Review:

In Lecture 2 we described two simple finite difference methods to solve a second order differential equation: The Euler method, with error of the order of $O(h^2)$, and the Euler Cromer method with error of order $O(h^3)$. As an example we applied it to the equation for the angle $\theta(t)$ of a simple pendulum as a function of the time $\bar{t}$, $\theta'' = d^2\theta/d\bar{t}^2 = -\sin(\theta)$. Time $\bar{t}$ is a dimension-less quantity, that measures time in units of $\sqrt{\ell/g}$, as described in the homework assignment, where ℓ is the length of the pendulum, and g the acceleration of gravity, 9.8 m/s^2.
It is interesting to note that the equation above is non-linear, but the finite difference methods are able to handle this. Instead of $\sin(\theta)$, the term could also have been $1/\theta^3$, for example. But in this case there would have been a singularity at the origin. So, why do people not use this type of method for non

linear equations in general?
Answer: the algorithm error $O(h^3)$ (or $O(h^6)$ for the Numerov method for linear eqs.) accumulates errors very fast. That in turn requires to make h very small, and as a result the accumulation of round-off errors soon overwhelms the total error. In addition, the method is numerically slow, and a different method, that expands the solution in a set of basis functions becomes preferable. This method is the subject of the present lecture.

Galerkin and Collocation Methods [4]

Assume that the equation to be solved for the function $u(x)$ is

$$Lu = f \tag{1}$$

where L is a linear operator, either in differential or integral form, the function $f(x)$ is given, and the independent variable x is contained in some interval $[a, b]$.

A common method to solve for u is to expand it in terms of a complete (and hopefully orthogonal) set of basis functions $\phi_i(x)$, $\quad i = 1, 2, ..N,\ N+1, ..\infty$, and solve for the expansion coefficients c_i. However, the expansion has to be truncated at some upper limit N, thus inntroducing an algorithm error, and hence the result, $u^{(N)}$

$$u^{(N)}(x) = \sum_{i=1}^{N} c_i\ \phi_i(x), \quad a \le x \le b \tag{2}$$

is only an approximation to the exact solution u. The aim is to minimize the error, called remainder $\mathcal{R}$

$$L\ u^{(N)}(x) - f(x) = \mathcal{R}^{(N)}(x). \tag{3}$$

In the limit $N \to \infty$, $u^{(N)} \to u$, and $\mathcal{R}^{(N)} \to 0$. Please note that at this point x is a continuous variable, and the size of $\mathcal{R}$ may not be uniform (i.e.,limited by an upper limit for all x within the interval $[a, b]$), and hence integrals over this interval are employed. This is however not the case for the collocation methods, where integrals are not performed explicitly, as described below.

0.1 The Galerkin Method

In one of the simplest forms of the Galerkin method, the overlap integral $< \chi_i|\mathcal{R} >$

$$< \chi_i|\mathcal{R} > = \int_a^b \chi_i(x)\rho(x)\mathcal{R}(x)\ dx \tag{4}$$

over any of the set of auxiliary basis functions $\chi_i\,(x)$ is considered, and is set to zero. In the integral above ρ is a positive weight function that depends on the type of integration being performed. By multiplying both sides of Eq. (3) with χ_j , making use of the expansion (2), remembering the linearity of the operator L, and after integrating the result over the interval $[a, b]$ one finds

$$\sum_{i=1}^{N} L_{ji}\ c_i - F_j = < \chi_j|\mathcal{R} > = 0, \quad j = 1, 2, ..N \tag{5}$$

where

$$L_{ji} = < \chi_j|L\phi_i >, \quad \text{and } F_j = < \chi_j|f > . \tag{6}$$

Eq. (5) is a matrix equation, and the whole expansion procedure (2) is a discretization of the operator L. That discretization depends on the choice of the basis set $\{\phi_i\}$. The matrix L_{ji} is a square matrix, and hopefully it admits an inverse, with not too large a numerical error. That error is described by a condition number C, which we may have time to discuss later on.

Some comments are useful here.

1. If the ϕ_i are solutions of a part $\bar{L}$ of the operator L, i.e., $\bar{L}\ \phi_i = \lambda_i\phi_i$, where λ_i are the discrete bound-state eigenvalues that depend on the appropriate boundary conditions of the ϕ_i, if one replaces the χ_i by the ϕ_i, and if one keeps only one expansion function ϕ_0 then one obtains the perturbation theory formulation that is very common in physics applications. In this case one finds an improved eigenvalue λ, close to λ_0 by successive iterations, and also finds an improved function ψ that is close to ϕ_0. But we can do much better, see below.

2. If the ϕ_i , $i = 1, 2, ..N$ are Sturmian functions, which are known eigenfunctions of some piece $\bar{L}$ of the operator L that obey the appropriate boundary conditons for the function u in Eq. (1), and if the χ_j are replaced by the Sturmian functions ϕ_j, then the sturmian expansion (2) may converge very rapidly. This approach is made use of in many applications to physics. The main challenge in this case is to obtain a practical method to calculate the sturmian functions and the eigenvalues λ [1].

3. A good choice of the set $\{\phi_i\}$ is crucial for the rapid convergence of the expansion (2), and theorems relating to the size of N required for a desired accuracy (or smallness of $\mathcal{R}$) will be presented further below for the case of spectral expansions. For the finite element procedure, where the whole domain of the independent variable is split into segments (called elements or partitions), and for each partition an expansion of the type (2) is performed, the quantity f may contain the requirement that the solution $u^{(N)}$ from an adjoining previous partition match smoothly to the solution in the next element [2].

4. An important point for the numerical implementation of the Galerkin method is that the choice of the discrete support points in the interval $[a, b]$ is not crucial, other than for the requirement that the integrals be as accurate as possible, if done numerically. For example, equidistant mesh points are needed if Simpson's integration rule is used. If the integrals can be done analytically, then of course no choice of mesh points is required. This is in contrast to the Collocation method, where the choice of mesh points becomes critical.

5. A very useful set of basis functions ϕ_i, $i = 1, 2, ..N$ are Lagrange functions. There are various types of Lagrange functions [3]. For each type a set of N support points ξ_j is defined in the interval $[a, b]$, and each Lagrange function ϕ_i goes through zero at all support points with the exception of ξ_i, where its value is unity. The advantage of these functions is that the integrals f_j (6) can be performed very accurately [3] using Gauss integration methods, requiring only the knowledge of f at ξ_j. These functions are also now used in finite element calculations [5], and an accuracy study is contained in Ref. [2].

0.2 Collocation Method.

In this case a choice of support points ξ_i, $i = 1, 2, ..N$ in $[a, b]$ is required. A direct connection with the Galerkin method can be established by choosing the set of functions χ_i that are used in Eqs. (5) and (6) as Dirac delta functions

$$\chi_i(x) = \delta(x - \xi_i), \tag{7}$$

in which case Eq. (5) becomes

$$\sum_{i=1}^{N} [L\phi_i]_{\xi_j} \; c_i - f(\xi_j) \; = 0, \quad j = 1, 2, ..N. \tag{8}$$

An advantage is that no integrals have to be carried out, and, once the coefficients c_i are obtained from the solution of the matrix equation (8), the value of $u^{(N)}$ can be calculated for any continuous value of x from Eq. (2). However, one difficulty is in finding a good method to establish the location and number of support points that is suitable for a given problem. One way to remedy this difficulty is to use special functions that vanish at a given set of mesh points. Two examples will now be given

Example 1: The equi-distant Fourier mesh [3]: For a given value of N, the mesh points are equidistant and given by

$$\xi_i = i, \quad i = -\frac{1}{2}(N-1), -\frac{1}{2}(N-2) ... \frac{1}{2}(N-1), \tag{9}$$

and the corresponding Lagrange-Fourier functions are [6]

$$\phi_i(x) = \frac{\sin(\pi(x - \xi_i))}{N \sin(\frac{\pi}{N}(x - \xi_i))}. \tag{10}$$

Example 2: a) Lagrange Interpolation functions [5], [2]: These are polynomials all of the same order $N-1$

$$\mathcal{L}_i(x) = \prod_{k=1}^{N} \frac{x-\xi_k}{\xi_i-\xi_k},\ k \neq i;\quad i = 1, 2, ...N, \tag{11}$$

and the mesh-points are Lobatto points.
b) An alternative choice [4] is to use a sequence of a particular orthogonal polynomial $P(x)$, for example Legendre or Chebyshev. In contrast to example a) the order of each polynomial increases from 0 to $N-1$, and $\phi_i(x) = P_{i-1}(x)$, $i = 1, 2, ..N$. The support points ξ_k, $k = 1, 2, ..N$ are the zeros of the polynomial $P_N(x)$.For the Chebyshev case, they can be obtained by means of simple trigonometric expressions, as will be seen.

For Chebyshev Polynomials $\phi_i(x) = T_{i-1}(x)$, $i = 1, 2, ..N$, the discrete orthogonality holds

$$\frac{\pi}{N}\sum_{k=1}^{N} T_n(\xi_k)\ T_m(\xi_k) = \frac{\pi}{2}\delta_{n.m}(1+\delta_{0n}) \quad n < N,\ m < N \tag{12}$$

where the ξ_k are the zero's of T_N, given by

$$\xi_k = \cos[\frac{\pi}{N}(k-1/2)],\quad k = 1, 2, ...N. \tag{13}$$

For examples 2a and 2b the support points are not equi-spaced, which, as we will see during the discussion of spectral methods, gives rise to a higher accuracy in the expansion (2) than if the points are equi-spaced. An important additional feature is the use of the Gauss integration expression

$$\int_a^b \psi(x)\rho(x)dx = \sum_{k=1}^{N} w_k\psi(\xi_k) \tag{14}$$

where $\rho(x)$ is the weight function appropriate for each type of orthogonal polynomial, and the weights w_k are obtained by solving the linear system

$$\sum_{k=1}^{N} w_k(\xi_k)^n = \int_a^b x^n\rho(x)dx,\ n = 0, 1, ..(N-1). \tag{15}$$

The relation (14) is exact if $\psi(x)$ is a polynomial whose order is not greater than $2N+1$.

In order to obtain the coefficients c_i in Eqs. (5) in the collocation method, one can proceed in two ways: In the first method one uses both for the functions χ_j and the functions ϕ_j the Lagrange functions $\mathcal{L}_j$ defined in Eq. (11). Further, using the vanishing of the $\mathcal{L}_i$ at all mesh points other than ξ_i, together with the Gauss integration expression (14) one obtains again Eq. (8), with the difference that the basis functions and support points are now well defined. In the second

method one can use for both the χ_j and the ϕ_j one of the set of orthogonal polynomials P_{j-1} described in example 2b, and after using Gauss's quadrature, one obtains

$$\sum_{i=1}^{N} M_{ji} c_i = F_j, \tag{16}$$

where

$$M_{ji} = \sum_{k=1}^{N} w_k \phi_j(\xi_k) [L\phi_i]_{\xi_k} \tag{17}$$

and

$$F_j = \sum_{k=1}^{N} w_k \phi_j(\xi_k) f(\xi_k). \tag{18}$$

The difference between Eqs. (8) and Eqs. (16)-(18) is that the former requires the values of the functions at only one support point, line by line, while the latter contain sums over all support points in each line. For that reason the latter is much more "spectral" than the former.

References

[1] G.H. Rawitscher, " Positive energy Weinberg states for the solution of scattering problems", Phys. Rev. C **25**, 2196-2213 (1982); Rawitscher, G., "Iterative solution of integral equations on a basis of positive energy Sturmian functions", Phys. Rev. E **85**, 026701(2012);

[2] J. Power and G. Rawitscher, Phys. Rev. E 86 (2012) 066707.

[3] D. Baye, phys. stat. sol. 243 (2006)1095-1109; D. Baye "The Lagrange-mesh method", Phys. Rep. **565** (2015) 1-107

[4] A. Deloff, Ann. Phys. (NY) **322** (2007) 1373–1419.

[5] T. N. Rescigno and C. W. McCurdy, Phys. Rev. A **62**, 032706 (2000);

[6] Eq. (2.8) on p. 13 of L. N. Trefethen, *Spectral Methods in MATLAB*, (SIAM, Philadelphia, PA, 2000);