INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
MIRAMARE - P.O.B. 586 - 34100 TRIESTE (ITALY) - TELEPHONES: 224281/2/3/4/5/6 - CABLE: CENTRATOM

SMR/23 /28

AUTUMN COURSE ON APPLICATIONS OF ANALYSIS
TO MECHANICS

22 September - 26 November 1976

RESIDUALS AND DIFFERENTIAL EQUATIONS

E. L. Ortiz
Imperial College
London, U.K.

These are preliminary lecture notes intended for participants only.
Copies are available outside the Publications Office (T-floor) or from room 112.

II - RESIDUALS AND DIFFERENTIAL EQUATIONS

2.1.- Residuals

Let us consider now the elliptic problem :

Problem 1 :
$$\begin{cases} \Delta z = 0 & \text{in } \mathcal{D} \in \mathbb{R}^2 \\ z|_{\partial\mathcal{D}} = \varphi_0 \end{cases}$$

We take $\{\varphi_i(\bar{x})\}$, $i = 1(1)n$: $\varphi_i|_{\partial\mathcal{D}} = 0$ as a coordinate system for z in $\mathcal{P}_n$ and construct the trial solution

$$\hat{z} = \varphi_0 + \sum_1^n c_i \varphi_i \qquad (1)$$

Def 2.1 : We call
$$R = R(\bar{c}, \bar{x}) = \Delta \hat{z} = \Delta\varphi_0 + \sum_1^n c_i \Delta\varphi_i \qquad (2)$$
where $\bar{c} = \{c_i\}$, $i = 1(1)n$, the residual function corresponding to Problem 1

Remark 1.2 : R is also called the error in the equation.

We shall now try to determine the coefficients $\bar{c}$ in the expression of $\hat{z}$ by forcing R to satisfy conditions given by a functional f_j, such that R is orthogonal to a certain vector $\bar{w} = \{w_j\}$:

$$(w_j, R) = 0 \qquad j = 1(1)n.$$

From (1) and (2)

$$\sum c_i (w_j, \Delta\varphi_i) = -(w_j, \Delta\varphi_0) \quad , \quad i = 1(1)n \ ,$$

or

$$\underline{B}\ \bar{c} = \bar{d},$$

if we call

$$\underline{B} = ((w_j, \Delta\varphi_i)) \ ; \ \bar{c} = \{c_i\} \text{ and } \bar{d} = -(f_j, \Delta\varphi_0).$$

Finally, if B^{-1} exists

$$\bar{c} = \underline{B}^{-1} \bar{d}$$

solves our problem.

2.2.- Realizations of $\{f_j\}$

We shall discuss now a number of examples of particular realizations of the f_j. These examples are well known methods for the numerical solution of operator equations. A combination of the characteristics of some of these methods may be useful when a method for the numerical solution of a particular operator equation is required.

1) The full domain method :

We take $D = \bigcup_i D_i$ and impose on R the requirement that it must be equal to zero in some averaged sense in each of the subdomains D_i of D. To this end we take

$$w_j(\bar{x}) = \begin{cases} 1 & \text{on } D_j \subset D \\ 0 & \text{otherwise} \end{cases}$$

The coefficients $\bar{c}$ are determined out of the conditions

$$(w_j, R) = \int_D R(\bar{c}, \bar{x})\, w_j(\bar{x})\, d\bar{x} = \int_{D_j} R(\bar{c}, \bar{x})\, d\bar{x} = 0$$

for $j = 1(1)n$.

This method was introduced by Biezeno and Koch in 1923 in connection with the He stud

equilibrium of rods, beams and plates. It has been reconsidered (with the name of integral relations method) by Belotserkovskii and Shovshin (1965), where some suggestions are made for the numerical calculation of $\int_{\mathcal{D}_i} R \, d\bar{x}$.

2) The least squares method

We now require that the quadratic average of R be minimized on $\mathcal{D}$:

$$\int_{\mathcal{D}} R^2(C, \bar{x}) \, d\bar{x} = \text{min}$$

In order to satisfy this condition we take

$$w_j = \frac{\partial R}{\partial C_j}$$

and require that

$$(w_j, R) = \int_{\mathcal{D}} R \frac{\partial R}{\partial C_j} \, d\bar{x} = 0$$

The method of least squares has a long history, which goes back to Gauss and Legendre, in the beginning of last century. The mathematical foundation of this method for operator equations is due to Picone (1928) and Mikhlin (1955).

3) The method of moments

R must now be orthogonal to the first n coordinate functions $\psi_1 \dots \psi_n$, which we take as the w_j's:

$$(w_j, R) = \int_D R(\bar{c}, \bar{x})\, \psi_j(\bar{x})\, d\bar{x} \qquad j = 1(1)n$$

A first approximation of this method is known as the von Karman - Pohlhausen integral method: the requirement is that $\int R\, d\bar{x} = 0$ over the total domain D.

4) The Galerkin method

We now put the condition that the residual function R be orthogonal to the subspace generated by the trial functions $\{\varphi_j\}$, $j = 1(1)n$. The coefficients $\bar{c}$ follow from the algebraic system

$$\int_D R(\bar{c}, \bar{x})\, \varphi_j(\bar{x})\, d\bar{x} = 0 \qquad J = 1(1)n$$

Remark 2.2: in this method the projection system and the coordinate system are identical.

5) The collocation method

The solution $\hat{z}$ in the collocation method is obtained by a classical process of linear

interpolation: it is required that the solution be exact at n points $\bar{x}_j$:

$$R(\bar{x}_j) = 0 \qquad j = 1(1)n$$

This is achieved if (f_j, R) is defined by a $\delta(\bar{x} - \bar{x}_j)$ acting on R, ie with $w_j = \delta(\bar{x} - \bar{x}_j)$:

$$\int_{\mathcal{D}} R(\bar{c}, \bar{x})\, \delta(\bar{x} - \bar{x}_j)\, d\bar{x} = R(\bar{c}, \bar{x}_j) = 0$$

for $j = 1(1)n$.

This method was developed by Fraser [1937] in connection with problems on the stability of aeronautical structures. The nodes $\bar{x}_j$ were not clearly specified.

6) The Tau method

In this method we look for a solution $\hat{z}$ which minimizes the residual $R(\bar{c}, \bar{x})$ in the topology of $C[\mathcal{D}]$:

$$\min_{\mathcal{P}_n} \max_{\mathcal{D}} |R(\bar{c}, \bar{x})|$$

To this end we perturbe the original equation in such a way that the new perturbed equation has an exact polynomial solution and the perturbation satisfies a minimum condition.

If we consider the case of an homogeneous ordinary differential equation,

$$D y = 0 \qquad y(x_0) = y_0$$

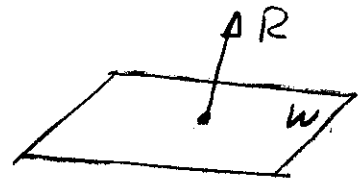
in a compact $[x_0, x_1]$, the problem is equivalent to solve exactly $D y_n^* = h_n$

where h_n is the best uniform approximation of zero in $[x_0, x_1]$ (or an extension of it).

Clearly, the collocation method is a particular (not optimal) case of the Tau method.

Remark 3.2 In the last methods, where we require that R be orthogonal to W:

$$W = \operatorname*{sp}_{j=1(1)n} \{w_j\}$$



we try to push the residual R outside a certain subspace in which the approximation is required. The idea of the proof of the convergence of these methods is based on the completeness of the w_j: R ends up by being orthogonal to all the $\{w_j\}_1^\infty$, then must be equal to zero.

In the case of collocation a careful analysis must be made of the position of the nodes

$\bar{x}_j$ where the error is zero. An simple increase in the number of points at which the $R = 0$ does not necessarily guarantees that the error will go to zero as $n \to \infty$.

In the sub domain method the proof of convergence is based on the fact that $\text{mes}(D_i) \to 0$ as $n \to \infty$.