

**THE EFFECT OF NUMERICAL INTEGRATION ON THE
FINITE ELEMENT APPROXIMATION OF LINEAR
FUNCTIONALS**

IVO BABUŠKA, UDAY BANERJEE, AND HENGGUANG LI

ABSTRACT. In this paper, we have studied the effect of numerical integration on the Finite Element Method (FEM) based on piecewise polynomials of degree k , in the context of approximating linear functionals, which are also known as “quantities of interest”. We have obtained the optimal order of convergence, $\mathcal{O}(h^{2k})$, of the error in the computed functional, when the integrals in the stiffness matrix and the load vector are computed with a quadrature rule of algebraic precision $2k - 1$. However, this result was obtained under an increased regularity assumption on the data, which is more than required to obtain the optimal order of convergence of the energy norm of the error in the finite element solution with quadrature. We have obtained a lower bound of the error in the computed functional for a particular problem, which indicates the necessity of the increased regularity requirement of the data. Numerical experiments have been presented indicating that over-integration may be necessary to accurately approximate the functional, when the data lack the increased regularity.

1. INTRODUCTION

Determination of various quantities of interest is one of the major goals in scientific computation. For example in elasticity computations, typical quantities of interest are various resultants, stress intensity factors, etc. These quantities are values of functionals evaluated at the solution of the underlying problems, e.g., the solution of the system of partial differential equations modeling elasticity (Lamé equations).

Suppose the solution of the problem is characterized by the solution $u \in V \subset H^1(\Omega)$ of the variational problem

$$a(u, v) = f(v), \quad \forall v \in V,$$

where $a(\cdot, \cdot)$ is a continuous bilinear form satisfying the inf-sup condition on $V \times V$ and $f(\cdot)$ is a continuous linear form on V . Then the quantity of interest is given by $G(u)$, where $G(\cdot)$ is a bounded linear functional on V , characterized by $g \in V$ satisfying $G(v) = a(v, g)$, $\forall v \in V$. $G(u)$ is approximated by the quantity $G(u_h)$, where $u_h \in S_h \subset V$ is the Galerkin approximation of u . Using the Galerkin orthogonality $a(u - u_h, v) = 0$, $\forall v \in S_h$, the error in the computed quantity of interest $G(u_h)$ is given by

$$(1) \quad G(u) - G(u_h) = a(u - u_h, g - v), \quad \forall v \in S_h.$$

The research of the first author was partially supported by NSF grant DMS 0611094. The research of the second author was partially supported by NSF grant DMS 0610778.

In particular, if u_h is the solution of the finite element method based on piecewise polynomials of degree k , and if $u, g \in H^{k+1}(\Omega)$, then

$$|G(u) - G(u_h)| \approx \mathcal{O}(h^{2k}) \approx \|u - u_h\|_{H^1(\Omega)}^2.$$

We note that the enhanced accuracy (i.e., $\|u - u_h\|_{H^1(\Omega)}^2$) in the computed quantity of interest is the consequence of the Galerkin orthogonality.

In many applications, the quantity of interest, e.g., stress intensity factors or resultants as mentioned before, is given as $F(u)$, where the functional $F(\cdot)$ is *not* bounded on V but $F(u)$ is finite. In such situations, $F(\cdot)$ could be expressed in terms of a bounded linear functional $G(\cdot)$ on V , and $F(u)$ is approximated by a quantity F_{u_h} , which is computed using the values of $G(u_h)$, such that

$$F(u) - F_{u_h} \approx G(u) - G(u_h).$$

We will illustrate a procedure to write $F(u)$ in terms of a bounded linear functional $G(u)$ in Section 4. These ideas were first systematically studied in [1, 2]. A detailed discussion on computing various quantities of interest can also be found in Chapter 11 of [15]. We note however that F_{u_h} will have the enhanced accuracy again as a consequence of the Galerkin orthogonality.

In a finite element method, the definite integrals in the elements of the stiffness matrix and the load vector are approximated by numerical integration; as a consequence, the Galerkin orthogonality is violated and (1) does not hold. In this paper, we will address the consequences of the violation of the Galerkin orthogonality and study the effect of numerical integration on the accuracy of the computed quantity of interest. We will show that the loss of accuracy in the computed quantity of interest due to numerical integration could be more than the loss of accuracy in the energy norm of the finite element solution due to numerical integration. We will also give sufficient conditions on the quadrature as well as on the data of the problem that will yield accurate values of the computed quantity of interest.

The effect of numerical integration on the finite element solution u_h have been studied by various authors; we mention [12], [13], [10], [11], [5], and [7] (for the p -version). An excellent exposition of this problem is given in [9] and we will use various ideas from [9] in this study. It is well-established (see Theorem 4.1.6 in [9]) that the error in u_h in the energy norm yields optimal order of convergence, $\mathcal{O}(h^k)$ provided

- the numerical integration rule used on each triangle of the finite element triangulation is exact for all polynomials of degree $2k - 2$,
- the exact solution $u \in H^{k+1}(\Omega)$ and the data of the underlying problem, i.e., the variable coefficients and the load function of the PDE, have sufficient regularity depending on k ,

where the finite element method is based on piecewise polynomials of degree k .

In this paper, we have considered the finite element method based on a quasi-uniform mesh on a polygonal domain and have obtained the following results:

- We have shown that the error in the computed quantity of interest, under numerical integration, is $\mathcal{O}(h^{2k})$. We obtained this result under the assumption that $u, g \in H^{k+1}(\Omega)$, the numerical integration rule on each triangle is exact for all polynomials of degree $2k - 1$, and the data of the underlying problem have more regularity than is required to obtain the optimal order

of convergence of the energy norm of the error in u_h , as mentioned in the second bullet of the last paragraph.

- We have obtained a lower bound of the error in the computed quantity of interest for a particular problem. This result indicates that increased regularity assumptions on the data of the problem, as mentioned in the last bullet, is in general necessary.

We have also presented computational results in this paper that show that “over-integration” may be required to obtain accurate value of the quantity of interest, when the data of the problem do not have the required increased regularity.

We mention that the effect of numerical integration in the context of eigenvalue approximation by the finite element method, based on piecewise polynomials of degree k , was addressed in [6] and the optimal order of convergence, $\mathcal{O}(h^{2k})$, was obtained for the error in approximate eigenvalues. It is interesting to note that the optimal order of convergence of the error in the computed quantity of interest, as we show in this paper, is also $\mathcal{O}(h^{2k})$. We further mention that a lower norm error estimate for the finite element solution was also obtained in [6]; the order of convergence in this estimate is comparable to the result of Theorem 3.6 of this paper.

We note that the use of a numerical integration rule that is exact for the polynomials of degree $2k - 1$ is not restrictive, as the quadrature rules used in most finite element codes satisfy this condition. But the increased regularity requirement on the data of the problem may have serious consequences. This requirement indicates that the Jacobian of the mapping for a curved element in a finite element triangulation (in the case of a non-polygonal domain) should be sufficiently smooth – a requirement that is restrictive especially in the h - p version of the finite element method, where a curved element could be large and the Jacobian of the associated mapping may not have the required increased regularity. In such situations, over-integration may become necessary to control the energy norm of the error in the finite element solution; but even more over-integration may be required to control the error in the computed quantity of interest. We have not studied curved elements or h - p version of finite element method in this paper, but the results presented in this paper could be helpful in the analysis of such problems.

The paper is organized as follows. In Section 2, we have given the notation and the model problem, and defined the finite element method with numerical integration. We then defined the quantity of interest as a bounded linear functional $G(u)$, and presented the optimal error estimate for the computed value of $G(u)$ in the case of exact integration. In Section 3, we presented one of our main results, where we obtained the optimal error estimate for the computed value of $G(u)$ in the presence of numerical integration. We derived a lower bound of the error in the computed value of a linear functional for a particular problem in the presence of numerical integration in Section 4, which is another main result of this paper. In Section 5, we gave some computational results that illuminate the results obtained in Sections 3 and 4. We presented our concluding remarks in Section 6.

2. PRELIMINARIES AND NOTATION

The model problem: Let $\Omega \subset \mathbb{R}^n$ be a bounded n -dimensional domain with piecewise straight faces. We define the standard Sobolev spaces as follows.

$$H^m(\Omega) = \{v \in L^2(\Omega), \partial^\alpha v \in L^2(\Omega), \forall |\alpha| \leq m\},$$

with the multi-index $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$, $|\alpha| := \sum_{i=1}^n \alpha_i$, and $\partial^\alpha := \partial^{\alpha_1} \partial^{\alpha_2} \dots \partial^{\alpha_n}$. The H^m -norm and semi-norm of any $v \in H^m(\Omega)$ is given by

$$\begin{aligned} \|v\|_{H^m(\Omega)}^2 &:= \sum_{|\alpha| \leq m} \|\partial^\alpha v\|_{L^2(\Omega)}^2, \\ |v|_{H^m(\Omega)}^2 &:= \sum_{|\alpha|=m} \|\partial^\alpha v\|_{L^2(\Omega)}^2, \end{aligned}$$

while the L^2 -norm is

$$\|v\|_{L^2(\Omega)}^2 = \int_{\Omega} |v|^2 dx.$$

In addition, we denote by $H_0^1(\Omega) \subset H^1(\Omega)$ the subspace consisting of zero-trace functions.

We consider the following elliptic boundary value problem:

$$(2) \quad \begin{cases} -\operatorname{div} A \nabla u = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega, \end{cases}$$

where $A = (a_{ij}(x))$ is a symmetric matrix, and $a_{ij}(x)$ are smooth functions. We will refer to the functions $a_{ij}(x)$ and $f(x)$ as the *data of the problem*. We assume that A satisfies the uniform ellipticity condition, i.e., there exists a constant $C > 0$, such that

$$\sum_{i,j=1}^n a_{ij}(x) \xi_i \xi_j \geq C \sum_{i=1}^n \xi_i^2, \quad \forall \xi \in \mathbb{R}^n, \quad \forall x \in \bar{\Omega}.$$

The variational form of the problem (2) is given as

$$(3) \quad \begin{cases} u \in H_0^1(\Omega), \\ a(u, v) = \int_{\Omega} \sum_{i,j=1}^n a_{ij} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} dx = \int_{\Omega} f v dx, \quad \forall v \in H_0^1(\Omega). \end{cases}$$

Based on the Poincaré inequality, it is clear that the bilinear form $a(\cdot, \cdot)$ is coercive, namely, for a constant $C > 0$,

$$a(v, v) \geq C \|v\|_{H^1(\Omega)}^2, \quad \forall v \in H_0^1(\Omega).$$

Moreover, the bilinear form is bounded, i.e., there is constant $C > 0$, such that

$$|a(w, v)| \leq C \|w\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)}, \quad \forall w, v \in H_0^1(\Omega).$$

Therefore, $\|v\|_a^2 := a(v, v)$ defines an equivalent norm on $H_0^1(\Omega)$, and the solution of equation (3) is unique by the Lax-Milgram Lemma.

The finite element method: Let \mathcal{T} be a quasi-uniform triangulation of Ω , and let

$$S_h = \{v \in C^0(\bar{\Omega}), v|_{\partial\Omega} = 0, v|_K \in \mathcal{P}_k(K), \forall K \in \mathcal{T}\} \subset H_0^1(\Omega)$$

be the finite element subspace, where \mathcal{P}_k is the space containing all polynomials of degree k . The finite element approximation to the solution $u \in H_0^1(\Omega)$ of (3) is given by the finite element method (FEM),

$$\begin{cases} u_h \in S_h, \\ a(u_h, v_h) = \int_{\Omega} f v_h dx, \quad \forall v_h \in S_h, \end{cases}$$

where $f(v_h) = \int_{\Omega} f v_h dx$. It is well known that

$$(4) \quad \|u - u_h\|_{H^1(\Omega)} \leq Ch^k \|u\|_{H^{k+1}(\Omega)}.$$

Note that $a(u_h, v_h)$ and $f(v_h)$ contain definite integrals that are computed numerically. Consequently, the FEM with numerical integration is given by

$$(5) \quad \begin{cases} u_h^* \in S_h, \\ a^*(u_h^*, v_h) = f^*(v_h), \quad \forall v_h \in S_h, \end{cases}$$

where

$$a^*(u, v) = \sum_{K \in \mathcal{T}} \sum_{l=1}^L \omega_{l,K}^{(1)} \sum_{i,j=1}^n [a_{ij} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j}] (b_{l,K}^{(1)}),$$

and

$$f^*(v) = \sum_{K \in \mathcal{T}} \sum_{l=1}^L \omega_{l,K}^{(0)} [fv] (b_{l,K}^{(0)}).$$

The two sets $\{\omega_{l,K}^{(1)}, b_{l,K}^{(1)}\}_{l=1}^L$ and $\{\omega_{l,K}^{(0)}, b_{l,K}^{(0)}\}_{l=1}^L$ determine two quadrature rules (possibly different), on the triangle $K \in \mathcal{T}$.

We define the error functionals,

$$E_K^{(i)}(\varphi) = \int_K \varphi(x) dx - \sum_{l=1}^L \omega_{l,K}^{(i)} \varphi(b_{l,K}^{(i)}), \quad i = 0, 1.$$

Let $F_K : \hat{x} \in \hat{K} \rightarrow F_K(\hat{x}) := B_K \hat{x} + b_K$ be the affine mapping that maps the reference element \hat{K} onto K . Then the error functional on the reference element \hat{K} is,

$$\hat{E}^{(i)}(\hat{\varphi}) := \int_{\hat{K}} \hat{\varphi}(\hat{x}) d\hat{x} - \sum_{l=1}^L \hat{\omega}_{l,\hat{K}}^{(i)} \hat{\varphi}(\hat{b}_{l,\hat{K}}^{(i)}),$$

where $b_{l,K}^{(i)} = F_K(\hat{b}_{l,\hat{K}}^{(i)})$, $\omega_{l,K}^{(i)} = \det(B_K) \hat{\omega}_{l,\hat{K}}^{(i)}$, and $\hat{\varphi}(\hat{x}) = \varphi(x)$ for any $x = F_K(\hat{x})$, $\hat{x} \in \hat{K}$. Note that the sets $\{\hat{\omega}_{l,\hat{K}}^{(i)}, \hat{b}_{l,\hat{K}}^{(i)}\}_{l=1}^L$, $i = 0, 1$, define numerical quadrature rules (possibly different) on the reference element \hat{K} . Thus, by the standard scaling argument, it is clear that

$$E_K^{(i)}(\varphi) = \det(B_K) \hat{E}^{(i)}(\hat{\varphi}).$$

We define the *algebraic precision* of a quadrature rule to be m , if the rule is exact on all polynomials of degree $\leq m$. Throughout this paper we will assume that the algebraic precision of the quadrature rules given by $\{\hat{\omega}_{l,\hat{K}}^{(i)}, \hat{b}_{l,\hat{K}}^{(i)}\}_{l=1}^L$, $i = 0, 1$ is $2k - 1$, i.e., $\hat{E}^{(i)}(\hat{\varphi}) = 0$, $\forall \hat{\varphi} \in \mathcal{P}_{2k-1}$.

We now state a result similar to the Theorem 4.1.6 in [9] that we will use later.

Theorem 2.1. *Suppose $\hat{E}^{(i)}(\hat{\varphi}) = 0 \forall \hat{\varphi} \in \mathcal{P}_{2k-1}(\hat{K})$ and let $a_{ij} \in W^{k,\infty}(\Omega)$ and $f \in H^{k+1}(\Omega)$. If $u \in H^{k+1}(\Omega)$, then for $n = 1, 2, 3$ we have*

$$(6) \quad \|u - u_h^*\|_{H^1(\Omega)} \leq Ch^k \left(\sum_{i,j=1}^n \|a_{ij}\|_{W^{k,\infty}(\Omega)} \|u\|_{H^{k+1}(\Omega)} + \|f\|_{H^{k+1}(\Omega)} \right)$$

where the constant C is independent of u and h , but may depend on k .

The proof of this result can be obtained by following the arguments of the proof of Theorem 4.1.6 in [9] and we do not give the proof in this paper. We note that this result can be extended for $n \geq 4$ provided $f \in W^{k+1,q}(\Omega)$ with a suitable $q \geq 2$ depending on k and n .

The linear functionals: Let $G : H_0^1(\Omega) \rightarrow \mathbb{R}$ be a bounded linear functional on $H_0^1(\Omega)$. We are interested in approximating the quantity of interest, $G(u)$, where u is the solution of equation (3).

Recall that $\|\cdot\|_a$ is equivalent to $\|\cdot\|_{H^1(\Omega)}$ for any function in $H_0^1(\Omega)$. Then, by the Riesz representation theorem, there exists a unique $g \in H_0^1(\Omega)$, such that

$$(7) \quad G(v) = a(v, g), \quad \forall v \in H_0^1(\Omega).$$

We approximate $G(u)$ with $G(u_h)$, and the associated error bound is given in the following lemma.

Lemma 2.2. *Suppose that $g \in H^{s+2}(\Omega)$, $0 \leq s \leq k-1$, and let $u \in H^{k+1}(\Omega)$. Then, there is a constant $C > 0$ independent of g , u , and the mesh size h , such that*

$$(8) \quad |G(u) - G(u_h)| \leq Ch^{k+s+1} \|g\|_{H^{s+2}(\Omega)} \|u\|_{H^{k+1}(\Omega)}.$$

Proof. Let $g_h \in S_h$ be the projection of g with respect to $a(\cdot, \cdot)$, namely,

$$(9) \quad a(g - g_h, v_h) = 0, \quad \forall v_h \in S_h.$$

Now using the Galerkin orthogonality $a(u - u_h, v) = 0$ for all $v \in S_h$, the boundedness of $a(\cdot, \cdot)$, a standard approximation result, and (4), we get

$$\begin{aligned} |G(u) - G(u_h)| &= |a(g, u - u_h)| = |a(g - g_h, u - u_h)| \\ &\leq C \|g - g_h\|_{H^1(\Omega)} \|u - u_h\|_{H^1(\Omega)} \\ &\leq Ch^{k+s+1} \|g\|_{H^{s+2}(\Omega)} \|u\|_{H^{k+1}(\Omega)}, \end{aligned}$$

which is the desired result. \square

Remark 2.3. It is clear from (8) in Lemma 2.2 that when $s = k-1$, i.e., when $g \in H^{k+1}(\Omega)$, we have

$$|G(u) - G(u_h)| \leq Ch^{2k} \|g\|_{H^{k+1}(\Omega)} \|u\|_{H^{k+1}(\Omega)}.$$

This is the optimal order of convergence that can be obtained for the error in computing $G(u)$ by $G(u_h)$, where u_h is the finite element solution based on piecewise polynomials of degree k and $u \in H^{k+1}(\Omega)$. We note that the same order of convergence, i.e., $\mathcal{O}(h^{2k})$, is obtained in the context of approximation of eigenvalues by the same FEM (see [3]).

Remark 2.4. The result (8) of Lemma 2.2 also holds for problems with other boundary conditions (e.g., Robin boundary conditions), as long as the associated bilinear form $a(\cdot, \cdot)$ induces an equivalent norm $\|\cdot\|_a$ on a subspace $V \subset H^1(\Omega)$, $G : V \rightarrow \mathbb{R}$ is a bounded linear functional, and $S_h \subset V$. For the problem (2), $V = H_0^1(\Omega)$.

Remark 2.5. As mentioned in the introduction, often a quantity of interest $F(u)$ is finite, but $F : V \rightarrow \mathbb{R}$ is not a bounded linear functional on V . In such situations, $F(u)$ is written in terms of a bounded linear functional $G(u)$ on V , and $F(u)$ is approximated by a quantity F_{u_h} such that $|F(u) - F_{u_h}| \approx |G(u) - G(u_h)|$ (see [4]).

Let $g \in V$ be the unique solution of $a(v, g) = G(v)$, $\forall v \in V$. Then it is clear from (8) of Lemma 2.2 that

$$|F(u) - F_{u_h}| \approx \mathcal{O}(h^{k+s+1}),$$

provided $u \in H^{k+1}(\Omega)$ and $g \in H^{s+2}(\Omega)$, $0 \leq s \leq k-1$. In Section 4, we will consider a particular quantity of interest $F(u)$, where $F(\cdot)$ is not a bounded functional on V .

3. THE EFFECT OF NUMERICAL INTEGRATION

In almost all finite element computations, numerical integration is unavoidable. Consequently, u_h^* instead of u_h is available, and $G(u)$ is approximated by $G(u_h^*)$. We shall concentrate on estimating $|G(u) - G(u_h^*)|$ in this section, and start with the following Strang-type lemma.

Lemma 3.1. *Let $u_h^* \in S_h$ be the finite element solution with numerical integration, given in (5). Let G , g , $a^*(\cdot, \cdot)$, and $f^*(\cdot)$ be defined as in the last section. Then,*

$$(10) \quad G(u) - G(u_h^*) = a(g - g_h, u - u_h) + f(g_h) - f^*(g_h) + a^*(u_h^*, g_h) - a(u_h^*, g_h),$$

where $g_h \in S_h$ is the projection of g onto S_h , given by $a(g - g_h, v_h) = 0$, $\forall v \in S_h$.

Proof. Recall $u_h \in S_h$ is the finite element solution with exact integration. We first note that

$$(11) \quad \begin{aligned} G(u) - G(u_h^*) &= a(g, u - u_h^*) = a(g, u - u_h) + a(g, u_h - u_h^*) \\ &= a(g - g_h, u - u_h) + a(g, u_h - u_h^*), \end{aligned}$$

where we used the well-known fact that u_h is the projection of u on S_h with respect to $a(\cdot, \cdot)$.

Now, using the definition of g_h and (5), we have

$$\begin{aligned} a(g, u_h - u_h^*) &= a(g_h, u_h - u_h^*) = a(g_h, u_h) - a(g_h, u_h^*) \\ &= f(g_h) - a(g_h, u_h^*) \\ &= f(g_h) - f^*(g_h) + f^*(g_h) - a(g_h, u_h^*) \\ &= f(g_h) - f^*(g_h) + a^*(u_h^*, g_h) - a(u_h^*, g_h). \end{aligned}$$

Combining the above with (11) completes the proof. \square

It is clear from Lemma 3.1, that the effect of numerical integration on the approximation of $G(u)$, i.e., on the error $G(u) - G(u_h^*)$, depends on $f(g_h) - f^*(g_h)$ and $a^*(u_h^*, g_h) - a(u_h^*, g_h)$. In fact, if the integration is exact, these differences are zeros. To estimate these terms, we need the following two lemmas.

We now present the first result towards estimating the terms in (10). This result is similar to Lemma 6.2 in [6], where the analysis was given for the one-dimensional case; we prove it here for higher dimensions.

Lemma 3.2. *Suppose $\hat{E}^{(0)}(\hat{\chi}) = 0$, $\forall \hat{\chi} \in \mathcal{P}_{2k-1}(\hat{K})$ and let $0 \leq s \leq k-1$. Then, for $1 \leq n \leq 3$,*

$$|E_K^{(0)}(f\phi)| \leq Ch^{k+s+1} \|f\|_{H^{k+s+1}(K)} \|\phi\|_{H^{s+2}(K)}, \quad \forall \phi \in \mathcal{P}_k(K),$$

where the constant C is independent of h and K , but may depend on k .

Proof. We first prove it for $k > 1$. Let $\hat{\Pi}$ be the L^2 -projection onto $\mathcal{P}_{s+1}(\hat{K})$. We write

$$(12) \quad \hat{E}^{(0)}(\hat{f}\hat{\phi}) = \hat{E}^{(0)}(\hat{f}\hat{\Pi}\hat{\phi}) + \hat{E}^{(0)}(\hat{f}\hat{\phi} - \hat{f}\hat{\Pi}\hat{\phi}).$$

Since $k + s + 1 \geq k + 1 > n/2$, we use Sobolev's inequality (see 1.4.6 in [8]) to get

$$|\hat{E}^{(0)}(\hat{f}\hat{\Pi}\hat{\phi})| \leq C \|\hat{f}\hat{\Pi}\hat{\phi}\|_{L^\infty(\hat{K})} \leq C \|\hat{f}\hat{\Pi}\hat{\phi}\|_{H^{k+s+1}(\hat{K})}.$$

We recall that $\hat{E}^{(0)}(\hat{\chi}) = 0$, $\forall \hat{\chi} \in \mathcal{P}_{k+s}(\hat{K})$. Therefore,

$$(13) \quad \begin{aligned} |\hat{E}^{(0)}(\hat{f}\hat{\Pi}\hat{\phi})| &\leq C \|\hat{f}\hat{\Pi}\hat{\phi}\|_{H^{k+s+1}(\hat{K})} \leq C \sum_{i=0}^{s+1} |\hat{f}|_{H^{k+s+1-i}(\hat{K})} |\hat{\Pi}\hat{\phi}|_{W^{i,\infty}(\hat{K})} \\ &\leq C \sum_{i=0}^{s+1} |\hat{f}|_{H^{k+s+1-i}(\hat{K})} |\hat{\Pi}\hat{\phi}|_{H^i(\hat{K})}, \end{aligned}$$

where we used the Bramble-Hilbert Lemma and the equivalence of norms on finite dimensional spaces. Since $\hat{\Pi}v = v$, $\forall v \in \mathcal{P}_{i-1}(\hat{K})$, $1 \leq i \leq s+1$, by the inverse inequality and the Bramble-Hilbert Lemma, we get

$$|\hat{\phi} - \hat{\Pi}\hat{\phi}|_{H^i(\hat{K})} \leq C \|\hat{\phi} - \hat{\Pi}\hat{\phi}\|_{L^2(\hat{K})} \leq C |\hat{\phi}|_{H^i(\hat{K})},$$

where C may depend on k . Therefore,

$$|\hat{\Pi}\hat{\phi}|_{H^i(\hat{K})} \leq C |\hat{\phi}|_{H^i(\hat{K})} + |\hat{\phi} - \hat{\Pi}\hat{\phi}|_{H^i(\hat{K})} \leq C |\hat{\phi}|_{H^i(\hat{K})}.$$

Also since $\hat{\Pi}$ is the L^2 -projection, the above inequality is true for $i = 0$. Thus from (13) and using a standard scaling argument, we obtain

$$(14) \quad \begin{aligned} |\hat{E}^{(0)}(\hat{f}\hat{\Pi}\hat{\phi})| &\leq C \sum_{i=0}^{s+1} |\hat{f}|_{H^{k+s+1-i}(\hat{K})} |\hat{\phi}|_{H^i(\hat{K})} \\ &\leq C \det(B_K)^{-1} h^{k+s+1} \sum_{i=0}^{s+1} |f|_{H^{k+s+1-i}(K)} |\phi|_{H^i(K)} \\ &\leq C \det(B_K)^{-1} h^{k+s+1} \|f\|_{H^{k+s+1}(K)} \|\phi\|_{H^{s+1}(K)}. \end{aligned}$$

We now estimate the term $\hat{E}^{(0)}(\hat{f}\hat{\phi} - \hat{f}\hat{\Pi}\hat{\phi})$ in (12). Note that $\hat{f}\hat{\phi} - \hat{f}\hat{\Pi}\hat{\phi} = 0$ if $s = k - 1$. So assume $0 \leq s < k - 1$. Since $k > n/2$, again using the Sobolev inequality, we obtain

$$\begin{aligned} |\hat{E}^{(0)}(\hat{f}\hat{\phi} - \hat{f}\hat{\Pi}\hat{\phi})| &\leq C \|\hat{f}\hat{\phi} - \hat{f}\hat{\Pi}\hat{\phi}\|_{L^\infty(\hat{K})} \leq C \|\hat{f}\|_{L^\infty(\hat{K})} \|\hat{\phi} - \hat{\Pi}\hat{\phi}\|_{L^\infty(\hat{K})} \\ &\leq C \|\hat{f}\|_{H^k(\hat{K})} \|\hat{\phi} - \hat{\Pi}\hat{\phi}\|_{L^\infty(\hat{K})}. \end{aligned}$$

Hence, for a fixed $\hat{\phi} \in \mathcal{P}_k(\hat{K})$, the linear functional $\hat{f} \in H^k(\hat{K}) \rightarrow \hat{E}^{(0)}(\hat{f}\hat{\phi} - \hat{f}\hat{\Pi}\hat{\phi})$ is bounded and vanishes over the space $\mathcal{P}_{k-1}(\hat{K})$. Now using the Bramble-Hilbert Lemma, the equivalence of norms on finite dimensional spaces, the scaling argument, and the fact that $\hat{\Pi}$ leaves $\mathcal{P}_{s+1}(\hat{K})$ invariant, we obtain

$$(15) \quad \begin{aligned} |\hat{E}^{(0)}(\hat{f}\hat{\phi} - \hat{f}\hat{\Pi}\hat{\phi})| &\leq C \|\hat{f}\|_{H^k(\hat{K})} \|\hat{\phi} - \hat{\Pi}\hat{\phi}\|_{L^2(\hat{K})} \leq C \|\hat{f}\|_{H^k(\hat{K})} |\hat{\phi}|_{H^{s+2}(\hat{K})} \\ &\leq C \det(B_K)^{-1} h^{k+s+2} |f|_{H^k(K)} |\phi|_{H^{s+2}(K)} \\ &\leq C \det(B_K)^{-1} h^{k+s+2} \|f\|_{H^k(K)} \|\phi\|_{H^{s+2}(K)}. \end{aligned}$$

Finally combining (12), (14), and (15), we have

$$|E_K^{(0)}(f\phi)| = |\det(B_K)\hat{E}^{(0)}(\hat{f}\hat{\phi})| \leq Ch^{k+s+1}\|f\|_{H^{k+s+1}(K)}\|\phi\|_{H^{s+2}(K)},$$

which completes the proof for $k > 1$.

Now we prove the result of this lemma for $k = 1$. In this case, $s = 0$ and $\hat{\phi} \in \mathcal{P}_1$. Using the Sobolev inequality, we get

$$|\hat{E}^{(0)}(\hat{f}\hat{\phi})| \leq C\|\hat{f}\hat{\phi}\|_{L^\infty(\hat{K})} \leq C\|\hat{f}\hat{\phi}\|_{H^2(\hat{K})}.$$

Since $\hat{E}^{(0)}(\hat{\chi}) = 0, \forall \hat{\chi} \in \mathcal{P}_1(\hat{K})$, using the Bramble-Hilbert Lemma, the equivalence of norms on finite dimensional spaces, and the scaling argument, we have

$$\begin{aligned} |\hat{E}^{(0)}(\hat{f}\hat{\phi})| &\leq C|\hat{f}\hat{\phi}|_{H^2(\hat{K})} \leq C\sum_{i=0}^1|\hat{f}|_{H^{2-i}(\hat{K})}|\hat{\phi}|_{W^{i,\infty}(\hat{K})} \\ &\leq C\sum_{i=0}^1|\hat{f}|_{H^{2-i}(\hat{K})}|\hat{\phi}|_{H^i(\hat{K})} \\ &\leq C\det(B_K)^{-1}h^2\sum_{i=0}^1|f|_{H^{2-i}(K)}|\phi|_{H^i(K)} \\ &\leq C\det(B_K)^{-1}h^2\|f\|_{H^2(K)}\|\phi\|_{H^1(K)}. \end{aligned}$$

Therefore, we have

$$|E_K^{(0)}(f\phi)| = |\det(B_K)\hat{E}^{(0)}(\hat{f}\hat{\phi})| \leq Ch^2\|f\|_{H^2(K)}\|\phi\|_{H^2(K)},$$

which completes the proof for $k = 1$. \square

Remark 3.3. The result in Lemma 3.2 has been obtained for $1 \leq n \leq 3$. A similar result could be obtained for $n \geq 4$ provided $f \in W^{k+s+1,q}(K)$, where $q > 2$ is a real number satisfying $k > n/q$. We do not prove the result in this paper.

We now state the next result, which is similar to Lemma 6.1 in [6] (it was also proved for the one-dimensional case). We do not give a proof as it could be obtained by following exactly the arguments in the proof of Lemma 6.1 in [6].

Lemma 3.4. *Suppose $\hat{E}^{(1)}(\hat{\chi}) = 0, \forall \hat{\chi} \in \mathcal{P}_{2k-1}(\hat{K})$ and for $0 \leq s \leq k-1$, let $a(x) \in W^{k+s+1,\infty}(\Omega)$. Then,*

$$|E_K^{(1)}(a\varphi\phi)| \leq Ch^{k+s+1}\|a\|_{W^{k+s+1,\infty}(\Omega)}\|\varphi\|_{H^k(K)}\|\phi\|_{H^{s+1}(K)}, \quad \forall \varphi, \phi \in \mathcal{P}_k(K),$$

where the constant C is independent of K .

Remark 3.5. We note that in Lemma 6.1 of [6], it was assumed that $a(x) \in W^{2k,\infty}(\Omega)$. A careful reading of the proof of Lemma 6.1 of [6] reveals that $a(x) \in W^{k+s+1,\infty}(\Omega), 0 \leq s \leq k-1$, is only needed; we have used this assumption on $a(x)$ in the above Lemma 3.4.

With Lemma 3.1, 3.2 and 3.4, we are now ready to give the upper bound of $|G(u) - G(u_h^*)|$ in the following theorem.

Theorem 3.6. *Suppose $\hat{E}^{(i)}(\hat{\chi}) = 0, \forall \hat{\chi} \in \mathcal{P}_{2k-1}(\hat{K}), i = 0, 1$. For $0 \leq s \leq k-1$, let $f \in H^{k+s+1}(\Omega), g \in H^{s+2}(\Omega)$, and $a_{ij} \in W^{k+s+1,\infty}(\Omega)$. Then for $1 \leq n \leq 3$, we have*

$$(16) \quad |G(u) - G(u_h^*)| \leq Ch^{k+s+1}(\|u\|_{H^{k+1}(\Omega)} + \|f\|_{H^{k+s+1}(\Omega)})\|g\|_{H^{s+2}(\Omega)},$$

where the constant C does not depend on the mesh size h . In particular, for $s = k - 1$, we have

$$(17) \quad |G(u) - G(u_h^*)| \leq Ch^{2k}(\|u\|_{H^{k+1}(\Omega)} + \|f\|_{H^{2k}(\Omega)})\|g\|_{H^{k+1}(\Omega)}.$$

Proof. We first note from Lemma 3.1 that

$$(18) \quad G(u) - G(u_h^*) = a(g - g_h, u - u_h) + f(g_h) - f^*(g_h) + a^*(u_h^*, g_h) - a(u_h^*, g_h),$$

where $g_h \in S_h$ satisfies $a(g - g_h, v_h) = 0$, $\forall v_h \in S_h$. It is clear from (4) and a approximation result that

$$(19) \quad \begin{aligned} |a(g - g_h, u - u_h)| &\leq C\|g - g_h\|_{H^1(\Omega)}\|u - u_h\|_{H^1(\Omega)} \\ &\leq Ch^{k+s+1}\|g\|_{H^{s+2}(\Omega)}\|u\|_{H^{k+1}(\Omega)}. \end{aligned}$$

We will now obtain upper bounds for $|f(g_h) - f^*(g_h)|$ and $|a^*(u_h^*, g_h) - a(u_h^*, g_h)|$. For $v \in S_h$, we define

$$\|v\|_{i,\mathcal{T}} := \left(\sum_{K \in \mathcal{T}} \|v\|_{H^i(K)}^2 \right)^{1/2}, \quad i = 0, 1, \dots, k.$$

Clearly, $\|v\|_{1,\mathcal{T}} = \|v\|_{H^1(\Omega)}$ which is also true for any $v \in H^1(\Omega)$. Then from Lemma 3.2, we get

$$(20) \quad \begin{aligned} |f(g_h) - f^*(g_h)| &\leq \sum_{K \in \mathcal{T}} |E_K^{(0)}(f g_h)| \\ &\leq Ch^{k+s+1} \sum_{K \in \mathcal{T}} \|f\|_{H^{k+s+1}(K)} \|g_h\|_{H^{s+2}(K)} \\ &\leq Ch^{k+s+1} \|f\|_{H^{k+s+1}(\Omega)} \|g_h\|_{s+2,\mathcal{T}}. \end{aligned}$$

Let $\mathcal{I}_h v \in S_h$ be the S_h -interpolant of $v \in H^{s+2}(\Omega)$. Then it is well known (see [9]) that

$$(21) \quad \|v - \mathcal{I}_h v\|_{i,\mathcal{T}} \leq Ch^{s+2-i} \|v\|_{H^{s+2}(\Omega)}, \quad i = 0, 1, \dots, s+2.$$

Therefore from the triangle inequality, the inverse inequality, and a standard approximation result, we have

$$\begin{aligned} \|g_h\|_{s+2,\mathcal{T}} &\leq \|g\|_{H^{s+2}(\Omega)} + \|g - \mathcal{I}_h g\|_{s+2,\mathcal{T}} + \|\mathcal{I}_h g - g_h\|_{s+2,\mathcal{T}} \\ &\leq C\|g\|_{H^{s+2}(\Omega)} + Ch^{-(s+1)} \|\mathcal{I}_h g - g_h\|_{1,\mathcal{T}} \\ &\leq C\|g\|_{H^{s+2}(\Omega)} + Ch^{-(s+1)} (\|\mathcal{I}_h g - g\|_{H^1(\Omega)} + \|g - g_h\|_{H^1(\Omega)}) \\ &\leq C\|g\|_{H^{s+2}(\Omega)}. \end{aligned}$$

Hence from (20), we get

$$(22) \quad |f(g_h) - f^*(g_h)| \leq Ch^{k+s+1} \|f\|_{H^{k+s+1}(\Omega)} \|g\|_{H^{s+2}(\Omega)}.$$

For the last term $|a^*(u_h^*, g_h) - a(u_h^*, g_h)|$, we use Lemma 3.4 and follow the argument leading to (20) to get

$$(23) \quad \begin{aligned} &|a^*(u_h^*, g_h) - a(u_h^*, g_h)| \\ &\leq Ch^{k+s+1} \sum_{i,j=1}^n \|a_{i,j}\|_{W^{k+s+1,\infty}(\Omega)} \|u_h^*\|_{k+1,\mathcal{T}} \|g_h\|_{s+2,\mathcal{T}}. \end{aligned}$$

Again by the triangle inequality, the inverse inequality, (21) with $v = u$, $s = k - 1$, and (6), we obtain

$$\begin{aligned}
\|u_h^*\|_{k+1,\mathcal{T}} &\leq \|u\|_{H^{k+1}(\Omega)} + \|u - \mathcal{I}_h u\|_{k+1,\mathcal{T}} + \|\mathcal{I}_h u - u_h^*\|_{k+1,\mathcal{T}} \\
&\leq C\|u\|_{H^{k+1}(\Omega)} + Ch^{-k}\|\mathcal{I}_h u - u_h^*\|_{1,\mathcal{T}} \\
&\leq C\|u\|_{H^{k+1}(\Omega)} + Ch^{-k}(\|\mathcal{I}_h u - u\|_{H^1(\Omega)} + \|u - u_h^*\|_{H^1(\Omega)}) \\
&\leq C\left(\sum_{i,j=1}^n \|a_{ij}\|_{W^{k,\infty}(\Omega)}\|u\|_{H^{k+1}(\Omega)} + \|f\|_{H^{k+1}(\Omega)}\right).
\end{aligned}$$

Hence from (23) we have,

$$\begin{aligned}
&|a^*(u_h^*, g_h) - a(u_h^*, g_h)| \\
(24) \quad &\leq Ch^{k+s+1}\left[\sum_{i,j=1}^n \|a_{ij}\|_{W^{k+s+1,\infty}(\Omega)}\right]\left(\sum_{i,j=1}^n \|a_{ij}\|_{W^{k,\infty}(\Omega)}\|u\|_{H^{k+1}(\Omega)}\right. \\
&\quad \left. + \|f\|_{H^{k+1}(\Omega)}\right)\|g\|_{H^{s+2}(\Omega)}.
\end{aligned}$$

Finally, combining (18), (19), (22), and (24), we obtain

$$\begin{aligned}
&|G(u) - G(u_h^*)| \\
&\leq Ch^{k+s+1}\left[\|g\|_{H^{s+2}(\Omega)}\|u\|_{H^{k+1}(\Omega)} + \|f\|_{H^{k+s+1}(\Omega)}\|g\|_{H^{s+2}(\Omega)}\right. \\
&\quad \left. + \left(\sum_{i,j=1}^n \|a_{ij}\|_{W^{k+s+1,\infty}(\Omega)}\right)\left(\sum_{i,j=1}^n \|a_{ij}\|_{W^{k,\infty}(\Omega)}\|u\|_{H^{k+1}(\Omega)}\right.\right. \\
&\quad \left. \left. + \|f\|_{H^{k+1}(\Omega)}\right)\|g\|_{H^{s+2}(\Omega)}\right] \\
&\leq Ch^{k+s+1}(\|u\|_{H^{k+1}(\Omega)} + \|f\|_{H^{k+s+1}(\Omega)})\|g\|_{H^{s+2}(\Omega)},
\end{aligned}$$

which is the desired result. \square

Remark 3.7. We note that we get the optimal order of convergence (17) under an increased regularity requirement on the data, i.e., $f \in H^{2k}(\Omega) \cap \mathcal{C}^0(\Omega)$ and $a_{ij} \in W^{2k,\infty}(\Omega)$; this increased regularity is not necessary to obtain (8) – an optimal-order error estimate with exact integration. We mention that an increased regularity of the data was also assumed (see [6]) for the optimal order of convergence, $\mathcal{O}(h^{2k})$, of approximate eigenvalues, obtained from finite element method under numerical integration.

Remark 3.8. It is well known that to obtain the optimal order of convergence of the H^1 -norm of the error in the finite element solution under numerical integration, one needs an increased regularity of the data, namely, $a_{ij} \in W^{k,\infty}(\Omega)$ and $f \in H^k(\Omega) \cap \mathcal{C}^0(\Omega)$ for $1 \leq n \leq 3$ (see [9]). Moreover, the algebraic precision of the numerical integration rule is required to be $2k - 2$. In a standard finite element software, often the algebraic precision of the numerical quadrature is $2k - 1$, which is the same as the assumption used in Theorem 3.6. In Theorem 3.6, however, we required a higher regularity of the data, i.e., $a_{ij} \in W^{2k,\infty}(\Omega)$ and $f \in H^{2k}(\Omega) \cap \mathcal{C}^0(\Omega)$ for the optimal order of convergence of $G(u_h^*)$. Thus, the error in approximating linear functionals may be more sensitive to numerical integration than the error (in H^1 -norm) in the finite element solution, when the data is not sufficiently smooth.

Remark 3.9. The result of Theorem 3.6 also holds when the exact solution u satisfies other boundary conditions, provided the functions in S_h satisfy the essential boundary conditions (see Remark 2.4).

4. A LOWER BOUND ON THE ERROR IN APPROXIMATING FUNCTIONALS

In this section, our main goal is to show that the smoothness assumption on the data of the problem (see Theorem 3.6) is necessary, in general, to obtain the optimal order of convergence of the computed quantity of interest. Also in this section, we will consider a quantity of interest $F(u)$, which is finite but $F(\cdot)$ is not a bounded linear functional on the energy space. We will first write $F(u)$ in terms of a bounded linear functional $G(u)$ and use the results in Section 3 to obtain an upper bound of the error in the computed quantity of interest, when the data has enough smoothness. We will then obtain a lower bound of the error in computed quantity of interest for a particular problem. This result will indicate that the increased regularity of the data is, in general, necessary to obtain the optimal order of convergence of the computed quantity of interest.

Consider the one-dimensional problem on $\Omega = (0, 1)$ with the Robin boundary condition

$$(25) \quad \begin{cases} -u''(x) = f(x) & x \in \Omega, \\ u(0) = 0, \quad u(1) + u'(1) = 0, \end{cases}$$

Suppose we want to approximate the quantity of interest $F(u) = u'(0)$.

Let

$$H_D^1 := \{v \in H^1(\Omega), \quad v(0) = 0\}.$$

The variational formulation of (25) is given by

$$(26) \quad \begin{cases} u \in H_D^1, \\ a(u, v) := \int_0^1 u'v' dx + u(1)v(1) = \int_0^1 f v dx, \quad \forall v \in H_D^1, \end{cases}$$

where we assume that $f \in L^2(\Omega)$. Using the Sobolev embedding theorem and the Poincaré inequality, it is clear that there exist positive constants C_1, C_2 , such that

$$(27) \quad C_1 \|v\|_{H^1(\Omega)}^2 \leq a(v, v) \leq C_2 \|v\|_{H^1(\Omega)}^2, \quad \forall v \in H_D^1,$$

Therefore, $\|v\| := a(v, v)$ is an equivalent norm in H_D^1 with the inner product $a(\cdot, \cdot)$.

We first note that $F(u) = u'(0)$ is finite, but $F(\cdot)$ is not a bounded linear functional on H_D^1 , and therefore we cannot use the framework developed in Section 3. Moreover, it is well known that $F(u_h)$ does not give an accurate approximation of $F(u)$, where u_h is the finite element solution. To address this issue (see Section 6.2 in [4]), we choose a function $\psi(x) \in H^1(\Omega)$, such that $\psi(0) = 1$ and $\psi(1) = 0$. Then using integration by parts, we have

$$\begin{aligned} F(u) &= u'(0) = -u'\psi|_0^1 = -\int_0^1 \psi u'' dx - \int_0^1 \psi' u' dx \\ &= \int_0^1 f \psi dx - \int_0^1 \psi' u' dx. \end{aligned}$$

Let

$$G(u) := \int_0^1 \psi' u' dx.$$

Then by Schwartz's inequality, we get

$$\begin{aligned} |G(v)| &= \left| \int_0^1 \psi' v' dx \right| \leq |\psi|_{H^1(\Omega)} |v|_{H^1(\Omega)} \\ &\leq C \|v\|_{H^1(\Omega)}, \quad \forall v \in H_D^1, \end{aligned}$$

where $C = |\psi|_{H^1(\Omega)}$. Thus $G(\cdot)$, which depends on $\psi(x)$, is a bounded linear functional on H_D^1 , and $F(u)$ can be written as

$$(28) \quad F(u) = \int_0^1 f \psi dx - G(u).$$

We note that this expression of $F(u)$ is meaningful, since $f \in L^2(\Omega)$. We further note that $F(u)$ can be written in this form for any $\psi(x) \in H^1(\Omega)$ with $\psi(0) = 1$ and $\psi(1) = 0$. We consider a particular $\psi(x)$ given by

$$(29) \quad \psi(x) := 1 - x + \epsilon \sin(\pi x),$$

for a fixed $0 < \epsilon \leq 1$. In fact, we could have considered $\epsilon = 0$ in the definition of $\psi(x)$, but we chose $\epsilon > 0$ to show a particular feature at the end of this section.

It is clear from (28) that, if $\int_0^1 f \psi dx$ is computed accurately, we can use the framework developed in Section 3 to accurately approximate $G(u)$, which in turn will give an accurate approximation of $F(u)$. We further note that by the Riesz representation theorem, there exists a unique $g \in H_D^1$ such that

$$(30) \quad a(v, g) = G(v), \quad \forall v \in H_D^1.$$

An easy calculation using integration by parts yields

$$g(x) = \epsilon \sin(\pi x) - 0.5x.$$

To approximate the solution of (26), we consider the points $x_j = jh$, $j = 0, 1, \dots, n$, where n is a natural number and $h = 1/n$. Let $S_h \subset H_D^1$ be the piecewise quadratic ($k = 2$) finite element space with uniform mesh of size h . The FEM to approximate the solution $u \in H_D^1$ of (26) is given by

$$(31) \quad \begin{cases} u_h \in S_h, \\ a(u_h, v_h) = \int_0^1 f v_h dx, \quad \forall v_h \in S_h. \end{cases}$$

We define the quantity

$$(32) \quad F_{u_h} := \int_0^1 f \psi dx - G(u_h).$$

If $\int_0^1 f \psi dx$ is computed exactly, then the quantity of interest $F(u)$ can be approximated by F_{u_h} and from (28), we get

$$F(u) - F_{u_h} = G(u) - G(u_h).$$

Clearly $g(x) = \epsilon \sin(\pi x) - 0.5x \in H^3(\Omega)$. If $u \in H^3(\Omega)$, then from (8) of Lemma 2.2 with $k = 2$ and $s = k - 1 = 1$, we get

$$(33) \quad |F(u) - F_{u_h}| = |G(u) - G(u_h)| \leq Ch^4 \|g\|_{H^3(\Omega)} \|u\|_{H^3(\Omega)}.$$

But u_h is often not available, as the integrals in (31) are computed by numerical integration, and therefore we cannot compute F_{u_h} to approximate $F(u)$.

We use the 2-point Gaussian quadrature, determined by the set $\{\omega_{l,i}, b_{l,i}\}_{l=1}^2$, on the interval (x_{i-1}, x_i) , $i = 1, 2, \dots, n$, to approximate the integrals in (31).

This quadrature rule is exact for polynomials of degree $2k - 1 = 3$, as required in Theorem 3.6. The FEM under numerical integration is given by

$$(34) \quad \begin{cases} u_h^* \in S_h, \\ a(u_h^*, v_h) = f^*(v_h), \quad \forall v_h \in S_h, \end{cases}$$

where

$$f^*(v) = \sum_{i=1}^n \sum_{l=1}^2 \omega_{l,i} f(b_{l,i}) v(b_{l,i}).$$

We note that with the 2-point Gaussian quadrature, we have $a^*(u, v) = a(u, v)$ for all $u, v \in S_h$.

We approximate $F(u)$ by the quantity

$$(35) \quad F_{u_h^*} := \int_0^1 f \psi dx - G(u_h^*).$$

If $\int_0^1 f \psi dx$ is computed exactly, then from (28) we get

$$(36) \quad F(u) - F_{u_h^*} = G(u) - G(u_h^*).$$

Suppose $f \in H^4(\Omega)$. Then from (17) of Theorem 3.6 (considering $k = 2$ and $s = k - 1 = 1$), we have

$$(37) \quad |F(u) - F_{u_h^*}| = |G(u) - G(u_h^*)| \leq Ch^4 (\|u\|_{H^3(\Omega)} + \|f\|_{H^4(\Omega)}) \|g\|_{H^3(\Omega)}.$$

We note however that in practice, the definite integral $\int_0^1 f \psi dx$ in (35) is also computed by numerical integration. Therefore this definite integral has to be computed accurately to obtain $|F(u) - F_{u_h^*}| = \mathcal{O}(h^4)$. We further note that the framework developed in Section 3 for bounded linear functionals could be used to analyze the accuracy of $F_{u_h^*}$, when the quantity of interest is $F(u)$ and $F(\cdot)$ is not a bounded linear functional.

Remark 4.1. It is possible to approximate $F(u)$ directly without using the function $\psi(x)$ as in (32). In particular, $F(u) = u'(0)$ in our example could be well approximated by directly evaluating $F(u_h) = u_h'(0)$ (assuming exact integration), as it is well known (see [14, 16]) that $|F(u) - F(u_h)| \leq \|u - u_h\|_{W^{1,\infty}(\Omega)} = \mathcal{O}(h^k)$. But computing $F(u_h)$ in terms of the function $\psi(x)$ as in (32) enables us to get $|F(u) - F(u_h)| \approx \mathcal{O}(h^{2k}) \approx \|u - u_h\|_{H^1(\Omega)}^2$. We mention however that computing $F(u_h)$ using (32) (or $F(u_h^*)$ using (35)) is costlier than a direct computation, since an appropriate $\psi(x)$ has to be chosen and the integrals $\int_0^1 f \psi dx$ and $G(u_h)$ (or $G(u_h^*)$) have to be computed accurately. On the other hand, computing $F(u_h)$ using (32) (or $F(u_h^*)$ using (35)) will allow the use of a coarser mesh to approximate $F(u)$ within a given tolerance.

Remark 4.2. In many problems $F(u)$ could be infinite and thus meaningless, e.g., say when $F(u)$ is the value of the stresses at a re-entrant corner in the elasticity problem. But computation of $F_{u_h^*}$ in such a situation could yield a finite value; however, $|F(u) - F(u_h^*)|$ is meaningless and any conclusion based on the value of $F_{u_h^*}$ could be misleading.

We will now obtain a lower bound of the error $|F(u) - F_{u_h^*}|$ for a particular problem, where the data does not have the increased regularity assumed in Theorem 3.6. This result will indicate that the increased regularity assumption on the data

in Theorem 3.6 may be necessary to obtain (17). We consider a particular problem (25) with

$$f(x) = (1-x)^{5/3}.$$

Clearly, $f \in H^2(\Omega)$, but $f \notin H^3(\Omega)$. Thus f has less regularity than required in Theorem 3.6 with $k = 2$ and $s = k - 1 = 1$, and $|F(u) - F_{u_h^*}|$ may not have the optimal order of convergence $\mathcal{O}(h^4)$ in contrast to (37) (see also Remarks 3.7 and 3.8). We note however that the regularity $f \in H^2(\Omega)$ is sufficient to obtain the optimal order of convergence, $\mathcal{O}(h^2)$, for $\|u - u_h^*\|_{H^1(\Omega)}$ (see [9]).

To analyze the error $G(u) - G(u_h^*)$, and consequently the error $F(u) - F_{u_h^*}$ (see (36)), we first consider the projection $g_h \in S_h$ of g given by

$$(38) \quad a(g - g_h, v_h) = 0, \quad \forall v_h \in S_h.$$

Since $g_h \in S_h$, it is piecewise quadratic and can be written as

$$(39) \quad g_h(x) = \sum_{i=1}^n [c_i \phi_i(x) + b_i B_i(x)],$$

where $\phi_i(x)$ is the usual ‘‘hat function’’ centered at x_i , and $B_i(x) = \sqrt{6}h^{-2}(x - x_{i-1})(x - x_i)$ is the quadratic ‘‘bubble function’’ on the interval (x_{i-1}, x_i) . It is well known (in one-dimension) that

$$c_i = g_h(x_i) = g(x_i),$$

and therefore, $g_h(x)$, for $x \in (x_{i-1}, x_i)$, can be written as

$$(40) \quad g_h(x) = \frac{g(x_i) - g(x_{i-1})}{h}(x - x_{i-1}) + g(x_{i-1}) + b_i B_i(x).$$

Remark 4.3. We note that if we define $\psi(x)$ in (29) with $\epsilon = 0$, then $g(x) = -0.5x$. Consequently, $g_h(x)$, which is a projection of $g(x)$ onto S_h , is the same as $g(x)$, i.e., $g_h(x) = g(x)$.

In the following lemma, we estimate the constant b_i in (40).

Lemma 4.4. *Let $g_h(x) = \sum_{i=1}^n [c_i \phi_i(x) + b_i B_i(x)]$ be the projection of g onto S_h given by (38). Then,*

$$b_i = \frac{\int_{x_{i-1}}^{x_i} \psi' B_i' dx}{\int_{x_{i-1}}^{x_i} (B_i')^2 dx} = -\frac{\pi^2 \sqrt{6}}{12} \epsilon h^2 \sin\left(\frac{\pi}{2}(x_{i-1} + x_i)\right) + \mathcal{O}(\epsilon h^4).$$

Proof. We first recall from (30) that

$$a(g_h, v_h) = a(g, v_h) = \int_0^1 \psi' v_h' dx, \quad \forall v_h \in S_h.$$

By the definition of $a(\cdot, \cdot)$ and the orthogonality of Legendre polynomials, we have

$$\begin{aligned} \int_0^1 \psi'(x) B_i'(x) dx &= a(g_h, B_i(x)) \\ &= \int_0^1 B_i'(x) \sum_{i=1}^n (c_i \phi_i'(x) + b_i B_i'(x)) dx = \int_{x_{i-1}}^{x_i} b_i [B_i'(x)]^2 dx. \end{aligned}$$

Therefore,

$$b_i = \frac{\int_{x_{i-1}}^{x_i} \psi'(x) B_i'(x) dx}{\int_{x_{i-1}}^{x_i} [B_i'(x)]^2 dx}.$$

Since $B_i(x) = \sqrt{6}h^{-2}(x - x_{i-1})(x - x_i)$, a direct computation gives

$$\int_{x_{i-1}}^{x_i} [B_i'(x)]^2 dx = \frac{2}{h}.$$

We now focus on the integral of $\int_{x_{i-1}}^{x_i} \psi'(x) B_i'(x) dx$. Clearly,

$$\begin{aligned} \int_{x_{i-1}}^{x_i} \psi'(x) B_i'(x) dx &= \sqrt{6}h^{-2} \int_{x_{i-1}}^{x_i} (2x - x_{i-1} - x_i) [\epsilon\pi \cos(\pi x) - 1] dx \\ &= \epsilon\sqrt{6}h^{-2} \left(2x_i \sin(\pi x_i) - 2x_{i-1} \sin(\pi x_{i-1}) \right. \\ &\quad \left. - (x_i + x_{i-1}) [\sin(\pi x_i) - \sin(\pi x_{i-1})] - 2 \int_{x_{i-1}}^{x_i} \sin(\pi x) dx \right) \\ &= \epsilon\sqrt{6}h^{-2} \left((x_i - x_{i-1}) [\sin(\pi x_i) + \sin(\pi x_{i-1})] \right. \\ &\quad \left. + \frac{2}{\pi} [\cos(\pi x_i) - \cos(\pi x_{i-1})] \right) \\ &= \epsilon\sqrt{6}h^{-2} \left(h [\sin(\pi x_{i-1}) + \sin(\pi x_i)] \right. \\ &\quad \left. - \frac{4}{\pi} \sin\left(\frac{\pi h}{2}\right) \sin\left(\frac{\pi(x_{i-1} + x_i)}{2}\right) \right). \end{aligned}$$

Using the Taylor series,

$$\sin(x) = x - \frac{x^3}{3!} + \mathcal{O}(x^5),$$

for x small, we have

$$(41) \quad \int_{x_{i-1}}^{x_i} \psi'(x) B_i'(x) dx = \epsilon\sqrt{6}h^{-1} \left(\sin(\pi x_{i-1}) + \sin(\pi x_i) - 2 \sin\left(\frac{\pi(x_{i-1} + x_i)}{2}\right) \right. \\ \left. + \frac{\pi^2 h^2}{12} \sin\left(\frac{\pi(x_{i-1} + x_i)}{2}\right) + \mathcal{O}(h^4) \right).$$

Now,

$$\begin{aligned} &\epsilon\sqrt{6}h^{-1} \left(\sin(\pi x_{i-1}) + \sin(\pi x_i) - 2 \sin\left(\frac{\pi(x_{i-1} + x_i)}{2}\right) \right) \\ &= \epsilon\sqrt{6}h^{-1} \left(2 \cos\left(\frac{3\pi}{4}x_i + \frac{\pi}{4}x_{i-1}\right) \sin\left(\frac{\pi}{4}h\right) - 2 \cos\left(\frac{3\pi}{4}x_{i-1} + \frac{\pi}{4}x_i\right) \sin\left(\frac{\pi}{4}h\right) \right) \\ &= \frac{\sqrt{6}\epsilon\pi}{2} \left(\left(1 - \frac{\pi^2 h^2}{96}\right) \left[\cos\left(\frac{3\pi}{4}x_i + \frac{\pi}{4}x_{i-1}\right) - \cos\left(\frac{3\pi}{4}x_{i-1} + \frac{\pi}{4}x_i\right) \right] + \mathcal{O}(h^4) \right) \\ &= \frac{\sqrt{6}\epsilon\pi}{2} \left(\left(1 - \frac{\pi^2 h^2}{96}\right) \left[-2 \sin\left(\frac{\pi}{4}h\right) \sin\left(\frac{\pi}{2}(x_i + x_{i-1})\right) \right] + \mathcal{O}(h^4) \right) \\ (42) \quad &= \frac{\sqrt{6}\epsilon\pi}{2} \left(\left(1 - \frac{\pi^2 h^2}{96}\right) \left[-\frac{\pi}{2}h + \frac{\pi^3}{192}h^3 + \mathcal{O}(h^5) \right] \sin\left(\frac{\pi}{2}(x_i + x_{i-1})\right) + \mathcal{O}(h^4) \right). \end{aligned}$$

Combining (41) and (42), we obtain

$$\int_{x_{i-1}}^{x_i} \psi'(x) B_i'(x) dx = -\frac{\pi^2 \sqrt{6}}{6} \epsilon \sin\left(\frac{\pi}{2}(x_{i-1} + x_i)\right) h + \mathcal{O}(\epsilon h^3),$$

where the absolute value of the coefficient of ϵh^3 is less than 1, and independent of h . Therefore,

$$b_i = \frac{\int_{x_{i-1}}^{x_i} \psi'(x) B_i'(x) dx}{\int_{x_{i-1}}^{x_i} (B_i'(x))^2 dx} = -\frac{\pi^2 \sqrt{6}}{12} \epsilon \sin\left(\frac{\pi}{2}(x_{i-1} + x_i)\right) h^2 + \mathcal{O}(\epsilon h^4).$$

□

We are now ready to provide a lower bound of $|G(u) - G(u_h^*)|$.

Theorem 4.5. *Let u be the solution of (26), and let u_h^* be the solution of (34) with $k = 2$. Then,*

$$|G(u) - G(u_h^*)| \geq \frac{0.75}{81} C_g h^{8/3} + \mathcal{O}(h^{11/3}).$$

where C_g is a fixed constant depending on the numerical integration rule.

Proof. We recall that $a^*(u, v) = a(u, v)$ for all $u, v \in S_h$. Therefore, from Lemma 3.1, we have

$$G(u) - G(u_h^*) = a(g - g_h, u - u_h) + f(g_h) - f^*(g_h),$$

where the bilinear form $a(\cdot, \cdot)$ is as defined in (26). Hence,

$$(43) \quad |G(u) - G(u_h^*)| \geq |f(g_h) - f^*(g_h)| - |a(g - g_h, u - u_h)|.$$

We first address the term $|f(g_h) - f^*(g_h)|$.

We recall that g_h is piecewise quadratic. Therefore, by the standard error formula for the Gaussian quadrature, there is $\xi_i \in (x_{i-1}, x_i)$, such that

$$\begin{aligned} f(g_h) - f^*(g_h) &= \sum_{i=1}^n \left(\int_{x_{i-1}}^{x_i} [f g_h] dx - \sum_{l=1}^2 \omega_{i,l} [f g_h](b_{i,l}) \right) = C_g h^5 \sum_{i=1}^n [f g_h]^{(4)}(\xi_i) \\ &= C_g h^5 \sum_{i=1}^n \{ [f^{(4)} g_h](\xi_i) + [f^{(3)} g_h'](\xi_i) + [f^{(2)} g_h^{(2)}](\xi_i) \} \\ (44) \quad &= C_g h^5 \sum_{i=1}^n \left[\frac{40}{81} (1 - \xi_i)^{-7/3} g_h(\xi_i) + \frac{10}{27} (1 - \xi_i)^{-4/3} g_h'(\xi_i) \right. \\ &\quad \left. + \frac{10}{9} (1 - \xi_i)^{-1/3} g_h''(\xi_i) \right], \end{aligned}$$

where $C_g > 0$ is a fixed constant depending on the numerical integration. We will now estimate each term in the above equation.

We first estimate the term

$$(45) \quad C_g h^5 \sum_{i=1}^n \frac{40}{81} (1 - \xi_i)^{-7/3} g_h(\xi_i).$$

in (44). Recall that $g_h(x_i) = g(x_i)$, and note that for the bubble function $B_i(x_{i-1}) = B_i(x_i) = 0$. Thus, from the definition of g_h in (40), we have

$$(46) \quad \begin{aligned} & (1 - \xi_i)^{-7/3} g_h(\xi_i) \\ &= (1 - \xi_i)^{-7/3} \left[\frac{g(x_i) - g(x_{i-1})}{h} (\xi_i - x_{i-1}) + g(x_{i-1}) + b_i B_i(\xi_i) \right]. \end{aligned}$$

Recall $g(x) = \epsilon \sin(\pi x) - 0.5x$. For the term $\frac{g(x_i) - g(x_{i-1})}{h} (\xi_i - x_{i-1})$, using the Taylor series, we have

$$\begin{aligned} g(x_i) - g(x_{i-1}) &= \epsilon \sin(\pi x_i) - \epsilon \sin(\pi x_{i-1}) - 0.5(x_i - x_{i-1}) \\ &= 2\epsilon \sin\left(\frac{\pi}{2}h\right) \cos\left(\frac{\pi}{2}(x_{i-1} + x_i)\right) - 0.5h \\ &= \left[\epsilon \pi \cos\left(\frac{\pi}{2}(x_{i-1} + x_i)\right) - 0.5 \right] h + \mathcal{O}(\epsilon h^3). \end{aligned}$$

Therefore,

$$(47) \quad \begin{aligned} & \left| \frac{g(x_i) - g(x_{i-1})}{h} (\xi_i - x_{i-1}) \right| \\ & \leq h \left| \epsilon \pi \cos\left(\frac{\pi}{2}(x_{i-1} + x_i)\right) - 0.5 + \mathcal{O}(\epsilon h^2) \right| \leq C_1 h. \end{aligned}$$

Meanwhile, by Lemma 4.4,

$$(48) \quad \begin{aligned} |b_i B_i(\xi_i)| &= \left| \frac{\pi^2 \sqrt{6}}{12} \epsilon \sin\left(\frac{\pi}{2}(x_{i-1} + x_i)\right) h^2 + \mathcal{O}(\epsilon h^4) \right| \sqrt{6} h^{-2} (\xi_i - x_{i-1}) (\xi_i - x_i) \\ &= \left| \frac{\pi^2}{2} \epsilon \sin\left(\frac{\pi}{2}(x_{i-1} + x_i)\right) + \mathcal{O}(\epsilon h^2) \right| |(\xi_i - x_{i-1})(\xi_i - x_i)| \\ &\leq \left| \frac{\pi^2}{2} \epsilon \sin\left(\frac{\pi}{2}(x_{i-1} + x_i)\right) + \mathcal{O}(\epsilon h^2) \right| h^2 \leq C_2 h^2. \end{aligned}$$

We will now find upper bounds for the term $g(x_{i-1})$ in (46) for $1 \leq i \leq \lceil \frac{7n}{8} \rceil - 1$ and $\lceil \frac{7n}{8} \rceil \leq i \leq n$, where $\lceil x \rceil$ denotes the largest integer that is less than or equal to x . We first note that $g(x) < 1$ for $0 \leq x \leq 1$ and for all $0 < \epsilon \leq 1$; consequently

$$(49) \quad g(x_{i-1}) < 1, \quad \text{for } 1 \leq i \leq \lceil \frac{7n}{8} \rceil - 1.$$

Furthermore, an easy calculation shows that for all $0 < \epsilon \leq 1$, $g(x) < -0.035$ for $0.87 \leq x \leq 1$. Since, for h small (in fact $h < 0.0025$), $x_{i-1} > 0.87$ for $i \geq \lceil \frac{7n}{8} \rceil$, it is clear that

$$(50) \quad g(x_{i-1}) < -0.035, \quad \text{for } \lceil \frac{7n}{8} \rceil \leq i \leq n.$$

We next note that the function $(1 - x)^{-7/3}$ is positive and increasing on $(0, 1)$. Therefore, recalling that $x_{i-1} \leq \xi_i \leq x_i$ and using the Riemann sum ideas, we easily see that for h small,

$$(51) \quad h^5 \sum_{i=1}^{\lceil \frac{7n}{8} \rceil - 1} (1 - \xi_i)^{-7/3} \leq h^4 \int_0^{7/8} (1 - x)^{-7/3} dx = \frac{3}{4} h^4 (8^{4/3} - 1),$$

and

$$(52) \quad h^5 \sum_{i=\lceil \frac{7n}{8} \rceil}^n (1 - \xi_i)^{-7/3} \geq h^4 \int_{7/8}^{1-h} (1 - x)^{-7/3} dx = \frac{3}{4} h^4 (h^{-4/3} - 8^{4/3}).$$

We now write the summation in (45) as

$$\begin{aligned}
& C_g h^5 \sum_{i=1}^n \frac{40}{81} (1 - \xi_i)^{-7/3} g_h(\xi_i) \\
&= \frac{40}{81} C_g h^5 \sum_{i=1}^n (1 - \xi_i)^{-7/3} \left[\frac{g(x_i) - g(x_{i-1})}{h} (\xi_i - x_{i-1}) + g(x_{i-1}) + b_i B_i(\xi_i) \right] \\
(53) \quad &= \frac{40}{81} C_g h^5 \sum_{i=[7n/8]}^n (1 - \xi_i)^{-7/3} \left[\frac{g(x_i) - g(x_{i-1})}{h} (\xi_i - x_{i-1}) + g(x_{i-1}) \right. \\
&\quad \left. + b_i B_i(\xi_i) \right] + \frac{40}{81} C_g h^5 \sum_{i=1}^{[7n/8]-1} (1 - \xi_i)^{-7/3} \left[\frac{g(x_i) - g(x_{i-1})}{h} (\xi_i - x_{i-1}) \right. \\
&\quad \left. + g(x_{i-1}) + b_i B_i(\xi_i) \right].
\end{aligned}$$

For $i \geq [7n/8]$ and h small, using (47), (48) and (50), we get

$$\frac{g(x_i) - g(x_{i-1})}{h} (\xi_i - x_{i-1}) + g(x_{i-1}) + b_i B_i(\xi_i) \leq -0.035 + C_1 h + C_2 h^2 \leq -0.025.$$

Similarly using (47), (48) and (49), we have for $1 \leq i \leq [7n/8] - 1$ and h small,

$$\frac{g(x_i) - g(x_{i-1})}{h} (\xi_i - x_{i-1}) + g(x_{i-1}) + b_i B_i(\xi_i) \leq 1 + C_1 h + C_2 h^2 \leq 2.$$

Therefore, from (53), we have

$$\begin{aligned}
(54) \quad C_g h^5 \sum_{i=1}^n \frac{40}{81} (1 - \xi_i)^{-7/3} g_h(\xi_i) &\leq \frac{40}{81} C_g h^5 \left[-0.025 \sum_{i=[7n/8]}^n (1 - \xi_i)^{-7/3} \right. \\
&\quad \left. + 2 \sum_{i=1}^{[7n/8]-1} (1 - \xi_i)^{-7/3} \right] \\
&\leq -\frac{0.75}{81} C_g h^4 (h^{-4/3} - 8^{4/3}) + \frac{60}{81} C_g h^4 (8^{4/3} - 1).
\end{aligned}$$

For the term

$$C_g h^5 \sum_{i=1}^n \frac{10}{27} (1 - \xi_i)^{-4/3} g'_h(\xi_i),$$

in (44), we follow a similar argument as above and get, for all $0 < \epsilon \leq 1$ and h small,

$$\begin{aligned}
& g'_h(\xi_i) < 3.45 \quad \text{for } i \leq \left[\frac{7n}{8} \right] - 1 \\
& \text{and } g'_h(\xi_i) < -0.5 \quad \text{for } i \geq \left[\frac{7n}{8} \right],
\end{aligned}$$

and consequently obtain, for h small,

$$\begin{aligned}
(55) \quad C_g h^5 \sum_{i=1}^n \frac{10}{27} (1 - \xi_i)^{-4/3} g'_h(\xi_i) &\leq -\frac{5}{9} C_g h^4 (h^{-1/3} - 8^{1/3}) \\
&\quad + \frac{35}{9} C_g h^4 (8^{1/3} - 1).
\end{aligned}$$

For the term

$$C_g h^5 \sum_{i=1}^n \frac{10}{9} (1 - \xi_i)^{-1/3} g_h''(\xi_i),$$

in (44), a similar and even simpler argument with $0 < \epsilon \leq 1$ and h small yields,

$$(56) \quad \begin{aligned} C_g h^5 \sum_{i=1}^n \frac{10}{9} (1 - \xi_i)^{-1/3} g_h''(\xi_i) \\ \leq -\frac{5}{6} C_g \epsilon h^5 (8^{-2/3} - h^{2/3}) + \frac{5}{3} C_g h^4 (1 - 8^{-2/3}). \end{aligned}$$

Hence, combining (44), (54), (55), and (56), we conclude that

$$\begin{aligned} f(g_h) - f^*(g_h) &= C_g h^5 \sum_{i=1}^n \left[\frac{40}{81} (1 - \xi_i)^{-7/3} g_h(\xi_i) + \frac{10}{27} (1 - \xi_i)^{-4/3} g_h'(\xi_i) \right. \\ &\quad \left. + \frac{10}{9} (1 - \xi_i)^{-1/3} g_h''(\xi_i) \right] \\ &\leq -\frac{0.75}{81} C_g h^{8/3} - \frac{5}{9} C_g h^{11/3} + \mathcal{O}(h^4). \end{aligned}$$

Consequently,

$$|f(g_h) - f^*(g_h)| \geq \frac{0.75}{81} C_g h^{8/3} + \mathcal{O}(h^{11/3}).$$

Now, since

$$\begin{aligned} |a(g - g_h, u - u_h)| &\leq C \|g - g_h\|_{H^1(\Omega)} \|u - u_h\|_{H^1(\Omega)} \\ &\leq C h^4 \|g\|_{H^3(\Omega)} \|u\|_{H^3(\Omega)}, \end{aligned}$$

we have

$$\begin{aligned} |G(u) - G(u_h^*)| &\geq \left| |f(g_h) - f^*(g_h)| - |a(g - g_h, u - u_h)| \right| \\ &\geq \frac{0.75}{81} C_g h^{8/3} + \mathcal{O}(h^{11/3}), \end{aligned}$$

which completes the proof of the desired result. \square

We recall the definition of $F_{u_h^*}$ in (35), and we assume that the definite integral $\int_0^1 f\psi dx$ is evaluated exactly. We approximate the quantity of interest $F(u)$ by $F_{u_h^*}$. A lower bound of the error $|F(u) - F_{u_h^*}|$ is easily obtained using (36) and Theorem 4.5, which we state below.

Theorem 4.6. *Let u be the solution of (25) with $f(x) = (1 - x)^{5/3}$, and u_h^* be the finite element solution ($k = 2$) obtained with the 2-point Gaussian quadrature as above. Suppose the integral $\int_0^1 f\psi dx$ in the definition of $F_{u_h^*}$ is computed exactly. Then, for the quantity of interest $F(u) = u'(0)$, we have*

$$|F(u) - F_{u_h^*}| = |G(u) - G(u_h^*)| \geq \frac{0.75}{81} C_g h^{8/3} + \mathcal{O}(h^{11/3}),$$

for h small.

Remark 4.7. If we choose $f(x) = (1 - x)^{2/3}$ in our example (25), then $f(x) \in H^1(\Omega)$ but $f(x) \notin H^2(\Omega)$. Using the FEM with $k = 1$ and 1-point or the 2-point Gaussian rule, we can again show that $|F(u) - F_{u_h^*}| \geq C h^{5/3}$. For the FEM with $k = 2$, we can also obtain the same lower bound $\mathcal{O}(h^{5/3})$ for $|F(u) - F_{u_h^*}|$ with the 2-point

Gaussian rule. We do not present the analysis here, but will illuminate this in the numerical examples of the next section.

Remark 4.8. We note that the use of the Gaussian quadrature with more integration points will result into a smaller value of C_g , and thus the computed value of $|G(u) - G(u_h^*)|$ will be smaller. More Gaussian points, however, will not increase the order of convergence as $h \rightarrow 0$. We show this in our examples in the next section.

We recall that $g(x) = \epsilon \sin(\pi x) - 0.5x \in H^3(\Omega)$. Moreover, since $f(x) = (1-x)^{5/3} \in H^2(\Omega)$, it is well known that $u \in H^4(\Omega)$. Therefore (33) holds for our particular problem, i.e., $|F(u) - F_{u_h}| \leq Ch^4$. Then, we can easily conclude from Theorem 4.6, that the ratio

$$(57) \quad R_h(u) := \frac{|F(u) - F_{u_h^*}|}{|F(u) - F_{u_h}|} = \frac{|G(u) - G(u_h^*)|}{|G(u) - G(u_h)|} \geq Ch^{-4/3} \rightarrow \infty, \quad \text{as } h \rightarrow 0.$$

Thus the ratio $R_h(u)$, which could be interpreted as a relative error in the computed quantity of interest due to numerical integration, becomes unbounded as $h \rightarrow 0$. In contrast, the ratio of $\|u - u_h^*\|_{H^1(\Omega)}$ to $\|u - u_h\|_{H^1(\Omega)}$ is bounded as $h \rightarrow 0$. We remark that a careful analysis with $\epsilon = 0$ in (29) shows (in fact, the analysis is simpler) that the results in Theorem 4.5 and Theorem 4.6 are still true (with different constants). However, from the definition of $g(x)$ in (30) with $\epsilon = 0$, we have

$$F(u) - F_{u_h} = G(u) - G(u_h) = a(g, u - u_h) = a(g - g_h, u - u_h) = 0,$$

since $g(x) = g_h(x)$ (see Remark 4.3). Therefore, we will not be able to draw a meaningful conclusion about the ratio $R_h(u)$ as $h \rightarrow 0$. This is one of the reasons we used $\epsilon \neq 0$ in (29).

5. NUMERICAL ILLUSTRATIONS

In this section, we will present numerical results that will illuminate the results obtained in the last section. We will show that the optimal order of convergence in the approximation of functionals, i.e., the quantities of interest, is not obtained when the data do not have the increased regularity as specified in Theorem 3.6. We will further show that over-integration may yield the optimal order of convergence, at least in the pre-asymptotic range, i.e., when the mesh parameter h is not too small.

We consider two boundary value problems, namely,

$$(58) \quad \begin{aligned} -u'' &= (1-x)^{5/3} \quad \text{in } \Omega, \\ u(0) &= 0, \quad u(1) + u'(1) = 0, \end{aligned}$$

and

$$(59) \quad \begin{aligned} -u'' &= (1-x)^{2/3} \quad \text{in } \Omega, \\ u(0) &= 0, \quad u(1) + u'(1) = 0, \end{aligned}$$

where $\Omega = (0, 1)$. We note that in (58), the data $f(x) = (1-x)^{5/3} \in H^2(\Omega)$, whereas in (59), the data $f(x) = (1-x)^{2/3} \in H^1(\Omega)$ but $\notin H^2(\Omega)$. The goal in both problems is to approximate the bounded linear functional $G(u) = \int_0^1 \psi' u' dx$, where $\psi(x) = 1 - x + \sin \pi x$. We recall that the functional $G(u)$ is related to the functional $F(u) = u'(0)$ as discussed in the last section.

n	$ G(u) - G(u_h) $	rate	$ G(u) - G(u_{h,1}^*) $	rate
10	1.833E-03		1.663E-03	
20	4.569E-04	2.004	4.122E-04	2.012
40	1.142E-04	2.001	1.027E-04	2.005
80	2.853E-05	2.000	2.565E-05	2.002
160	7.133E-06	2.000	6.407E-06	2.001
320	1.783E-06	2.000	1.513E-06	2.001
640	4.458E-07	2.000	4.002E-07	2.000

TABLE 1. u_h and $u_{h,1}^*$ are piecewise linear ($k = 1$) finite element solutions with exact integration and with 1-point Gaussian quadrature, respectively, approximating the solution of (58). The data of the problem has the increased regularity for $k = 1$.

In this section, we will denote the finite element solution with numerical integration by $u_{h,\ell}^*$, where the integrals in the load vector and in the stiffness matrix are approximated by the ℓ -point Gaussian integration rule. We will also approximate the order of convergence of $u_{h,\ell}^*$, for a fixed ℓ , using the expression

$$\text{rate} = \frac{\ln(|e_n/e_{2n}|)}{\ln(2)},$$

where e_n is the error in the computed quantity for $h = 1/n$, i.e., $e_n = |G(u) - G(u_{h,\ell}^*)|$.

We first consider the problem (58). We computed the piecewise linear (i.e., $k = 1$) finite element solutions u_h and $u_{h,1}^*$ to approximate the solution u of (58); we recall that u_h is the finite element solution (piecewise linear in this case) with exact integration. We note that the algebraic precision of 1-point Gauss rule is 1, as required in Theorem 3.6 for $k = 1$. Also $f(x)$ has the increased regularity for the case $k = 1$, i.e., $f \in H^2(\Omega)$. We have presented the values of $|G(u) - G(u_h)|$ and $|G(u) - G(u_{h,1}^*)|$ for several values of h in Table 1, where $n = 1/h$; As expected, we observe the optimal order of convergence $\mathcal{O}(h^2)$ for $|G(u) - G(u_h)|$ and $|G(u) - G(u_{h,1}^*)|$, which illuminates the results in Lemma 2.2 and Theorem 3.6 respectively.

We next computed the piecewise quadratic (i.e., $k = 2$) finite element solutions u_h , $u_{h,2}^*$ and $u_{h,6}^*$ to approximate the solution u of the same problem (58). We note that the algebraic precision of the 2-point Gauss rule is 3, as required in Theorem 3.6 for $k = 2$. But the data $f \notin H^4(\Omega)$, and thus does not have the increased regularity required in Theorem 3.6 for the case $k = 2$. We mention that $u_{h,6}^*$ is the finite element solution with “over-integration”. In Table 2, we presented the computed values of $|G(u) - G(u_h)|$, $|G(u) - G(u_{h,2}^*)|$ and $|G(u) - G(u_{h,6}^*)|$.

1. We observe that the values $|G(u) - G(u_h)|$ are converging with order $\mathcal{O}(h^4)$, which is optimal. But clearly, the values $|G(u) - G(u_{h,2}^*)|$ are not converging with the optimal order. In fact, the order of convergence is monotonically decreasing with increasing values of n . It appears that the order of convergence is getting closer to $\mathcal{O}(h^{8/3})$, which is the order of the lower bound of the error given in Theorem 4.5.

n	$ G(u) - G(u_h) $	rate	$ G(u) - G(u_{h,2}^*) $	rate	$ G(u) - G(u_{h,6}^*) $	rate
10	1.293E-06		4.264E-06		1.297E-06	
20	8.124E-08	3.992	3.840E-07	3.473	8.188E-08	3.985
40	5.086E-09	3.998	4.110E-08	3.224	5.186E-09	3.981
80	3.180E-10	3.999	5.154E-09	2.995	3.337E-10	3.958
160	1.987E-11	4.000	7.209E-10	2.838	2.234E-11	3.901
320	1.194E-12	4.000	1.072E-10	2.750	1.581E-12	3.820

TABLE 2. u_h , $u_{h,2}^*$, and $u_{h,6}^*$ are piecewise quadratic ($k = 2$) finite element solutions with exact integration, with 2-point Gaussian quadrature, and with 6-point Gaussian quadrature (over-integration), respectively, approximating the solution of the problem (58). The data of the problem does not have the increased regularity for $k = 2$.

2. We observe that the order of convergence of $|G(u) - G(u_{h,6}^*)|$ is approximately $\mathcal{O}(h^4)$ initially for smaller values of n , but the convergence rate is slowing down for larger values of n , i.e., for small h .
3. We observe that the values of $|G(u) - G(u_{h,6}^*)|$ are smaller than the values of $|G(u) - G(u_{h,2}^*)|$, as indicated in Remark 4.8. Thus, over-integration yields smaller absolute values of the error, but does not affect the asymptotic convergence rate.

We remark that for $k = 2$, the data $f(x)$ of the problem (58) has adequate regularity for the optimal order of convergence of the finite element solution in the energy norm (see Theorem 4.1.6 in [9]) and both $\|u - u_{h,2}^*\|_{H^1(\Omega)}$ and $\|u - u_{h,6}^*\|_{H^1(\Omega)}$ yield the optimal order of convergence $\mathcal{O}(h^4)$. We have not presented numerical results to illuminate this fact in this paper.

We now consider the problem (59). First, we computed the piecewise linear (i.e., $k = 1$) finite element solutions u_h , $u_{h,1}^*$, and $u_{h,2}^*$ to approximate the solution u of the problem (59). Note that $u_{h,2}^*$ is the finite element solution with over-integration (only one extra Gauss point). We have summarized the results in Table 3. We then computed piecewise quadratic (i.e., $k = 2$) finite element solutions u_h , $u_{h,2}^*$, and $u_{h,50}^*$ to approximate the solution u of the problem (59); $u_{h,50}^*$ is the finite element solution with over-integration (lots of extra Gauss points). We have summarized the results for $k = 2$ in Table 4. We note that the Gauss rules employed to compute $u_{h,1}^*$ in Table 3 and $u_{h,2}^*$ in Table 4 have the algebraic precision, as required in Theorem 3.6. But the data $f(x)$ of problem (59) does not have the increased regularity required in Theorem 3.6, either for $k = 1$ or for $k = 2$.

1. We observe from both Table 3 and Table 4 that the values of $|G(u) - G(u_h)|$ converge with the optimal order $\mathcal{O}(h^{2k})$, $k = 1, 2$, as expected in Lemma 2.2, since $u \in H^3(\Omega)$.
2. It is clear from Table 3 that the rate of convergence of $|G(u) - G(u_{h,1}^*)|$ is decreasing with increasing values of n . It is important to recall that the lower bound of the error is $\mathcal{O}(h^{5/3})$, as mentioned in Remark 4.7. The rate of convergence of $|G(u) - G(u_{h,2}^*)|$ in Table 3 is close to the optimal rate of convergence, $\mathcal{O}(h^2)$, for smaller values of n , but the rate slows down monotonically with increasing values of n . This indicates over-integration does not yield the asymptotic rate of convergence. The second and the third

n	$ G(u) - G(u_h) $	rate	$ G(u) - G(u_{h,1}^*) $	rate	$ G(u) - G(u_{h,2}^*) $	rate
10	3.220E-03		4.270E-03		3.274E-03	
20	8.047E-04	2.001	1.114E-03	1.938	8.203E-04	1.997
40	2.011E-04	2.000	2.924E-04	1.930	2.059E-04	1.994
80	5.028E-05	2.000	7.725E-05	1.920	5.175E-05	1.992
160	1.257E-05	2.000	2.060E-05	1.907	1.303E-05	1.990
320	3.413E-06	2.000	5.549E-06	1.892	3.286E-06	1.987
640	7.856E-07	2.000	1.513E-06	1.875	8.306E-07	1.984

TABLE 3. u_h , $u_{h,1}^*$, and $u_{h,2}^*$ are piecewise linear ($k = 1$) finite element solutions with exact integration, with 1-point Gaussian quadrature, and with 2-point Gaussian quadrature (over-integration), respectively, approximating the solution of the problem (59). The data of the problem does not have the increased regularity for $k = 1$.

n	$ G(u) - G(u_h) $	rate	$ G(u) - G(u_{h,2}^*) $	rate	$ G(u) - G(u_{h,50}^*) $	rate
10	8.647E-07		5.492E-05		8.665E-07	
20	5.695E-08	3.925	1.575E-05	1.802	5.751E-08	3.913
40	3.670E-09	3.956	4.750E-06	1.730	3.847E-09	3.902
80	2.337E-10	3.973	1.465E-06	1.697	2.895E-10	3.732
160	1.478E-11	3.982	4.566E-07	1.682	3.237E-11	3.161

TABLE 4. u_h , $u_{h,2}^*$, and $u_{h,50}^*$ are piecewise quadratic ($k = 2$) finite element solutions with exact integration, with 2-point Gaussian quadrature, and with 50-point Gaussian quadrature (over-integration), respectively, approximating the solution of the problem (59). The data of the problem does not have the increased regularity for $k = 2$.

columns also show that over-integration reduces the absolute value of the error.

3. The second column of Table 4 shows that the rate of convergence of $|G(u) - G(u_{h,2}^*)|$ decreases with increasing values of n and indicates that the order of convergence is approaching $\mathcal{O}(h^{5/3})$, the lower bound of the error as mentioned in Remark 4.7. Also the third column of Table 4 clearly indicates that over-integration does not yield optimal order of convergence, but certainly, over-integration reduces the absolute value of the error.

These experiments indicate that the increased regularity assumptions to obtain the optimal order of convergence in Theorem 3.6 is, in general, necessary. Also, over-integration reduces the absolute values of the error, and may even show the optimal order of convergence in the pre-asymptotic range. But over-integration does not yield the asymptotic optimal rate of convergence.

6. CONCLUSIONS

In this paper, we have presented sufficient conditions on the data of the problem and on the numerical integration rule employed in the finite element method to obtain the optimal order of convergence of the error in the computed quantity of interest. We have also shown that the sufficient conditions on the data of the

problem are, in general, also necessary. Our results clearly show that a finite element solution with numerical integration may converge optimally in the energy norm, but a quantity of interest, computed using the same finite element solution, may not yield optimal order of convergence. Moreover, our computational results indicate that over-integration may be necessary to obtain an accurate value of the computed quantity of interest.

REFERENCES

- [1] I. Babuška and A. Miller. The post-processing approach in the finite element method. Part I: Calculation of displacements, stresses and higher order derivatives of displacement. *Int. J. Numer. Meth. Engng*, 20:1085–1109, 1984.
- [2] I. Babuška and A. Miller. The post-processing approach in the finite element method. Part II: Calculation of stress intensity factors. *Int. J. Numer. Meth. Engng*, 20:1111–1129, 1984.
- [3] I. Babuška and J. Osborn. Eigenvalue problems. In *Handbook of numerical analysis, Vol. II*, pages 641–787. North-Holland, Amsterdam, 1991.
- [4] I. Babuška and T. Strouboulis. *The finite element method and its reliability*. Numerical Mathematics and Scientific Computation. Oxford University Press, New York, 2001.
- [5] I. Babuška and A.K. Aziz. Survey lectures on the mathematical foundation of the finite element method. In A.K. Aziz, editor, *The mathematical foundations of the finite element method with applications to partial differential equations*, pages 3–359, New York, 1972. Academic Press.
- [6] U. Banerjee and J. E. Osborn. Estimation of the effect of numerical integration in finite element eigenvalue approximation. *Numer. Math.*, 56:735–762, 1990.
- [7] U. Banerjee and M. Suri. The effect of numerical quadrature in the p -version of the finite element method. *Math. Comp.*, 59:1–20, 1992.
- [8] S.C. Brenner and L.R. Scott. *The Mathematical Theory of Finite Element Methods*. Texts in Applied Mathematics. Springer, New York, 2002.
- [9] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*, volume 4 of *Studies in Mathematics and Its Applications*. North-Holland, Amsterdam, 1978.
- [10] P.G. Ciarlet and P.-A. Raviart. The combined effect of curved boundaries and numerical integration in isoparametric finite element methods. In A.K. Aziz, editor, *The mathematical foundations of the finite element method with applications to partial differential equations*, pages 404–474, New York, 1972. Academic Press.
- [11] G.J. Fix. Effects of quadrature errors in finite element approximation of steady state, eigenvalue and parabolic problems. In A.K. Aziz, editor, *The mathematical foundations of the finite element method with applications to partial differential equations*, pages 525–556, New York, 1972. Academic Press.
- [12] R.J. Herbold, M.H. Schultz, and R.S. Varga. The effect of quadrature errors in the numerical solution of boundary value problems by variational techniques. *Aequationes Math.*, 3:247–270, 1969.
- [13] R.J. Herbold and R.S. Varga. The effect of quadrature errors in the numerical solution of two-dimensional boundary value problems by variational techniques. *Aequationes Math.*, 7:36–58, 1972.
- [14] A. H. Schatz and L. B. Wahlbin. Interior maximum-norm estimates for finite element methods, Part II. *Math. Comp.*, 64(221):907–928, 1995.
- [15] B Szabó and I. Babuška. *Finite Element Analysis*. John Wiley & Sons, INC, New York, 1991.
- [16] L. B. Wahlbin. *Superconvergence in the Galerkin Finite Element Methods*, volume 1605 of *Lecture Notes in Mathematics*. Springer, New York, 1991.

IVO BABUŠKA, INSTITUTE FOR COMPUTATIONAL ENGINEERING AND SCIENCES, UNIVERSITY OF TEXAS AT AUSTIN, ACE 6.412, AUSTIN, TX 78712, USA

UDAY BANERJEE, DEPARTMENT OF MATHEMATICS, SYRACUSE UNIVERSITY, SYRACUSE, NY 13244, USA

HENGGUANG LI, DEPARTMENT OF MATHEMATICS, SYRACUSE UNIVERSITY, SYRACUSE, NY 13244, USA