A generalized diagonal mass matrix spectral element method for non-quadrilateral elements

Mark A. Taylor [a] B. A. Wingate [b]

[a] *National Center Atmospheric Research, PO Box 3000, Boulder CO 80307, (taylorm@ucar.edu)*

[b] *MS B256, Los Alamos National Laboratory, Los Alamos New Mexico, (wingate@lanl.gov)*

Abstract

We introduce a Fekete point spectral element method. This is a generalization of the traditional quadrilateral based spectral element method to any general element such as triangles. It retains the exponential convergence and the diagonal mass matrix of the original method. We first solve a Sturm-Liouville problem in the square and the triangle to determine the correct functional space used for approximation. Once the functional space is known, we use the Fekete criterion to compute near optimal grids for these spaces which have the same number of points as the dimension of the functional space. This allows the construction of a well-behaved cardinal function basis which leads to a diagonal mass matrix.

Key words: triangle, Lebesgue constant, spectral elements, h-p finite elements, Fekete points

1 Introduction

The spectral finite element method [9] is a spectrally accurate algorithm for solving differential equations on unstructured grids. Typically the computational domain is broken up into quadrilateral elements. Within each of these elements all variables are approximated by high degree polynomial expansions. The discrete equations are then derived using an integral form of the equations to be solved. When used with conforming elements and a clever choice of test functions and collocation points, the resulting mass matrix is diagonal [8], which leads to a computationally efficient method. This property allows initial value problems like the shallow water and primitive equations used in geophysical modeling to be treated fully explicitly.

At present, the diagonal mass matrix spectral element method is only available with conforming quadrilateral grids. This is because the method relies on Gauss-like quadrature formula where there are the same number of quadrature points as the dimension of the functional space. It is generally believed that such optimal quadrature formulas do not exist in the triangle. In this work we propose a more general derivation of the diagonal mass matrix spectral element method based on Fekete points. For quadrilateral elements, this new derivation is identical to the conventional spectral element method. But it also allows many other types of elements like triangles, hexagons and tetrahedra.

2 The spectral finite element method

The diagonal mass matrix spectral element method is a standard finite element method with particular choices of basis functions, test functions, and high order quadrature for the evaluations of integrals. We now follow [8] and repeat the standard spectral element discretization, but in a more general way which will allow us to extend the method to other domains like triangles. We start with the simple linear advection equation

$$u_t = -\mathbf{v} \cdot \nabla u.$$

We solve this equation for u, in a channel D with inflow boundary conditions as in [12]. The velocity field $\mathbf{v}$ is given, with no flow through the boundary. We first write the weak or integral form of the equations. We multiply by a test function ψ and integrate over the entire domain D

$$\int\limits_D u_t \psi \, dA = -\int\limits_D \psi \mathbf{v} \cdot \nabla u \, dA,$$

where dA is the usual area measure. The next step is to decompose the domain D into elements D_k. We will be interested in the case were each D_k can easily be mapped to a reference element Ω. We will consider the case where Ω is either a square or triangle. The integral form of the equation can then be written

$$\sum_k \int\limits_\Omega u_t \psi J^k \, dA = -\sum_k \int\limits_\Omega (\mathbf{v} \cdot \nabla u) \, \psi J^k \; dA, \tag{1}$$

where J^k is the Jacobian of the mapping from our reference element Ω to D_k. To complete the spatial discretization, we must determine a set of basis functions to represent our unknown variables within each element, a set of global test functions for the integral equations and a procedure for computing the resulting integrals. In a diagonal mass matrix spectral element method, these functions and quadrature procedure are completely determined by two choices:

(1) The functional space $\mathcal{P}_N$ which is spanned by the basis functions within an element. Let $N = \dim \mathcal{P}_N$.
(2) Collocation points within each element. There must by exactly N points, with their convex hull forming the boundary of our element.

The choice of functional space $\mathcal{P}_N$ will be covered in Section 4, while the choice of collocation points will be covered in Section 5.

Once the functional space $\mathcal{P}_N$ and associated collocation points are chosen, the diagonal mass matrix spectral element discretization proceeds as follows. Since we use the same number of collocation points as the dimension of $\mathcal{P}_N$, we can construct a cardinal function basis for $\mathcal{P}_N$. If $\mathcal{P}_N$ is a polynomial truncation, then the cardinal functions will be the Lagrange interpolating polynomials. With these basis functions, it is possible to construct continuous, piecewise polynomial (assuming $\mathcal{P}_N$ is a polynomial space) global test functions defined over all the elements. At the collocation points, these test functions will be discrete delta functions. The final step in the derivation of the method is to evaluate the integrals with inexact, but high order quadrature at our collocation points. When we allow for this error in the computation of the integrals, we get a remarkable simplification: a diagonal mass matrix. This simple derivation is in [8] and for arbitrary space/collocation points in [11]. Finally, we note that the cardinal function basis is used only in the derivation of the method. In a numerical method, we would use a well conditioned and easy to compute basis.

3 Interpolation

Diagonal mass matrix spectral element methods rely heavily on interpolation and thus the interpolation properties of our discretization are very important. This is because, as mentioned above, they require the same number of collocation points as the dimension of the functional space $\mathcal{P}_N$ being used to represent functions. Spectral transforms which respect this restriction (such as the FFT, Chebyshev and Legendre transforms) are equivalent to interpolation at the collocation points by the space spanned by the underlying basis functions. In this section we derive the classic bound on the max-norm interpolation error, which is then used to motivate our choice of $\mathcal{P}_N$ and collocation points in the following sections.

Let $\mathcal{P}_N$ be an N dimensional vector space made up of polynomials, and let Ω be some domain such as the square or triangle. Take N points, z_i, which are solvable in $\mathcal{P}_N$. That is, given an arbitrary function f, there is a unique function $g \in \mathcal{P}_N$ where $g(z_i) = f(z_i)$. We will denote the interpolating polynomial g by $I_N(f(z))$. We define the max norm of functions and operators in the usual way,

$$\|f\| = \max_{z\in\Omega} |f(z)|, \qquad \|I_N\| = \max_{\|f\|=1} \|I_N(f)\|.$$

The quantity $\|I_N\|$ is usually called the Lebesgue constant.

One can then ask how well does $I(f)$ approximate f? To see this, take a function $h \in \mathcal{P}_N$ which best approximates f in the max norm. The function h is not necessarily $I_N(f)$, but $h = I_N(h)$. Thus,

$$\begin{aligned}
\|f - I_N(f)\| &= \|f - h + I_N(h) - I_N(f)\| \\
&\le \|f - h\| + \|I_N\|\,\|h - f\| \\
&\le (1 + \|I_N\|)\,\|f - h\|,
\end{aligned}$$

and we see that our max-norm error has two components:

(1) The first component, $(1 + \|I_N\|)$ is controlled by the Lebesgue constant. Finding points in the triangle with a small Lebesgue constant will be addressed in Section 5.
(2) The second component, $\|f-h\|$ is determined solely by how well our finite dimensional functional space $\mathcal{P}_N$ can approximate smooth functions f. Finding such functional spaces will be done using Sturm-Liouville theory in Section 4.

We conclude this section with two examples in the well understood domain of the interval $[-1, 1]$. Take $\mathcal{P}_N$ to be the span of the first N trigonometric polynomials, and take our N collocation points to be equally spaced in $[-1, 1]$. In this example, the interpolating trigonometric polynomial can be computed very efficiently with the FFT. Also, the Lebesgue constant is known to be optimally small. However, if our functions of interest f are not periodic, the max norm $\|f - h\|$ will be $O(1)$ because of Gibbs oscillations and we will not have convergence as N is increased. For the other extreme, take the same equally spaced collocation points but let $\mathcal{P}_N$ be the space of polynomials of up to degree $N - 1$. For smooth functions f, $\|f - h\|$ converges quite rapidly. But in this case the Lebesgue constant is known to grow exponentially in N, and again our interpolation error will not converge. This is manifested by the well known Runge phenomena, where the interpolating polynomial has wild oscillations between the interpolation points.

4 Choosing a functional space

Most classical basis functions used in numerical work are given by eigenfunctions of Sturm-Liouville problems. Thus we are motivated to find basis functions by looking at appropriate Sturm-Liouville problems for the square and the triangle. The space $\mathcal{P}_N$ can then be taken as the span of these eigenfunctions. We follow [13] and define a natural basis for a domain Ω as one in which the convergence of the expansion of an arbitrary function f is dictated only by the regularity of f. Furthermore the expansion will converge at a rate faster than any polynomial for $f \in C^\infty$. Such a basis can be constructed by solving a self adjoint (over C^∞) Sturm-Liouville problem in Ω.

In the square $[-1,1] \times [-1,1]$, the classical Sturm-Liouville problem is given by

$$\frac{\partial}{\partial x}\left[(1-x^2)\frac{\partial u}{\partial x}\right] + \frac{\partial}{\partial y}\left[(1-y^2)\frac{\partial u}{\partial y}\right] + \lambda u = 0.$$

The eigenfunctions of this equation $g_{mn}(x,y)$ are given by a tensor product of Legendre polynomials, $g_{mn}(x,y) = P_m(x)P_n(y)$, with eigenvalues $\lambda_{mn} = m(m+1)+n(n+1)$. The g_{mn} are orthogonal in the square with respect to the area measure dA. The tensor product nature of g_{mn} and λ_{mn} suggests taking $\mathcal{P}_N$ to be the span of the eigenfunctions g_{mn} with $m(m+1) < C$ and $n(n+1) < C$, for some constant C. This leads to the standard diamond polynomial truncation used in most non-periodic quadrilateral spectral methods. We call this the diamond truncation because of its shape when pictorially represented in Pascal's triangle [13]. It is given by the tensor product of polynomials of x with polynomials in y of at most degree d. The dimension of this space is $N = (d+1)^2$.

There are three properties of this equation we wish to maintain when constructing a Sturm-Liouville problem for other domains. First, the equation (and the resulting eigenspaces) is invariant under the symmetry group of the square, D_4 (permutations of x and y). Secondly, it is self-adjoint since the coefficients of each directional derivative vanish along boundaries not tangent to that direction. Finally, we have chosen a uniform weighting so that our eigenfunctions will be orthogonal with respect to the area measure dA.

For the triangle, we seek the simplest possible equation which has all these properties. This is most easily constructed in a local coordinate system with variables a, b and c representing directions tangent to each edge of an equilateral triangle. The equation is

$$\frac{\partial}{\partial a}\left[(a_\perp^2 - 3a^2)\frac{\partial u}{\partial a}\right] + \frac{\partial}{\partial b}\left[(b_\perp^2 - 3b^2)\frac{\partial u}{\partial b}\right] + \frac{\partial}{\partial c}\left[(c_\perp^2 - 3c^2)\frac{\partial u}{\partial c}\right] + 3\lambda u = 0.$$

In this coordinate system it is easy to see that the equation is invariant under D_3, the symmetry group of the traingle (permutations of a, b, c). Furthermore, we have chosen quadratic coefficients of each directional derivative which vanish along boundaries not tangent to the directional derivative. This guarantees the equation will be self adjoint.

This equation can be solved after transforming to Cartesian coordinates. Remarkably, the eigenfunctions g_{mn} turn out to be the Koornwinder-Dubiner polynomials, and the eigenvalues are $\lambda_{mn} = (m+n)(m+n+2)$ [13]. The coupling between m and n for these eigenvalues leads us to suggest $\mathcal{P}_N$ to be the span of all g_{mn} with $\lambda_{mn} < C$ for some constant C. When viewed in Pascal's triangle, this space is a triangular truncation of polynomials, given by the space of all polynomials in two variables of up to a certain degree d. The dimension of this space is $N = (d+1)(d+2)/2$. These results generalize to tetrahedron and higher dimensional simplexes [13].

5 Choosing collocation points

We now must find a set of collocation points in our domain which have a small Lebesgue constant for $\mathcal{P}_N$. Lebesgue points (the points with the smallest Lebesgue constant) would be an obvious choice, however almost nothing seems to be known about Lebesgue points in more than one dimension. Nor are we aware of a feasible method for computing them numerically. Choices such as equally spaced points (in the square or triangle), or a product Gaussian grid (restricted to the triangle) lead to disasters. The Lebesgue constants for these points grow exponentially fast with N. If the points are not chosen very carefully, the interpolating polynomial will have wild oscillations between the collocation points, a fundamental problem which can not be overcome by after-the-fact preconditioning.

For the triangle, there are other tractable alternatives to Lebesgue points. The ones we prefer are Fekete points. References for Fekete points can be found in the recent paper [1], which summarizes many of the known results and gives some open questions. Fekete points are characterized by the points which maximize the determinant of a generalized Vandermonde matrix. They are defined for any reasonable domain and space $\mathcal{P}_N$. They have been computed for the triangle in [2,5,11]. If the steepest descent method from [11] is used, the ill-conditioned Vandermonde determinant need not be calculated.

Other approaches include directly minimizing a variant of the Lebesgue constant [5] and using collocation points based on a minimum energy electrostatic problem [7]. These points have the best Lebesgue constant of any points known in the triangle for degree $d \leq 10$. For degree $d > 10$, the points with the smallest yet known Lebesgue constant are computed in [11]. Exponential convergence of interpolation and differentiation on these grids is shown in [11].

In addition to a small Lebesgue constant in the triangle, there are several other reasons for the choice of Fekete points. For the interval $[-1, 1]$, with $\mathcal{P}_N$ polynomials up to degree d, Fejér showed that Fekete points are the Gauss-Lobatto quadrature points [6]. More recently, Bos has shown that Fekete points in the square with $\mathcal{P}_N$ the usual diamond polynomial truncation are a tensor product of the Gauss-Lobatto quadrature points [4]. This proof is more difficult than it seems because one is not allowed to assume the Fekete points have any structure (such as a tensor product). Thus a Fekete point spectral element method is identical to the conventional Gauss-Lobatto quadrature point spectral element method. Also, theoretical work [3] and numerical evidence [11] suggests that Fekete points along the boundary of the triangle are the Gauss-Lobatto points, making Fekete point triangular and quadrilateral elements naturally conform. Finally, cardinal functions defined at Fekete points are guaranteed to have a maximum value of 1 over the domain, and thus have no wild oscillations between the points.

6 Preliminary results and Summary

Preliminary results for Equation 1 using a filtered Adams-Bashforth time stepping scheme are presented in Figure 1. This figure shows that for a given degree the method is not as accurate as the Dubiner method [10,12] on triangles. But the two methods converge exponentially fast at about the same rate. The tremendous savings gained by not having to invert a mass matrix allows the Fekete method to achieve errors comparable to those from the Dubiner method at less computation cost.

In conclusion, we have outlined a Fekete point spectral element method which can be extended to non-quadrilateral domains such as the triangle. We have given preliminary results which show exponential convergence. The performance of this method for non-linear equations and stability considerations is a topic of current research.

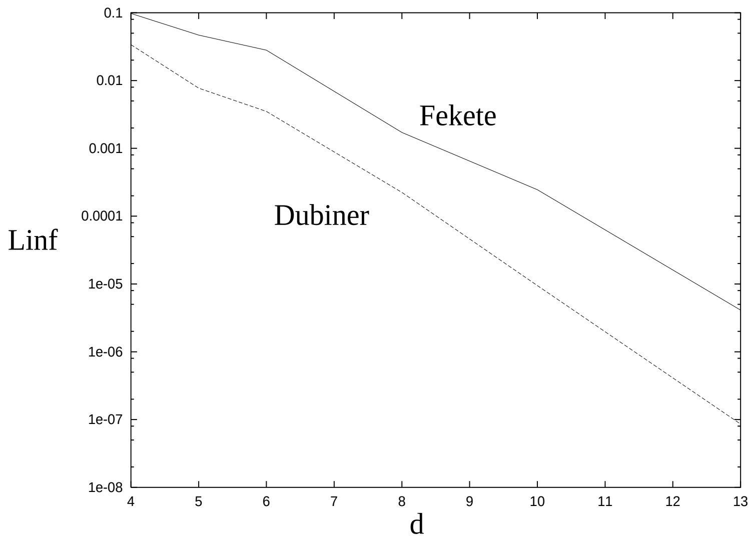

Fig. 1. *For a Gaussian initial condition from [12] we show the L_∞ error from Equation 1 after one period. The grid has 30 right triangles with polynomial degree d within each triangle. The fully explicit Fekete method converges at about the same rate as the Dubiner method.*

References

[1] T. Bloom, L. Bos, C. Christensen and N. Levenberg, *Rocky Mountain J. Math.* **22** (1992) 441.

[2] L. Bos, *J. Approx. Theory* **38** (1983) 43.

[3] L. Bos, *J. Approx. Theory* **64** (1991) 271.

[4] L. Bos, personal communication.

[5] Q. Chen and Ivo Babuška, *Comput. Methods Appl. Mech. Engrg.* **128** (1995) 405.

[6] L. Fejér, *Ann. Scuola Norm. Sup. Pisa Sci. Fis. Mt. Ser. II* **1** (1932) 263.

[7] J. Hesthaven, *SIAM J. Numer. Anal.* **35** (1998) 655.

[8] Y. Maday and A. T. Patera, in *State of the Art Surveys in Computational Mechanics.* A. K. Noor ed. ASME, New York, 1988, p. 71.

[9] A. T. Patera, *J. Comput. Phys.* **54,** (1984) 468.

[10] S. J. Sherwin and G.E. Karniadakis, *Comp. Meth. Appl. Mech. Eng.* **123** (1995) 189.

[11] M. A. Taylor and B. A. Wingate, *SIAM J. Numer. Anal.* submitted.

[12] B. Wingate and J. P. Boyd in: *Proceedings of the Third International Conference on Spectral and High Order Methods* (1995) p 305.

[13] B. A. Wingate and M. A. Taylor, *SIAM J. Numer. Anal.* submitted.