leqq 4.$$

We can now formulate the following theorem:

THEOREM 1. *Let $u \in C^{(5)}(\bar{R})$ be the solution of* (1.5) *and suppose that a function ϕ satisfying* (4.8) *exists. Then we have*

$$\text{(4.14)} \qquad \max_{P} |\epsilon(P)| \leqq kh^3$$

with $\epsilon(P) = u(P) - U(P)$, $P \in R_h + C_{1h}^* + C_{2h}^* + C_{1h} + C_{2h}$, *U being the solution of* (4.1). *The constant k used in* (4.14) *depends only on u and* ϕ *but not on h.*

Proof. In (4.6), take $V(P) = \epsilon(P)$, then

$$\epsilon(P) = h^2 \sum_{Q \in R_h + C_{1h}^* + C_{2h}^*} G_h(P, Q)[-\Delta_h \epsilon(Q)] + h \sum_{Q \in C_{1h}} G_h(P, Q)[\delta_n \epsilon(Q) + \alpha(Q)\epsilon(Q)].$$

Since $G_h(P, Q) \geqq 0$, we have

$$\begin{aligned} |\epsilon(P)| \leqq & \left[h^2 \sum_{Q \in R_h} G_h(P, Q)\right] \cdot \max_{Q \in R_h} |\Delta_h \epsilon(Q)| \\ & + \left[h \sum_{Q \in C_{1h}^*} G_h(P, Q)\right] \cdot \max_{Q \in C_{2h}^*} |h \Delta_h \epsilon(Q)| \\ & + \left[\sum_{Q \in C_{2h}^*} G_h(P, Q)\right] \cdot \max_{Q \in C_{2h}^*} |h^2 \Delta_h \epsilon(Q)| \\ & + \left[h \sum_{Q \in C_{1h}} G_h(P, Q)\right] \cdot \max_{Q \in C_{1h}} |\delta_n \epsilon(Q) + \alpha(Q)\epsilon(Q)|. \end{aligned} \tag{4.15}$$

This immediately yields (4.14) using (4.9), (4.12) and (4.13) together with (1.9), (3.3), (1.11) and (2.23).

For examples of an application of the theorems of these and the following sections, we refer to [4].

V. The Neumann Problem. In this section, we shall consider an $O(h^3|\ln h|)$ approximation for the Neumann problem (1.3). The solution of (1.3), when it exists, is only unique up to an additive constant. This constant is usually determined by a normalization relation such as

$$\int_R u \, d\sigma = 0, \qquad \int_C u \, ds = 0 \quad \text{or} \quad u(x_0, y_0) = 0. \tag{5.1}$$

We shall consider the problem solved once we have found one of the solutions to (1.3).

We shall approximate the solution of (1.3) by

$$\begin{aligned} -\Delta_h U(P) &= f^*(P), & P &\in R_h' + C_h^*, \\ \delta_n U(P) &= g^*(P), & P &\in C_h, \\ U(0) &= u_0. \end{aligned} \tag{5.2}$$

The point 0 shall be a mesh point well in the interior of R, and we define R_h' as $R_h - 0$. The sets R_h, C_h^* and C_h are as in Section I. The operator Δ_h is defined by (1.6) for $P \in R_h'$ and by the appropriate operator defined in Section III for $P \in C_h^*$. δ_n is defined by (2.22), which formula can now of course be written as

$$\delta_n V(P) = \sum_{i=1}^{4} a_i \{ V(P) - V(P_i)\}. \tag{5.3}$$

The functions $f^*(P)$ and $g^*(P)$ are, as in Section IV, defined as

$$
\begin{aligned}
f^*(P) &= f(P) + \frac{h^2}{12}\,\Delta f(P), && P \in R_h', \\
f^*(P) &= f(P) - hF(P), && P \in C_h^*, \\
g^*(P) &= g(P) + F_1(P), && P \in C_h,
\end{aligned} \tag{5.4}
$$

with $F(P)$ and $F_1(P)$ as in (4.2); u_0 is a given constant. It is easy to see that the coefficient matrix of (5.2) is of positive type.

We now define a discrete function $N(P, Q)$ by

$$
\begin{aligned}
-\Delta_h N(P, Q) &= h^{-2}\delta(P, Q), && P \in R_h' + C_h^*, \\
\delta_n N(P, Q) &= h^{-1}\delta(P, Q), && P \in C_h, \\
N(0, Q) &= \delta(0, Q),
\end{aligned} \tag{5.5}
$$

for $Q \in R_h + C_h{}^* + C_h$, with $\delta(P, Q)$ the Kronecker delta. Clearly, $N(P, Q) \geqq 0$, while, for any mesh function $V(P)$,

$$
\begin{aligned}
V(P) = h^2 &\sum_{Q \in R_h' + C_h^*} N(P, Q)[-\Delta_h V(Q)] \\
&+ h \sum_{Q \in C_h} N(P, Q)[\delta_n V(Q)] + N(P, 0)\,V(0).
\end{aligned} \tag{5.6}
$$

A relation similar to (5.6) was given in (4.6). Taking $V(P) = 1$ in (5.6) yields

$$
N(P, 0) = 1, \qquad P \in R_h + C_h^* + C_h,
$$

which makes it possible to rewrite (5.6) as

$$
V(P) - V(0) = h^2 \sum_{Q \in R_{h'} + C_h^*} N(P, Q)[-\Delta_h V(Q)] + h \sum_{Q \in C_h} N(P, Q)[\delta_n V(Q)]. \tag{5.7}
$$

The following theorem is proved in [4]:

THEOREM 2. *Let $u \in C^{(5)}(\bar{R})$ be the solution of* (1.3) *and let R be such that either its boundary has no corners, or a function $\phi \in C^{(3)}(\bar{R})$ exists, satisfying*

$$
\begin{aligned}
-\Delta\phi &\geqq 1 && \text{in } R, \\
\partial\phi/\partial n &\geqq 1 && \text{on } C - C_1, \\
|\partial\phi/\partial n| &< \delta_1 && \text{on } C_1,
\end{aligned} \tag{5.8}
$$

where C_1 is a smooth arc on C of nonzero length, whose endpoints are not corners. Then we have

$$
\max_P |\epsilon(P)| \leqq kh^3\,|\ln h|
$$

with $\epsilon(P) = u(P) - U(P)$, $U(P)$ being the solution of (5.2), *$P \in R_h + C_h{}^* + C_h$. The constant k depends only on u and ϕ but not on h.*

Since this proof follows the same course as the proof given in [1], apart from certain complications due to the different nature of the Green's function under consideration, we shall give no further details here but refer once more to [4].

VI. The Dirichlet Problem. In this section we shall consider an $O(h^4)$ approximation for the Dirichlet problem (1.2) in which we shall use the operators given in Section III. Bramble and Hubbard [1] have given several finite-difference approximations for this problem, the most accurate of which has an $O(h^4)$ discretization error. In the interior of the region under consideration, they use the nine-point formula (1.6), while near the boundary, where use of this formula is not possible, an approximation is used in which not all the coefficients of the points used in approximating Δ for the central point have the same sign. Thus the resulting coefficient matrix is not of positive type. As we have seen in Section I, this means that it has to be shown separately that the system is solvable by iterative methods. Rockoff [9] has shown that the Jacobi- and Gauss-Seidel methods for this approximation converge. For the point SOR method, a range of relaxation factors can be given for which convergence is also proved. However, no details of these proofs are given in [9].

The approximation we propose to give here has also an $O(h^4)$ discretization error. The error bound obtained here is in general smaller than that of Bramble and Hubbard and under no circumstances greater. It has, moreover, a coefficient matrix which is of positive type.

We approximate (1.2) by

$$\begin{aligned} -\Delta_h U(P) &= f^*(P), \qquad P \in R_h + C_h^*, \\ U(P) &= g(P), \qquad P \in C_h. \end{aligned} \tag{6.1}$$

The function $f^*(P)$ is defined as $f(P) + h^2 \Delta f(P)/12$ and the sets R_h, $C_h{}^*$ and C_h are as in Section I. The operator Δ_h is defined by (1.6) for $P \in R_h$, while for $P \in C_h{}^*$ the appropriate formula from Section III is chosen.

We then have the following theorem:

Theorem 3. *Let $u(P) \in C^{(7)}(\bar{R})$ be the solution of* (1.2) *and $U(P)$ that of* (6.1). *We then have the following inequality for the discretization error $\epsilon(P) = u(P) - U(P)$*:

$$\max_P |\epsilon(P)| \leqq h^4 \left\{ \frac{2}{3} M_4 + \frac{M_6 d^2}{480} + O(h) \right\} \tag{6.2}$$

for $P \in R_h + C_h{}^* + C_h$.

The $O(h^4)$ approximation of Bramble and Hubbard [1] cited above yields the discretization error

$$\max_P |\epsilon(P)| \leqq h^4 \left\{ \frac{2}{3} M_4 + \frac{M_6 d^2}{120} + O(h) \right\}. \tag{6.3}$$

Apart from the fact that this method has a coefficient matrix which is not of positive type, comparison of (6.2) and (6.3) shows that our method has an upper bound which is never greater than that of Bramble and Hubbard, and may be up to a factor four smaller.

Again, the details of the proof are similar to those in the earlier work of Bramble and Hubbard, and may be found in [4].

Rekencentrum der Rijksuniversiteit
Landleven 1
Postbus 800
Groningen, The Netherlands

J. H. Bramble & B. E. Hubbard, "On the formulation of finite difference analogues of the Dirichlet problem for Poisson's equation," *Numer. Math.*, v. 4, 1962, pp. 313–327. MR **26** #7157.

→ J. H. Bramble & B. E. Hubbard, "A finite difference analogue of the Neumann problem for Poisson's equation," *J. Soc. Indust. Appl. Math. Ser. B. Numer. Anal.*, v. 2, 1965, pp. 1–14. MR **32** #8516.

3. J. H. Bramble & B. E. Hubbard, "Approximation of solutions of mixed boundary value problems for Poisson's equation by finite differences," *J. Assoc. Comput. Mach.*, v. 12, 1965, pp. 114–123. MR **30** #1615.

4. H. van Linde, *High-Order Finite Difference Methods for Poisson's Equation*, Thesis, Groningen, 1971.

→ G. H. Shortley & R. Weller, "The numerical solution of Laplace's equation," *J. Appl. Phys.*, v. 9, 1938, pp. 334–348.

6. J. H. Bramble & B. E. Hubbard, "On a finite difference analogue of an elliptic boundary problem which is neither diagonally dominant nor of non-negative type," *J. Mathematical Phys.*, v. 43, 1964, pp. 117–132. MR **28** #5566.

7. R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1962. MR **28** #1725.

→ E. Batschelet, "Über die numerische Auflösung von Randwertproblemen bei elliptischen partiellen Differentialgleichungen," *Z. Angew. Math. Phys.*, v. 3, 1952, pp. 165–193. MR **15**, 747.

9. M. Rockoff, "On the numerical solution of finite difference approximations which are not of positive type," *Notices Amer. Math. Soc.*, v. 10, 1963, p. 108. Abstract #597-169.