THE WEIGHTED RESIDUAL METHOD AND VARIATIONAL TECHNIQUE IN THE SOLUTION OF DIFFERENTIAL EQUATIONS

by

K. K. Sen

University of Singapore

1. Introduction. It is well known that in dealing with problems of engineering and mathematical physics, one is required to build mathematical models of physical situations. These models involve differential equations or integro-differential equations as equations of change, constitutive relations (if any), boundary and/or initial conditions, which together give rise to problems quite often not amenable to exact solutions. In such situations, one is tempted to evolve approximate methods of solutions where there are slightest hopes of good results. In "weighted residual method" and "variational technique" one nurtures this hope. Extensive use has been made of these methods for solving linear and non-linear problems in continuum mechanics, the study of hydrodynamic stability, transport processes etc. of various complexities [1,2]. In what follows, we outline the main features of the above two methods.

2. The weighted residual method

The weighted residual method may be considered to be a unified version of a group of methods used to solve approximately boundary value, initial value and eigen value problems. In this, knowledge of a function of say space and time is sought, given (a) equations of change in the form of differential equations (ordinary or partial) or integro-differential equations (b) constitutive relations (if any) such as equations of state and (c) boundary conditions and/or initial conditions. Basic methodology for solution consists in (a) assuming a trial solution in which time dependence is

kept undetermined, but the functional dependence of the remaining independent variables (sometimes all) is pre-assigned and (b) obtaining the function of time by requiring the trial solution to satisfy the differential equation in some specified sense.

The main features of this method will be demonstrated in the case of initial value problems, though the extension of the technique to boundary value and eigenvalue problems is not beset with any undue difficulty.

Initial value problem

Let $U(\underset{\sim}{X}\ t)$ be the function to be studied. Let the equation of change be

$$L(U) - \frac{\partial U}{\partial t} = 0, \quad \underset{\sim}{X} \in V, \ t > 0, \tag{2.1}$$

where L is a differential operator involving spatial derivatives only and V, a three dimensional domain with boundary S.

For simplicity, we assume that there are no constitutive relations to take care of.

Boundary condition : $U(\underset{\sim}{X},t) = f_s(\underset{\sim}{X},t), \ \underset{\sim}{X}$ on $S \ \forall \ t$ (2.2)

Initial condition : $U(\underset{\sim}{X},0) = U_I(\underset{\sim}{X}), \ \underset{\sim}{X} \in V$. (2.3)

To solve this we assume a trial solution,

$$U_T(\underset{\sim}{X},t) = U_S(\underset{\sim}{X},t) + \sum_{i=1}^{N} C_i(t) U_i(\underset{\sim}{X},t) , \tag{2.4}$$

where the approximating function $U_i(\underset{\sim}{X},t)$ is specified in in such a way that the ordinary condition is satisfied, i.e.

$$U_S = f_S \text{ and } U_i(\underset{\sim}{X},t) = 0 \text{ for } \underset{\sim}{X} \text{ on } S \ \forall \ t. \tag{2.5}$$

Next step is to build up the residuals for differential equation and initial condition as a measure of the extent to which U_T satisfies the equation of change and initial

condition.

$$R(U_T) = L(U_T) - \frac{\partial U_T}{\partial t} \text{ : differential equation residual} \quad (2.6)$$

$$R_I(U_T) = U_I(\underset{\sim}{X}) - U_S(\underset{\sim}{X},0) - \sum_{i=1}^{N} C_i(0)U_i(\underset{\sim}{X},0) \text{ : initial condition residual.} \quad (2.7)$$

As the number of approximating functions N is increased, one hopes that the residual will become smaller and smaller. Residual equals to zero implies exact solution. When, however, the residual is not zero, a "weighted residual" is put equal to zero, and this in essence is the main feature of the method, from which its name is derived [1].

For this we define

$$\langle \omega, v \rangle = \int_V \omega v dv \text{ (spatial average or inner product),} \quad (2.8)$$

and state the approximation as

$$\langle \omega_j, R(U_T) \rangle = 0 \quad (2.9)$$

$$\langle \omega_j, R_I(U_T) \rangle = 0 \quad (2.10)$$

where ω_j (j=1,2...N) are chosen "weighted functions", N in number. Equation (2.9) gives N linear or non-linear simultaneous differential equations in $C_i(t)$ and (2.10) generates the $C_i(0)$ for the preceding differential equations. Once $C_i(t)$ are determined this way, one can obtain the approximate solution by substitution in equation (2.4).

In addition, we have in the field two modified versions of the method, which are given below.

(A) Boundary method: Here trial solutions are selected to identically satisfy the differential equations. R_B and R_I are treated as in (2.9) and (2.10), and in (2.9) spatial

average is replaced by an average over the boundary [3].

(B) Mixed method: In this it is not feasible to choose the trial solution U_T either to satisfy the boundary condition or the differential equations. In such cases, R, R_I and R_B are made orthogonal to different weighting functions [4]. There is some arbitrariness involved in this method in that the number of equations obtained exceeds the number of unknowns. A way of removing this arbitrariness has been suggested by Synder et al [5] through Galerkin method.

It is clear that in the weighted residual method, which we shall call WRM henceforward, the two main problems will be

(a) choice of weighting function ω_j

and (b) choice of approximating functions $U_i(\underset{\sim}{X},t)$.

Choice of weighting functions

Different choices of the weighting functions give different complexions of WRM. Some of these are listed below along with their proper characteristics.

(i) Collocation method [6]

This corresponds to the weighting function, $\omega_j = \delta(\underset{\sim}{X}_j - \underset{\sim}{X})$, where δ is the Kronecker delta. This implies that the differential equation holds exactly at collocation points $\underset{\sim}{X}_j$. As N increases, the residual R vanishes at more and more points and in the limit $N \to \infty$, R is zero throughout the domain.

(ii) Sub-domain method [7]

In this the differential equation is satisfied on the average in each of the N sub-domains V_j, i.e.

$$\omega_j = \begin{cases} 1 & x \in V_j \\ 0 & x \notin V_j \end{cases} \qquad j = 1,2,..N .$$

(iii) Least square method [8]

Here $$\omega_j = \frac{\partial R(U_T)}{\partial C_j}$$

This amounts to the mean square residual

$$I = \int [R(U_T)]^2 dv$$

being minimised, i.e. $\delta I = 0$.

(iv) Galerkin method

This particular method first proposed in 1915 has been extensively used with necessary modifications and generalisations for solving problems of elasticity, hydrodynamic stability, transport processes etc. In this example of WRM, the weighting functions are taken as the approximating functions themselves $\omega_i = U_i(\underset{\sim}{X},t)$ and these $U_i(\underset{\sim}{X},t)$ are the subsets of a complete set of hopefully orthogonal functions. $R(U_T)$ which usually is continuous can vanish in (2.9) only if it is orthogonal to each and every member of the complete system of which U_i are a part. In practice, $R(U_T)$ is made orthogonal only to a finite number of members of the complete set. The generalisation of the method has taken place in various directions. To name a few,

(i) the form of U_T has been generalised to $U_T = f(\underset{\sim}{X},\{C_i(t)\})$ with weighting functions $\omega_j = \frac{\partial f}{\partial C_j}$,

(ii) ω_j has been taken as $\omega_j = K(U_j)$ where K is a specified differential operator,

(iii) residual is made orthogonal to members of a complete set of functions which need not be the same as the approximating function (this is sometimes called moment method).

Allowing for the generalisation and modifications, Galerkin

method seems to be most versatile in applicability amongst the various classes of WRM.

Choice of approximating functions

The choice of approximating functions is indeed the most crucial problem of WRM. For this one needs to exploit every bit of knowledge one has about the physical problem at hand. Information like symmetry properties, decay properties or oscillating behaviour of the system are to be fully utilised. Sometimes the approximating functions U_i are chosen to satisfy the boundary conditions or derived boundary conditions (Krylov and Kantorovich [20]).
It is not possible to give an overall recipe for the selection of approximating functions. Usually several sets of approximating functions are available and an optimal choice depends very much on the foresight and intuition of the user.

Convergence problem: The proof of convergence of the approximating functions is indeed a difficult one for WRM. In fact for Galerkin method, the study of convergence was taken up seriously as late as 1940, while the method was proposed as early as 1915. Since then good progress has been made in the proof of convergence of Galerkin method for eigenvalue problems involving ordinary differential equations, second order elliptic differential equations, hyperbolic differential equations occuring in hydrodynamic stability problems and Navier-Stokes equations with time dependence. For other WRM, the convergence proofs are rather rare except perhaps for some cases of least square method and to a lesser extent for the collocation method. For non-linear problems, very little is known about the convergence.

However, as Ames has pointed out, for most physical and engineering problems, computation of error bound is more useful than the actual proof of convergence of the approximating function. In fact more attention is paid to this aspect of the problem.

3. Variational Methods

In the variational description of a physical system, it is stipulated that the state of the system is specified by a statement that variation or differential of a particular functional is equal to some fixed value (usually zero). This description is completed by giving (a) the functions with respect to which variation is taken and (b) constraints as auxiliary conditions. The Euler-Lagrange equations to which the stationary value of functional leads to are the equations of change.

The method consists in expressing the functional in the stationary state (sometimes called dynamical potential) in terms of a number of adjustable parameters. Then it is varied with respect to these parameters and finally they are so evaluated as to make the variations vanish. We take an example to demonstrate this scheme which is now known as Ritz method. Let the functional or dynamical potential be defined by

$$I = \int_0^t \mathcal{F}(x_i(t), \dot{x}_i(t), t)dt, \tag{3.1}$$

then $\delta I = 0$, leads to Euler-Lagrange equations

$$\frac{d}{dt}\frac{\partial \mathcal{F}}{\partial \dot{x}_i} - \frac{\partial \mathcal{F}}{\partial x_i} = 0 \quad . \tag{3.2}$$

The equations (3.2) are the equations of change.

To solve the problem we assume a trial solution,

$$x_T = x_S(t) + \sum_{i=1}^{N} C_i x_i(t) \tag{3.3}$$

where x_S satisfies the non-zero boundary specification, and x_i are the members of a complete set which vanish at any boundary where x(t) is specified. Now substituting (3.3) in (3.1) and integrating, I becomes a function of C_i

only.

Now putting $\frac{\partial I}{\partial C_i} = 0$, (3.4),

we obtain N algebraic equations for determining C_i. Substituting these in (3.3), an approximate solution is obtained, the accuracy of which can be augmented by successive approximations.

In Galerkin method, one deals directly with Euler-Lagrange equations. One substitutes the approximate solution of type (3.3) in the Euler-Lagrange equation and obtains the residual of the equation

$$\frac{\partial \mathcal{F}_{ap}}{\partial x} - \frac{d}{dt}\frac{\partial \mathcal{F}_{ap}}{\partial \dot{x}} = R_T(C_1, C_2, \ldots, C_N, t) \quad (3.5)$$

for each value of t.

Then according to the scheme suggested by Galerkin, one can write

$$\int_0^\theta R_T \, x_i(t)dt = 0, \; i = 1,2\ldots N \quad (3.6)$$

i.e. R_T is made orthogonal to $x_i(t)$.
This gives us the necessary number of equations to determine C_i.

4. Comments

In the above, we have demonstrated the use of Ritz method and Galerkin method for solving a particular class of problems. We have seen that in actuality, Galerkin method starts a step later than the variational method (Ritz method) in dealing with Euler-Lagrange equations rather than dynamic potentials. Then the question arises: Given the equations of change and boundary and/or initial

conditions

(a) Will it always be possible to have a dynamic potential to generate the variation principle?

(b) If not, is it worthwhile to attempt the construction of some non-unique, restricted potential?

In answer to (a), we can make the following deposition [9].

(i) The construction of dynamical potential is always possible only when the descriptive equations are linear and self-adjoint.

(ii) For linear non self-adjoint problems, a variational formulation is possible where the original problem and their adjoints are inextricably coupled. On analysis, this method appears to be a generalised version of Galerkin method.

(iii) By applying an adjoint operator, the order of the differential equations could be doubled and the functional be constructed out of the square of the residual, which is minimised. This in essence is the method of least squares and shares with it the limited applicability - the limitation almost breaking down in the case of problems of transient performance.

On the other hand, if the problem falls under the category (b), i.e. if no functional or potential exists whose stationary property leads to the description of the dynamical system, one proposes to use [9]

(i) quasi variational principle

or (ii) restricted ariational principle.

In (i) [10], rather than a functional, a functional differential or a variation is defined, the vanishing of which gives the equation or equations of change

$$\hat{\delta} I = \int_0^t \int_V \left[L(U) - \frac{\partial U}{\partial t} \right] \delta U \, dV = 0 \qquad (4.1)$$

Not uncommonly, integration over time is omitted.

In (ii), i.e. in restricted variational principle, u and $\frac{\partial u}{\partial t} = y$ are treated as independent functions and

$$\overline{\delta_y J} = \int_V \left[L(u) - y \right] \delta u \, dv = 0 \, , \qquad (4.2)$$

y is kept constant during the variation.

In both (4.1) and (4.2), one adheres to the equation of change and the question of existence of a dynamical potential is thrown into the background. Variational integral need not be stationary here. As Serrin [12] has pointed out, the construction of the functional in this way is at best a reformulation of the equation of change. In such situations, the adaptation of WRM like Galerkin method may be an easier alternative. When the equations of change are given, the process of going back to the construction of a functional looks like an unnecessary exercise. Examples of such pseudo-variational principle lie in the so-called "Method of local potentials" put forward by Glansdorff and Prigogine [13] or Lagrangian Thermodynamics advocated by Biot [14]. Both schemes are being vigorously followed during the last two decades or so. The basis of both these methods is the principle of minimum entropy production during irreversible processes. However, this minimum principle has not been firmly established up to this day. Onsäger hypothesis on which this principle is based gives the linear relation between flux and force and can at best give us a constitutive relation. This cannot play the role of the equation of change. Thus the hypothesis stands on a rather uncertain ground as Truesdel [15] has suggested.

However, the variational principle may have some additional advantage over WRM under the following circumstances.

(a) The variational integral represents a well defined physical quantity like total energy or entropy whose behaviour during the change is well known.

(b) The variational principle represents an actual maximum or minimum principle giving the upper or the lower bound.

(c) A dual variational principle exists i.e. both upper and lower bounds can be formulated. The closeness of the two can be exploited to give the solution.

(d) A mathematical model is obtained in the form of an integral, which forms the fundamental problem of the calculus of variation. Sometimes, in such cases, one can prove the existence of a solution.

(e) One is interested in the study of the stability of systems which deals with perturbed equations of change, constitutive relation and boundary and/or initial condition.

5. Some applications in radiative transfer

Finally we draw attention to some applications of the methods in which we were interested.

(a) About fifteen years ago, faced with the problems of solving the integro-differential equation of transfer under appropriate boundary conditions, we developed what is called a double-interval spherical harmonic method [16,17]. Part of this, on close analysis, can be classified as a modified Galerkin method.

The transfer equation giving the variation of specified intensity $I(\tau,\mu)$ with optical depth τ in a plane parallel isotropic scattering atmosphere is given by

$$\frac{\partial I(\tau,\mu)}{\partial \tau} = I(\tau,\mu) - J(\tau) \qquad (5.1)$$

where $J(\tau)$, the mean intensity, $= \frac{1}{2}\int_{-1}^{+1} I(\tau,\mu)d\mu$. (5.2)

$\mu = \cos\theta$, θ being the inclination of $I(\tau,\mu)$ to the outward drawn normal to the plane parallel atmosphere.

Equation (5.1) is to be solved under the boundary conditions

$$\left.\begin{aligned} I(\tau,0) &\equiv 0 \qquad \text{for } -1 \leq \mu \leq 0 \\ I(\tau,\mu)e^{-\tau} &\to 0 \qquad \text{as } \tau \to \infty \end{aligned}\right\} \tag{5.3}$$

Modified Galerkin scheme was utilised as follows.

$I(\tau,\mu)$ was represented as $I^{+}(\tau,\mu)$ for $\mu \in (0,1)$ and $I^{-}(\tau,\mu)$ for $\mu \in (-1,0)$, and the trial solutions were taken in the form of an expansion in finite series in Legendre polynomials as

$$\begin{aligned} I^{+}(\tau,\mu) &= A\tau + \sum_{m=0}^{m}(2m+1)I_m^{+}(\tau)\mu P_m(2\mu-1) \\ I^{-}(\tau,\mu) &= A\tau + \sum_{m=0}^{m}(2m+1)I_m^{-}(\tau)\mu P_m(2\mu+1) \end{aligned} \tag{5.4}$$

Substitution of (5.4) in (5.1) gave us residuals R^{+} and R^{-}. Using the recurrence formula for $\mu P_m(2\mu\pm1)$ and choosing the weighting function ω_m orthogonal to R^{+} and R^{-} (in this case $\omega_m = P_m(2\mu\pm1)$ in the respective ranges) and integrating over the relevant range of μ, one could obtain a finite number of ordinary differential equations in $I_m^{+}(\tau)$ and $I_m^{-}(\tau)$. These equations whose number depended on the truncation point of the trial solution could be solved exactly. The results obtained were comparable in accuracy with exact solutions even with $m = 2$.

(b) Recently in an attempt to study the dispersion relation and stability for a time dependent radiative transfer problem in plane parallel homogeneous medium in local thermodynamic equilibrium, we used a variational technique [18].

The physical process involved was described by the transfer equation and cooling rate equation containing space and time variations of specific intensity $I(x,\mu,t)$ and temperature $T(x,t)$ given by

$$\frac{1}{c}\frac{\partial I(x,\mu,t)}{\partial t} + \mu\frac{\partial I(x,\mu,t)}{\partial x} = \chi(x,t)\left[\beta T^4 - I(x,\mu,t)\right] \tag{5.5}$$

and

$$c_v\frac{\partial T(x,t)}{\partial t} = 4\pi\chi(x,t)\left[J(x,t) - \beta T^4\right] \tag{5.6}$$

where $\chi(x,t)$ is the absorption coefficient, c_v the coefficient of specific heat per unit volume and $J(x,t)$ the mean intensity defined by (5.2) and c the velocity of light.

With $\chi dx = d\tau$ and $\chi c\, dt = dy$, (5.7)

$$\frac{\partial I}{\partial y} = -\mu \frac{\partial I}{\partial \tau} + (\beta T^4 - I) \tag{5.8}$$

$$\alpha \frac{\partial T}{\partial y} = J - \beta T^4 \tag{5.9}$$

where $\alpha = (cC_v/4\pi)$,

variational formulation was developed by defining

$$\begin{aligned} T &= T^*(\tau,y) + \delta T(\tau,y) \\ \text{and} \qquad & \qquad (5.10) \\ I &= I^*(\tau,\mu,y) + \delta I(\tau,\mu,y) \end{aligned}$$

where I^*, T^* define an unperturbed dynamical equilibrium state and δI and δT are arbitrary variations.

Then dealing with the perturbed equations and retaining only the first order terms in δI and δT, we obtain a set of two linear equations in terms of which one can build a functional L given by

$$L = \int_y \int_\tau \left[\left\{ \frac{\partial I^*}{\partial y} + \mu \frac{\partial I^*}{\partial \tau} - \beta T^{*4} \right\} I + \tfrac{1}{2} I^2 + \left\{ \alpha \frac{\partial T^*}{\partial y} - J^* \right\} \alpha T + (\alpha\beta/5) T^5 \right] d\tau dy \tag{5.11}$$

The extremal of this functional with the subsidiary conditions $I = I^*$, $T = T^*$ gives the time-dependent equations of change for the unperturbed state. The study of this functional L by a procedure suggested by Schechter and Himmelblau [19] leads us to the dispersion equations and the conditions of stability.

References

1. Crandall S. H. Engineering Analysis. McGraw Hill, (1956).

2. Finlayson B. A. & Scriven L. E. App. Mech. Rev. 19, 735, (1966) 187 references.

3. Collatz L. The Numerical Treatment of Differential Equations. Springer-Verlag, Berlin Gottingen - Heidelberg, 1960 (third edition).

4. Schuleshko P. Austr. Jour. Appl. Sc. 10, 1-7, 8-16, (1959).

5. Snyder L. J., Spriggs T. W. and Stewart W. E. AICHE J. 10, 535 (1964).

6. Frazer R. A., Jones W. P. and Skan W. Approximations to functions and to the solutions of differential equations. Gt. Brit. Aers. Res. Council Report and Memo. 1799 (1937).

7. Biezeno C. B. and Koch J. J. Ing. Grav. 38, 25 (1923).

8. Picone M. Rend. Circul. Mat. Palerno 52, 225 (1928).

9. Finlayson B. A. & Scriven L. E. Int. Jour. Heat & Mass Transf 10, 799 (1967) 50 references.

10. Robert P. H. Non-equilibrium Thermodynamics, Variational Techniques and Stability. University of Chicago Press, p130, (1965).

11. Rosen P. J. Chem. Phys. 21, 1220, (1953)
J. App. Phys. 25, 336, (1954).

12. Serrin J. Handbook der Physik, Vol. 8, Part I, Springer-Verlag, Berlin (1959).

13. Glansdorff P. and Prigogine I. Physica 30, 351 (1964).

14. Biot M. A. Seventh Anglo-American Aeronautreal Conference, Inst. Aero. Sc. p 418, (1960).

15. Truesdel C. Rational Thermodynamics. McGraw Hill Book Company (1969)

16. Wilson S. J. and Sen K. K. Publ. Astro. Soc. Japan, 15, 351 (1962)

17. Wilson S. J. and Sen K. K. Ann. d'Astrophysique 27, 46 (1964); 28, 348 (1965) 28, 855 (1965)

18. Wilson S. J. and Sen K. K. Astron. & Astrophys. 44, 377, 1975

19. Schechter R. S. and Himmelblau, D. M. Phys. Fluids 8, 1431 (1965)

20. Krulov and Kantorovich. Approximate Methods of Higher Analysis. Interscience Publishers. Inc. (1958).