Lecture 6
Boundary Value Problems of ODE

Many important problems in mathematical physics reduce to second order ODEs of the form

$$Lu = -\frac{d}{dx}\left(p(x)\frac{du}{dx}\right) + q(x)u = f(x), \tag{1}$$

that must be integrated in an interval $a \leq x \leq b$ under the boundary conditions

$$h_1 u'(a) - h_2 u(a) = 0, \qquad H_1 u'(b) + H_2 u(b) = 0, \tag{2}$$

where $h_1, \quad h_2, \quad H_1, \quad H_2$ are constants. This is to be said a boundary value problem for the second order ODE. The simplest problem of this type is the problem of the static sag of a wire that, in the equilibrium state, occupies the segment $a \leq x \leq b$ of the x axis. The element $(x, x + dx)$ is under the influence of a restoring force $q(x)udx$ and an external force $f(x)dx$, and tensions $p(x)$ and $p(x+dx)$ are applied to the ends x and $x+dx$ of the element, respectively. In the general case, the boundary conditions (2) correspond to elastically fixed wire ends. In the special case in which, for example, $h_1 \neq 0$, $h_2 = 0$, the end $x = a$ of the wire is free; if, however, $h_1 = 0$, $h_2 \neq 0$, this end is rigidly fixed.
This equation also describes small stationary longitudinal vibrations of elastic rod. In this case $p = ES$ and $q = -\omega^2 \rho S$, where $E(x)$ is the Young modulus, $S(x)$ is the square of the cross section of the rod, ω is frequency, ρ is density, $h_1 = H_1 = E$ and $h_2 = H_2$ is the stiffness coefficient according to Hook low.
The general form of the boundary value problem looks like this

$$y'' = f(x, y, y'), \qquad y(a) = A, \qquad y(b) = B, \qquad a \leq x \leq b. \tag{3}$$

In general, we cannot assure the existence of a solution to (3) simply assuming that $f(x_1, x_2, x_3)$ has good properties. However, for the following case

$$-y'' = f(x, y), \qquad y(0) = A, \qquad y(1) = B, \qquad 0 \leq x \leq 1, \tag{4}$$

we can formulate the **Theorem**: the problem (4) has a unique solution in $C[0, 1]$ if $f(x, y)$ satisfies the following inequality

$$|f(x, y_1) - f(x, y_2)| \leq k|y_1 - y_2|, \qquad 0 < k < 1,$$

in the infinite strip $0 \leq x \leq 1$, $-\infty < y < +\infty$.

The proof is based on the Green's function integral representation

$$y(x) = c_1 + c_2 x + \int_0^1 g(x,t) f(t, y(t)) dt,$$

where c_1 and c_2 are chosen in such a way so that $y_0(x) = c_1 + c_2 x$ satisfies the boundary conditions in (4), and $g(x,t)$ is the Green's function

$$g(x,t) = \begin{cases} x(1-t), & \text{if} \quad 0 < x < t < 1, \\ t(1-x), & \text{if} \quad 0 < t < x < 1. \end{cases}$$

Shooting Method

The boundary value problem being considered is

$$y'' = f(x, y, y'), \qquad y(a) = A, \qquad y(b) = B, \qquad a \leq x \leq b. \tag{5}$$

One natural way to attack this problem is to solve the related initial-value problem, with a guess as to the appropriate initial value $y'(a)$. Then we can integrate the equation to obtain the approximate solution, hoping that $y(b) = B$. If not, then the guessed value of $y'(a)$ can be altered and we can try again. The process is called **shooting**, and there are ways of doing it systematically.

Let us denote the guessed value of $y'(a)$ by z, so that the corresponding initial value problem is

$$y'' = f(x, y, y'), \qquad y(a) = A, \qquad y'(a) = z, \qquad a \leq x \leq b. \tag{6}$$

The solution of this problem is $y = y(x, z)$. The objective is to select z so that $y(b, z) = B$. We put

$$\phi(z) = y(b, z) - B,$$

so that our objective is simply to solve the equation $\phi(z) = 0$ for z. The Newton's method is applicable, and for iterations of z we have

$$z_{n+1} = z_n - \frac{\phi(z_n)}{\phi'(z_n)}.$$

To determine $\phi'(z)$, we introduce

$$u(x, z) = \frac{\partial y}{\partial z}$$

and differentiate with respect to z all the equations in (6). This becomes

$$u'' = f_y(x, y, y')u + f_{y'}(x, y, y')u', \qquad u(a) = 0, \qquad u'(a) = 1.$$

The last differential equation is called the first variational equation. It can be solved , for example, by multistep method and, then,

$$\phi'(z) = u(b, z).$$

This enables us to use the Newton's method to find a root of ϕ. The corresponding algorithm works in the following way. First, we obtain the solution $y = y(x, z_0)$ of

$$y'' = f(x, y, y'), \qquad y(a) = A, \qquad y'(a) = z_0, \qquad a \leq x \leq b.$$

Then, substituting $y = y(x, z_0)$ and $y' = y'(x, z_0)$ into the differential equation

$$u'' = f_y(x, y, y')u + f_{y'}(x, y, y')u',$$

we solve the corresponding initial value problem with $u(a) = 0, \qquad u'(a) = 1$. Thus, we have $\phi(z_0) = y(b, z_0) - B$ and $\phi'(z_0) = u(b, z_0)$. Hence, we know z_1, and so on.
An elaboration of the shooting method is called **multiple shooting**. The basic strategy here is to divide the given interval $[a, b]$ into subintervals and attempt to solve the global problem in pieces. Let us describe what would be done if the interval were divided into just two parts, $[a, c]$ and $[c, b]$. On the two subintervals, we solve initial value problem to obtain two functions y_1 and y_2:

$$y'' = f(x, y, y'), \qquad y(a) = A, \qquad y'(a) = z_1, \qquad a \leq x \leq c,$$

$$y'' = f(x, y, y'), \qquad y(b) = B, \qquad y'(b) = z_2, \qquad c \leq x \leq b.$$

Notice that z_1 and z_2 are parameters at out disposal. The function y_1 is required only on the interval $[a, c]$, and y_2 is required only on the interval $[c, b]$. The parameters z_1 and z_2 are now to be adjusted until the piecewise function

$$y(x) = \begin{cases} y_1(x), & \text{if } \ a < x < c, \\ y_2(x), & \text{if } \ c < x < b \end{cases}$$

becomes a solution to the problem. Thus, we require continuity in $y(x)$ and $y'(x)$ at the point c: $y_1(c) - y_2(c) = 0$, $y_1'(c) - y_2'(c) = 0$. These two conditions are to be fulfilled by an adroit choice of z_1 and z_2. This would be done by Newton's method in dimension 2. Multiple shooting with k subintervals will involve k sub-functions.

Finite-Difference Method

We now consider a boundary value problem for second order linear ODE:

$$y'' + p(x)y' + q(x)y = f(x), \qquad y(a) = A, \qquad y(b) = B, \qquad a \le x \le b. \tag{7}$$

Another approach to handle the problem consists of the discretisation of the x-interval followed by the use of approximate formulas of finite differences for derivatives

$$\begin{aligned} y'(x_i) &= \frac{y_{i+1} - y_{i-1}}{2h} - \frac{h^2}{6} y'''(\xi_1), \\ y''(x_i) &= \frac{y_{i+1} + y_{i-1} - y_i}{h^2} - \frac{h^2}{12} y^{(4)}(\xi_2), \end{aligned} \tag{8}$$

where $x_i = a + ih$, $i = 0, 1, 2, ..., n+1$, $h = (b-a)/(n+1)$. Thus, the discrete version of (7) is then the system of linear algebraic equations

$$y_{i-1}(hp_i - 2) + y_i(4 - 2q_i h^2) + y_{i+1}(-hp_i - 2) = -2f_i h^2, \qquad y_0 = A, \qquad y_{n+1} = B, \tag{9}$$

where $p_i = p(x_i)$, $q_i = q(x_i)$, $f_i = f(x_i)$. For example, if $n = 4$, the corresponding matrix form is given by

$$\begin{pmatrix} 4 - 2q_1h^2 & -hp_2 - 2 & 0 & 0 \\ hp_1 - 2 & 4 - 2q_2h^2 & -hp_3 - 2 & 0 \\ 0 & hp_2 - 2 & 4 - 2q_3h^2 & -hp_4 - 2 \\ 0 & 0 & hp_3 - 2 & 4 - 2q_4h^2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \end{pmatrix}$$

with

$$F_1 = -2f_1h^2 - A(hp(a) - 2), \qquad F_2 = -2f_2h^2,$$
$$F_3 = -2f_3h^2, \qquad F_4 = -2f_4h^2 - B(hp(b) - 2).$$

This system is tridiagonal and can be solved by a special Gaussian algorithm. In general case, we have $n \times n$ system $Ty = F$, where T is the tridiagonal $n \times n$ matrix, $y = (y_1, ..., y_n)^T$ and $F = (F_1, ..., F_n)^T$.

In the error analysis, using approximations (8), for the error vector $(e_1, ..., e_n)$, where $e_i = y(x_i) - y_i$, $i = 1, ..., n$ and $y(x)$ is assumed to be the exact solution, we obtain the system

$$Te = h^4 G$$

with constant vector $G = (g_1, ..., g_n)^T$ in the right-hand side. If $\det T \neq 0$, we have

$$e = h^4 T^{-1} G,$$

and hence,

$$||e|| \leq h^4 ||T^{-1}|| \cdot ||G||.$$

Thus, if $h \to 0$, then $||e|| \to 0$.

Ritz method

Ritz, Galerkin, Square Least methods are used widely on problems in which it is required to determine an unknown function. Of course, boundary value problems for differential equations are in this category. Suppose we are confronted with a problem of the form

$$Lu(x) = f(x) \tag{10}$$

in which L is the linear operator

$$Lu = -\frac{d}{dx}\Big(p(x)\frac{du}{dx}\Big) + q(x)u = f(x),$$

$f(x)$ is a given function, and $u(x)$ is is the function to be determined from the equation and boundary conditions

$$u(a) = 0, \qquad u(b) = 0.$$

We assume that $p(x) \geq p_0 > 0$, $q(x) \geq 0$ for $x \in [a, b]$. This means that operator L is symmetric and positive definite in the real Hilbert space $L_2[a, b]$ since, using the integration by parts, we have

$$< Lu, u >= \int_a^b Lu \cdot u dx = -\int_a^b (pu')' u dx + \int_a^b qu^2 dx = \int_a^b p(u')^2 dx + \int_a^b qu^2 dx \geq 0$$

for arbitrary u. The quantity $< Lu, u >$ is called the energy of the element u relative the operator L.

Consider a linear quadratic functional

$$J(u) =< Lu, u > -2 < f, u >= \int_a^b p(u')^2 dx + \int_a^b qu^2 dx - 2\int_a^b f u dx. \tag{11}$$

A theorem states that if the equation $Lu(x) = f(x)$ has a solution u_0, this solution minimize the functional $J(u)$. Conversely, if there exists an element u_0 that minimizes the functional

$J(u)$, this element satisfies the equation $Lu(x) = f(x)$. The proof is based on the possibility to reduce the functional $J(u)$ to the form

$$J(u) =< L(u - u_0), u - u_0 > - < Lu_0, u_0 >,$$

and analysing of the function

$$g(t) = J(u_0(x) + tv(x)) =< Lu_0, u_0 > +2t < Lu_0, v > +t^2 < Lv, v >$$

$$-2 < f, u_0 > -2t < f, v > .$$

The basic idea of Ritz method consists in replacing the boundary value problem for $Lu(x) = f(x)$ by the problem of minimazing the functional $J(u)$. Suppose we select basis functions $u_1, u_2, ..., u_n, ...$ in $L_2[a, b]$. Hence, we seek solution in the form

$$u_n(x) = \sum_{m=1}^{n} c_m u_m(x), \tag{12}$$

where $c_1, c_2, ..., c_n$ are unknown constants. Substituting this sum into the functional $J(u) =< Lu, u > -2 < f, u >$, we obtain

$$J(u_n) = J(c_1, c_2, ..., c_n) = \sum_{m,k=1}^{n} c_m c_k < Lu_m, u_k > -2 \sum_{m=1}^{n} c_m < f, u_m >, \tag{13}$$

which is quadratic form with respect to the unknown constants $c_1, c_2, ..., c_n$. Looking for a minimum element of $J(u_n) = J(c_1, c_2, ..., c_n)$, we require that

$$\frac{\partial}{\partial c_m} J(c_1, c_2, ..., c_n) = 0, \qquad m = 1, 2, ..., n.$$

Thus, we come to the linear $n \times n$ system of algebraic equations for the unknown constants $c_1, c_2, ..., c_n$

$$\sum_{m=1}^{n} c_m < Lu_m, u_k >=< f, u_k >, \qquad k = 1, 2, ..., n, \tag{14}$$

which is called Ritz system. After we have found the unknown constants $c_1, c_2, ..., c_n$, the expression (12) gives us an approximate solution.

Consider the following example:

$$-u'' = \cos x, \qquad u(0) = 0, \qquad u(\pi) = 0.$$

This boundary value problem has the exact solution $u = \cos x + 2x/\pi - 1$. Let us solve this problem using Ritz method. First we choose the basis functions

$$\sin 2x, \quad \sin 4x, ..., \quad \sin 2nx$$

as the solution is symmetric with respect to $x = \pi/2$. Hence, we seek solution in the form

$$u_n(x) = \sum_{m=1}^{n} c_m \sin 2mx,$$

For the matrix elements $< Lu_m, u_k >$ in the system (14) we have

$$< Lu_m, u_k >= \int_0^{\pi} 4m^2 \sin 2mx \sin 2kx dx = \begin{cases} 0, & \text{if} \quad m \neq k, \\ 2\pi m^2, & \text{if} \quad m = k. \end{cases}$$

Then,

$$< f, u_m >= \int_0^{\pi} \cos x \sin 2mdx = \frac{1}{2}\int_0^{\pi} \sin(2m-1)xdx +$$
$$\frac{1}{2}\int_0^{\pi} \sin(2m+1)xdx = \frac{4m}{4m^2-1}.$$

Hence, the Ritz system is given by

$$\begin{pmatrix} 2\pi 1^2 c_1 & 0 & \dots & 0 \\ 0 & 2\pi 2^2 c_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 2\pi n^2 c_n \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \dots \\ c_n \end{pmatrix} = \begin{pmatrix} \frac{4}{4\cdot 1^2-1} \\ \frac{4\cdot 2}{4\cdot 2^2-1} \\ \dots \\ \frac{4n}{4\cdot n^2-1} \end{pmatrix}$$

Thus, we obtain

$$c_m = \frac{2}{\pi m(4m^2-1)},$$

and approximation to the solution is given by

$$u_n = \frac{2}{\pi}\sum_{m=1}^{n} \frac{\sin 2mx}{m(4m^2-1)}.$$

In the case $n \to +\infty$, we obtain the infinite Fourier series which converges to the exact solution $u = \cos x + 2x/\pi - 1$ in the real Hilbert space $L_2[a, b]$.