

## Diffusion with Reaction (nonsymmetric)

Consider reaction and diffusion in a porous slab. This problem is analogous to heat conduction in a slab with a heat source which is dependent on temperature and position. The governing equations are:

$$\frac{d^2 y}{dx^2} + r(y, x) = 0 \quad (9)$$

with:

$$y = 0 \text{ at } x = 0, 1$$

For a simple  $k^{\text{th}}$  order reaction with the reactivity independent of position, the reaction term is:

$$r(y, x) = 4\phi^2(1 - y)^k \quad (10)$$

where  $\phi$  is the Thiele modulus. It is defined using half the thickness of the slab, so the factor of 4 is required when the full slab is considered.

To solve the problem using a Method of Weighted residuals we start with a trial solution, following Eq. (6):

$$y \cong \tilde{y} = \sum_{i=0}^{n+1} \tilde{y}_i \ell_i(x) \quad (11)$$

The interpolation points include the endpoints,  $x_0 = 0$  and  $x_{n+1} = 1$ , which drop out due to the boundary conditions. The interior points are either Gauss or Lobatto quadrature base points. The residual is formed by substitution of the approximate solution into the equation:

$$\sum_{i=1}^n \tilde{y}_i \frac{d^2 \ell_i}{dx^2} + r(\tilde{y}, x) = R(x, \tilde{y}) \quad (12)$$

**Orthogonal Collocation Method:** With the collocation method the residual is set to zero at the collocation points,  $x_j$ :

$$\sum_{i=1}^n \tilde{y}_i \left. \frac{d^2 \ell_i}{dx^2} \right|_{x_j} + r \left[ \left( \sum_{i=1}^n \tilde{y}_i \ell_i(x_j) \right), x_j \right] = 0 \quad (13)$$

Since  $\ell_i(x_j) = \delta_{ij}$ , this simplifies to:

$$\sum_{i=1}^n B_{ji} \tilde{y}_i + r(\tilde{y}_j, x_j) = 0 \quad (14)$$

where:

$$B_{ji} = \left. \frac{d^2 \ell_i}{dx^2} \right|_{x_j}$$

To solve the problem, we need only the collocation points, i.e. Gauss or Lobatto quadrature base points, and the **B** matrix, the second derivative of the Lagrange interpolating polynomials. Appendix A describes how these quantities are calculated. Eq. (13) is a set of algebraic equations that are nonlinear if the reaction is not first order or  $0^{\text{th}}$  order. A Newton-Raphson procedure works well for the nonlinear solution. The nonlinear reaction terms appear only on the diagonal, which simplifies the calculations. We shall see

that for a full moments or Galerkin method, the reaction terms are distributed throughout the matrix.

**Results:** Fig. 1 shows solutions for a first order reaction, Eq. (10) with  $k = 1$  and  $\phi = 5$ . Approximate solutions are shown with  $n = 4$  for orthogonal collocation at both Gauss and Lobatto points. With this relatively high reaction rate most of the reaction occurs near the boundary. The fifth order polynomial can only approximate the sharp profile by oscillating about the exact solution. The Lobatto points produce a more accurate solution than the Gauss points. Since this problem is symmetric about  $x = 0.5$ , it can be solved more efficiently using symmetric trial functions. We will come back to this problem when we consider methods for symmetric problems.

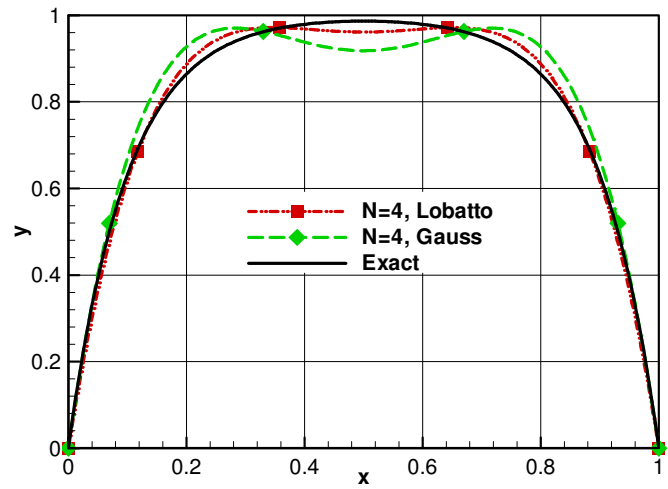


Fig. 1 First order reaction, Eq. (10),  $\phi = 5$

To make the problem nonsymmetric and more interesting, consider the case of a first order reaction, but with a rate constant which varies across the slab according to:

$$r(y, x) = \phi^2(1 - y)(0.2 + 1.6x^2(3 - 2x)) \quad (15)$$

The spatial variation in parenthesis varies from 0.2 on the left edge to 1.8 on the right edge, with an average value of 0.5 on the left half and 1.5 on the right half giving an overall average of 1.0. For a given value of  $\phi$ , the average rate constant is the same as for Eq. (10). Fig. 2 shows solutions to this problem for  $\phi = 5$  with Gauss and Lobatto points. The solution with Lobatto points is again more accurate than with Gauss points.

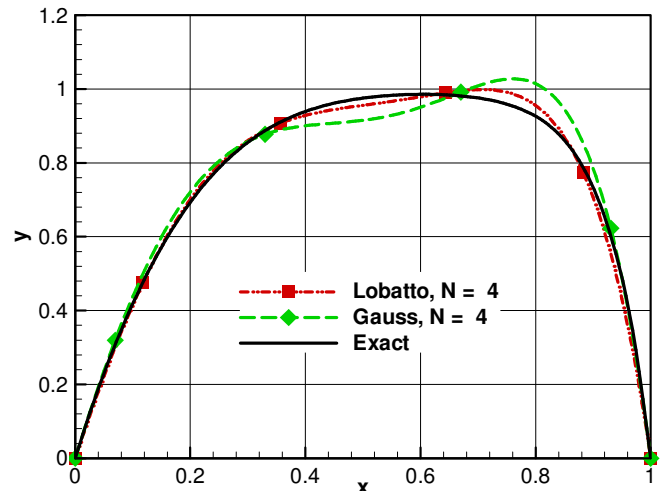


Fig. 2 First order reaction, spatial variation, Eq. (15),  $\phi = 5$

**Moments Method:** To solve the problem by the method of moments, the residual is weighted by  $x^{(i-1)}$  for  $i = 1, \dots, n$ ; however, weighting by any linearly independent set of  $n$  polynomials through degree  $n - 1$  will give identical results. One set of linearly independent polynomials are the Lagrange interpolating polynomials through only the  $n$  interior points.

These polynomials are related to those in Eq. 11 by:

$$l_i^*(x) x(1 - x) = l_i(x) x_i(1 - x_i) \quad (16)$$

where the asterisk indicates the reduced polynomial. Using these weight functions in Eq. (4) together with the residual function, Eq. (12) and integrating numerically, the problem becomes:

$$\sum_{k=1}^m \left[ \sum_{i=1}^n \tilde{y}_i \frac{d^2 \ell_i}{dx^2} \Big|_{x_k} + r \left( \left( \sum_{i=1}^n \tilde{y}_i \ell_i(x_k) \right), x_k \right) \right] W_k \ell_j^*(x_k) = 0 \quad (17)$$

where the  $x_k$  in Eq. (17) designate the quadrature base points. For  $m > n$  the quadrature base points are different from the nodal interpolation points used to define the trial functions, Eq. (11). If  $m = n$ , and we let the interpolation points correspond to the quadrature points, some wonderful simplifications occur. For this case, Eq.(17) simplifies to:

$$\sum_{i=1}^n W_j B_{ji} \tilde{y}_i + W_j r(\tilde{y}(x_j), x_j) = 0 \quad (18)$$

Eq.(18) is identical to Eq.(14) multiplied by the quadrature weight,  $W_j$ . Now we ask, does  $m = n$  provide enough accuracy so that Eq. (18) is a good approximation to Eq. (17)?

First consider integration of the diffusion terms in Eq. (17) with  $n$  point Gaussian quadrature. The trial functions,  $\ell_i(x)$ , are polynomials of degree  $n + 1$ , so the second derivative is of degree  $n - 1$ . The weight functions,  $\ell_j^*(x)$  are of degree  $n - 1$ . Since Gaussian quadrature is exact for polynomials through degree  $2n - 1$ , the diffusion terms in Eq. (18) are exact.

Now, consider the source terms in Eq. (17). With Eq. (15) the integrands for the reaction terms in Eq. (17) are of degree  $2n+3$ , so  $n$  point Gaussian quadrature misses exact integration by 4 degrees. Gaussian quadrature with  $m = n + 2$  is required for exact integration. If there were no cubic spatial variation of reactivity, it would miss exact integration by only one degree.

A common method for treating nonlinearities is to interpolate nonlinear terms into the trial space. With this approach, the reaction term is approximated by:

$$r(y, x) \cong \sum_{i=0}^{n+1} \ell_i(x) r(y(x_i), x_i) \quad (19)$$

Substitution of Eq. (19) into Eq. (17) shows that  $n$  point Gaussian quadrature misses exact integration by only one degree.

**Table 1 Integration requirements for reaction terms**

	with spatial variation, Eq. (15)	without spatial variation	interpolated source term, Eq. (19)
$m$ required for exact	$n + 2$	$n + 1$	$n + 1$
$m=n$ , degrees error	4	1	1

Table 1 summarizes the quadrature requirements for the reaction terms when Gaussian quadrature is used. From this analysis, we conclude that  $n$  point Gaussian quadrature, or orthogonal collocation at Gauss points, gives an accurate approximation to the moments method. The diffusion term is integrated exactly and the source term is accurately approximated.

If Eq.(19) is used to represent the source term, Eq. (17) reduces to:

$$\sum_{i=1}^n W_j B_{ji} \tilde{y}_i + \sum_{i=1}^n M_{ji} r(\tilde{y}_i, x_i) = 0 \quad (20)$$

where,  $\mathbf{M}$  is:

$$M_{ji} = \int_0^1 \ell_j^*(x) \ell_i(x) dx$$

We observe that for the full moments method, Eq. (17) with  $m > n$ , the reaction terms are distributed throughout the equations and are evaluated at points other than the interpolation points. This feature complicates solution of the equations, especially for nonlinear rate terms. Interpolation of the rate terms simplifies the equations some, Eq. (20), but the rate terms are still distributed throughout the equations. With orthogonal collocation, Eq. (14) or (18), the rate terms appear only on the diagonal, which simplifies the equations significantly.

We also note that the matrix  $\mathbf{WB}$  is symmetric while  $\mathbf{M}$  in Eq.(20) is not. If Eq. (14) were multiplied by  $\mathbf{W}$ , making it identical to Eq. (18), the solution matrix would be symmetric. A symmetric matrix problem can be solved more quickly than a nonsymmetric matrix problem.

If we were to use Lobatto quadrature with  $n$  interior points and the two end points, it would not reduce to a collocation method because of the end points. orthogonal collocation at Lobatto quadrature base points bears no direct relationship to the moments method.

**Results:** For the same case as in Fig. 2, Fig. 3 shows a comparison of solutions with the orthogonal collocation method at Gauss points and the full moments method. It appears that the additional complexity of the moments method does not produce much improvement in the approximate solution for this problem.

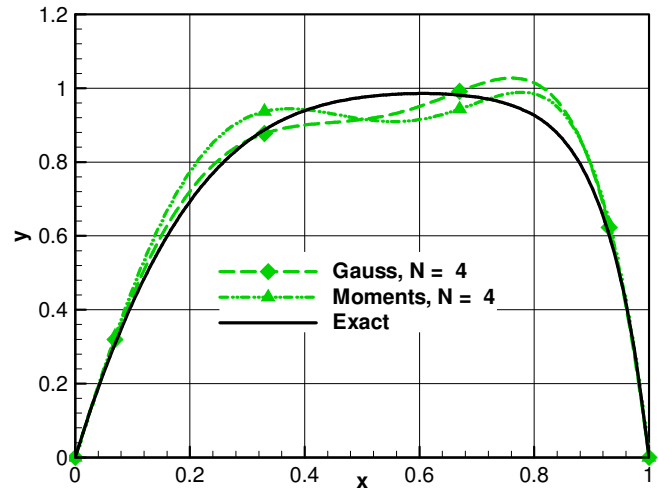


Fig. 3 Orthogonal Collocation & Moments methods, Eq. (15)

**Galerkin Method:** To solve the problem with the Galerkin method, the residual is weighted by the trial functions  $\ell_j(x)$  for  $j = 1, \dots, n$ . With the Galerkin method, it is customary to integrate the second derivative term by parts, so the problem becomes:

$$\sum_{i=1}^n \ell_j \frac{d\ell_i}{dx} \tilde{y}_i \Big|_0^1 - \int_0^1 \left( \sum_{i=1}^n \frac{d\ell_j}{dx} \frac{d\ell_i}{dx} \tilde{y}_i - \ell_j(x) r(\tilde{y}, x) \right) dx = 0 \quad (21)$$

The first term drops out because  $\ell_j(0) = \ell_j(1) = 0$  for all  $j$ . Using quadrature to perform the integration, the equation becomes:

$$\sum_{k=1}^m W_k \left( \sum_{i=1}^n \frac{d\ell_j}{dx} \Big|_{x_k} \frac{d\ell_i}{dx} \Big|_{x_k} \tilde{y}_i - \ell_j(x_k) r(\tilde{y}(x_k), x_k) \right) = 0 \quad (22)$$

The  $x_k$  in Eq. (22) designate the quadrature base points, which differ from the nodal interpolation points for  $m > n$ . Since the trial functions are polynomials of degree  $n+1$ , the diffusion term is of degree  $2n$ . For Eq. (10) with  $k = 1$  or if the reaction term is interpolated using Eq. (19), the reaction term is of degree  $2n+2$ . With Eq.(15) the reaction term is of degree  $2n+5$ . An  $n$  point Gaussian quadrature is exact through degree  $2n - 1$ , so neither term would be integrated exactly. On the other hand, Lobatto quadrature with  $n$  interior points gives exact integration through degree  $2n+1$ , so it gives exact integration for the diffusion term, but misses exact integration of the reaction term by one degree (4 degrees for Eq. (15)). With Lobatto quadrature we get the two additional degrees of accuracy for free, because  $\ell_j(0) = \ell_j(1) = 0$  for all  $j$  in Eq. (20). With Lobatto quadrature Eq. (22) reduces to:

$$\sum_{i=1}^n C_{ji} \tilde{y}_i - W_j r(\tilde{y}_j, x_j) = 0 \quad (23)$$

where:

$$C_{ji} = \sum_{k=0}^{n+1} W_k A_{kj} A_{ki} \quad \text{and} \quad A_{ki} = \frac{d\ell_i}{dx} \Big|_{x_k}$$

Since Lobatto quadrature is sufficiently accurate to perform the integration by parts exactly, it follows that:

$$C_{ji} = \delta_{j,n+1} A_{n+1,i} - \delta_{i,0} A_{0i} - W_j B_{ji} \quad (24)$$

Given this relationship, it is clear that Eq. (23) is equal to Eq. (14) multiplied by the negative of the quadrature weights,  $W_j$ . Although the equations are equivalent, Eq. (23) is amenable to more efficient solution techniques due to the symmetry of the  $\mathbf{C}$  matrix. From this discussion we conclude that Orthogonal Collocation with Lobatto quadrature base points is an accurate approximation to the Galerkin method.

As with the Moments method, the implementation of the full Galerkin method requires that the source terms be integrated more accurately. If the reaction terms are interpolated using Eq. (19) and the resulting equations are integrated exactly, Eq. (21) becomes:

$$\sum_{i=1}^n [C_{ji} \tilde{y}_i - M_{ji} r(\tilde{y}_i, x_i)] = 0 \quad (25)$$

where  $\mathbf{M}$  is given by:

$$M_{jk} = \int_0^1 \ell_k(x) \ell_j(x) dx$$

The matrix,  $\mathbf{M}$ , is symmetric, but full. This causes the reaction terms to be distributed throughout the matrix rather than being isolated to the diagonals as for Eq. (14) or (23). The distributed reaction terms add considerable complexity especially for nonlinear rate terms.

In finite element terminology, the matrix  $\mathbf{C}$  is called the stiffness matrix and  $\mathbf{M}$  is called the mass matrix. These names reflect the roots of the finite elements method in structural mechanics. To simplify the method, it is a common practice to do an *ad hoc* lumping of the off diagonal elements of  $\mathbf{M}$  onto the diagonal. Lumping of Eq. (25) produces Eq. (23).

**Results:** For the same case as in Figs. 2 and 3, Fig. 4 shows a comparison of solutions with the Orthogonal Collocation method using Lobatto points and the full Galerkin Method. It appears that the Galerkin method offers a small improvement to the solution for this problem.

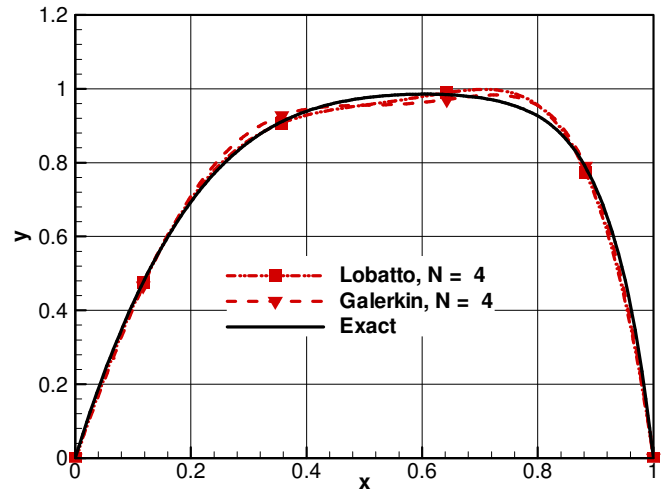


Fig. 4 Orthogonal Collocation & Galerkin methods, Eq. (15)

### Mass Conservation and Fluxes

For this type of problem, one is usually interested in the overall affect of the reaction. For example, if a fluid were flowing on both sides of the slab we would want to know the flux of components from the slab. For a heat transfer problem the flux of heat would be of interest. For the symmetric problem, this is quantified by the effectiveness factor which is the ratio of the average reaction rate to the rate with no diffusion resistance:

$$\eta = \frac{\int_0^1 r(y, x) dx}{\int_0^1 r(0, x) dx} \quad (26)$$

For a first order reaction, Eq. (10) with  $k = 1$  and no spatial variation, the effectiveness factor from the analytical solution is:

$$\eta = \tanh(\phi) / \phi \quad (27)$$

Eq. (26) is of limited interest for a nonsymmetric problem, since the breakdown of left and right side fluxes is normally of interest. The boundary fluxes are related to the average reaction rate by the overall balance:

$$-\left. \frac{dy}{dx} \right|_0^1 = \int_0^1 r(y, x) dx \quad (28)$$

The two outward normal fluxes can be calculated to get the desired breakdown. The sum of the fluxes will give the average reaction *provided the method conserves mass*.

Eq. (28) was developed by integrating Eq.(9). In general, a method will be conservative if Eq. (4) holds for  $w_i(x) = 1$ . If the method includes unity as a weight function or if unity can be obtained from a linear combination of the weight functions, the method is conservative. The method of moments and orthogonal collocation at Gauss points are both conservative because for our modified weight functions,  $\sum \ell_i^*(x) = 1$ . Since these methods are conservative, Eq. (28) is obeyed when the boundary fluxes are calculated by differentiation of the approximate solution:

$$\left. \frac{dy}{dx} \right|_{x=0} = \sum_{i=0}^{n+1} A_{0i} \tilde{y}_i \quad \text{and} \quad \left. \frac{dy}{dx} \right|_{x=1} = \sum_{i=0}^{n+1} A_{n+1,i} \tilde{y}_i \quad (29)$$

For the Galerkin method the sum of all the Lagrange interpolating polynomials is also unity, but the first and last of these are not used as weight functions. The method appears not to be conservative due to the left over terms on the right side below:

$$\left. \frac{dy}{dx} \right|_0^1 + \int_0^1 r(y, x) dx = \int_0^1 (\ell_0 + \ell_{n+1}) R(x, \tilde{y}) dx \quad (30)$$

This problem is related to the weak treatment of flux boundary conditions which will be discussed later. The left over terms correspond to Eq. (21) for  $j = 0$  and  $j = n+1$ . Since we cannot simultaneously satisfy Eq. (28) and Eq. (29), we choose to satisfy only Eq. (28), the overall balance. From Eqs. (21) and (23) this is accomplished by:

$$\left. \frac{dy}{dx} \right|_{x=0} = - \sum_{i=1}^n C_{0i} \tilde{y}_i + W_0 r(0,0) \quad \text{and} \quad (31)$$

$$\left. \frac{dy}{dx} \right|_{x=1} = \sum_{i=1}^n C_{n+1,i} \tilde{y}_i - W_{n+1} r(0,1)$$

For a full Galerkin method, the right hand sides of Eq. (31) are replaced by Eq. (25) evaluated at 0 and  $n+1$ . It is often useful and more intuitive to replace  $\mathbf{C}$  in Eq. (31) by the equivalent expression from Eq. (24):

$$\left. \frac{dy}{dx} \right|_{x=0} = \sum_{i=1}^n A_{0i} \tilde{y}_i + W_0 \left( \sum_{i=1}^n B_{0i} \tilde{y}_i + r(0,0) \right) \quad \text{and} \quad (32)$$

$$\left. \frac{dy}{dx} \right|_{x=1} = \sum_{i=1}^n A_{n+1,i} \tilde{y}_i - W_{n+1} \left( \sum_{i=1}^n B_{n+1,i} \tilde{y}_i + r(0,1) \right)$$

This equivalent expression shows the flux is given by Eq. (29), but with a correction term equal to the residual evaluated at the boundary and multiplied by the endpoint quadrature weight. These correction terms correspond to the right hand side of Eq. (30). Since the endpoint weights are zero for Gauss points, Eq. (32) is valid for either Lobatto or Gauss points.

Table 2 shows the normalized fluxes calculated for the examples shown in Figs. 2 – 4, i.e. Eq.(15) with  $\varphi = 5$ .

**Table 2 Calculated Fluxes for Reaction and Diffusion,  $n = 4$**

	Flux left	Flux right	Flux Total	Error left	Error right	Error total
Exact	0.05062	0.13368	0.18429			
Gauss	0.05013	0.12097	0.17110	-0.0096	-0.0952	-0.0716
Moments	0.04902	0.12598	0.17500	-0.0316	-0.0576	-0.0504
Moments, Eq. (20)	0.05553	0.11577	0.17129	0.0969	-0.1340	-0.0706
Lobatto, Eq. (31)	0.05073	0.13742	0.18814	0.0021	0.0280	0.0209
Galerkin	0.05053	0.13561	0.18614	-0.0017	0.0145	0.0100
Galerkin, Eq. (25)	0.05070	0.13601	0.18671	0.0016	0.0175	0.0131
Lobatto, Eq. (29)	0.04666	0.10497	0.15263	-0.0782	-0.2147	-0.1773
Galerkin, Eq. (29)	0.04191	0.11105	0.15396	-0.1721	-0.1693	-0.1700
Galerkin, Eqs. (25) & (29)	0.03953	0.10974	0.14927	-0.2190	-0.1791	-0.1900

The relative accuracy of the fluxes is in general agreement with observations of the accuracy of the various methods in Figs. 2 through 4. The Galerkin method is the most accurate method, with errors of less than 2% for the individual fluxes. This error increases a small amount if the reaction terms are interpolated according to Eq. (19). For orthogonal collocation at Lobatto points, the errors are less than 3%. With all of these methods, the fluxes were calculated using Eq. (31) or the equivalent for the Galerkin method. If the fluxes are calculated using Eq. (29) the errors are generally 15% to 20%. With the full moments methods the errors are as large as 6%, while orthogonal collocation at Gauss points gives errors as large as 10%. According to our calculations, the moments method with interpolation of the reaction terms, Eq. (20), gives even greater error.

Fig. 5 shows the error in the right side flux calculations with the various methods. The results are consistent with those in Table 2, i.e. the Galerkin method and orthogonal collocation at Lobatto points are the most accurate methods, provided the fluxes are properly calculated. These methods give errors of less than 1% with  $n=5$  and less than 0.1% with  $n=6$ . The accuracy of these methods is lost completely if fluxes are calculated from Eq. (29). Accuracy with the moments method and orthogonal collocation at Gauss points is also quite good, giving errors of less than 1% with  $n=6$  and less than 0.1% with  $n=7$ . The flux calculations with these methods are simpler, Eq. (29), than with the Lobatto/Galerkin methods, Eq. (31 or 32); however, a larger  $n$  is required for equivalent accuracy.

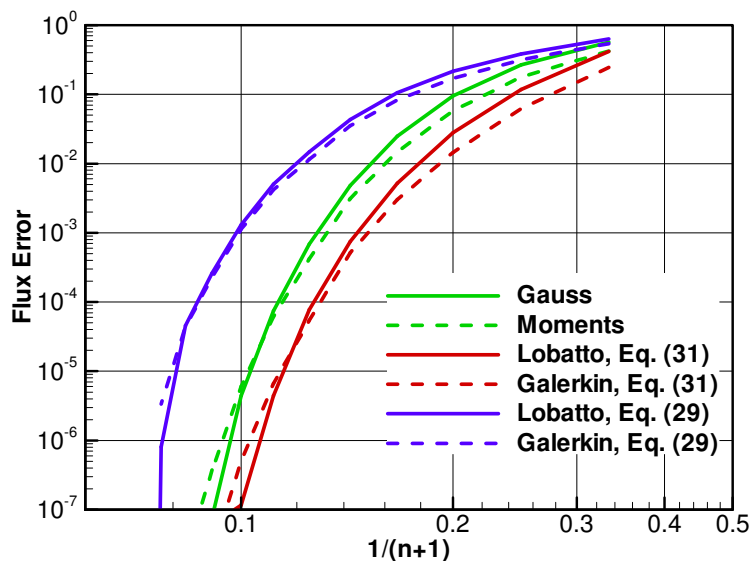


Fig. 5 Errors of flux calculated at  $x = 1$  for various methods

It has long been observed that orthogonal collocation at Lobatto points gives superior accuracy compared to the same number of Gauss points for problems with Dirichlet boundary conditions. Lobatto points were observed to give better accuracy for the solution and flux quantities if they can be calculated by integration, e.g. Eq. (26). As we found here, fluxes calculated from Eq. (29) were found to give poor accuracy long ago. For these reasons, the use of Lobatto points was restricted to symmetric problems with specified values on the boundaries. For a nonsymmetric problem, integration, e.g. Eq. (28), gives the sum of the fluxes, whereas the individual flux values are normally required. It is presumably for this reason that Lobatto points for nonsymmetric problems has not previously been discussed. However, the Figures and Table 2 clearly show that accurate solutions and accurate fluxes can be calculated with Lobatto points for this nonsymmetric problem. To our knowledge, the correct method for flux calculations, e.g. Eq.(31), has not been described previously. We will find that Lobatto points give similar advantages for problems with flux boundary conditions, provided the boundary conditions are correctly implemented.