

Error Analysis of p-Version Discontinuous Galerkin Method for Heat Transfer in Built-Up Structures

Hideaki Kaneko*

Department of Mathematics and Statistics

Old Dominion University

Norfolk, Virginia 23529-0077

Kim S. Bey

Thermal Structures Branch

Structure Division

NASA Langley Research Center

Hampton, VA 23681

*This author is supported by NASA- Grant NAG-1-01092

**Error Analysis of p -version discontinuous Galerkin method
for heat transfer in built-up structures**

H. Kaneko and K. S. Bey

Abstract

The purpose of this paper is to provide an error analysis for the p -version of the discontinuous Galerkin finite element method for heat transfer in built-up structures. As a special case of the results in this paper, a theoretical error estimate for the numerical experiments recently conducted by James Tomey [7] is obtained.

Key words: Discontinuous Galerkin Method, Parabolic Equations, p -Finite Element Method.

1 Introduction.

The purposes of this paper are to report the state of the art information on time discretization techniques in the discontinuous Galerkin method for parabolic problems (this section) and to establish error analysis for p -version of the finite element method for such problems (section 2). Also a discussion of various time discretization techniques are included (section 3). The discontinuous Galerkin method is applied to the following standard model problem of parabolic type:

Find u such that

$$\begin{aligned} u_t(x, t) - \Delta u(x, t) &= f(x, t), & x \in \Omega, t \in R^+, \\ u(x, t) &= 0, & x \in \partial\Omega, t \in R^+, \\ u(x, 0) &= u_0(x), & x \in \Omega, \end{aligned} \tag{1.1}$$

where Ω is a closed and bounded set in R^3 with boundary $\partial\Omega$, $R^+ = (0, \infty)$, $\Delta u = \partial^2 u / \partial x^2 + \partial^2 / \partial y^2 + \partial^2 u / \partial z^2$, $u_t = \partial u / \partial t$, and the functions f and u_0 are given data. In this paper, region Ω is assumed to be a thin body in R^3 , such as a panel on the wing or fuselage of an aerospace vehicle. p -version of the finite element method is considered in all directions including time variable. Because of the special characteristics of the region, it is assumed that, through the thickness, only one element is taken. This allows us to avoid construction of finite elements

that are too thin to violate the quasiuniform condition (see (1.11) below). The recent paper [5] addresses the similar issue under the framework of the ‘modified’ hp -finite element scheme. For a Banach space X and $I = (0, T)$ indicating a time interval, we denote by $L^p(I; X)$, $1 \leq p \leq \infty$, and $H^k(I; X)$, $0 \leq k \in \mathbb{R}$, the Lebesgue and Sobolev spaces. Also $P^p(I; X)$ denotes the set of all polynomials of degree $\leq p$ with coefficients in X , i.e., $q(t) \in P^p(I; X)$ if and only if $q(t) = \sum_{j=0}^p x_j t^j$ for some $x_j \in X$ and $t \in I$. Let T_I be a partition of I into $M(I)$ subintervals $\{I_n = (t_{n-1}, t_n)\}_{n=1}^{M(I)}$. We set $k_n = t_n - t_{n-1}$. Denote by u_n^+ and u_n^- the right and the left limits of u at t_n respectively. We set $[u]_n = u_n^+ - u_n^-$, $n = 1, \dots, M(I) - 1$. For each time interval I_n , an approximation order $p_n \geq 0$ is assigned and they are stored in the vector $\bar{p} = \{p_n\}_{n=1}^{M(I)}$. The semidiscrete space is then given by

$$V^{\bar{p}}(T_I; X) = \{u: I \rightarrow X: u|_{I_n} \in P^{p_n}(I_n; X), 1 \leq n \leq M(I)\}.$$

If \bar{p} is a constant vector, i.e., $p_n = p$ for all $1 \leq n \leq M(I)$, then we write $V^p(T_I; X)$. The number of degrees of freedom of the time discretization will be denoted by $\text{NRDOF}(V^{\bar{p}}(T_I; X))$ and it is, of course, equal to $\sum_{n=1}^{M(I)} (p_n + 1)$. The semidiscrete solution $U \in P^{p_n}(I_n; X)$ of the problem (1.1), if U is already determined on I_k , $1 \leq k \leq n - 1$, is found by solving the following problem:

Find $U \in P^{p_n}(I_n; X)$ such that

$$\int_{I_n} \{(U_t, V)_2 + (\nabla U, \nabla V)_2\} dt + (U_{n-1}^+, V_{n-1}^+)_2 = \int_{I_n} (g, V)_{X^* \times X} dt + (U_{n-1}^-, V_{n-1}^+)_2 \quad (1.2)$$

for all $V \in P^{p_n}(I_n; X)$ and $U_0^- = u_0$.

Here we assume that $L^2(\Omega)$ is densely embedded in a Banach space X . The following theorem and its subsequent corollary are reported in [10]. Theorem describes the error estimate for the semidiscrete solution explicitly in terms of the time steps, the approximation orders and the local regularities of the solution.

Theorem 1.1 *Let u be the solution of (1.1) and U the semidiscrete solution in $V^{\bar{p}}(T_I; X)$. Assume that $u|_{I_n} \in H^{s_n+1}(I_n; X)$ for $1 \leq n \leq M(I)$ and s_n , $1 \leq n \leq M(I)$, nonnegative integers. Then*

$$\|u - U\|_{L^2(I; X)}^2 \leq C \sum_{n=1}^{M(I)} \left(\frac{k_n}{2}\right)^{2(s_n+1)} \max(1, p_n)^{-2} \frac{\Gamma(p_n + 1 - s_n^*)}{\Gamma(p_n + 1 + s_n^*)} \|u\|_{H^{s_n^*+1}(I_n; X)}^2$$

for any $0 \leq s_n^* \leq \min(p_n, s_n)$.

In the p -version of the discontinuous Galerkin finite element method, a temporal partition T_I is fixed and convergence is obtained by $p_n \rightarrow \infty$. For $p_n = p$ for each $n = 1, \dots, M(I)$, and for a smooth solution u , we obtain the following:

Corollary 1.2 *Let $p_n = p$ for each $n = 1, \dots, M(I)$ and $k_{max} = \max\{k_n\}$. Let $u \in H^{s_0+1}(I; X)$, for a nonnegative integer s_0 , be the exact solution of (1.1) and $U \in V^p(T_I; X)$ the semidiscrete solution. Then*

$$\|u - U\|_{L^2(I; X)}^2 \leq C \frac{k_{max}^{2\{\min(s_0, p)+1\}}}{\max(1, p)^{2(s_0+1)}} \|u\|_{H^{s_0+1}(I; X)}^2,$$

where C depends on s_0 , but is independent of k_{max} and p .

Remark 1.1: As pointed out in [10], Corollary 1.2 shows that for smooth solutions where s_0 is large, it is better to increase p than to reduce k_{max} at a fixed, often low p . Since $N \equiv \text{NRDOF}(V^{\bar{p}}(T_I; X)) \sim p$, we see that for p -version of the finite element method,

$$\|u - U\|_{L^2(I; X)} \leq CN^{-s_0-1}. \quad (1.3)$$

Using the standard approximation theory for analytic functions (1.3) reduces to

$$\|u - U\|_{L^2(I; X)} \leq Ce^{-bp}, \quad (1.4)$$

for some $b > 0$ independent of p . If the solution is not smooth in time, it is still possible to approximate it in exponential orders by a hp -finite element method which combines a certain geometric partition with the semidiscrete space $V^{\bar{p}}(T_I; X)$ where \bar{p} is linearly distributed, (see [10] for detail). Using the h -finite element method with non-uniform graded time partitions, such non-smooth solutions can be approximated in an algebraically optimal order, but not exponentially, using different approaches, (see, e.g., [5], [10]). The standard p -finite element method does not perform well in this context. Therefore, since we aim to establish the p -version of the finite element method for (1.1), we assume for the remaining of this paper that solutions are smooth in time. This assumption allows us to establish the order of approximation in time variable that is compatible with the approximation orders in the spatial variables.

Now we consider the problem of discretizing the space Ω . For simplicity, we assume that

$$\Omega = \omega \times \left[-\frac{d}{2}, \frac{d}{2}\right] = \sum_{l=1}^{M(h)} K^l \times \left[-\frac{d}{2}, \frac{d}{2}\right],$$

where ω is divided into $M(h)$ number of triangular elements, each denoted by K^l . Let $K_{\xi,\eta}$ denote the master triangular element defined by

$$K_{\xi,\eta} = \{(\xi,\eta) \in R^2: \begin{array}{ll} 0 \leq \eta \leq (1+\xi)\sqrt{3} & -1 \leq \xi \leq 0 \text{ or} \\ 0 \leq \eta \leq (1-\xi)\sqrt{3} & 0 \leq \xi \leq 1 \end{array}\}.$$

Let $S^p(K_{\xi,\eta})$ denote the space of polynomials of degree $\leq p$ on $K_{\xi,\eta}$, -i.e.,

$$S^p(K_{\xi,\eta}) = \text{span}\{\xi^i \eta^j: i, j = 0, 1, \dots, p; i + j \leq p\}.$$

First, the shape functions for the master element $K_{\xi,\eta}$ are formed. To accomplish this, the barycentric coordinates are introduced via

$$\lambda_1 = (1 - \xi - \eta/\sqrt{3})/2, \quad \lambda_2 = (1 + \xi - \eta/\sqrt{3})/2, \quad \lambda_3 = \eta/\sqrt{3}.$$

λ_i 's form a partition of unity and λ_i is identically equal to one at a vertex of $K_{\xi,\eta}$ and vanishes on the opposite side of $K_{\xi,\eta}$. The hierarchical shape functions on $K_{\xi,\eta}$ consists of internal as well as external functions. The normalized antiderivatives of the Legendre polynomials are defined by

$$\bar{\psi}_i(\zeta) = \sqrt{\frac{2i+1}{2}} \int_{-1}^{\zeta} P_i(t) dt, \quad i = 1, 2, 3, \dots$$

Now, the external shape functions consist of 3 nodal shape functions

$$N_i(\xi, \eta) = \lambda_i, \quad i = 1, 2, 3,$$

and $3(p-1)$ side shape functions $N_i^{[j]}(\xi, \eta)$, $i = 1, \dots, p-1$, $j = 1, 2, 3$. The index j indicates one of three sides of $K_{\xi,\eta}$. Noting that $\bar{\psi}_i(\pm 1) = 0$,

$$\bar{\psi}_i(\eta) = \frac{1}{4}(1 - \eta^2)\varphi_i(\eta), \quad i = 1, 2, 3, \dots \quad (1.4)$$

where $\varphi_i(\eta)$ is a polynomial of degree $i-1$. For instance, $\varphi_1(\eta) = -\sqrt{6}$, $\varphi_2(\eta) = -\sqrt{10}\eta$ and $\varphi_3(\eta) = \frac{\sqrt{14}}{4}(1 - 5\eta^2)$, etc. The side shape functions are constructed as follows:

$$\begin{aligned} N_i^{[1]}(\xi, \eta) &= \lambda_2 \lambda_3 \varphi_i(\lambda_3 - \lambda_2) \\ N_i^{[2]}(\xi, \eta) &= \lambda_3 \lambda_1 \varphi_i(\lambda_1 - \lambda_3), \quad i = 1, \dots, p-1, \\ N_i^{[3]}(\xi, \eta) &= \lambda_1 \lambda_2 \varphi_i(\lambda_2 - \lambda_1). \end{aligned} \quad (1.5)$$

From (1.4) and (1.5), there are $3 + 3(p-1)$ shape functions. As $\dim(S^p(K_{\xi,\eta})) = \frac{(p+1)(p+2)}{2}$, the remaining $\frac{(p-1)(p-2)}{2}$ basis elements are constructed in terms of internal shape functions.

Clearly, nontrivial internal shape functions on $K_{\xi,\eta}$ exists only if $p \geq 3$. For $p = 3$, the bubble function on $K_{\xi,\eta}$ defined below serves as the internal function;

$$b_{K_{\xi,\eta}}(\xi, \eta) = \lambda_1 \lambda_2 \lambda_3 = \frac{\eta}{4\sqrt{3}} \left(\left(1 - \frac{\eta}{\sqrt{3}}\right)^2 - \xi^2 \right). \quad (1.6)$$

Moreover, the collection $I^p(K_{\xi,\eta})$ of higher-order internal shape functions can be constructed from

$$I^p(K_{\xi,\eta}) = \{b_{K_{\xi,\eta},v}: v \in S^{p-3}(K_{\xi,\eta})\} = \{b_{K_{\xi,\eta}}\} \otimes S^{p-3}(K_{\xi,\eta}), \quad p \geq 3.$$

Let T_h , $h > 0$, be a triangulation of ω . let $x = Q_x^l(L_1, L_2, L_3)$ and $y = Q_y^l(L_1, L_2, L_3)$ be the element mappings of the standard triangle $K_{\xi,\eta}$ to the l th triangular element $K^l \in T_h$, e.g., the linear mapping onto K^l with vertices $\{(x_i^l, y_i^l)\}_{i=1}^3$,

$$Q_x^l(L_1, L_2, L_3) = \sum_{i=1}^3 x_i^l L_i, \quad Q_y^l(L_1, L_2, L_3) = \sum_{i=1}^3 y_i^l L_i.$$

The space of all polynomials of degree $\leq p$ on K^l is denoted by $S^p(K^l)$ and its basis can be formed from the shape functions of $S^p(K_{\xi,\eta})$ described above by transforming them under Q_x^l and Q_y^l . The finite element space $S^{p,l}(\omega, T_h)$ is now defined. For ω , $p \geq 0$ and $k \geq 0$,

$$S^{p,k}(\omega, T_h) = \{u \in H^k(\omega): u|_K \in S^p(K), K \in T_h\}. \quad (1.7)$$

Assume that a triangulation $\{T_h\}$, $h > 0$, of ω consists of $\{K_h^l\}_{l=1}^{M(h)}$ and that $h_l = \text{diam}(K_h^l)$, for $l = 1, \dots, M(h)$.

In the z -variable for through the thickness approximation, the local variable τ is defined in the reference element $[-1, 1]$ and Γ is mapped onto the reference element by Q_z , i.e.,

$$\Gamma = Q_z([-1, 1]), \quad z = Q_z(\tau).$$

Clearly, Q_z is a linear function defined by

$$z = Q_z(\tau) = \frac{1}{2}(1 - \tau)\left(-\frac{d}{2}\right) + \frac{1}{2}(1 + \tau)\frac{d}{2}, \quad \tau \in [-1, 1]$$

The Jacobian of Q_z is constant

$$\frac{dz}{d\tau} = \frac{d}{2}.$$

In this paper, the basis functions of $P_p([-1, 1])$ are taken to be the one-dimensional hierarchical shape functions. See [8] for a complete discussion of the basis elements used in the p and hp -finite element methods.

For example, in approximating an element in $H^l[-1, 1]$, with $l = 0$, $\psi_i(\tau) = P_{i-1}(\tau)$, $1 \leq i \leq p + 1$, where P_{i-1} is the Legendre polynomial of degree $i - 1$, form the hierarchical basis functions. With $l = 1$, the external (ψ_1 and ψ_2) and internal (ψ_i , $i \geq 3$) shape functions are defined by

$$\begin{aligned}\psi_1(\tau) &= \frac{1-\tau}{2}, & \psi_2(\tau) &= \frac{1+\tau}{2} \\ \psi_i(\tau) &= \left(\frac{2i-3}{2}\right)^{1/2} \int_{-1}^{\tau} P_{i-2}(t) dt, & 3 \leq i \leq p+1.\end{aligned}\tag{1.8}$$

Note that ψ_i 's form an orthogonal family with respect to the energy inner product $(\cdot, \cdot)_E$,

$$(\psi_i, \psi_j)_E \equiv \int_{-1}^1 \psi_i'(t) \psi_j'(t) dt = \int_{-1}^1 P_i(t) P_j(t) dt = \delta_{ij}.$$

Also note that the internal shape functions satisfy

$$\psi_i(\pm 1) = 0, \quad \text{for } 3 \leq i \leq p + 1.$$

For the case $l = 2$ and $p \geq 3$, the four nodal shape functions and the remaining $p - 3$ internal shape functions given by

$$\begin{aligned}\psi_1(\tau) &= \frac{1}{4}(1-\tau)^2(1+\tau), & \psi_2(\tau) &= \frac{1}{4}(1-\tau)^2(2+\tau) \\ \psi_3(\tau) &= -\frac{1}{4}(1+\tau)^2(1-\tau), & \psi_4(\tau) &= \frac{1}{4}(1+\tau)^2(2-\tau) \\ \psi_i(\tau) &= \left(\frac{2i-5}{2}\right)^{1/2} \int_{-1}^{\tau} (\tau-\eta) P_{i-3}(\eta) d\eta, & i &= 5, \dots, p+1.\end{aligned}\tag{1.9}$$

In this case, the internal shape functions satisfy

$$\frac{d^j \psi_i}{d\tau^j}(\pm 1) = 0, \quad \text{for } 5 \leq i \leq p + 1 \text{ and } j = 0, 1.$$

For example, using the shape functions in (1.8), any element $u \in H^1[-1, 1]$ can be approximated by $u_p \in P_p([-1, 1])$, in the form

$$u_p(\tau) = \frac{1-\tau}{2} u(-1) + \frac{1+\tau}{2} u(1) + \sum_{i=3}^{p+1} a_i \psi_i(\tau).\tag{1.10}$$

For approximating the solutions of parabolic problems with the homogeneous Dirichlet boundary condition, the first two terms will be dropped, as $u(-1) = u(1) = 0$. A sequence of triangulations $\{T_h\}_{h>0}$ is called the quasiuniform mesh if

$$\frac{h}{\text{diam}(K)} \leq \gamma, \quad \text{for all } h > 0,\tag{1.11}$$

with $h = \max_{K \in T_h} \text{diam}(K)$, and for some $\gamma > 0$. $P_p(\Gamma)$ denotes the space of all polynomials of degree $\leq p$ defined on Γ .

The following is proved by Babuška, Szabo and Katz in [1]. See also [2] by Babuška and Suri on a related discussion. Here Ω_0 denotes a bounded polygonal domain in R^2 .

Theorem 1.3 *Let $u \in H^k(\Omega_0)$. Then there exists a sequence $z_p \in P^p(\Omega_0)$, $p = 1, 2, \dots$, such that, for any $0 \leq l \leq k$,*

$$\|u - z_p\|_{l, \Omega_0} \leq Cp^{-(k-l)} \|u\|_{k, \Omega_0},$$

where C is independent of u and p .

The parameters k and l are not necessarily integral. Their proof relies heavily on the approximation power of the trigonometric polynomials.

With $l = 0$ in Theorem 1.3 and using the usual duality argument, the results in Theorem 1.3 are further improved by Babuška and Suri in [2] (theorem 2.9), (see also a series of papers by Gui and Babuška [9]), to the hp -finite element setting as follows:

Theorem 1.4 *Let T_h be a quasiuniform partition of Ω_0 . Then for $k \geq 1$, $u \in H^k(\Omega_0)$,*

$$\inf_{v \in S^{p,k}(\omega, T_h)} \|u - v\|_{L_2(\Omega_0)} \leq Ch^\nu p^{-k} \|u\|_{H^k(\Omega_0)}$$

where $\nu = \min(k, p + 1)$.

Note that Corollary 1.2 can be derived also from Theorem 1.4, and they establish the same result in terms of time variable approximation and spatial variable approximation, respectively. The corresponding error estimate in the $\|\cdot\|_{H^k(\Omega_0)}$ is also available in [2]. Now assume that Ω_0

in Theorems 1.3 and 1.4 is $\Omega = \omega \times [-\frac{d}{2}, \frac{d}{2}]$ and consider the problem of approximating elements in $H^k(\omega \times [-\frac{d}{2}, \frac{d}{2}])$ by the tensor product space $S^{p_1, k}(\omega, T_h) \otimes P^{p_2}[-\frac{d}{2}, \frac{d}{2}]$. Using Theorems 1.3 and 1.4, the following is proved in [5]:

Theorem 1.5 *Let $u \in H^k(\omega \times [-\frac{d}{2}, \frac{d}{2}])$. Then there exists $u^* \in S^{p_1, k}(\omega, T_h) \otimes P^{p_2}([-\frac{d}{2}, \frac{d}{2}])$,*

$$\|u - u^*\|_{L_2(\omega \times [-\frac{d}{2}, \frac{d}{2}])} = O(h^{\nu_1} p_1^{-k} + d^{\nu_2} p_2^{-k}),$$

where $\nu_i = \min(k, p_i + 1)$ for $i = 1, 2$ and $h = \max_{K \in T_h} \text{diam}(K)$, with T_h a triangulation of ω .

A straightforward extension of Theorem 1.5 provides the following which describes a total error estimate for approximating functions in $H^k(\omega \times [-\frac{d}{2}, \frac{d}{2}] \times I)$ by elements from $S^{p_1, k}(\omega, T_h) \otimes PP^2[-\frac{d}{2}, \frac{d}{2}] \otimes PP^3(I)$.

Theorem 1.6 *Let $u \in H^k(\omega \times [-\frac{d}{2}, \frac{d}{2}] \times I)$. Then there exists $\bar{u} \in S^{p_1, k}(\omega, T_h) \otimes PP^2([-\frac{d}{2}, \frac{d}{2}]) \otimes PP^3(I)$,*

$$\|u - \bar{u}\|_{L_2(\omega \times [-\frac{d}{2}, \frac{d}{2}] \times I)} = O(h^{\nu_1} p_1^{-k} + d^{\nu_2} p_2^{-k} + k_{max}^{\nu_3} p_3^{-k}),$$

where $\nu_i = \min(k, p_i + 1)$, $i = 1, 2, 3$ and $h = \max_{K \in T_h} \text{diam}(K)$, with T_h a triangulation of ω and $k_{max} = \max_{1 \leq n \leq N} k_n$.

Proof: Take X in $V^p(I; X)$ used in Corollary 1.2 to be $H^k(\omega \times [-\frac{d}{2}, \frac{d}{2}])$. Let $u^* \in V^{p_3}(I; H^k(\omega \times [-\frac{d}{2}, \frac{d}{2}]))$ be the element approximating u . From Corollary 1.2,

$$\|u - u^*\|_{L_2(I; H^k(\omega \times [-\frac{d}{2}, \frac{d}{2}]))} = O(k_{max}^{\nu_3} p_3^{-k}).$$

Approximate $u^*|_{\omega \times [-\frac{d}{2}, \frac{d}{2}]}$ by \bar{u} by Theorem 1.5. Then,

$$\begin{aligned} \|u - \bar{u}\|_{L_2(\omega \times [-\frac{d}{2}, \frac{d}{2}] \times I)} &= \|u - u^*\|_{L_2(\omega \times [-\frac{d}{2}, \frac{d}{2}] \times I)} + \|u^* - \bar{u}\|_{L_2(\omega \times [-\frac{d}{2}, \frac{d}{2}] \times I)} \\ &= C_1 h^{\nu_1} p_1^{-k} + C_2 d^{\nu_2} p_2^{-k} + C_3 k_{max}^{\nu_3} p_3^{-k}, \end{aligned}$$

where C_1, C_2 and C_3 are constants independent of h, k_{max}, d, p_1, p_2 and p_3 .

Using (1.4), we obtain

Corollary 1.7 *Let $u \in H^\infty(\omega \times [-\frac{d}{2}, \frac{d}{2}] \times I)$. Then there exists $u^* \in S^{p_1, k}(\omega, T_h) \otimes PP^2([-\frac{d}{2}, \frac{d}{2}]) \otimes PP^3(I)$,*

$$\|u - u^*\|_{L_2(\omega \times [-\frac{d}{2}, \frac{d}{2}] \times I)} \leq C e^{-b \min(p_1, p_2, p_3)},$$

where C and b are independent of p_1, p_2 and p_3 .

2 p -Version of Discontinuous Galerkin Finite Element Method

In this section, p -version of discontinuous Galerkin finite element method for the parabolic problem (1.1) is described. The main goal here is to provide an error analysis for the p -finite element method using the results from Section 1. The semidiscrete approximation equation (1.2) is now upgraded to a fully discretized equation below. It is assumed as in Theorem 1.5

and Corollary 1.6 that the degree vectors in space, through-the-thickness and time are assumed to be p_1 , p_2 and p_3 respectively. Define

$$W^{(p_1, p_2, p_3)} = \{V: R^+ \rightarrow S^{p_1, k} \otimes P^{p_2}[-\frac{d}{2}, \frac{d}{2}]: V|_{I_n} \in P^{p_3}(I_n), n = 1, 2, \dots, N\}, \quad (2.1)$$

where

$$P^{p_3}(I_n) = \{V(t) = \sum_{i=0}^{p_3} V_i t^i: V_i \in S^{p_1, k} \otimes P^{p_2}[-\frac{d}{2}, \frac{d}{2}]\}.$$

Then the fully discretized discontinuous Galerkin method can be described as follows:

Find $U \in W^{(p_1, p_2, p_3)}$ such that for $n = 1, 2, \dots$,

$$\int_{I_n} \{(U_t, V) + (\nabla U, \nabla V)\} dt + ([U]_{n-1}, V_{n-1}^+) = \int_{I_n} (f, v) dt \quad \text{for all } V \in W^{(p_1, p_2, p_3)}, \quad (2.2)$$

We consider in (1.1) only the case of isotropic materials along with Dirichlet boundary conditions. However, extensions to anisotropic materials as well as mixed boundary conditions where Neumann boundary conditions are incorporated are possible and the present analysis carries over to these cases. The thesis by Tomey [7] treats transversely anisotropic materials as well as isotropic materials along with a mixed boundary condition.

The notation of [7] is followed. The solution u is approximated over $K^l \times (-\frac{d}{2}, \frac{d}{2}) \times I_n$ using the outer tensor products:

$$u|_{K^l \times (-\frac{d}{2}, \frac{d}{2}) \times I_n} \simeq (\phi \otimes \psi \otimes \theta)^T \mathbf{a}^n \equiv \chi^T \mathbf{a}^n, \quad (2.3)$$

which is

$$u|_{K^l \times (-\frac{d}{2}, \frac{d}{2}) \times I_n} \simeq \sum_{\alpha=0}^{p_1} \sum_{\beta=0}^{p_2} \sum_{\gamma=0}^{p_3} \varphi_{\alpha}^l(x, y) \psi_{\beta}(z) \theta_{\gamma}(t) a_{\alpha\beta\gamma}^n.$$

Equation (2.2) becomes

$$\begin{aligned} & \sum_{K^l \in \mathcal{T}_h} \int_{K^l} \int_{\Gamma} [\int_{I_n} \chi(\frac{\partial \chi}{\partial t})^T + \nabla \chi (\nabla \chi)^T dt] + \chi(t_{n-1}^+) \chi(t_{n-1}^+)^T dzd\omega \cdot \mathbf{a}^n \\ & = \sum_{K^l \in \mathcal{T}_h} \int_{K^l} \int_{\Gamma} \int_{I_n} f \chi dt dzd\omega + \int_{K^l} \int_{\Gamma} \chi(t_{n-1}^-) \chi(t_{n-1}^-)^T dzd\omega \cdot \mathbf{a}^{n-1}. \end{aligned} \quad (2.4)$$

Equation (2.4) can be written in the following matrix form [7]

$$\sum_{K^l \in T_h} (\mathbf{C}_{K^l} + \mathbf{K}_{K^l} + \mathbf{M}_{K^l}) \cdot \mathbf{a}^n = \sum_{K^l \in T_h} (\mathbf{H}_{K^l} + \mathbf{M}_{K^l} \cdot \mathbf{a}^{n-1}), \quad (2.5)$$

where \mathbf{C}_{K^l} , \mathbf{K}_{K^l} and \mathbf{M}_{K^l} are, respectively, the element capacitance, conductance (stiffness) and mass matrices, whereas \mathbf{H}_{K^l} is the element load vector. Thesis by Tomey [7] describes in detail as to how these matrices should be assembled. In the next section, we consider different bases in time variable, delineating an advantage of each choice of basis elements. The following theorem can be proved by minor modifications to the proof of theorem 1.1, [3] and by making use of Theorem 1.6.

Theorem 2.1 *Let $u \in H^k(\omega \times [-\frac{d}{2}, \frac{d}{2}] \times I)$. Suppose that there is a constant γ such that the time steps k_n satisfy $k_n \leq \gamma k_{n+1}$, $n = 1, \dots, N-1$ and let U^n denote the solution of (2.2) approximating u at t_n . Then there is a constant C depending only on γ and a constant β , where $\rho_K \geq \beta h_K$ and ρ_K is the diameter of the circle inscribed in K for all $K \in T_h$, such that for $n = 1, 2, \dots, N$,*

$$\|u - U^n\|_{L_2(\omega \times [-\frac{d}{2}, \frac{d}{2}] \times I)} \leq C(1 + \log \frac{t_n}{k_n})^{1/2} \{C_1 h^{\nu_1} p_1^{-k} + C_2 d^{\nu_2} p_2^{-k} + C_3 k_{\max}^{\nu_3} p_3^{-k}\},$$

where $\nu_i = \min(k, p_i + 1)$, $i = 1, 2, 3$; C_1 , C_2 and C_3 are constants independent of h , k_{\max} , d , p_1 , p_2 and p_3 .

Similarly, Corollary 1.7 implies the following

Corollary 2.2 *Let $u \in H^\infty(\omega \times [-\frac{d}{2}, \frac{d}{2}] \times I)$. Suppose that there is a constant γ such that the time steps k_n satisfy $k_n \leq \gamma k_{n+1}$, $n = 1, \dots, N-1$ and let U^n denote the solution of (2.2) approximating u at t_n . Then there is a constant C depending only on γ and a constant β , where $\rho_K \geq \beta h_K$ and ρ_K is the diameter of the circle inscribed in K for all $K \in T_h$, such that for $n = 1, 2, \dots, N$,*

$$\|u - U^n\|_{L_2(\omega \times [-\frac{d}{2}, \frac{d}{2}] \times I)} \leq C e^{-b \min(p_1, p_2, p_3)},$$

where C and b are independent of p_1, p_2 and p_3 .

Remark 2.1: Maintaining throughout computation a certain accuracy of numerical solution obtained from (2.1) is always desirable and Theorem 2.1 gives an insight to the following adaptive

scheme. Suppose that a tolerance of δ is required. Then, as h , d and k_{\max} are known *a priori*, the minimum approximation degrees' requirement in spatial, through the thickness and time variable are obtained from

$$h^{\nu_1} p_1^{-k} < \delta, \quad d^{\nu_2} p_2^{-k} < \delta, \quad k_{\max}^{\nu_3} p_3^{-k} < \delta. \quad (2.6)$$

Moreover, for the case of infinitely differentiable u , with $p = \min(p_1, p_2, p_3)$ where p_i 's satisfy inequalities in (2.6), Corollary 2.2 implies that

$$\|u - U^n\|_{L_2(\omega \times [-\frac{d}{2}, \frac{d}{2}] \times I)} \leq C e^{-bp} < \delta.$$

Remark 2.2: Thus far, the case for the constant degree vectors was considered. Nonconstant degree vectors can also be incorporated easily from Theorem 1.1. For instance, nonconstant degree vector $\bar{p} = (p_n)$ can be derived from the inequality in Theorem 1.1. Note that the bound given in Theorem 1.1 combines all errors in time approximation over $M(I)$ intervals. Thus, for each $n = 1, 2, \dots, M(I)$, the error in discretization in time variable over I_n is given by

$$\|u - U\|_{L^2(I_n; X)}^2 \leq C \left(\frac{k_n}{2}\right)^{2(s_n+1)} \max(1, p_n)^{-2} \frac{\Gamma(p_n + 1 - s_n^*)}{\Gamma(p_n + 1 + s_n^*)} \|u\|_{H^{s_n^*+1}(I_n; X)}^2, \quad (2.6)$$

for any $0 \leq s_n^* \leq \min(p_n, s_n)$. Since $\frac{\Gamma(p_n+1-s_n^*)}{\Gamma(p_n+1+s_n^*)} \sim p_n^{-2s_n^*}$ as $p_n \rightarrow \infty$, (2.6) becomes

$$\|u - U\|_{L^2(I_n; X)} \leq C k_n^{s_n+1} p_n^{-(s_n+1)}.$$

As k_n 's are known, construct the degree vector \bar{p} by requiring each component p_n to satisfy

$$k_n^{s_n+1} p_n^{-(s_n+1)} < \delta.$$

A construction of nonconstant degree vector corresponding to a triangulation T_h for the region ω is similar.

3 Discretizations in Time Variable.

In this section, effects of the use of different basis elements to approximate the solution in time variable are considered. The solution u is approximated over $K^l \times (-\frac{d}{2}, \frac{d}{2}) \times I_n$ using the outer tensor products:

$$u|_{K^l \times (-\frac{d}{2}, \frac{d}{2}) \times I_n} \simeq (\phi \otimes \psi \otimes \theta)^T \mathbf{a}^n \equiv \chi^T \mathbf{a}^n. \quad (3.1)$$

To better illustrate the choice of basis elements for time variable, we write a semidiscrete solution $U_n \in P^{p_n}(I_n, X)$ by

$$U_n = \sum_{j=0}^{p_n} u_{j,n} \theta_{j,n}. \quad (3.2)$$

Here $u_{j,n} \in X$ are unknown coefficients to be determined. V_n is defined similarly. For convenience, we let $I = I_n$ and drop the time step index n . Substituting U_n of (3.2) and the corresponding V_n into (1.2), we obtain the following:

Find $\{u_j\}_{j=0}^p \subset X$ such that

$$\begin{aligned} & \sum_{i,j=0}^p \{ [\int_I \theta'_j \theta_i dt + \theta_j^+(t_{n-1}) \theta_i^+(t_{n-1})] (u_j, v_i)_X + \int_I \theta_j \theta_i dt (\nabla u_j, \nabla v_i)_X \} \\ & = \sum_{i=0}^p \{ \int_I (g, v_i \theta_i)_{X^* \times X} dt + (U_{n-1}^-, v_i \theta_i^+(t_{n-1})) \} \quad \text{for all } \{v_i\}_{i=0}^p \subset X. \end{aligned} \quad (3.3)$$

Transforming the integrals into the reference interval $[-1, 1]$ under F^{-1} and letting (see [11])

$$\begin{aligned} \hat{A}_{ij} & \equiv \int_{-1}^1 \hat{\theta}'_j \hat{\theta}_i d\hat{t} + \hat{\theta}_j^+(-1) \hat{\theta}_i^+(-1), & \hat{B}_{ij} & \equiv \int_{-1}^1 \hat{\theta}_j \hat{\theta}_i d\hat{t}, \\ \hat{f}_i^1(v_i) & \equiv \left(\int_{-1}^1 (g \circ F) \hat{\theta}_i d\hat{t}, v_i \right)_X, & \hat{f}_i^2(v_i) & \equiv (U_{n-1}^- \hat{\theta}_i(-1), v_i)_X, \end{aligned}$$

equation (3.3) becomes:

Find the coefficient $\{u_j\}_{j=0}^p \subset X$ such that for all $\{v_j\}_{j=0}^p \subset X$

$$\sum_{i,j=0}^p \hat{A}_{ij} (u_j, v_i)_X + \frac{k}{2} \hat{B}_{ij} (\nabla u_j, \nabla v_i)_X = \sum_{i=0}^p \frac{k}{2} \hat{f}_i^1(v_i) + \hat{f}_i^2(v_i). \quad (3.4)$$

Here $k = k_n$. The strong form of equation (3.4) is

$$\sum_{j=0}^p \hat{A}_{ij} u_j + \frac{k}{2} \hat{B}_{ij} \Delta u_j = \frac{k}{2} \hat{l}_i^1 + \hat{l}_i^2, \quad i = 0, 1, \dots, p, \quad (3.5)$$

where $\hat{l}_i^1 = \int_{-1}^1 (g \circ F) \hat{\theta}_i d\hat{t}$ and $\hat{l}_i^2 = U_{n-1}^- \hat{\theta}_i(-1)$. Hence, in order to execute the p -version of the finite element method, the following integrals must be computed for assembly of the global matrix.

$$\int_{t_{n-1}}^{t_n} \theta \left(\frac{d\theta}{dt} \right)^T dt, \quad \int_{t_{n-1}}^{t_n} \theta \theta^T dt, \quad \int_{t_{n-1}}^{t_n} \theta dt.$$

The set of the canonical polynomials $\theta_{\nu+1}(t) = t^\nu$ were used in [7] and the components of the matrices which represent \hat{A}_{ij} and \hat{B}_{ij} were computed exactly. We consider two other alternatives for θ .

First, we consider $\theta_\nu(\mathbf{t}) = \left(\frac{t-t_{n-1}}{k_n}\right)^\nu$: Let $U|_{I_n} = \sum_{\nu=0}^p \left(\frac{t-t_{n-1}}{k_n}\right)^\nu \Phi_\nu^n(\bar{x})$. Since $U^{n-1,+} = \Phi_0^n(\bar{x})$, the advantage of this basis is that the term $(U^{n-1,+}, v^{n-1,+})$ in (2.1) is simplified. Also,

$$\int_{t_{n-1}}^{t_n} \theta \left(\frac{d\theta}{dt}\right)^T dt = \begin{bmatrix} 0 & 1 & 1 & \dots & 1 \\ 0 & \frac{1}{2} & \frac{2}{3} & \dots & \frac{p}{p+1} \\ 0 & \frac{1}{3} & \frac{2}{4} & \dots & \frac{p}{p+2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \frac{1}{p+1} & \frac{2}{p+2} & \dots & \frac{p}{2p} \end{bmatrix},$$

$$\int_{t_{n-1}}^{t_n} \theta \theta^T dt = k_n \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \dots & \frac{1}{p+1} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \dots & \frac{1}{p+2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \frac{1}{p+1} & \frac{1}{p+2} & \frac{1}{p+3} & \dots & \frac{1}{2p+1} \end{bmatrix},$$

and

$$\int_{t_{n-1}}^{t_n} \theta dt = k_n \begin{bmatrix} 1 \\ \frac{1}{2} \\ \frac{1}{3} \\ \vdots \\ \frac{1}{p+1} \end{bmatrix}.$$

As stated in [11], the ideal choice for $\hat{\theta}_i$ would be the one under which the matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ diagonalize simultaneously. If the diagonalizations of $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ are possible, then equation (3.4) decouples into $p + 1$ independent equations, reducing the size of computation. The canonical basis and the basis just considered generate the full matrices for $\hat{\mathbf{A}}$ and for $\hat{\mathbf{B}}$. In order to select basis functions in time variable which takes into account of the structures of $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$, we now consider the Legendre polynomials for θ .

Second: The translated Legendre polynomials for $\theta_\nu(t)$ are considered. This is essentially the same as the normalized Legendre polynomials used in [11]. We extend the discussion in [11] by exhibiting the general form for $\hat{\mathbf{A}}$ and state its characteristics. The orthogonal nature of the Legendre polynomials guarantees the matrix $\hat{\mathbf{B}}$ to be diagonal. Hence it remains to analyze the matrix $\hat{\mathbf{A}}$. In [11], it is stated that ‘...this (diagonalization) seems not to be possible with time shape functions in \mathbf{R} , but numerical experiments show that $\hat{\mathbf{A}}$... is diagonalizable in \mathbf{C} at least for $0 \leq p \leq 100$ ’. We are unable to provide a mathematical proof of their

statement at this point, but our investigation thus far indicates that $\hat{\mathbf{A}}$ is non-defective (see, [4]), -i.e., the algebraic multiplicity and the geometric multiplicity of each eigenvalue are the same. Hence, $\hat{\mathbf{A}}$ is diagonalizable over the complex field \mathbf{C} . In this paper, we propose to use the real Schur decomposition of the matrix $\hat{\mathbf{A}}$, rather than the diagonalization technique in [11], to establish a solution scheme for (3.4). The real Schur decomposition leads to a modified backward substitution scheme. The method is mathematically justifiable for any degree p . Moreover, the current method avoids complex number arithmetic which was necessary in [11], and thus the cost of computation is approximately the same as [11].

The advantage of this choice as basis elements in time variable lies in the formations of $\int_{t_{n-1}}^{t_n} \theta \theta^T dt$, $\int_{t_{n-1}}^{t_n} \theta dt$ as well as $\int_{t_{n-1}}^{t_n} \theta (\frac{d\theta}{dt})^T dt$, all of which are of banded structures as seen below. Recall that $P_i(x)$ denote the Legendre polynomial of degree i defined over $[-1, 1]$. Let $x^n(t) = \frac{2}{t_n - t_{n-1}}t - \frac{t_n + t_{n-1}}{t_n - t_{n-1}}$ and $L_i^n(t) = P_i(x^n(t))$ for each $n = 1, 2, \dots, M(I)$ and $i = 0, 1, \dots$. Obviously,

$$\frac{2}{t_n - t_{n-1}} \int_{t_{n-1}}^{t_n} L_i^n(t) L_j^n(t) dt = \delta_{ij} \quad \text{for each } i, j = 0, 1, \dots,$$

Thus,

$$\int_{t_{n-1}}^{t_n} \theta \theta^T dt = \begin{bmatrix} \frac{k_n}{2} & & & \\ & \frac{k_n}{2} & & \\ & & \ddots & \\ & & & \frac{k_n}{2} \end{bmatrix},$$

and

$$\int_{t_{n-1}}^{t_n} \theta dt = \begin{bmatrix} k_n \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Also

$$\int_{t_{n-1}}^{t_n} \theta \left(\frac{d\theta}{dt}\right)^T dt = \begin{bmatrix} 0 & k_n & 0 & k_n & 0 & k_n & \cdots \\ 0 & 0 & k_n & 0 & k_n & 0 & \cdots \\ 0 & 0 & 0 & k_n & 0 & k_n & \cdots \\ 0 & 0 & 0 & 0 & k_n & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

The formation of the last matrix $\int_{t_{n-1}}^{t_n} \theta (\frac{d\theta}{dt})^T dt$ requires some tedious but straightforward calculations which are not so obvious. Thus, we include them below. Clearly, it is sufficient to derive

the following:

$$\int_{-1}^1 \theta \left(\frac{d\theta}{dt} \right)^T dt = \begin{bmatrix} 0 & 2 & 0 & 2 & 0 & 2 & \cdots \\ 0 & 0 & 2 & 0 & 2 & 0 & \cdots \\ 0 & 0 & 0 & 2 & 0 & 2 & \cdots \\ 0 & 0 & 0 & 0 & 2 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \quad (3.6)$$

For the Legendre polynomials of the first kind, $P_i(x)$, $i \geq 0$, $x \in [-1, 1]$, first note that

$$P_{2i}(-1) = P_{2i}(1) = 1, \quad P_{2i+1}(-1) = -1, \quad \text{and } P_{2i+1}(1) = 1 \text{ for } i \geq 0.$$

To derive (3.6), we must show that

$$\int_{-1}^1 P_i(t) P_j'(t) dt = 0, \quad \text{for all } i \text{ and } j \text{ with } i \geq j \geq 0, \quad (3.7)$$

and

$$\int_{-1}^1 P_i(t) P_{i+2l}'(t) dt = 0, \quad \int_{-1}^1 P_i(t) P_{i+2l+1}'(t) dt = 2, \quad \text{for all } i \geq 0 \text{ and } l \geq 0. \quad (3.8)$$

Equations (3.7) and (3.8) are verified by induction. As $P_0(t) = 1$, (3.7) is verified with $i = 0$.

Also

$$\int_{-1}^1 P_0(t) P_{2l}'(t) dt = P_0(t) P_{2l}(t) \Big|_{-1}^1 - \int_{-1}^1 P_0'(t) P_{2l}(t) dt = 0, \quad \text{for all } l \geq 0.$$

and similarly

$$\int_{-1}^1 P_0(t) P_{2l+1}'(t) dt = P_0(t) P_{2l+1}(t) \Big|_{-1}^1 - \int_{-1}^1 P_0'(t) P_{2l+1}(t) dt = 2, \quad \text{for all } l \geq 0.$$

Now assume that (3.7) and (3.8) are satisfied for some $i = i^* - 1$ and for all j such that $i^* - 1 \geq j \geq 0$ and for all $l \geq 0$. First, for the diagonal element,

$$\int_{-1}^1 P_{i^*}(t) P_{i^*}'(t) dt = P_{i^*}(t) P_{i^*}(t) \Big|_{-1}^1 - \int_{-1}^1 P_{i^*}'(t) P_{i^*}(t) dt$$

implies that $\int_{-1}^1 P_{i^*}(t) P_{i^*}'(t) dt = 0$. Also, for $i^* > j \geq 0$,

$$\int_{-1}^1 P_{i^*}(t) P_j'(t) dt = P_{i^*}(t) P_j(t) \Big|_{-1}^1 - \int_{-1}^1 P_{i^*}'(t) P_j(t) dt = \begin{cases} 2 - 2 = 0 & \text{for } i^* - j \text{ odd,} \\ 0 - 0 = 0 & \text{for } i^* - j \text{ even} \end{cases}$$

where the inductive assumption was used in computing the second integral. This shows (3.7) for all i and $i \geq j \geq 0$. It remains to prove that $\int_{-1}^1 P_{i^*}(t) P_{i^*+2l}'(t) dt = 0$ and $\int_{-1}^1 P_{i^*}(t) P_{i^*+2l+1}'(t) dt = 2$ for all $l \geq 0$. The case for $l = 0$ is done. For $l > 0$,

$$\int_{-1}^1 P_{i^*}(t) P_{i^*+2l}'(t) dt = P_{i^*}(t) P_{i^*+2l}(t) \Big|_{-1}^1 - \int_{-1}^1 P_{i^*}'(t) P_{i^*+2l}(t) dt.$$

The last integral is 0 by (3.6) and the term $P_{i^*}(t)P_{i^*+2l}(t)|_{-1}^1$ is 0 regardless of i^* even or odd. Hence $\int_{-1}^1 P_{i^*}(t)P'_{i^*+2l}(t)dt = 0$. $\int_{-1}^1 P_{i^*}(t)P'_{i^*+2l+1}(t)dt = 2$ is similar. Thus, (3.6) is verified.

Now, from (3.6), it is clear that

$$\int_{-1}^1 \theta \left(\frac{d\theta}{dt} \right)^T dt + \theta_{n-1}^+ \theta_{n-1}^+ = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & \dots \\ -1 & 1 & 1 & 1 & 1 & 1 & \dots \\ 1 & -1 & 1 & 1 & 1 & 1 & \dots \\ -1 & 1 & -1 & 1 & 1 & 1 & \dots \\ 1 & -1 & 1 & -1 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \quad (3.9)$$

Formula (3.9) provides a general construction method for the assembly of the p -finite element matrix with any order p when the Legendre polynomials are used in time variable. For $p = 5$, with normalizing factors incorporated, (3.9) can be used to derive the matrix $\hat{\mathbf{A}}$ in [11], (eq. (4.11)). Two observations on the matrix $\hat{\mathbf{A}}_{p+1}$ are as follows:

Theorem 3.1 *Let $\hat{\mathbf{A}}_{p+1}$ denote the $(p+1) \times (p+1)$ leading principle matrix of (3.6) for each $p = 0, 1, \dots$. Then $\hat{\mathbf{A}}_{p+1}$ is invertible for each p .*

Proof: This follows immediately by noting that $\det(A_{p+1}) = 2^p$.

The matrix $\hat{\mathbf{A}}_{p+1}$ is not positive definite for some p . However:

Theorem 3.2 *The matrix $\hat{\mathbf{A}}_{p+1}$ is non-negative definite for all $n = 0, 1, \dots$*

As in (3.2), denote the semidiscrete solution U as well as V in (1.2) as

$$U = \sum_{j=0}^p u_j \theta_j, \quad V = \sum_{j=0}^p v_j \theta_j \quad (3.10)$$

where the subscript n is dropped. Let $\theta_i = L_i$, the transformed Legendre polynomial of degree i and $\hat{\theta}_i = P_i$, the Legendre polynomial of degree i defined over $[-1, 1]$ in equations (3.3) and (3.4). Also, denote a basis for the finite element space $SP^{1,k}(\omega, T_h) \otimes PP^2[-\frac{k}{2}, \frac{k}{2}]$ for X by $\{s_j\}_{j=1}^D$ with $D = \dim(SP^{1,k}(\omega, T_h) \otimes PP^2[-\frac{k}{2}, \frac{k}{2}])$. The trial and test functions u_j and v_i in the semi-discrete system (3.3) above are further approximated by

$$u_j \sim \sum_{l=1}^D u_j^l s_l(x), \quad v_i \sim \sum_{k=1}^D v_i^k s_k(x). \quad (3.11)$$

Substituting (3.11) into (3.4), the fully discrete p -finite element system can be written as (ref. [11]), for the unknown coefficient vectors $\vec{u}_j = (u_j^1, u_j^2, \dots, u_j^D)^T \in R^D$,

$$\begin{bmatrix} \hat{A}_{00}M + \frac{k}{2}S & \cdots & \hat{A}_{0p}M \\ \vdots & \ddots & \vdots \\ \hat{A}_{p0}M & \cdots & \hat{A}_{pp}M + \frac{k}{2}S \end{bmatrix} \begin{bmatrix} \vec{u}_0 \\ \vdots \\ \vec{u}_p \end{bmatrix} = \frac{k}{2} \begin{bmatrix} \vec{f}_0^1 \\ \vdots \\ \vec{f}_p^1 \end{bmatrix} + \begin{bmatrix} \vec{f}_0^2 \\ \vdots \\ \vec{f}_p^2 \end{bmatrix}, \quad (3.12)$$

where

$$M = \{(s_l, s_k)_X\}_{l,k=1}^D, \quad S = \{(\nabla s_l, \nabla s_k)_X\}_{l,k=1}^D,$$

and

$$\begin{aligned} \vec{f}_j^1 &= (\hat{f}_j^1(s_1), \hat{f}_j^1(s_2), \dots, \hat{f}_j^1(s_D))^T \\ \vec{f}_j^2 &= (\hat{f}_j^2(s_1), \hat{f}_j^2(s_2), \dots, \hat{f}_j^2(s_D))^T. \end{aligned}$$

Note that the use of the translated Legendre polynomial served well because $\hat{\mathbf{B}}_{ij} = \delta_{ij}$ in (3.5). Equation (3.5) can be written in vector form as

$$\hat{\mathbf{A}}\vec{u} + \frac{k}{2}[\delta_{ij}]\Delta\vec{u} = \vec{F}, \quad (3.13)$$

where $\vec{F} = (\frac{k}{2} \int_{-1}^1 (g \circ x^n) \hat{\theta}_0 dt + U_{n-1}^- \hat{\theta}_0(-1), \dots, \frac{k}{2} \int_{-1}^1 (g \circ x^n) \hat{\theta}_p dt + U_{n-1}^- \hat{\theta}_p(-1))^T$. In [11], $\hat{\mathbf{A}}$ is diagonalized and equation (3.13) is solved for $\vec{w} = Q^T \vec{u}$ from

$$Q^T \hat{\mathbf{A}} Q \vec{w} + \frac{k}{2} \Delta \vec{w} = Q^T \vec{F}. \quad (3.14)$$

It is reported in [11] that, as the result of their numerical experiments, the matrix \hat{A}_{p+1} is diagonalizable up to its order $p = 100$. Subsequently, equation (3.13) is decoupled via (3.14) into $p + 1$ independent scalar equations, each of which requires complex arithmetic to solve. A new approach which uses the real Schur decomposition is now presented. The new approach does not require the complex arithmetic.

Theorem 3.3 (*Real Schur Decomposition, [4] p.219*) *If $A \in R^{n \times n}$, then there exists an orthogonal matrix $Q \in R^{n \times n}$ such that*

$$Q^T A Q = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ 0 & R_{22} & \cdots & R_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{mm} \end{bmatrix} \quad (3.15)$$

where each R_{ii} is either a 1×1 matrix or a 2×2 matrix having complex conjugate eigenvalues of A .

By Theorem 3.1, it is guaranteed that no 1×1 matrix in Schur decomposition for $\hat{\mathbf{A}}$ is 0. Also, it is worth noting that the Schur decomposition is an orthogonal similarity transformation and thus avoid the computation of Q^{-1} as required in the diagonalization process done in [11]. Equation (3.14) is then solved by the backward substitution using block matrices. More specifically, \vec{w}_{p-1} and \vec{w}_p are found from

$$(R_{mm}M + \frac{k}{2}S) \begin{bmatrix} \vec{w}_{p-1} \\ \vec{w}_p \end{bmatrix} = \begin{bmatrix} \vec{F}_{p-1} \\ \vec{F}_p \end{bmatrix}, \quad (3.16)$$

or from

$$(R_{mm}M + \frac{k}{2}S)\vec{w}_p = \vec{F}_p, \quad (3.17)$$

according to R_{mm} being 2×2 or 1×1 matrix respectively. Here, we used the standard convention that if $R_{mm} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $R_{mm}M = \begin{bmatrix} aM & bM \\ cM & dM \end{bmatrix}$. Computation proceeds to find \vec{w}_{p-2} as well as \vec{w}_{p-3} if R_{m-1m-1} is 2×2 . Namely, assuming that R_{mm} is 2×2 , if R_{m-1m-1} is a scalar, then \vec{w}_{p-2} is found by solving

$$\left[R_{m-1m-1}M + \frac{k}{2}S \right] \vec{w}_{p-2} = \vec{F}_{p-2} - [R_{m-1m}M] \begin{bmatrix} \vec{w}_{p-1} \\ \vec{w}_p \end{bmatrix}.$$

If R_{m-1m-1} is 2×2 , then \vec{w}_{p-3} and \vec{w}_{p-2} are found from

$$\left[R_{m-1m-1}M + \frac{k}{2}S \right] \begin{bmatrix} \vec{w}_{p-3} \\ \vec{w}_{p-2} \end{bmatrix} = \begin{bmatrix} \vec{F}_{p-3} \\ \vec{F}_{p-2} \end{bmatrix} - [R_{m-1m}M] \begin{bmatrix} \vec{w}_{p-1} \\ \vec{w}_p \end{bmatrix}.$$

The case for 1×1 R_{mm} is similar. It is important to recall that the computation thus described can be completed because of Theorem 3.1.

4 Start-up Singularities:

In this final section, we make some comments on the start-up singularities normally associated with the parabolic problems. The regularity assumptions in Theorem 2.1 and Corollary 2.2 were taken so that the current fully p -finite element method could provide numerical solutions where

the discretization error associated in time can be made consistent with the discretization error associated in space. However, as indicated earlier, time singularities arise due to various types of incompatible initial data. To capture such singularities, hp -version of finite element method must be considered. In [10], a nonuniform time discretization is determined by considering the conditions on f as well as the initial function u_0 in (1.1). More specifically, the function f is assumed to be piecewise analytic as a function on $[0, T]$ with values in H , i.e.,

$$\|f^{(l)}\|_H \leq C d^l \Gamma(l+1), \quad t \in [0, T], \quad l \in N_0$$

with constants C and d independent of l and t . Also, u_0 is assumed to be in $H_\theta = (H, X)_{\theta, 2}$, $0 \leq \theta \leq 1$, where $(H, X)_{\theta, 2}$ is a space between H and X determined by the K -method of interpolation, (see [8]). An h -version of the discontinuous Galerkin finite element method developed in [5] establishes a nonuniform time discretization scheme which is based upon the behavior of $\|u^{(1)}(t)\|_X$. The direct inspection into the smoothness of $\|u^{(1)}(t)\|_X$ in determining time partition takes into account of various possibilities under which the start-up singularities associated with parabolic problems arise. Analysis used in determining the nonuniform graded partition points in [5] is distinct from the one used in [10] and an example is provided in [5], which demonstrates that the method of Kaneko, Bey and Hou gives more sparse time partition points than the ones given in [10].

References

- [1] I. Babuška, B. A. Szabo and I. N. Katz, *The p-version of the finite element method*, SIAM JI. Num. Anal., Vol. 18, No.3, (1981), pp. 515-545.
- [2] I. Babuška and M. Suri, *The p and hp versions of the finite element method, basic principles and properties*, SIAM Review, Vol. 36, No. 4, (1994), pp. 578-632.
- [3] Kenneth Eriksson and Claes Johnson, *Adaptive finite element methods for parabolic problems I: a linear model problem*, SIAM JI. Num. Anal., Vol. 28, No. 1, (1991), pp. 43-77.
- [4] Gene H. Golub and Charles F. Van Loan, *Matrix Computations*, Johns Hopkins Univ. Press, (1985).
- [5] Hideaki Kaneko, Kim S. Bey and Gene J. W. Hou, *Discontinuous Galerkin finite element method for parabolic problems with non-smooth data*, (preprint, NASA)
- [6] Hideaki Kaneko, Kim S. Bey and Gene J. W. Hou, *Discontinuous Galerkin finite element method for parabolic problems with modified hp finite element approximation technique*, (preprint, NASA)
- [7] J. P. Tomey, *p-version discontinuous Galerkin method for heat transfer in built-up structures*, M.S. thesis, George Washington University, May, 2001.
- [8] Ch. Schwab, *p- and hp- Finite Element Methods*, Numerical Mathematics and Scientific Computation, Oxford Science Publication, Oxford University Press, (1998).
- [9] W. Gui and I. Babuška, *The h, p and hp versions of the finite element method in 1 dimension, Part I. The error analysis of the p-version*, Numer. Math., Vol. 49, (1986), pp. 577-612. *Part II. The error analysis of the h and hp versions*, Numer. Math., Vol. 49, (1986), pp. 613-657. *Part III. The adaptive hp version*, Numer. Math., Vol. 49, (1986), pp. 659-683.
- [10] D. Schötzau and C. Schwab, *Time discretization of parabolic problems by the hp-version of the discontinuous Galerkin finite element method*, SIAM JI. Num. Anal., Vol. 38, No. 3, (2000), pp. 837-875.

- [11] T. Werber, K. Gerdes, D. Schötzau and C. Schwab, *hp discontinuous Galerkin time stepping for parabolic problems*, Research Report No. 2000-01, April 2000, Seminar für Angewandte Mathematik, ETH, Zürich.