J. Greenstadt

On the Reduction of Continuous Problems To Discrete Form

Abstract: A continuous problem, defined as one involving derivatives or integrals, is to be reduced to a discrete problem, involving only algebraic or evaluative operations. An approach involving cells instead of points is taken, and the unknown function is approximated by functional representations, each associated with one cell and an associated set of parameters. Suitable operations are then defined, each associated with a particular cell. These operations remove the configuration coordinates from the problem, leaving only the parameters. Similar operations are defined which link the approximations in adjacent cells, and which translate certain interface conditions to relations between parameters associated with cells. The entire set of relations is then the equivalent of the usual difference equations.

A variational algorithm is introduced in order to circumvent certain difficulties associated with matching equations and unknowns. This also permits the convenient retention of certain "exact conditions" associated with the continuous problem. Some illustrative examples are given.

1. Correspondence between continuous and discrete problems

The most important class of problems in applied mathematics is that consisting in what we may call the *continuous* type. In this class are included those problems containing equations which involve the operations of classical analysis, viz., differentiation and integration. One might also include any other operations in which limiting procedures play a role.

The one thing common to all of these limiting operations of analysis is that none can be performed on a digital computer, which does not deal with continuously varying quantities and hence is incapable of taking limits by finite numerical processes.

For this reason, the universal practice in solving continuous problems by digital computation is to stop short of the limit by solving, not the original continuous problem, but another *discrete* problem which approximates, in some appropriate sense, to the continuous problem, and whose discrete solution approximates to the continuous solution of the continuous problem. By "appropriate," we mean, in general, that if one were to allow a characteristic parameter (or parameters) of the discrete problem to approach its limiting value, then the discrete problem (and its solution) would go over—in the sense of a limit—into the continuous problem (and its solution). Examples of characteristic parameters would be mesh size, number of terms in a series, time increment, et cetera.

We may think of the relationship between the corresponding problems as a mapping of the continuous into the discrete, and vice versa. The difficulty is that there is (for the class of "proper" limiting procedures) a unique continuous problem corresponding to an infinite set of discrete problems, i.e., the mapping from discrete problem to continuous problem is many-to-one. This is advantageous in theoretical derivations of continuous formulations of physical problems, but it leads to an enormous ambiguity when, conversely, one wishes to map the continuous problem into a discrete one. For this reason it is possible to write several difference equations corresponding to the same differential equation.

It would be very desirable to find some principle, or procedural guide, which would enable one to map continuous problems into discrete ones without such a great ambiguity. Thus by setting certain general requirements on the discrete formulation, one would be led automatically to the proper set of discrete equations. The examination of one such general procedure is the subject of this paper.

2. General features of "discretization"

The essence of *discretization* is the replacement of variables which have a continuous range, by variables which have a discrete range, i.e., which are functions of an integral index of some sort. We shall briefly outline the two principal methods in use by which this is done.

• *A. The Difference Method*

The correspondence established by this method, which is the one most widely used, is as follows:[1,2]

(1) The domain D of the independent variables (e.g., $x^\alpha, \alpha=1,\cdots,N$), originally a continuous-point set, is replaced by a finite set of suitably chosen points $\{x_k^\alpha\}$ with $k=1,\cdots,K$.

(2) The dependent variable (e.g., $\psi(x)$), originally defined for an infinity of points $\{\mathbf{x}\}$ is replaced by a discrete dependent variable ψ_k, each of whose values is associated appropriately with a point $\mathbf{x}_k$. (The symbol $\mathbf{x}_k$ stands for the set $\{x_k^\alpha\}$, $\alpha=1,\cdots,N$.)

(3) The derivatives of ψ with respect to the x^α are replaced by corresponding finite differences, (e.g., $\partial\psi/\partial x^\alpha$ becomes $\Delta\psi/\Delta x^\alpha$).

(4) The integrals of ψ over any domain of the $\mathbf{x}$-space are replaced by appropriately weighted sums over the discrete finite set of values ψ_k, whose arguments $\mathbf{x}_k$ are contained in the domain, (e.g., $\int_V \psi d^N x$ becomes $\sum_k w_k\psi_k$, where $\sum_k w_k = V$).

(5) The continuous equations defining the problem are replaced by a finite set of algebraic equations, equal in number to the number of unknown values $\{\psi_k\}$.

(6) The solution value ψ_k is presumed to represent the value of the original function $\psi(\mathbf{x}_k)$ in an approximate sense, i.e., by suitably increasing the density of points $\mathbf{x}_k$, the former values may be brought as close to the latter ones as desired. Accordingly, the various differences and weighted sums will approach the corresponding derivatives and integrals, (e.g., $\Delta\psi/\Delta x^\alpha \to \partial\psi/\partial x^\alpha$, $\sum_k w_k\psi_k \to \int \psi d^N x$).

• *B. The Fourier Method*

In this method, the correspondence is established as follows:[3]

(1) The (continuous) variables $\mathbf{x}$ are replaced by a set of indices $\nu_1, \nu_2, \cdots, \nu_N$ (written for short as ν), each of which takes on the values 0, 1, 2, $\cdots$ (these sequences all terminate).

(2) The dependent variable $\psi(x)$ is replaced by a discrete set of coefficients (or parameters) θ^ν, each of which is a function of the indices ν. The mathematical relation connecting ψ and θ is in terms of a set of known functions $\phi_\nu(x)$, which are linearly independent and usually orthonormal over the domain of $\mathbf{x}$. The corresponding equation for ψ is:

$$\psi(x) = \sum_\nu \theta^\nu \phi_\nu(\mathbf{x}) . \tag{2.1}$$

(3) The derivatives of ψ with respect to the x^α are computed directly from Eq. (2.1). If we expand the derivative $\partial\phi_\nu/\partial x^\alpha$ in terms of the ϕ's, to obtain:

$$\frac{\partial\phi_\nu}{\partial x^\alpha} = \sum_\sigma C_{\alpha\nu}^\sigma \phi_\sigma , \tag{2.2}$$

then we can express the derivative $\partial\psi/\partial x^\alpha$ in terms of the ϕ's:

$$\frac{\partial\psi}{\partial x^\alpha} = \sum_{\nu,\sigma} \theta^\nu C_{\alpha\nu}^\sigma \phi_\sigma . \tag{2.3}$$

(4) The integrals of ψ with respect to the x^α are computed in an entirely analogous manner. If we write:

$$\int_D \phi_\nu(x)\, d^N x \equiv I_\nu \tag{2.4}$$

then

$$\int_D \psi(x)\, d^N x = \sum_\nu \theta^\nu I_\nu . \tag{2.5}$$

(5) The continuous equations are replaced by a finite set of algebraic equations, equal in number to the number of parameters.

There are difficulties with both of these methods. For this discussion, the noteworthy ones are as follows:[1,2]

Method A:

(1) There is an ambiguity in the replacement of derivatives with differences, particularly where the coefficients in the differential equation are not constant.

(2) There is no systematic way to form difference equations near irregular boundaries or interfaces.

(3) It is difficult to carry over into the discrete formulation the proper analogues for certain integral relations holding in the continuous problem.

Method B:

(1) It is difficult to find basis functions which satisfy the boundary conditions.

(2) If the function ψ has strong fluctuations over the domain D, the convergence of the expansion (2.1) is, in general, slow.

In some cases, efforts have been made to overcome the limitations of both of these methods. Some of the variations from the pure procedures are as follows:

Method A:

(1) Near irregular boundaries, the unknown function is approximated by a polynomial and made to fit exactly a number of mesh nodes. This polynomial is then differentiated analytically, and the derivative is then expressed in terms of the values at the nearby points. However, there is still considerable ambiguity as regards the number

of neighboring points, the accuracy of the expansion, the matching of numbers of coefficients with numbers of points, et cetera. This technique often makes use of certain "fictitious points" in order to make the differencing more straightforward.

(2) The usual naive methods of integration, e.g., trapezoidal and Simpson's rules, are modified near boundaries so as to accord with the difference equations which are finally adopted. In this way, integral identities ("conservation rules") are maintained.

Method B:

(1) Neither the differential equation nor the boundary conditions are required to be satisfied exactly, but only to the "best extent possible," i.e., in the sense that some measure of the discrepancy is minimized. We shall adopt this point of view later in our "method of cells."[4]

3. The method of cells

We shall adopt a point of view which in many respects embodies features of both Methods *A* and *B*. We shall divide the domain D into many cells, viz., small subdomains, generally as many in number as the mesh nodes of Method *A*. However, we shall treat the cells in all other respects as in Method *B*.

Let us first consider some preliminary notational necessities. First the domain D is divided into subdomains $D_k (k=1, \cdots, K)$. These subdomains (cells) may be of fairly general shape, and need not be rectangular, triangular, et cetera. In any case, given two cells, they will or will not share a common interface. If they do, we shall call them *contiguous neighbors*. If the contiguous cells happen to be numbered k and m, we shall denote their common interface by B_{km}.

In many cases, we shall wish to refer to the contiguous neighbors of D_k. If we denote the "contiguous neighbor function" by $n(k)$, then $\{D_{n(k)}\}$ is the set of cells contiguous to D_k. Similarly, $\{B_{kn(k)}\}$ is the set of interfaces between D_k and its contiguous neighbors; hence the perimeter of D_k.

Let the continuous equation of the problem be denoted by:

$$\mathfrak{L}(\psi) = 0 . \tag{3.1}$$

We assume that Eq. (3.1), together with the boundary conditions:

$$\mathfrak{B}(\psi) = 0 \tag{3.2}$$

specify the solution ψ uniquely.

Now, in order to accomplish the reduction to a discrete problem, we shall approximate $\psi(x)$ individually within each cell D_k by a known function of $\mathbf{x}$ and certain parameters θ_k^ν:

$$\psi(x) \cong f_k(\mathbf{x}, \theta_k^\nu) . \tag{3.3}$$

In particular (as in Method *B*), we may take a linear function:

$$\psi_k(\mathbf{x}) = \sum_\nu \theta_k^\nu \phi_{k\nu}(\mathbf{x}) . \tag{3.4}$$

We note that the set of θ's is different for each cell. The basis functions $\phi_{k\nu}$ may also be different, although in many cases they may be the same for each cell. In all cases we shall assume that sums over ν cover the appropriate ranges of ν (e.g., $\nu = 1, \cdots, M_k$, where M_k is the number of θ's in D_k).

Clearly, we cannot represent exactly *any* function of the general class which satisfy Eq. (3.1), by a formula such as (3.3), wherein $f_k(\mathbf{x}, \boldsymbol{\theta}_k)$ is predetermined in form so that our only freedom lies in the *finite* set of parameter values θ_k^ν. In this fact lies the approximate nature of the replacement (3.3).

Our ultimate intention, of course, is somehow to replace the function $\psi(x)$ and its *continuously* ranging arguments x^α by the function θ_k^ν and its *discretely* ranging arguments k and ν. The analytic equation (3.1) will then be replaced by a finite set of equations for the parameter values θ^ν.

We next apply the operator $\mathfrak{L}$ to f_k in cell D_k and obtain:

$$\mathfrak{L}\{f_k(\mathbf{x}, \boldsymbol{\theta}_k)\} = \rho_k(\mathbf{x}, \boldsymbol{\theta}_k) . \tag{3.5}$$

If, by accident, $f_k(\mathbf{x}, \boldsymbol{\theta})$ were an exact solution of (3.1) in D_k, ρ_k would vanish, but in general this is not the case. Further, we cannot in general select values of θ_k^ν which will make ρ_k vanish identically. We can, however, ask that ρ_k shall: (a) vanish "to some order"; (b) be "as small as possible." Requirement (a) is the one usually drawn upon in Method *A*, while requirement (b) is largely used in Method *B*. The great advantage of (b) in the cell method is that one may derive in an automatic manner a number of equations just equal to the number of parameters.

4. Variational formulation of cell equations

We shall now consider the way in which we might set up a variational functional so as to obtain the proper set of equations for the θ_k^ν. These would then be our discrete equations corresponding to the continuous equation (3.1).

Our method is to set up a *functional*, i.e., a function of the θ_k^ν, derived from the operator $\mathfrak{L}$ and the approximation $f_k(\mathbf{x}, \boldsymbol{\theta}_k)$, which is to be minimized by the appropriate selection of the values for θ_k^ν, and which would in fact take on its absolute minimum for the true solution $\psi(x)$.

There are various ways of doing this, depending on the form of $\mathfrak{L}$. If $\mathfrak{L}$ is a self-adjoint, definite operator, then (3.1) is derivable as the Euler equation of a variational problem. An example is Laplace's equation:

$$\nabla^2\psi \equiv \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} = 0 , \tag{4.1}$$

which is the Euler equation for the $\psi(x, y)$ which minimizes the integral:

$$I = \tfrac{1}{2} \int_D \left\{ \left(\frac{\partial\psi}{\partial x} \right)^2 + \left(\frac{\partial\psi}{\partial y} \right)^2 \right\} dxdy . \tag{4.2}$$

Hence, we might take for our functional the expression I in this case.

In the general case, we can always seek to minimize the integrated square of ρ_k itself:

$$I_k \equiv \tfrac{1}{2}\int_{D_k} \{\rho_k(\mathbf{x}, \boldsymbol{\theta}_k)\}^2 dD_k = \text{minimum, all } k\,. \qquad (4.3)$$

There are some important points to be observed here. One is that the object of an integration over a domain is to eliminate $\mathbf{x}$ from the problem, leaving only the indices ν, k. Further, $\psi(x)$ is eliminated also, leaving only θ_k^ν. A second point is that we really should minimize the sum of all the I_k simultaneously, i.e.,

$$J = \sum_k I_k = \sum_k \tfrac{1}{2}\int_{D_k} \{\rho_k(\mathbf{x}, \boldsymbol{\theta}_k)\}^2 dD_k = \text{minimum}\,. \qquad (4.4)$$

However, we now notice that we have carried the separation of D into $\{D_k\}$ too far, i.e., we have a separate representation for each cell, and a separate functional to minimize independently for each cell. Thus (4.3) and (4.4) are equivalent.

In order to find a natural means for tying together the various ψ_k so as to form a ψ which represents the solution to (3.1) and (3.2) in D, we draw upon the requirements laid down by the existence theorems for partial differential equations.

For the purposes of this argument, we shall assume that ψ is an elliptic partial differential operator of the second order. For this type of operator, the appropriate uniqueness requirements are that ψ should be continuous and have continuous first partial derivatives.

Clearly, we cannot require these rather stringent conditions to hold exactly for our approximation $\{\psi_k\}$, so we shall set up the following discrete analogue of these conditions and incorporate it into our variational principle.

The continuity of $\{\psi_k\}$ fails on the interfaces B_{km}. This failure is measured by the difference

$$C_{km} = [\psi_k(\mathbf{x}, \boldsymbol{\theta}_k) - \psi_m(\mathbf{x}, \boldsymbol{\theta}_m)]_{B_{km}}\,. \qquad (4.5)$$

The discontinuity in the tangential derivative of ψ along B_{km} is included in C_{km}, and is expressed simply as the tangential derivative of C_{km}, viz., $\partial C_{km}/\partial\tau$. However, the discontinuity in the normal derivative must be computed separately as follows:

$$D_{km} = \left[\frac{\partial\psi_k}{\partial n}(\mathbf{x}, \boldsymbol{\theta}_k) - \frac{\partial\psi_m}{\partial n}(\mathbf{x}, \boldsymbol{\theta}_m)\right]_{B_{km}}. \qquad (4.6)$$

(The normal derivatives are both taken in the same direction.)

In order to form the proper functional with these ingredients, we square C_{km} and D_{km} and integrate the squares over B_{km} as follows:

$$I_{km} = \frac{\lambda}{2}\int_{B_{km}} \{\psi_k - \psi_m\}^2 dB_{km} + \frac{\eta}{2}\int_{B_{km}} \left\{\frac{\partial\psi_k}{\partial n} - \frac{\partial\psi_m}{\partial n}\right\}^2 dB_{km}\,. \qquad (4.7)$$

The multipliers λ and η express the weights we wish to attach to continuity and to continuity of the derivative.

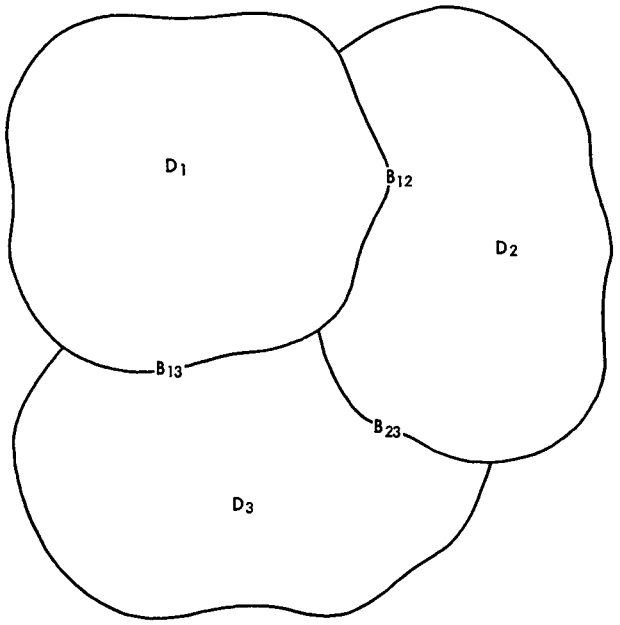

Figure 1

Now, when we sum I_{km} over all k and m (without repetitions) we find that the ψ_k are coupled together. Figure 1 shows how ψ_1, ψ_2, and ψ_3, for example, are linked together by their interface conditions on B_{12}, B_{23} and B_{13}.

We add $\sum_{km} I_{km}$ to J (of (4.4)) to obtain the complete functional (also denoted by J):

$$J = \tfrac{1}{2}\sum_k \int_{D_k} \{\rho_k(\mathbf{x}, \boldsymbol{\theta}_k)\}^2 dD_k + \tfrac{1}{2}\lambda\sum_{km}\int_{B_{km}} \{\psi_k - \psi_m\}^2 dB_{km} + \tfrac{1}{2}\eta\sum_{km}\int_{B_{km}} \left\{\frac{\partial\phi_k}{\partial n} - \frac{\partial\psi_m}{\partial n}\right\}^2 dB_{km}\,. \qquad (4.8)$$

We notice now the important fact that J is a function of the θ_k^ν only:

$$J = J(\theta_1^\nu, \theta_2^\nu, \cdots, \theta_K^\nu)\,. \qquad (4.9)$$

To obtain the equations for the parameters θ_k^ν, we simply differentiate J with respect to each θ_k^ν, and set the result equal to zero:

$$\frac{\partial J}{\partial\theta_k^\nu} = 0\,; \quad \nu = 1, \cdots, M_k\,; \quad k = 1, \cdots, K\,. \qquad (4.10)$$

Eq. (4.10) are now the equations pertaining to the discrete problem, and they correspond to Eq. (3.1).

It should be noted, finally, that the assigned conditions on the boundary of D itself require somewhat different expressions for C_{km} or D_{km}. If the boundary condition is of Dirichlet type, we must replace (4.5) by:

$$[C_k]_{B_k} = [\psi_k(\mathbf{x}, \boldsymbol{\theta}_k) - A(\mathbf{x})]_{B_k}\,. \qquad (4.11)$$

If it is of Neumann type, we must have:

$$[\mathrm{D}_k]_{B_k}=\left[\frac{\partial\psi_k}{\partial n}(\mathbf{x},\boldsymbol{\theta}_k)-A(\mathbf{x})\right]_{B_k}, \tag{4.12}$$

where the subscript B_k indicates the interface between cell D_k and the region outside D. Only those values of k are included, for which D_k is a boundary cell.

5. Incorporation of auxiliary conditions

In many problems of physical interest, it is of great importance to preserve certain integral relations (or other auxiliary conditions) exactly, even though the solution function ψ, which is required to fulfill these relations, satisfies the basic equations of the problem only approximately.

An example of such a relation is Gauss' Theorem applied to the Laplacian operator. Let ψ satisfy Eq. (4.1) in the interior of D. Then by Gauss' Theorem:

$$\int_D \nabla^2\psi dD=\int_B \frac{\partial\psi}{\partial n}dB=0\,, \tag{5.1}$$

where B is the boundary of D. Hence we may require, in the discrete problem, that

$$\int_{B_k}\frac{\partial\psi_k}{\partial n_k}dB_k=0 \tag{5.2}$$

exactly.

We may actually alter the discrete equations (4.10) for the θ_κ^ν in a systematic manner, so that if ψ_k satisfy these altered discrete equations, they will also automatically satisfy the auxiliary relations which have influenced the alteration.

Let us first write such relations in a more general form. We first associate with each cell D_k an exact condition which must be fulfilled. Denote this by $E_k(\boldsymbol{\theta}_k)$. An example would be the following:

$$E_k(\boldsymbol{\theta}_k)\equiv\int_{D_k}\mathfrak{L}\{f_k(\mathbf{x},\boldsymbol{\theta}_k)\}dD_k=\int_{D_k}\rho_k(\mathbf{x},\boldsymbol{\theta}_k)dD_k\,, \tag{5.3}$$

which is distinct from condition (4.4), insofar as we did not require the integral appearing therein to vanish.

Another possible exact relation would be the following:

$$F_{km}(\boldsymbol{\theta}_k,\boldsymbol{\theta}_m)\equiv\int_{B_{km}}\left\{\frac{\partial\psi_k}{\partial n}-\frac{\partial\psi_m}{\partial n}\right\}dB_{km}=0\,. \tag{5.4}$$

We may easily incorporate these conditions into the variational principle by using Lagrange's Method of Multipliers. We simply add to the functional J a linear combination of the E_k and F_{km}, with multipliers α_k and β_{km}. Hence, we have:

$$J=\sum_k I_k+\sum_{km}\int_{B_{km}}\{\tfrac{1}{2}\lambda C_{km}^2+\tfrac{1}{2}\eta D_{km}^2\}dB_{km}$$
$$+\sum_k\alpha_k E_k+\sum_{km}\beta_{km}F_{km}\,. \tag{5.5}$$

Note that we may have several conditions of each of the types represented by C, D, E, F, instead of only one of each. However, we should never require more exact relations than there are parameters θ_k^ν.

Our discrete equations corresponding to the functional J are as follows:

$$\frac{\partial J}{\partial\theta_k^\nu}=0\,;\quad\frac{\partial J}{\partial\alpha_k}=E_k=0\,;\quad\frac{\partial J}{\partial\beta_{km}}=F_{km}=0 \tag{5.6}$$

for all ν, k and m.

6. Remarks on the discrete equations

If we focus our attention on one of the cells, say D_{k_0}, we notice that the set of discrete equations corresponding to k_0 is as follows:

$$\frac{\partial J}{\partial\theta_{k_0}^\nu}=0\,;\qquad E_{k_0}=0\,;\qquad F_{k_0}m=0\,. \tag{6.1}$$

Now we notice an interesting property: Because $\theta_{k_0}^\nu$ enters into J in connection with its *contiguous neighbors only*, via C_{k_0m}, D_{k_0m} and F_{k_0m}, the differentiations in (6.1) will cause to drop out of these equations all θ_ν^κ which are not associated with contiguous neighbor cells $D_{n(k_0)}$ (see notational note in Section 3). Hence, after eliminating α_{k_0} and $\beta_{k_0n(k_0)}$ from these equations, the resulting set may be written:

$$W_{k_0}^\nu(\boldsymbol{\theta}_{k_0},\boldsymbol{\theta}_{n(k_0)};\lambda,\eta)=0\,;\qquad\nu=1,\cdots,M_{k_0}\,, \tag{6.2}$$

and we see therewith that the "principal" equations for $\theta_{k_0}^\nu$ involve only the parameters associated with contiguous neighbor cells. Hence, these discrete equations resemble difference equations insofar as the latter also connect nearest neighbors.

We may regard Eq. (6.2) as a discrete analogue to (3.1) in the following sense: A differential equation may be thought of as the expression of a certain connection between the characteristics of ψ at "infinitely near" points. In (6.2), the elements being connected are not "infinitely near," but only "near."

7. Application to Laplace's equation—regular cell

We shall now consider a simple example to illustrate the application of the cell method. We shall consider the typical cluster of cells as shown in Fig. 2 (overleaf).

For convenience, we have labeled the cells as illustrated. Without loss of generality, we can assume the squares to be 2 units on each side. The coordinate system for each square will have its origin at the center, and the x and y coordinates vary between -1 and $+1$.

We shall expand $\psi(x,y)$ in a series of functions which form an orthonormal set in each square. These shall be denoted by $\{\phi_\mu(x)\phi_\nu(y)\}$. Later, we shall actually use the sine and cosine functions, but for notational simplicity, we shall defer this. First, we have:

$$\psi_k(x,y)=\sum_{\mu\nu}\theta_k^{\mu\nu}\phi_\mu(x)\phi_\nu(y)\,;\qquad k=0,\cdots,4\,. \tag{7.1}$$

For the functional J, we shall use expression (5.5), except that I_0 shall be given by:

$$I_0=\tfrac{1}{2}\int_{D_0}\left\{\left(\frac{\partial\psi_0}{\partial x}\right)^2+\left(\frac{\partial\psi_0}{\partial y}\right)^2\right\}dxdy\,. \tag{7.2}$$

The computation of I_0, with the help of the orthonormality of the ϕ's, yields the following:

$$I_0=\tfrac{1}{2}\sum_{\sigma\tau\rho}\{\theta_0^{\sigma\rho}\,\theta_0^{\tau\rho}\int_{-1}^{1}\phi'_\sigma\phi'_\tau dx+\theta_0^{\rho\sigma}\,\theta_0^{\rho\tau}\int_{-1}^{1}\phi'_\sigma\phi'_\tau dy\}$$

$$\left(\phi'\equiv\frac{d\phi}{dx}\text{ or }\frac{d\phi}{dy}\right). \tag{7.3}$$

The expression representing the degree of discontinuity of ψ between cells is as follows:

$$I_c=\tfrac{1}{2}\sum_\nu\{[\sum_\mu(\theta_0^{\mu\nu}\overset{+}{\phi}_\mu-\theta_1^{\mu\nu}\overset{-}{\phi}_\mu)]^2+[\sum_\mu(\theta_0^{\mu\nu}\overset{-}{\phi}_\mu-\theta_3^{\mu\nu}\overset{+}{\phi}_\mu)]^2\}$$

$$+\tfrac{1}{2}\sum_\mu\{[\sum_\nu(\theta_0^{\mu\nu}\overset{+}{\phi}_\nu-\theta_2^{\mu\nu}\overset{-}{\phi}_\nu)]^2+[\sum_\nu(\theta_0^{\mu\nu}\overset{-}{\phi}_\nu-\theta_4^{\mu\nu}\overset{+}{\phi}_\nu)]^2\}\,. \tag{7.4}$$

where $\overset{+}{\phi}$ represents $\phi(+1)$ and $\overset{-}{\phi}$ represents $\phi(-1)$.

The expression representing the degree of discontinuity of $\partial\psi/\partial n$ between cells is as follows:

$$I_D=\tfrac{1}{2}\sum_\nu\{[\sum_\mu(\theta_0^{\mu\nu}\overset{+}{\phi}'_\mu-\theta_1^{\mu\nu}\overset{-}{\phi}'_\mu)]^2+[\sum_\mu(\theta_0^{\mu\nu}\overset{-}{\phi}'_\mu-\theta_3^{\mu\nu}\overset{+}{\phi}'_\mu)]^2\}$$

$$+\tfrac{1}{2}\sum_\mu\{[\sum_\nu(\theta_0^{\mu\nu}\overset{+}{\phi}'_\nu-\theta_2^{\mu\nu}\overset{-}{\phi}_\nu)]^2+[\sum_\nu(\theta_0^{\mu\nu}\overset{-}{\phi}'_\nu-\theta_4^{\mu\nu}\overset{+}{\phi}'_\nu)]^2\}\,. \tag{7.5}$$

For exact auxiliary relations, we shall require:

$$\int_{D_0}\nabla^2\psi dxdy=0 \tag{7.6}$$

and

$$\int_{B_{0i}}\left(\frac{\partial\psi_0}{\partial n_i}-\frac{\partial\psi_i}{\partial n_i}\right)dB_{0i}=0\,;\qquad i=1,\cdots,4\,. \tag{7.7}$$

In terms of the series (7.1), condition (7.6) becomes:

$$\sum_\mu\theta_0^{\mu0}\int_{-1}^{1}\phi''_\mu dx+\sum_\nu\theta_0^{0\nu}\int_{-1}^{1}\phi''_\nu dy=0\,, \tag{7.8}$$

and for (7.7), we have:

$$\sum_\mu(\theta_0^{\mu0}\overset{+}{\phi}'_\mu-\theta_1^{\mu0}\overset{-}{\phi}'_\mu)=0 \tag{7.9a}$$

$$\sum_\nu(\theta_0^{0\nu}\overset{+}{\phi}'_\nu-\theta_2^{0\nu}\overset{+}{\phi}'_\nu)=0 \tag{7.9b}$$

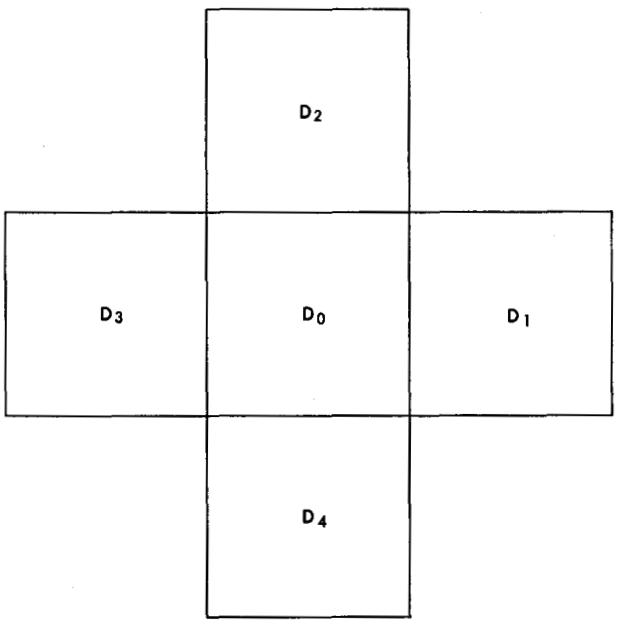

Figure 2

$$\sum_\mu(\theta_0^{\mu0}\overset{-}{\phi}'_\mu-\theta_3^{\mu0}\overset{+}{\phi}'_\mu)=0 \tag{7.9c}$$

$$\sum_\nu(\theta_0^{0\nu}\overset{-}{\phi}'_\nu-\theta_4^{0\nu}\overset{+}{\phi}'_\nu)=0\,. \tag{7.9d}$$

The first of Eq. (6.1) becomes:

$$\frac{\partial J}{\partial\theta_0^{\mu\nu}}=A_{\mu\sigma}\theta_0^{\sigma\nu}+\theta_0^{\mu\sigma}A_{\sigma\nu}+B_{\mu\sigma}\theta_0^{\sigma\nu}+\theta_0^{\mu\sigma}B_{\sigma\nu}-C_{\mu\sigma}\theta_1^{\sigma\nu}-\theta_2^{\mu\sigma}C_{\sigma\nu}-D_{\mu\sigma}\theta_3^{\sigma\nu}-\theta_4^{\mu\sigma}D_{\sigma\nu}+E_{\mu\nu}\,, \tag{7.10}$$

where

$$A_{\mu\sigma}=\int_{-1}^{1}\phi'_\mu\phi'_\sigma\left\{\begin{matrix}dx\\dy\end{matrix}\right\} \tag{7.11}$$

$$B_{\mu\sigma}=\lambda(\overset{+}{\phi}_\mu\overset{+}{\phi}_\sigma+\overset{-}{\phi}_\mu\overset{-}{\phi}_\sigma)+\eta(\overset{+}{\phi}'_\mu\overset{+}{\phi}'_\sigma+\overset{-}{\phi}'_\mu\overset{-}{\phi}'_\sigma) \tag{7.12}$$

$$C_{\mu\sigma}=\lambda\overset{+}{\phi}_\mu\overset{-}{\phi}_\sigma+\eta\overset{+}{\phi}'_\mu\overset{-}{\phi}'_\sigma \tag{7.13a}$$

$$D_{\mu\sigma}=\lambda\overset{-}{\phi}_\mu\overset{+}{\phi}_\sigma+\eta\overset{-}{\phi}'_\mu\overset{+}{\phi}'_\sigma \tag{7.13b}$$

$$E_{\mu\nu}=\alpha\left\{\delta_{0\nu}\int_{-1}^{1}\phi''_\mu dx+\delta_{\mu0}\int_{-1}^{1}\phi''_\nu dy\right\}+\beta_{01}\delta_{0\nu}\overset{+}{\phi}'_\mu+\beta_{02}\delta_{\mu0}\overset{+}{\phi}'_\nu+\beta_{03}\delta_{0\nu}\overset{-}{\phi}'_\mu+\beta_{04}\delta_{\mu0}\overset{-}{\phi}'_\nu\,. \tag{7.14}$$

The additional equations are, of course, (7.8) and (7.9).

We shall now assign, for the $\{\phi_\mu\}$, the sines and cosines as follows:

$$\phi_0(x)=1/\sqrt{2} \tag{7.15a}$$

$$\phi_\mu(x)=\sin p\pi x\,,\qquad \mu=2p-1 \tag{7.15b}$$

$$\phi_\mu(x)=\cos p\pi x\,,\qquad \mu=2p\,. \tag{7.15c}$$

Notice that, in all cases, we may write $p=\left[\frac{\mu+1}{2}\right]$, where the bracket means "the integral part of." With this notation, we find for $A_{\mu\sigma}$ the following result:

$$A_{\mu\sigma}=\pi^2p^2\delta_{\mu\sigma}\,;\qquad p=\left[\frac{\mu+1}{2}\right]. \tag{7.16}$$

We find further that Eq. (7.8) is automatically satisfied, and that the coefficient of α in (7.14) vanishes. Hence, α no longer appears in the problem. Further, if we denote the parity of μ by $\mathcal{P}(\mu)$, we have:

$$B_{\mu\sigma}=0\text{ , if }\mathcal{P}(\mu)\neq\mathcal{P}(\sigma) \tag{7.17a}$$

$$B_{2p,2q}=2\lambda(-1)^{p+q} \tag{7.17b}$$

$$B_{2p-1,2p-1}=2\eta\pi^2pq(-1)^{p+q} \tag{7.17c}$$

$$C_{\mu\sigma}=D_{\mu\sigma}=\tfrac{1}{2}B_{\mu\sigma} \tag{7.18}$$

and

$$E_{\mu\nu}=0\text{ if }\mu,\nu\neq0,\text{ or }\mathcal{P}(\mu)=\mathcal{P}(\nu) \tag{7.19a}$$

$$E_{2p-1,0}=(\beta_{01}+\beta_{03})p\pi(-1)^p \tag{7.19b}$$

$$E_{0,2q-1}=(\beta_{02}+\beta_{04})q\pi(-1)^q\,. \tag{7.19c}$$

We shall assume a simple approximation for ψ_0, viz.:

$$\psi_0=\sum_{\substack{\mu=0\\ \nu=0}}^{1}\theta_0^{\mu\nu}\phi_\mu(x)\phi_\nu(y)\,, \tag{7.20}$$

which contains the 4 parameters θ_0^{00}, θ_0^{10}, θ_0^{01} and θ_0^{11}. The equations (7.10) and (7.9) reduce to:

$$4\theta_0^{00}-(\theta_1^{00}+\theta_2^{00}+\theta_3^{00}+\theta_4^{00})=0 \tag{7.21a}$$

$$\theta_0^{10}=\theta_1^{10}=\theta_3^{10} \tag{7.21b}$$

$$\theta_0^{01}=\theta_2^{01}=\theta_4^{01} \tag{7.21c}$$

$$4\left(1+\frac{1}{2\eta}\right)\theta_0^{11}-(\theta_1^{11}+\theta_2^{11}+\theta_3^{11}+\theta_4^{11})=0\,. \tag{7.21d}$$

The first of these equations is the familiar Laplace difference equation (in this case, for the *mean values* θ_k^{00}). In order to maintain consistency, it is necessary to require, as indicated in (7.21b) and (7.21c), that $\theta_1^{10}=\theta_3^{10}$ and $\theta_2^{01}=\theta_4^{01}$. Hence, these parameters are completely uncoupled from the interior of D, but are determined by the boundary conditions alone. In (7.21d), if $\eta\to\infty$, which would impose complete continuity of $\partial\psi/\partial n$ (to the approximation as indicated), we see that $\{\theta_k^{11}\}$ also would satisfy Laplace's difference equation.

8. Laplace's equation—irregular cell

We shall now analyze a more difficult problem, viz., the case of an irregular (interface) cell. We shall consider an irregularly disposed interface between two regions, in each of which Laplace's equation is to be satisfied. However, the interface condition is slightly more complicated than before. Figure 3 depicts the geometry.

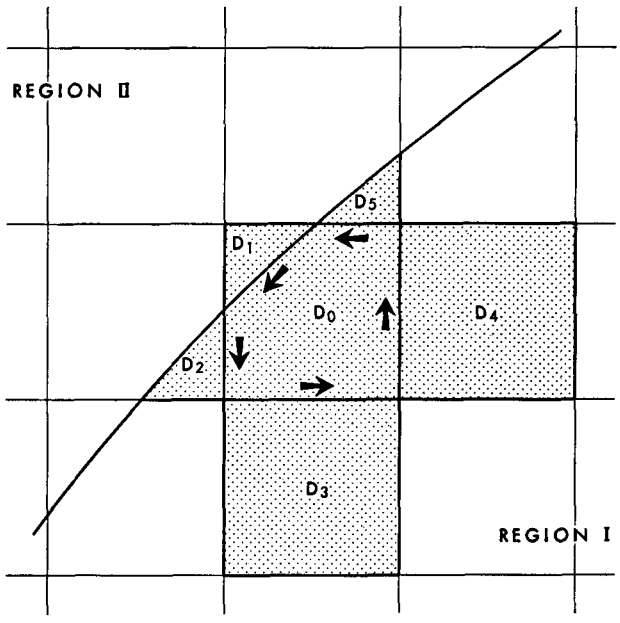

Figure 3

We notice that D_0 has five contiguous neighbors, instead of four, and that one of them (D_1) is not in the same "physical" region as D_0.

As far as the functional to be minimized is concerned, it shall be identical in content with (5.5), but now the various integrals will have to be computed differently. Starting with I_0, we shall have:

$$\begin{aligned}I_0=&\tfrac{1}{2}\int_{D_0}\{[\sum_{\mu,\nu}\theta_0^{\mu\nu}\phi'_\mu(x)\phi_\nu(y)]^2\\&+[\sum_{\mu,\nu}\theta_0^{\mu\nu}\phi_\mu(x)\phi'_\nu(y)]^2\}dxdy\\&=\tfrac{1}{2}\sum_{\mu\nu\sigma\tau}A_{\mu\nu\sigma\tau}\theta_0^{\mu\nu}\theta_0^{\sigma\tau}\,,\end{aligned} \tag{8.1}$$

where

$$\begin{aligned}A_{\mu\nu\sigma\tau}=\int_{D_0}&\{\phi'_\mu(x)\phi'_\sigma(x)\phi_\nu(y)\phi_\tau(y)\\&+\phi_\mu(x)\phi_\sigma(x)\phi'_\nu(y)\phi'_\tau(y)\}dxdy\,.\end{aligned} \tag{8.2}$$

In order to write the integrals over interfaces in a concise form, we must refine our representation of the interfaces. Let each boundary segment B_{0m} be represented as a parametrized curve as follows:

$$B_{0m}\colon\ x=u_m(s),\ y=v_m(s);\ -1\leq s\leq1\,. \tag{8.3}$$

The functions $u_m(s)$ and $v_m(s)$ are constructed in such a manner that the point (x, y) moves from one end of the interface segment to the other in the positive direction (as indicated by the arrows in Fig. 3) as s varies from -1 to $+1$. We may now write any line integral along B_{0m} in the form:

$$I_F = \int_{B_{0m}} F(x,y)\,dB_{0m},$$

$$= \int_{-1}^{1} F[u_m(s), v_m(s)]\sqrt{\dot{u}_m^2 + \dot{v}_m^2}\;ds\,, \tag{8.4}$$

where $\dot{u}_m \equiv \dfrac{du_m}{ds}$, $\dot{v}_m = \dfrac{dv_m}{ds}$.

To compute the normal derivative of ψ on a given interface, we make use of the well-known formula:

$$\frac{\partial\psi}{\partial n} = n_m^x \frac{\partial\psi}{\partial x} + n_m^y \frac{\partial\psi}{\partial y}\,, \tag{8.5}$$

where (n_m^x, n_m^y) is the unit normal to B_{0m}, and all quantities are calculated on B_{0m}.

The unit normal vector $\mathbf{n}_m$ may be calculated in terms of the vector $\boldsymbol{\tau}_m$ tangent to B_{0m}. The latter is given by:

$$\boldsymbol{\tau}_m = (\dot{u}_m, \dot{v}_m)\,. \tag{8.6}$$

The vector:

$$\hat{\mathbf{n}}_m = (\dot{u}_m^2 + \dot{v}_m^2)^{-\frac{1}{2}}(\dot{v}_m, -\dot{u}_m) \tag{8.7}$$

is clearly orthogonal to $\boldsymbol{\tau}$, and is of unit length. Hence:

$$\frac{\partial\psi}{\partial n} = (\dot{u}_m^2 + \dot{v}_m^2)^{-\frac{1}{2}}\left\{\dot{v}_m \frac{\partial\psi}{\partial x} - \dot{u}_m \frac{\partial\psi}{\partial y}\right\}_{B_{0m}}. \tag{8.8}$$

For convenience, we shall use the same functions $\phi_\mu(x)$ and $\phi_\nu(y)$ within an irregular cell, which were previously defined over the entire square of which it is a part. The coefficients $\theta^{\mu\nu}$ shall, of course, be different for the two cells composing a square.

For the interface discrepancies, we shall take:

$$C_{0m} = [\psi_0 - \psi_m]_{B_{0m}}\,; \qquad m = 1, \cdots, 5 \tag{8.9}$$

$$D_{01} = \left[\kappa_I \frac{\partial\psi_0}{\partial n_1} - \kappa_{II}\frac{\partial\psi_1}{\partial n_1}\right]_{B_{01}} \tag{8.10a}$$

$$D_{0m} = \left[\frac{\partial\psi_0}{\partial n_m} - \frac{\partial\psi_m}{\partial n_m}\right]_{B_{0m}}; \qquad m = 2, \cdots, 5\,. \tag{8.10b}$$

The coefficients κ_I and κ_{II} represent physical constants associated with Regions I and II respectively. (For example, in heat flow problems, these are conductivities; in neutron diffusion problems they are reciprocal transport cross-sections, et cetera.)

I_C is given by:

$$I_C = \tfrac{1}{2}\sum_{\mu\nu\sigma\tau}\sum_{m=1}^{5}\{B^{(m)}_{\mu\nu\sigma\tau}\theta^{\mu\nu}_0\theta^{\sigma\tau}_0 + C^{(m)}_{\mu\nu\sigma\tau}\theta^{\mu\nu}_0\theta^{\sigma\tau}_m + C^{(m)}_{\sigma\tau\mu\nu}\theta^{\sigma\tau}_0\theta^{\mu\nu}_m + D^{(m)}_{\mu\nu\sigma\tau}\theta^{\mu\nu}_m\theta^{\sigma\tau}_m\} \tag{8.11}$$

with:

$$B^{(m)}_{\mu\nu\sigma\tau} = \int_{B_{0m}} \phi_{0\mu}(x)\phi_{0\sigma}(x)\phi_{0\nu}(y)\phi_{0\tau}(y)\,dB_{0m} \tag{8.12a}$$

$$C^{(m)}_{\mu\nu\sigma\tau} = \int_{B_{0m}} \phi_{0\mu}(x)\phi_{m\sigma}(x)\phi_{0\nu}(y)\phi_{m\tau}(y)\,dB_{0m} \tag{8.12b}$$

$$D^{(m)}_{\mu\nu\sigma\tau} = \int_{B_{0m}} \phi_{m\mu}(x)\phi_{m\sigma}(x)\phi_{m\nu}(y)\phi_{m\tau}(y)\,dB_{0m}\,. \tag{8.12c}$$

The expression for D_{01} corresponding to Eq. (4.6) is as follows:

$$D_{01} = \sum_{\mu,\nu} E_{\mu\nu}(\kappa_I\theta^{\mu\nu}_0 - \kappa_{II}\theta^{\mu\nu}_1)\,, \tag{8.13}$$

where

$$E_{\mu\nu} = (\dot{u}_1^2 + \dot{v}_1^2)^{-\frac{1}{2}}\{\dot{v}_1\phi'_\mu(x)\phi_\nu(y) - \dot{u}_1\phi_\mu(x)\phi'_\nu(y)\}^{B_{01}} \tag{8.14}$$

D_{01} must be squared and integrated over B_{01}. We shall not carry this manipulation out in detail, since the calculation on C_{0m} illustrates the method. Further, we shall not write the expressions for $D_{0m}(m = 2, \cdots, 5)$, since these are quite analogous to (8.13), with the exception that the $\{\phi_\mu\}$ are now defined in different squares.

The end result of these calculations is a set of equations similar to (7.10); however, the formulas for the various coefficients are much more complicated. Nevertheless, there does not appear any difficulty of principle which would prevent the programming of the general procedure for an automatic computing machine.

9. Various generalizations

• *Nonlinear problems*

Two generalizations are possible here: (1) nonlinear equations with the retention of linear expansions of ψ; (2) nonlinear representation for ψ (i.e., nonlinear in $\{\theta^\nu\}$). Further, one may have both of these in one problem. In the case of (1), it is of course necessary to find a suitable variational functional.

• *Several dependent variables*

One simply appends an index to ψ, so that one deals with $\psi^p(p = 1, \cdots, P)$. Naturally, there must be P equations, but these can also be incorporated into an over-all functional without difficulty in principle.

• *Higher dimensions*

The only change necessary here is the modification in the expansion: e.g., $\psi = \sum_{\mu\nu\sigma}\theta^{\mu\nu\sigma}\phi_\mu(x)\phi_\nu(y)\phi_\sigma(z)$. Each interface B_{km} still separates two cells, et cetera.

10. Some questions for further investigation

Since the functional J is made to be positive, the resulting quadratic form in $\{\theta^{\mu\nu}_k\}$ has the same property. This means that the matrix of coefficients of the $\{\theta^{\mu\nu}_k\}$ is symmetric and positive definite. Hence, it should be possible to set up relatively simple iterative procedures for solution which may be guaranteed to converge. This, however, has not yet been carried out.

As far as the convergence of the approximation itself is concerned, one may expect it to be similar to that of the classical Rayleigh-Ritz method for a single domain or a small number of domains. No theorem has been formulated or proved, however, for the present case.

If it is feasible to prepare a program for an automatic computer to execute this method, the treatment of free boundaries would be facilitated by the computer's ability to set up new difference equations for the new (free) interface at each time step. It is also conceivable that, by introducing some $\phi_\mu(x, y)$ which themselves contain discontinuities, the free boundaries might be ignored.

Finally, it is possible that inappropriate difference equations, which violate "conservation laws" (or integral identities), may, by this characteristic, contain inherent tendencies to slow convergence when iterative methods of solution are attempted. The two most obvious defects of such difference equations are possible "fictitious" sources and sinks, or "fictitious" interfaces with low "error transmission" or "error diffusion" characteristics. It is possible that proper attention to the requirement of exact integral identities would remove these troubles.

Received December 26, 1958

Acknowledgment

Many thanks are due to Dr. P. H. Sterbenz for constructive criticism and discussion leading to the refinement of much of the theory. Thanks are also due to Mr. Frank Engel for a decisive criticism of a very early formulation, as well as to one of the referees, whose detailed criticisms of exposition led to a reorganization and simplification of the presentation.

References

1. W. E. Milne, *Numerical Solutions of Differential Equations,* John Wiley and Sons, New York, 1953. Chap. 8 et seq.
2. L. Collatz, *Numerische Behandlung von Differentialgleichungen,* Springer-Verlag, Berlin, 1955. Kap. IV.
3. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics,* McGraw-Hill Book Co., New York, 1953. Section 6.3.
4. Morse & Feshbach, *op. cit.,* Section 9.4.