Review of Gaussian Quadrature method

Nasser M. Abbasi

Spring 2006 compiled on — Sunday December 31, 2017 at 09:09 PM

1 The problem

To find a numerical value for the integral of a real valued function of a real variable over a specific range over the real line. This means to evaluate

$$I = \int_a^b f(x)\ dx$$

Geometrically, this integral represents the area under $f(x)$ from a to b

2 Solution

We can always approximate the area by dividing it in equal width strips and then sum the areas of all the strips.

In general, there will always be an error in the estimate of the area using this method. The error will become smaller the more strips we use (which implies a smaller strip width). Hence we can write

$$\int_a^b f(x)\ dx = \left(\sum_{i=1}^{N} \Delta x\ f(x_i) \right) + E$$

Where E is the error between the actual area and the approximated area using the above method of numerical integration. N above is the number of strips or can also be refereed to as the number of integration points.

Instead of keep referring to the 'width of the strip' all the time, we will call this quantity the weight w_i that we will multiply the value of the function with to obtain the area. Hence the above becomes

$$\int_a^b f(x)\ dx = \left(\sum_{i=1}^{N} w_i\ f(x_i) \right) + E$$

Using implied summation on indices the above becomes

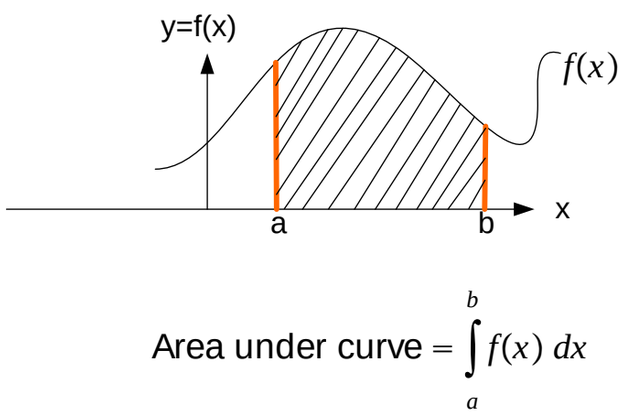


Figure 1: Integrating a function

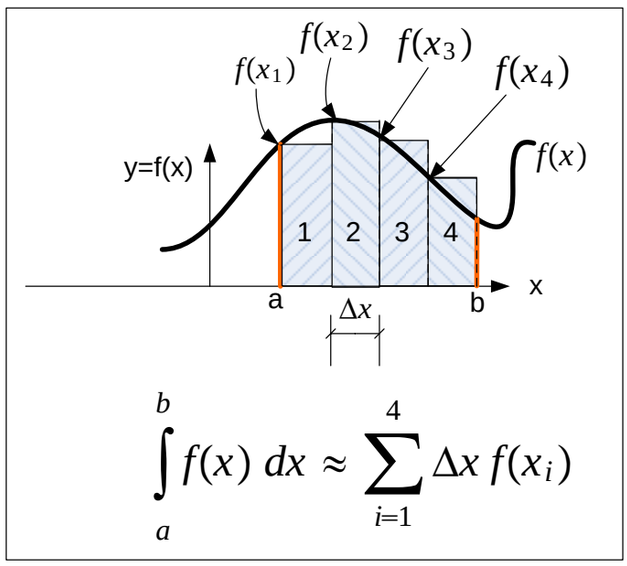


Figure 2: Numerical integration

$$\int_a^b f\left(x\right)\ dx = w_i\ f\left(x_i\right) + E$$

In the above we divided the range of the integration (the distance between the upper and lower limits of integration) into equal intervals. We can decide to evaluate $f(x_i)$ at the middle of the strip or at the start of the strip or at the end of the strip. In the diagram above we evaluated the $f(x)$ at the left end of the strip.

Our goal is to evaluate this integral such as the error E is minimum and using the smallest number of integration points. In a sense this can be considered an optimization problem with constraints: minimize the error of integration using the smallest possible number of points.

To be able to do this minimization, we need to consider what are the variables involved. We see that there are two degrees of freedom to this problem. One is the width of the strip or w_i. We do not have to use a fixed value of width, we can use different width for different strips if the resulting integral gives better approximation.

The second degree of freedom is the point at which to evaluate $f(x_i)$ associated with each strip. In the example above we choose to evaluate $f(x_i)$ at left end point of the strip. We can choose to select a different x_i point if this will result in a better approximation.

This is the main idea of Gauss Quadrature numerical integration. It is to be able to choose specific values for these two degrees of freedom, the w_i and the x_i.

It turns out that if the function $f(x)$ is a polynomial, then there is an optimal solution. There is an optimal $\{w_i, x_i\}$ for each polynomial of order n.

We can determine these degrees of freedom such that the error E is zero, and with the least possible number of integration points. We are able to tabulate these two degrees of freedom for each polynomial of specific order. In other words, if the function $f(x)$ is a polynomial of order n then we know before the computation starts what these 2 degrees of freedom should be. We know the locations of x_i and we know weight w_i that we need to multiply $f(x_i)$ with to obtain the area with minumum error.

You might ask how can this method of integration know the locations of the integration points x_i beforehand without being given the integration range of the function to integrate? It turns out that we will map $f\left(x\right)$ into a new known and specific range of integration (from -1 to $+1$) for the method that we will now discuss.

2.1 Gauss Quadrature

From now on we will assume the function $f(x)$ to be integrated is a polynomial in x of some order n.

Gauss quadrature is optimal when the function is a polynomial

The main starting point is to represent the function $f\left(x\right)$ as a combination of linearly independent basis.

Instead of using strips of equal width, we assume the width can vary from one strip to the next. Let us call the width of the strip w_i. Instead of taking the height of the i^{th} strip to be the value of the function at the left edge of the strip, let us also be flexible on the location of the x associated with strip w_i and call the height of the strip w_i as $f(x_i)$ where x_i is to be determined. Hence the above integration becomes

$$\begin{aligned}\int_a^b f(x)\ dx &= \left(\sum_{i=1}^{N} w_i\ f(x_i)\right) + E \\ &\approx \sum_{i=1}^{N} w_i\ f(x_i)\end{aligned}$$

So our goal is to determine w_i and x_i such as the error E is minimized in the above equation. We would really like to find w_i and x_i such that the error is zero with the smallest value for N.

It seems as if this is a very hard problem. We have $2N$ unknown quantities to determine. N different widths, and N associated x points to evaluate the height of each specific strip at. And we only have as an input $f(x)$ and the limits of integration, and we need to determine these $2N$ quantities such that the error in integration is zero.

In other words, the problem is to find w_i, x_i such that

$$I = \int_a^b f(x)\ dx = w_1 f(x_1) + w_2 f(x_2) + \cdots w_N f(x_N) \tag{1}$$

One way to make some progress is to expand $f(x)$ as a series. We can approximate $f(x)$ as convergent power series for example. If $f(x)$ happens to be a polynomial instead, we can represent it exactly using a finite sequence of Legendre polynomials. It is in this second case where this method makes the most sense to use due to the advantages we make from the second representation.

We show both methods below.

Expanding $f(x)$ as convergent power series over the range a, b gives

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots a_m x^m + \cdots \tag{2}$$

Substituting (2) into (1) gives

$$I = \int_a^b \left(a_0 + a_1 x + a_2 x^2 + \cdots a_m x^m + \cdots\right)\ dx = w_1 f(x_1) + w_2 f(x_2) + \cdots w_N f(x_N) \tag{3}$$

But

$$\begin{aligned}\int_a^b \left(a_0 + a_1 x + a_2 x^2 + \cdots a_m x^m + \cdots\right)\ dx &= \int_a^b a_0 dx + \int_a^b a_1 x\ dx + \int_a^b a_2 x^2\ dx + \cdots + \int_a^b a_m x^m\ dx + \cdots \\ &= a_0(b-a) + a_1\frac{\left(b^2-a^2\right)}{2} + a_2\frac{\left(b^3-a^3\right)}{3} + \cdots + a_m\frac{\left(b^{m+1}-a^{m+1}\right)}{m+1} + \cdots\end{aligned}$$

Substituting the above into (3) results in

$$\begin{aligned}a_0(b-a) + a_1\frac{\left(b^2-a^2\right)}{2} + a_2\frac{\left(b^3-a^3\right)}{3} + \cdots + a_m\frac{\left(b^{m+1}-a^{m+1}\right)}{m+1} + \cdots \\ = w_1 f(x_1) + w_2 f(x_2) + \cdots + w_N f(x_N)\end{aligned} \tag{4}$$

But from (2) we see that

$$\begin{aligned}f(x_1) &= a_0 + a_1 x_1 + a_2 x_1^2 + \cdots a_m x_1^m + \cdots \\ f(x_2) &= a_0 + a_1 x_2 + a_2 x_2^2 + \cdots a_m x_2^m + \cdots \\ &\cdots \\ f(x_N) &= a_0 + a_1 x_N + a_2 x_N^2 + \cdots a_m x_N^m + \cdots\end{aligned}$$

Substituting the above into (4) gives

$$\begin{aligned}
a_0 (b-a) + a_1 \frac{\left(b^2-a^2\right)}{2} + a_2 \frac{\left(b^3-a^3\right)}{3} + \cdots + a_m \frac{\left(b^{m+1}-a^{m+1}\right)}{m+1} + \cdots = \\
w_1 \left(a_0 + a_1 x_1 + a_2 x_1^2 + \cdots a_m x_1^m + \cdots\right) \\
+ w_2 \left(a_0 + a_1 x_2 + a_2 x_2^2 + \cdots a_m x_2^m + \cdots\right) \\
\cdots \\
+ w_N \left(a_0 + a_1 x_N + a_2 x_N^2 + \cdots a_m x_N^m + \cdots\right)
\end{aligned}$$

Rearranging gives

$$\begin{aligned}
a_0 (b-a) + a_1 \frac{\left(b^2-a^2\right)}{2} + a_2 \frac{\left(b^3-a^3\right)}{3} + \cdots + a_m \frac{\left(b^m-a^m\right)}{m} + \cdots = \\
a_0 \left(w_1 + w_2 + \cdots + w_N\right) \\
+ a_1 \left(w_1 x_1 + w_2 x_2 + \cdots + w_N x_N\right) \\
+ a_2 \left(w_1 x_1^2 + w_2 x_2^2 + \cdots + w_N x_N^2\right) \\
\cdots \\
+ a_m \left(w_1 x_1^m + w_2 x_2^m + \cdots + w_N x_N^m\right)
\end{aligned}$$

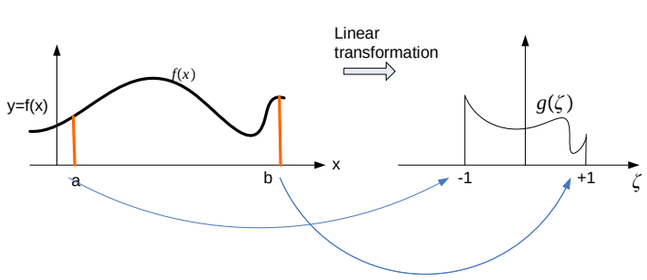
Equating coefficients on both sides results in

$$\begin{aligned}
w_1 + w_2 + \cdots + w_N &= b - a \\
w_1 x_1 + w_2 x_2 + \cdots + w_N x_N &= \frac{\left(b^2-a^2\right)}{2} \\
w_1 x_1^2 + w_2 x_2^2 + \cdots + w_N x_N^2 &= \frac{\left(b^3-a^3\right)}{3} \\
&\cdots \\
w_1 x_1^m + w_2 x_2^m + \cdots + w_N x_N^m &= \frac{\left(b^m-a^m\right)}{m} \\
&\cdots
\end{aligned} \tag{5}$$

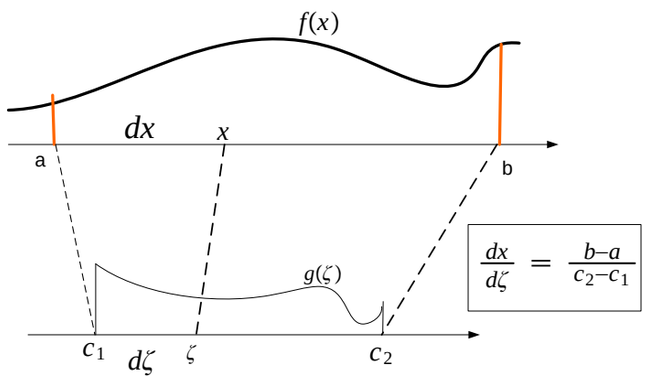
Since we have $2N$ unknown quantities to solve for (N weights w_i and N points x_i) we need $2N$ equations. In other words, we need to have $m = 2N$. The above set of simultaneous $2N$ equations can now be solved for the unknown w_i, x_i and this will give us the numerical integration we wanted.

The above assumed that the function $f(x)$ can be represented exactly by the power series expansion with m terms.

We now consider the representation of $f(x)$ as a series of Legendre polynomials. We do this since when $f(x)$ itself is a polynomial. We can represent $f(x)$ exactly by a finite number of Legendre polynomials. But since Legendre polynomials $P_n(x)$ are defined over $[-1, 1]$ we start by transforming $f(x)$ to this new range and then we can expand the mapped $f(x)$ (which we will call $g(\zeta)$) in terms of the Legendre polynomials.



An easy way to find this mapping is to align the ranges over each others and take the ratio between as the scale factor. This diagram shows this for a general case where we map $f(x)$ defined over $[a, b]$ to a new range defined over $[c_1, c_2]$



We see from the diagram that

$$\zeta = c_1 + d\zeta$$

But
$\frac{d\zeta}{dx}$ is the same ratio as $\frac{b-a}{c_2-c_1}$
Hence

$$\frac{dx}{d\zeta} = \frac{b-a}{c_2-c_1} \tag{6}$$

The above is called the Jacobin of the transformation. Now, From the diagram we see that

$$dx = x - a$$

And

$$d\zeta = \zeta - c_1$$

Hence (6) becomes

$$\frac{x-a}{\zeta - c_1} = \frac{b-a}{c_2-c_1}$$

Hence we obtain that

$$\zeta = \frac{x-a}{b-a}(c_2 - c_1) + c_1$$

And

$$x = \frac{b-a}{c_2-c_1}(\zeta - c_1) + a$$

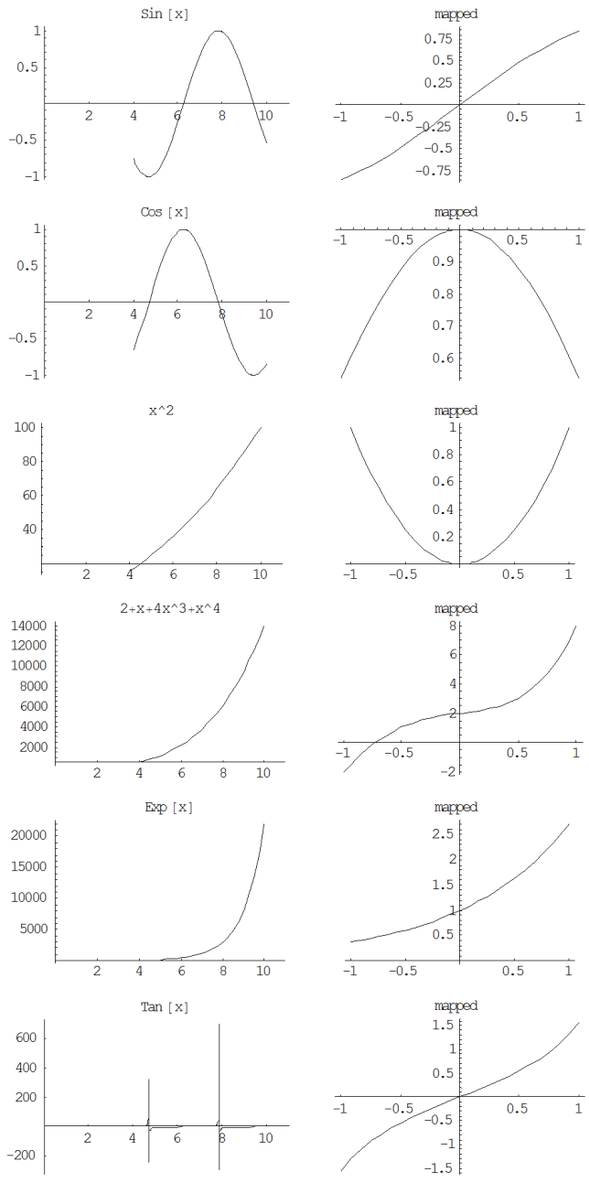
For the specific case when $c_1 = -1$ and $c_2 = +1$ the above expressions become

$$\begin{aligned}\zeta &= \frac{x-a}{b-a}(2) - 1 \\ &= \frac{2x - 2a - (b-a)}{b-a} \\ &= \frac{2x - a - b}{b-a}\end{aligned}$$

Hence

$$\begin{aligned}x &= \frac{b-a}{2}(\zeta + 1) + a \\ &= \frac{(b-a)\,\zeta + (a+b)}{2}\end{aligned}$$

Before we proceed further, It will be interesting to see the effect of this transformation on the shape of some functions. Below I plotted some functions under this transformation. The left plots are the original functions plotted over some range, in this case $[4, 10]$ and the right side plots show the new shape (the function $g(\zeta)$) over the new range $[-1, 1]$



We must remember that in the following analysis, we are integrating now the function $g(\zeta)$ over $[-1,1]$ and not $f(x)$ over $[a,b]$. Hence to obtain the required integral we need to transform back the final result. We will show how to do this below.

We can approximate any function $f(x)$ as a series expansion in terms of weighted sums of a set of basis functions. We do this for example when we use Fourier series expansion.

Hence we write

$$f(x) = \sum_{i}^{\infty} a_i \Phi_i(x) \tag{7}$$

We can give an intuitive justification to the above formulation as follows. If we think in terms of vectors. A vector $\mathbf{V}$ in an n-dimensional space is written in terms of its components as follows

$$\begin{aligned}\mathbf{V} &= a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots a_N\mathbf{e}_N \\ &= \sum_{i}^{N} a_i\mathbf{e}_i\end{aligned}$$

Where $\mathbf{e}_i$ is the basis vectors in that space.

If we now consider a general infinite dimensional vector space where each point in that space is a function, then we see that we can also represent that function as a weighted series of a basis functions just as we did for a normal vector.

There are many sets of basis functions we can choose to represent the function $f(x)$ with. The main requirements for the basis functions is that they are complete (This means they span the whole space) and there is defined an inner product on them.

For our purposes here, we are interested in the class of function $f(x)$ that are polynomials in x. The basis we will select to use are the Legendre basis as explained above. To do this, we transform $f(x)$ to $g(\zeta)$ as shown above and then now our integral becomes

$$\int_a^b f(x)\,dx = \int_{-1}^{1} f(x(\zeta))\left(\frac{(b-a)}{2}d\zeta\right)$$

This is because we found that $dx = \frac{(b-a)}{2}d\zeta$ from above.

If we call $f(x(\zeta))$ as $g(\zeta)$ the integral becomes

$$\int_a^b f(x)\,dx = \int_{-1}^{1} \frac{(b-a)}{2} g(\zeta)\ d\zeta$$

Since $\frac{(b-a)}{2}$ is the Jacobian of the transformation, we write the integral as

$$\int_a^b f(x)\,dx \rightarrow \int_{-1}^{1} J\ g(\zeta)\ d\zeta$$

Since the Jacobian is constant in this case, we will only worry about $\int_{-1}^{1} g(\zeta)\ d\zeta$ and we just need to remember to scale the result at the end by J. This is the integral we will now numerically integrate.

Equation (7) is now written as

$$g(\zeta) = \sum_i^{\infty} a_i P_i(\zeta)$$

Where $P_i(\zeta)$ is the Legendre polynomial of order i and $g(\zeta)$ is a polynomial in ζ or some order m.

Now we carry the same analysis we did when we expressed $f(x)$ as a power series. The difference now is that the limit of integration is symmetrical and the basis are now the Legendre polynomials instead of the polynomials x^n as the case was in the power series expansion. So now equation (1) becomes

$$I = \int_{-1}^{1} g(\zeta)\ d\zeta = w_1 g(\zeta_1) + w_2 g(\zeta_2) + \cdots + w_N g(\zeta_N) \tag{8}$$

And equation (2) becomes

$$g(\zeta) = a_0 P_0(\zeta) + a_1 P_1(\zeta) + a_2 P_2(\zeta) + \cdots a_m P_m(\zeta) + \cdots \tag{9}$$

Substituting (9) into (8) we get the equivalent of equation (3)

$$I = \int_{-1}^{1} \left(a_0 P_0(\zeta) + a_1 P_1(\zeta) + a_2 P_2(\zeta) + \cdots a_m P_m(\zeta) + \cdots\right)\ d\zeta \tag{10}$$

$$= w_1 g(\zeta_1) + w_2 g(\zeta_2) + \cdots + w_N g(\zeta_N) \tag{1}$$

$$\begin{aligned}\int_{-1}^{1} \left(a_0 P_0 + a_1 P_1 + a_2 P_2 + \cdots a_m P_m + \cdots\right)\ d\zeta &= \int_{-1}^{1} a_0 P_0 d\zeta + \int_{-1}^{1} a_1 P_1 d\zeta + \int_{-1}^{1} a_2 P_2\ d\zeta + \cdots + \int_{-1}^{1} a_m P_m d\zeta\ + \cdots \\ &= a_0(2) + a_1 0 + a_2 0 + \cdots + a0 + \cdots \\ &= 2a_0\end{aligned}$$

The above occurs since the integral of any Legendre polynomial of order greater than zero will be zero over $[-1, 1]$

Substituting the above into (10) we obtain

$$I = 2a_0 = w_1 g(\zeta_1) + w_2 g(\zeta_2) + \cdots + w_N g(\zeta_N) \tag{11}$$

But from (9) we see that

$$\begin{aligned} g(\zeta_1) &= a_0 P_0(\zeta_1) + a_1 P_1(\zeta_1) + a_2 P_2(\zeta_1) + \cdots a_m P_m(\zeta_1) + \cdots \\ g(\zeta_2) &= a_0 P_0(\zeta_2) + a_1 P_1(\zeta_2) + a_2 P_2(\zeta_2) + \cdots a_m P_m(\zeta_2) + \cdots \\ &\cdots \\ g(\zeta_N) &= a_0 P_0(\zeta_N) + a_1 P_1(\zeta_N) + a_2 P_2(\zeta_N) + \cdots a_m P_m(\zeta_N) + \cdots \end{aligned}$$

Substituting the above in (11) gives

$$\begin{aligned} 2a_0 = {} & w_1 (a_0 P_0(\zeta_1) + a_1 P_1(\zeta_1) + a_2 P_2(\zeta_1) + \cdots a_m P_m(\zeta_1) + \cdots) + \\ & + w_2 (a_0 P_0(\zeta_2) + a_1 P_1(\zeta_2) + a_2 P_2(\zeta_2) + \cdots a_m P_m(\zeta_2) + \cdots) + \\ & \cdots \\ & + w_N (a_0 P_0(\zeta_N) + a_1 P_1(\zeta_N) + a_2 P_2(\zeta_N) + \cdots a_m P_m(\zeta_N) + \cdots) \end{aligned}$$

Rearranging results in

$$\begin{aligned} 2a_0 = {} & a_0 (w_1 P_0(\zeta_1) + w_2 P_0(\zeta_2) + \cdots + w_N P_0(\zeta_N)) \\ & + a_1 (w_1 P_1(\zeta_1) + w_2 P_1(\zeta_2) + \cdots + w_N P_1(\zeta_N)) \\ & + \cdots \\ & + a_m (w_1 P_m(\zeta_1) + w_2 P_m(\zeta_2) + \cdots + w_N P_m(\zeta_N)) \end{aligned}$$

Equating coefficients gives

$$\begin{aligned} 2 &= w_1 P_0(\zeta_1) + w_2 P_0(\zeta_2) + \cdots + w_N P_0(\zeta_N) \\ 0 &= w_1 P_1(\zeta_1) + w_2 P_1(\zeta_2) + \cdots + w_N P_1(\zeta_N) \\ 0 &= w_1 P_2(\zeta_1) + w_2 P_2(\zeta_2) + \cdots + w_N P_2(\zeta_N) \\ &\cdots \\ 0 &= w_1 P_m(\zeta_1) + w_2 P_m(\zeta_2) + \cdots + w_N P_m(\zeta_N) \end{aligned}$$

If we select the points ζ_i to be the roots of P_{i-1} we can write the above as

$$\begin{aligned} 2 &= w_1 P_0(\zeta_1) + w_2 P_0(\zeta_2) + \cdots + w_N P_0(\zeta_N) \\ 0 &= w_1 P_1(\zeta_1) + w_2 P_1(\zeta_2) + \cdots + w_N P_1(\zeta_N) \\ 0 &= w_1 P_2(\zeta_1) + w_2 P_2(\zeta_2) + \cdots + w_N P_2(\zeta_N) \\ &\cdots \\ 0 &= w_1 P_m(\zeta_1) + w_2 P_m(\zeta_2) + \cdots + w_N P_m(\zeta_N) \end{aligned}$$

3 References

1. Mathematica Structural Mechanics help page
2. MIT course 16.20 lecture notes. MIT open course website.
3. Theory of elasticity by S. P. Timoshenko and J. N. Goodier. chapter 10