High-order finite-difference methods for Poisson's equation
van Linde, Hendrik Jan

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1971

Citation for published version (APA):
I. INTRODUCTION.

§ 1. Statement of the problem.

We shall consider the solution of the three boundary value problems for Poisson's equation

\[(1.1)\quad -\Delta u = f \quad \text{in } R\]

\(R\) is a bounded connected open set in the \((x,y)\) plane with boundary \(C\). The symbol \(\Delta\) denotes the Laplace operator \(\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\).

The Dirichlet problem for this equation is

\[(1.2)\quad -\Delta u = f \quad \text{in } R\]

\[u = g \quad \text{on } C\]

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

\[(1.3)\quad -\Delta u = f \quad \text{in } R\]

\[\frac{\partial u}{\partial n} = g \quad \text{on } C\]

\(\frac{\partial}{\partial n}\) denotes differentiation with respect to the outward-directed normal on \(C\). From Green's first identity follows that \(f\) and \(g\) must satisfy the relation

\[(1.4)\quad \int_R f \, ds + \int_C g \, ds = 0\]

Again under general assumptions a solution, unique up to an additive constant, exists.
Finally the third (or Robin) boundary value problem can be formulated as

\[-\Delta u = f \text{ on } R\]

\[\frac{\partial u}{\partial n} + \alpha u = g \text{ on } C_1\]

\[u = g_1 \text{ on } C_2\]

It is assumed here that the boundary \(C\) consists of the two parts \(C_1\) and \(C_2\). \(\frac{\partial}{\partial n}\) again denotes differentiation with respect to the outward-directed normal. We require that the function \(\alpha\) is 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 \(\alpha \geq \alpha_m > 0\), where \(\alpha_m\) is a constant (the set \(C_1^{(2)}\)).

We need only consider the cases where either \(C_2\) or \(C_1^{(2)}\) contains a non-empty 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 \(g_1\).

§ 2. Approximation by finite-difference methods.

In recent years, stimulated by the development of high-speed computers, much work has been done to solve partial differential equations by finite-difference methods, although the methods themselves are much older. Many different schemes have been proposed to get fast and accurate solutions to boundary value problems for elliptic equations of the above type.

The question of convergence of the methods was first discussed by Courant, Friedrichs and Lewy [1] (numbers between square brackets refer to the list of references). Gershgorin [2] first obtained an estimate for the order of convergence of a finite-difference approximation to the solution of the Dirichlet problem for a class of elliptic differential equa-
The most accurate 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 have formulated finite-difference analogues with an error estimate of $O(h^4)$ for the Dirichlet problem [3], $O(h^2|\ln h|)$ for the Neumann problem [4] and $O(h^2)$ for the Robin problem [5]. All these error estimates, as most of the estimates derived by others, contain certain higher derivatives of the solutions $u$ of the respective problems and are based on the assumption that these derivatives exist.

A priori error estimates based on the data (the functions given in the field and on the boundary) alone have been given for certain special cases by Wasow and others (see [6]) and later for the Dirichlet problem for Poisson's equation by Bramble and Hubbard [7].

In this work 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 $O(h^2|\ln h|)$. An $O(h^4)$ approximation for the Dirichlet problem will be given with certain advantages over the one proposed by Bramble and Hubbard [3].

§ 3. Approximation of the operators.

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 neighbours 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 $C_{7h}$ and
The mesh points of $R$ which are not in $R_h$ together form a set $C^*_h$. This set may be divided in 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 $\Delta_h$ to the Laplace operator $\Delta$ in $R_h$ and $C^*_h$ and in the case of the Neumann and Robin problems an analogue $\delta_n$ for the operator $\frac{\partial}{\partial n}$ on $C_h$. From the work of Bramble and Hubbard can be inferred that for the error estimates proposed in § 2 to be attained we need an approximation $\Delta_h$ to $\Delta$ 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 $\delta_n$ for $\frac{\partial}{\partial n}$ on $C_h$ with a truncation error of $O(h^3)$.

In [5] Bramble and Hubbard have given an approximation to the operator $\frac{\partial}{\partial n}$ with a truncation error of $O(h^2)$. In chapter 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 $\frac{\partial}{\partial n}$ 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 chapter 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 $\Delta$; if $(x,y)$ is a point of $R_h$, then

$$\Delta_h V(x,y) = \frac{1}{6h^2} \left[ 4[V(x,y+h) + V(x,y-h) + V(x+h,y) + V(x-h,y) + V(x+h,y+h) + V(x+h,y-h) + V(x-h,y+h) + V(x-h,y-h) - 20 V(x,y)] \right]$$ (1.6)
For $u \in C^{(7)}(\overline{R})$ the inequality

$$(1.7) \quad \left| \Delta_{h} u(P) - \Delta u(P) - \frac{h^2}{12} \Delta \Delta u(P) \right| \leq \frac{1}{30} M_{6} h^{4} + O(h^{5})$$

holds for $P \in R_{h}$, using the notation

$$(1.8) \quad M_{j} = \sup_{P \in R} \left\{ \left| \frac{\partial^{j} u(P)}{\partial x^{i} \partial y^{j-i}} \right| : i = 0, 1, \ldots, j \right\}$$

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

We shall also need the inequality

$$(1.9) \quad \left| \Delta_{h} u(P) - \Delta u(P) - \frac{h^2}{12} \Delta \Delta u(P) \right| \leq \frac{1}{5} M_{5} h^{3}$$

which holds for $P \in R_{h}$, if $u \in C^{(5)}(\overline{R})$.

In $C_{2h}^*$ we shall use the operator introduced by Shortley and Weller [8] for points like $(x,y)$ in fig. 1.1

We then approximate $\Delta$ by

![Fig. 1.1](attachment:image.png)
\[ \Delta_h V(x, y) = 2h^{-2} \left\{ \frac{1}{a(1+a)} V(x-ah, y) + \frac{1}{1+a} V(x+h, y) \right. \]
\[ + \frac{1}{\beta(1+\beta)} V(x, y+\beta h) + \frac{1}{1+\beta} V(x, y-h) - \left( \frac{1}{a} + \frac{1}{\beta} \right) V(x, y) \right\} \]

\( \alpha \) and \( \beta \) 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 \( u \in C^3(\bar{\Omega}) \) we have in a point \( P \) of \( C^*_h \) (or \( C^*_1h \) or \( C^*_{2h} \))
\[ (1.11) \quad |\Delta_h u(P) - \Delta u(P)| \leq \frac{\beta}{3} M_{\beta} h \]

§ 4. Preliminaries from matrix theory.

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.

We shall first give some definitions and theorems from matrix theory [9, 10].

Definition 1. A matrix \( A \) is monotone if \( Ax \geq 0 \) implies \( x \geq 0 \) for any vector \( x \).

Monotone matrices have an important property which may also be used as a definition:

Theorem 1 (Collatz [11]). A matrix \( A \) is monotone if and only
if $A$ is non-singular and $A^{-1} \geq 0$.

We further have the following definition:

**Definition 2.** An $n \times n$-matrix $A$ with elements $a_{ij}$ is of positive type if

(a) $a_{ij} \leq 0$, $i \neq j$

(b) $\sum_{k=0}^{n} a_{jk} \geq 0$ for all $j$, with strict inequality for $j \in J(A)$, and $J(A) \neq \emptyset$.

(c) for $i \notin J(A)$ there exists a finite sequence of non-zero elements of the form $a_{ik_1}, a_{k_1k_2}, \ldots, a_{k_{r-1}j}$, where $j \in J(A)$. Such a sequence is called a connection in $A$ from $i$ to $j$.

Bramble and Hubbard have proved the following theorem [10]:

**Theorem 2.** If a matrix $A$ is of positive type then it is monotone.

This theorem makes the matrices of positive type an easily identifiable subclass of the class of monotone matrices. It has already been remarked that the fact that the matrix of the system under consideration is of positive type can be used to give an upper bound for the discretization error, as we shall see in the following chapters.

The way in which the solution of an approximating system of linear equations, which for the problems under consideration will generally have a sparse coefficient matrix, can be obtained is of course of great importance. If this can not be done by iterative methods, the scheme is of little practical value. A coefficient matrix of positive type makes it easy to prove that the Jacobi- and Gauss-Seidel methods for the corresponding system of linear equations do indeed converge. Let us assume our original matrix, which is of positive type,
has the form $A = D - E - F$, where $D$ is a diagonal matrix and $E$ and $F$ are respectively strictly lower and upper triangular matrices. From the definition of matrices of positive type it then follows that the spectral radius $\rho(B)$ of the non-negative point Jacobi matrix $B = D^{-1}(E+F)$ associated with $A$ is $< 1$. From the Stein–Rosenberg theorem [12] then directly follows that the Jacobi- and Gauss–Seidel methods are both convergent, the latter method being asymptotically faster.

Recently error estimates were given for various approximation schemes which have coefficient matrices not of positive type [3, 9, 10]. For these methods the convergence of the various iterative methods has to be proved explicitly. Bramble and Hubbard [10] give an interesting example of a one-dimensional boundary value problem which is approximated by a difference method for which, with the mesh width $h$ sufficiently small, the Jacobi-method is divergent and the "forward-backward" Gauss–Seidel method convergent.