Applied Mathematics, 2016, 7, 1359-1407
Published Online July 2016 in SciRes. http://www.scirp.org/journal/am
http://dx.doi.org/10.4236/am.2016.712120

Scientific Research Publishing

Error Estimations, Error Computations, and Convergence Rates in FEM for BVPs

Karan S. Surana[1], A. D. Joy[1], J. N. Reddy[2]

[1]Department of Mechanical Engineering, Univeristy of Kansas, Lawrence, KS, USA
[2]Department of Mechanical Engineering, Texas A&M University, College Station, TX, USA
Email: kssurana@ku.edu

Received 3 June 2016; accepted 26 July 2016; published 29 July 2016

Abstract

This paper presents derivation of a priori error estimates and convergence rates of finite element processes for boundary value problems (BVPs) described by self adjoint, non-self adjoint, and nonlinear differential operators. A posteriori error estimates are discussed in context with local approximations in higher order scalar product spaces. A posteriori error computational framework (without the knowledge of theoretical solution) is presented for all BVPs regardless of the method of approximation employed in constructing the integral form. This enables computations of local errors as well as the global errors in the computed finite element solutions. The two most significant and essential aspects of the research presented in this paper that enable all of the features described above are: 1) ensuring variational consistency of the integral form(s) resulting from the methods of approximation for self adjoint, non-self adjoint, and nonlinear differential operators and 2) choosing local approximations for the elements of a discretization in a subspace of a higher order scalar product space that is minimally conforming, hence ensuring desired global differentiability of the approximations over the discretizations. It is shown that when the theoretical solution of a BVP is analytic, the a priori error estimate (in the asymptotic range, discussed in a later section of the paper) is independent of the method of approximation or the nature of the differential operator provided the resulting integral form is variationally consistent. Thus, the finite element processes utilizing integral forms based on different methods of approximation but resulting in VC integral forms result in the same a priori error estimate and convergence rate. It is shown that a variationally consistent (VC) integral form has best approximation property in some norm, conversely an integral form with best approximation property in some norm is variationally consistent. That is best approximation property of the integral form and the VC of the integral form is equivalent, one cannot exist without the other, hence can be used interchangeably. Dimensional model problems consisting of diffusion equation, convection-diffusion equation, and Burgers equation described by self adjoint, non-self adjoint, and nonlinear differential operators are considered to present extensive numerical studies using Galerkin method with weak form (GM/WF) and least squares process (LSP) to determine computed convergence rates of various

How to cite this paper: Surana, K.S., Joy, A.D. and Reddy, J.N. (2016) Error Estimations, Error Computations, and Convergence Rates in FEM for BVPs. *Applied Mathematics*, **7**, 1359-1407. http://dx.doi.org/10.4236/am.2016.712120

error norms and present comparisons with the theoretical convergence rates.

Keywords

Finite Element, Error Estimation, Convergence Rate, A Priori, A Posteriori, BVP, Variationally Consistent Integral Form, Variationally Inconsistent Integral Form, Differential Operator Classification, Self-Adjoint, Non-Self-Adjoint, Nonlinear

1. Introduction

It is now well recognized that in finite element computations there are three independent parameters: characteristic length of the discretization h, degree of approximation p, and the order k of the scalar product space. h and p have been well known for quite some time but introduction of k as an additional independent parameter in finite element computations is rather recent. Surana *et al*. [1]-[4] have shown the order k of the approximation space to be an independent parameter in all finite element computational processes in addition to h and p, hence k-version of finite element method in addition to h- and p-versions. The order k of the approximation space ensures global differentiability of order $k-1$ over the whole discretization. The appropriate choice of k is essential in ensuring that 1) the desired physics is preserved in the computational process and 2) the integrals are Riemann in the entire finite element process so that the equivalence of BVP with the integral form is preserved and the errors in the calculated solution can be computed correctly without knowledge of the theoretical solution. We elaborate more on some of these aspects in the following.

If the differential operator contains highest order derivatives of the dependent variables of orders $2m$, then the approximation of the solutions of the BVP must at least be of class C^{2m} *i.e*. of global differentiability of order $2m$ in order for this approximation to be admissible in the BVP in the pointwise sense. This requires that the order k of the approximation space must at least be $2m+1$ *i.e*. $k=2m+1$ is minimally conforming order of the approximation space. Clearly, the order k of the minimally conforming space is determined by the highest order of the derivatives of the dependent variable(s) in the BVP. When $k \geq 2m+1$, all integrals over the discretization $\bar{\Omega}^T$ remain Riemann. When $k=2m$, the integrals over $\bar{\Omega}^T$ are in Lebesgue sense and the corresponding approximation ϕ_h of the solution ϕ over $\bar{\Omega}^T$ is not admissible in the BVP $A\phi - f = 0$ in the pointwise sense. When $k \leq 2m-1$, the approximation ϕ_h of ϕ over $\bar{\Omega}^T$ is not admissible at all in the BVP. Choosing $k > 2m+1$ may be beneficial if the theoretical solution ϕ of the BVP is of higher order global differentiability than $2m$ as this choice incorporates higher order global differentiability aspects of ϕ in the computational process. Thus, now we have h-, p-, k-versions of the finite element processes and associated convergences and convergence rates.

The subject of a priori error estimation and a posteriori error estimation have been exhaustively studied and investigated with the objective that 1) perhaps a priori error estimates will help us in deciding the most prudent choices of h, p, and k so that the errors in the desired norms are reduced at the fastest rate during computations, 2) the a posteriori error estimates will guide us based on the current finite element solution in improving the accuracy of the subsequently computed solutions in the most prudent manner. The published literature on this subject is enormous and discussion of each writing on the subject in this paper is not feasible and is also of little benefit. Interested readers can refer to some selected publications [5]-[34] included here.

In the work presented in this paper, our objectives are:

a) To derive a priori error estimates for BVPs described by self adjoint, non-self adjoint, and nonlinear differential orperators when the theoretical solutions are analytic, thus establishing precise dependence of the chosen error norm on h, p, k, and the smoothness of the theoretical solution (for simplicity this is done using one dimensional BVPs).

b) To discuss the currently used a posteriori error estimation techniques, their shortcomings, and serious inadequacies when actual physics of the BVP is incorporated in the finite element computational process.

c) To demonstrate the need for a posteriori error computation and present a framework in which those computations can be performed without the knowledge of theoretical solutions.

d) To establish that higher order approximation spaces and variationally consistent integral forms are essential

for incorporating the desired physics of the BVP in the computational process and to ensure that the resulting finite element computational processes are unconditionally stable so that error estimations remain meaningful.

e) To perform numerical studies using one dimensional boundary value problem described by self adjoint, non-self adjoint, and nonlinear differential operators and to demonstrate exceptionally good agreement of the computed convergence rates with those established theoretically.

f) To establish that the a priori error estimates derived in (a) also hold for 2D and 3D BVPs when the integral forms in those BVPs are variationally consistent.

2. Preliminaries: Convergence and Convergence Rates, Convergence Behavior of Computations, Error Estimation, and Error Computations

In this section, we present some preliminary material and concepts that are essential in error estimation and error computations. Many of these are well known but are included in the following for completeness and for the sake of coherent continuation to the new work in this paper.

2.1. Convergence and Convergence Rate

Convergence of a finite element solution implies behavior of the error in the finite element solution (measured in some norm) as a function of the degrees of freedom or the characteristic length of the discretization. When the theoretical solution is known, the error in the finite element solution in some norm (L_2-norm, H^1-norm, etc.) can be computed and therefore we can study its behavior as a function of the degrees of freedom. When the theoretical solution is not known, perhaps estimating the error in some norm in the computed solution is a viable option. However, we shall see in a later section that this option only works in a restricted range of the behavior of error norm versus dofs. The third option is that if we are using minimally conforming spaces $V_h \subset H$ then residual functional $I(\phi_h)$ can be computed precisely as for minimally conforming spaces all integrals over the discretization $\bar{\Omega}^T$ of $\bar{\Omega}$, the domain of definition of the BVP, are Riemann. Proximity of $I(\phi_h)$ to zero is a measure of error due to the fact that when $\phi_h \simeq \phi$, $I(\phi_h) \simeq I(\phi) = 0$. Thus, $I(\phi_h)$ is in fact error measure in the solution ϕ_h over $\bar{\Omega}^T$. This option can always be used for any applications as it does not require theoretical solution but necessitates the approximation ϕ_h to be in a space of order $k \geq 2m+1$. In what follows we can use $I(\phi_h)$ as a measure of error over $\bar{\Omega}^T$, hence convergence of the computed solution ϕ_h to ϕ implies studying $I(\phi_h)$ versus dofs as more degrees of freedom are added to the discretization. When $I(\phi_h) \leq \Delta$, a predetermined tolerance of computed zero, we consider the finite element solution ϕ_h to be converged to the theoretical solution ϕ. We consider $\sqrt{I(\phi_h)}$ versus dofs or $\sqrt{I} = \sqrt{(E,E)} = \|E\|_{L_2}$, L_2-norm of residual E. We study $\sqrt{I}$ versus dofs using log-log scale, or more precisely we study $\log\|E\|_{L_2}$ versus log(*dofs*). $\log\|E\|_{L_2}$ and log(dofs) or log-log scale are necessary as the range of I could be $O(10^{1})$ - $O(10^{-20})$ and the range of dof could be $O(10^{1})$ - $O(10^{6})$ or higher.

2.2. Convergence Behavior of Computations

The material presented in this section is based on $\|E\|_{L_2}$ versus dof behavior, but the same concepts hold true for any other measure of error norm (*i.e.* $\|E\|_{L_2}$ can be replaced with any other error norm without affecting the basic behavior of the convergence graph). A typical convergence behavior of $\log\left(\sqrt{I}\right)$ or $\log\left(\|E\|_{L_2}\right)$ versus log(*dof*) is shown in **Figure 1**. This graph is generated using 1D convection-diffusion equation (a second order ODE) with $Pe = 1000$ and least squares finite element formulation based on residual functional. The progressively graded discretizations are generated beginning with two elements using a constant geometric ratio of 1.5. The smallest element is located at $x = 1.0$. $k = 3$ is used as it corresponds to the minimally conforming space. Minimum *p*-level of 5 (needed for $k = 3$) is considered for each progressively refined discretization. From **Figure 1**, we observe five distinct zones. In each one of these zones the behavior of $\sqrt{I}$ versus dofs is unique and distinct. The behavior of $\log\left(\sqrt{I}\right)$ versus log(dofs) shown in **Figure 1** illustrates the varying rate of convergence of the finite element solution (the slope of the curve) with varying dofs. In the middle portion represented by almost a straight line behavior the slope is almost constant, indicating constant convergence rate. We discuss the details related to the varying slope of the curve, associated rate of convergence, and its significance in the following.

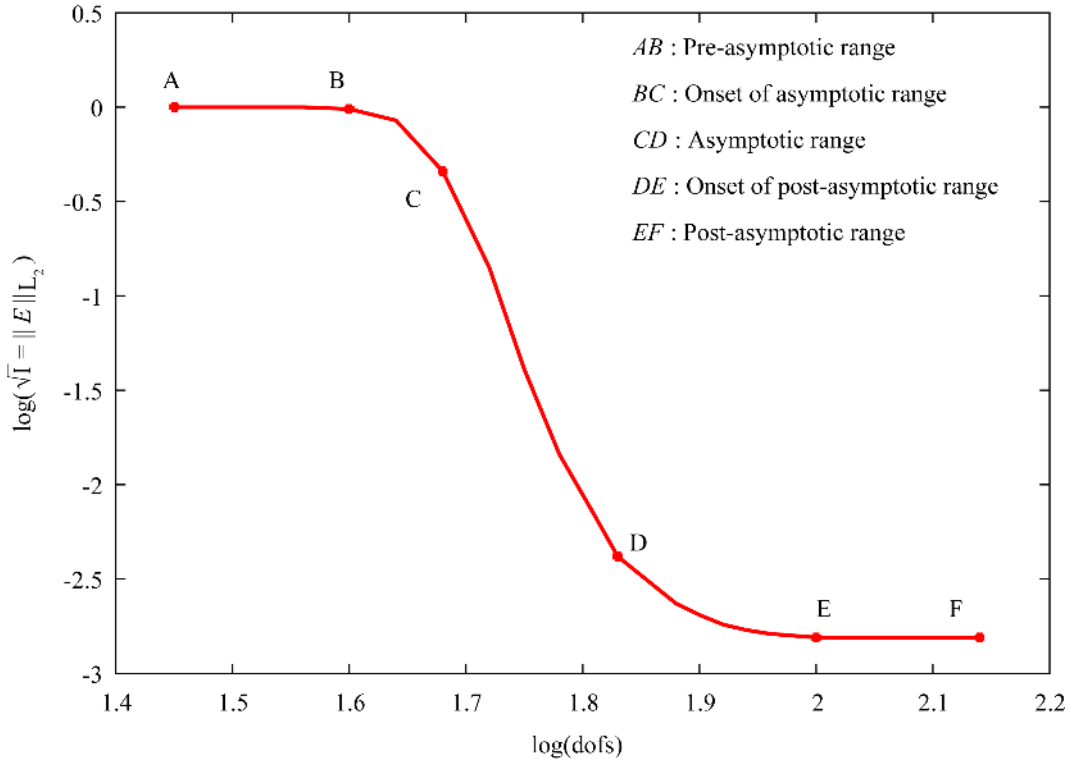

Figure 1. Typical convergence behavior of a finite element solution.

Pre-asymptotic range (***AB***): The range *AB* is called *pre-asymptotic range*. In this range as we move from location A toward location B additional degrees of freedom are added to the discretization but there is virtually no measurable reduction in the L_2-norm of E. The accuracy of the computed solution in this range is very poor (due to $\|E\|_{L_2}$ of the $O(1)$). Due to poor accuracy of the solution ϕ_h, hence $\sqrt{I}$, the I^e values for the elements are poor as well, hence these cannot be used to guide any form of adaptive refinement process. A posteriori error estimations in this range are not possible either as these require some regularity in the computed solution which is absent in ϕ_h in range *AB*. Thus, in this range adaptive processes are not possible as reliable indicators (either estimated or computed) based on ϕ_h are not possible.

Onset of asymptotic range (***BC***): The range *BC* is called *onset of asymptotic range*. In this range addition of degrees of freedom to the discretization results in measurable reduction in $\|E\|_{L_2}$ reflecting progressive improvement in accuracy of the computed solution ϕ_h from *B* to *C*. In this range I^e values or any other possible element error indicators are more accurate than range *AB*. In this range adaptive processes in *h*, *p*, or *hp* can be utilized keeping in mind that as we move closer to *C*, the values of I^e (or other indicators) for the elements of the discretization become more accurate, hence can be more effective in the adaptive process.

Asymptotic range (***CD***): In this range as more dofs are added to the discretization the improvement (reduction) in $\|E\|_{L_2}$ is most significant. This range on log-log scale is nearly linear, hence constant slope. Adaptive refinements in this range are most effective in reducing $\|E\|_{L_2}$. We observe that between *C* and *D* there are several orders of magnitude reduction in the value of $\|E\|_{L_2}$. Slope of the error norm versus dof graph in this range is called the *asymptotic convergence rate* of the finite element solution.

Onset of post-asymptotic range (***DE***): This range is almost reverse of the onset of asymptotic range. In this range reduction in $\|E\|_{L_2}$ progressively diminishes with the addition of degrees of freedom to the discretization indicating that substantial achievable reduction in $\|E\|_{L_2}$ has taken place up to point *D*. Computations in this range result in waste of significant resources (dofs) with very little gain in the objective of reducing $\|E\|_{L_2}$.

Post-asymptotic range (***EF***): In this range in spite of the addition of dofs to the discretization no measurable reduction is observed in $\|E\|_{L_2}$. This is generally due to the fact that within the accuracy of the computations (*i.e.* the word size on the computer we have reached a limit), hence the accuracy remains limited to the same number of decimal places in $\|E\|_{L_2}$ regardless of the increase in dofs.

2.3. Convergence Rates

In an abstract sense, the convergence rate of a finite element computational process is the rate at which the computed solution ϕ_h is approaching the theoretical solution ϕ as more degrees of freedom are added to the discretization through refining h or increasing p or changing k. That is it is the rate at which the error norm is approaching zero as more degrees of freedom are added. Thus, a measure of convergence rate of the finite element solution could be the slope of $\sqrt{I}$ (or $\|E\|_{L_2}$) versus dof behavior. Since dofs can be added through h, p, and k, the convergence rate of a finite element solution can be a function of h, p, k, and the smoothness of the theoretical solution at this stage of the discussion.

In range AB, the slope is almost zero. From B to C the slope increases as more dofs are added to the discretization thereby progressively increasing convergence rate from B to C. From C to D, the asymptotic range, the slope of $\|E\|_{L_2}$ versus dofs is almost constant and the reduction in $\|E\|_{L_2}$ is most significant as more dofs are added. Thus, in the asymptotic range the convergence rate is the highest (due to highest slope of $\log\left(\|E\|_{L_2}\right)$ versus $\log(dof)$) and is constant. In the onset of post-asymptotic range DE the convergence rate decreases and eventually becomes almost zero in the post-asymptotic range EF.

Remarks

I) Behavior of $\|E\|_{L_2}$ versus dofs shown in **Figure 1** is typical of other error norms as well, hence the discussion and conclusions related to **Figure 1** are applicable in the convergence behavior study using any other desired error norm.

II) Pre-asymptotic range AB, onset of post-asymptotic range DE, and post-asymptotic range EF should be avoided as in these ranges solution accuracy improvement is poor.

III) In range AB I^e values (or other measures) are not accurate enough to guide an adaptive process of any kind.

IV) Adaptive processes (h, p, k) can be initiated in the range BC as I^e values in this range are reasonable measure of error. Adaptive processes become more and more effective when we initiate them as we approach from B to C. In the range BC the slope of $\|E\|_{L_2}$ versus dof increases from B to C indicating improving convergence rate and eventually achieves the highest convergence rate value at C which remains almost constant in the asymptotic range CD.

V) A priori and a posteriori error estimates are only valid in the asymptotic range due to the fact it is only in this range that computed ϕ_h has desired regularity and the convergence rate is the highest, hence worth estimating a priori. The error estimates (a priori and a posteriori) can neither be derived accurately nor can be used meaningfully in regions other than BC.

2.4. Error Estimation and Error Computation

There are two types of error estimations generally considered: a priori error estimation and a posteriori error estimation. A priori error estimation refers to establishing dependence of some error norm on h, p, k, and the regularity of the theoretical solution before the computations are performed so that we have knowledge of the precise nature of the functional dependence of error norm on h, p, k, and the regularity of the theoretical solution. A posteriori error estimation refers to error estimates derived using a computed solution with specific choices of h, p, and k. The sole purpose of a posteriori error estimation is to use current finite element solution to derive element indicators that can perhaps be used to guide an adaptive process. Both of the error estimations require some regularity of the computed solution which only exists in the asymptotic range (range CD, **Figure 1**). This is a very significant restriction on the use of these estimates. For example, a priori error estimate cannot be used to predict convergence rate in the ranges AB, BC, DE, and EF as this is specifically derived using the regularity of ϕ_h that only exists in the asymptotic range. Likewise a posteriori estimate cannot be used for adaptivity in any ranges except CD.

Another point to note is that a posteriori error estimates are generally derived such that they quantify the weakness (es) in the finite element global approximation $\phi_h = \bigcup_e \phi_h^e$. Their derivations are largely based on C^0 local approximations which result in interelement discontinuity of the derivatives normal to the interelement boundaries. This may be quantified by establishing bounds that can be used for adaptivity. However if we use ϕ_h^e of class C^1 thereby ϕ_h of class C^1, then such bounds are meaningless. In k-version of finite element methods enabling higher order global differentiability approximations, majority of the a posteriori error esti-

mates based on interelement discontinuity of the derivatives are not meaningful. With the use of higher order approximations ϕ_h, the integrals can be maintained Riemann, $\|E\|_{L_2}$ and $\|E^e\|_{L_2}$ are true measures of the error in the finite element solution for $\bar{\Omega}^T$ and $\bar{\Omega}^e$ and can indeed be used in adaptive processes. These aspects are discussed in more details in later sections.

3. Variationally Consistent (VC) and Variationally Inconsistent (VIC) Integral Forms

The differential operators appearing in the totality of all BVPs can be mathematically classified in three categories: self-adjoint, non-self-adjoint, and nonlinear differential operators. The finite element processes for these operators can be derived by constructing integral forms using methods of approximation such as: Galerkin method (GM), Petrov-Galerkin method (PGM), weighted residual method (WRM), Galerkin method with weak form (GM/WF), and least squares method or process (LSP). The unconditional stability of the resulting computational process or lack thereof can be established by making a correspondence of these integral forms to the elements of the calculus of variations [1]-[4] [35]. The integral forms that result in unconditionally stable computational processes are termed variationally consistent (VC). The others are called variationally inconsistent (VIC). In VC integral forms the assembled coefficient matrices always remain positive-definite regardless of the admissible choices of *h*, *p*, and *k* whereas in VIC integral forms this can not always be ensured.

Definition 3.1 (consistent (VC) integral form of a BVP) *A variationally consistent integral form corresponding to the BVP* $A\phi - f = 0$ *consists of*

1) Existence of a functional $I(\phi)$ corresponding to the BVP $A\phi - f = 0$. This is generally by construction (or is assumed).

2) Necessary condition for the existence of an extremum of $I(\phi)$ is given by $\delta I(\phi) = 0$. The integral form $\delta I(\phi) = 0$ is used to determine ϕ. The Euler's equation resulting from $\delta I(\phi) = 0$ must be the BVP $A\phi - f = 0$.

3) $\delta^2 I > 0$, $= 0$, < 0 (minimum, saddle point, maximum of $I(\phi)$) is the sufficient condition or extremum principle. Extremum principle ensures that a ϕ obtained from $\delta I(\phi) = 0$ is unique. Extremum principle also establishes whether ϕ from $\delta I(\phi) = 0$ minimizes or maximizes $I(\phi)$ or yields a saddle point of $I(\phi)$.

When all these three elements are present in an integral formulation of the BVP $A\phi - f = 0$, then the integral form (resulting from $\delta I(\phi) = 0$ or otherwise) is called a variationally consistent integral form of the BVP $A\phi - f = 0$ (or simply VC integral process). VC integral form or process yields unique extremum of the functional $I(\phi)$ corresponding to $A\phi - f = 0$, hence a unique solution of the BVP $A\phi - f = 0$ (the Euler's equation resulting from $\delta I(\phi) = 0$).

Definition 3.2 (inconsistent integral form (VIC) of a BVP) If an integral form of a BVP (resulting from $\delta I(\phi) = 0$ or otherwise) is not variationally consistent, then it is variationally inconsistent. A variationally inconsistent integral form or process violates one or more of the three requirements needed for variational consistency of the integral form.

Remarks

1) Thus, we see that a variationally consistent integral form of a BVP $A\phi - f = 0$ emerges as a method of obtaining a unique solution of the BVP $A\phi - f = 0$.

2) The necessary condition (the integral form resulting from $\delta I(\phi) = 0$ or otherwise) provides a system of algebraic equations from which the solution ϕ is determined.

3) The sufficient condition or unique extremum principle ensures that a ϕ obtained from the integral form ($\delta I(\phi) = 0$ or otherwise) is unique, hence this ϕ yields a unique extremum of $I(\phi)$ as well as a unique solution of the Euler's equation which is the BVP under consideration.

4) Variationally consistent integral forms yield symmetric coefficient matrices in the algebraic systems and the coefficient matrices are positive-definite, hence have real, positive eigenvalues and real eigenvectors (basis). Such coefficient matrices are invertible, hence yield unique values of the unknowns in the corresponding algebraic systems.

5) When the integral form is variationally inconsistent, a unique extremum principle does not exist. In such cases the coefficient matrix in the algebraic system resulting from the integral form is not symmetric, hence is not ensured to be positive-definite. A unique solution of the unknowns in such algebraic systems is not ensured.

A consequence of the non-positive-definite coefficient matrix in the algebraic system is that such coefficient matrices may have zero or negative eigenvalues or the eigenvalues and eigenvectors may be complex. In summary, variationally inconsistent integral forms must be avoided at all cost due to the fact that when using such integral forms a unique solution of the BVP is not ensured. In other words when obtaining solution of BVPs, variationally consistent integral forms are essential to ensure unique solutions of the BVPs.

6) The definition stated above can be applied to any BVP provided we can show existence of a functional $I(\phi)$ corresponding to the BVP $A\phi - f = 0$ such that $\delta I = 0$ and $\delta^2 I$ are necessary and sufficient conditions for the existence of extremum of $I(\phi)$. A ϕ yielding unique extremum of $I(\phi)$ is also a unique solution of $A\phi - f = 0$.

7) We can show (see ref. [1]-[4] [35] for details) that a) the integral forms resulting from GM/WF are VC only for self-adjoint differential operators when the bilinear functional is symmetric, b) the integral form resulting from LSP is VC for all three classes of differential operators, and c) integral forms resulting from the other methods of approximation (GM, PGM, WRM) for all three clases of differential operators are VIC.

8) We show that VC integral forms in designing finite element processes are essential for the derivations of the a priori error estimates.

9) In the following, we only consider GM/WF and LSP, keeping in mind that the integral form from GM/WF is VC only for self-adjoint differential operators and for LSP the integral forms are VC for all three classes of operators.

4. Variational Consistency of the Integral Form and the Best Approximation Property

In this section, we present some theorems and their proofs regarding GM/WF and LSP for the three classes of differential operators and establish best approximation property of GM/WF for self-adjoint operators and LSP for all three classes of operators.

4.1. Galerkin Method with Weak Form (GM/WF): Self-Adjoint Operators

In this section, we revisit main steps of GM/WF for self-adjoint operators. Let

$$A\phi - f = 0 \quad \text{in } \Omega \tag{1}$$

be a boundary value problem in which the differential operator A is symmetric and its adjoint $A^* = A$ (*i.e.* the differential operator A is self adjoint). Based on fundamental lemma of calculus of variations we can write the following integral form [1] [35]:

$$(A\phi - f, v)_{\bar{\Omega}} = \int_{\bar{\Omega}} (A\phi - f) v \mathrm{d}\Omega = \int_{\bar{\Omega}} (A\phi) v \mathrm{d}\Omega - \int_{\bar{\Omega}} f v \mathrm{d}\Omega = 0 \tag{2}$$

in which $v = 0$ on Γ^* if $\phi = \phi_0$ (given) on Γ^*. v is called test function, hence $v = \delta\phi$ is admissible in (2). When $v = \delta\phi$ in (2), the integral form (2) is called integral form in Galerkin method. Since A is self adjoint, the BVP (1) only contains even order derivatives of ϕ. We transfer half of the differentiation from ϕ to v using integration by parts in the first term in (2) and collect those terms that contain both ϕ and v and define them collectively as $B(\phi, v)$ and those that contain only v and define them as $l(v)$, hence we can write the following.

$$B(\phi, v) = l(v) \tag{3}$$

Each term in $B(\phi, v)$ contains both ϕ and v but more importantly the orders of derivatives of ϕ and v in each term is same (*i.e.* $B(\phi, v)$ is symmetric), thus

$$B(\phi, v) = B(v, \phi) \tag{4}$$

and since A is linear, $B(\phi, v)$ is bilinear in ϕ, v and $l(v)$ is linear in v. Hence in this case quadratic functional $I(\phi)$ is possible and is given by

$$I(\phi) = \frac{1}{2} B(\phi, \phi) - l(\phi) \tag{5}$$

The integral form (3) is called weak form of (1). Due to the fact that (2) is integral form in Galerkin method, the weak form (3) is called integral form in Galerkin method with weak form (GM/WF). The quadratic functional $I(\phi)$ has physical significance as explained in reference [35]. If (1) represents a BVP associated with linear elasticity in solid mechanics, then $\frac{1}{2}B(\phi,\phi)$ is strain energy, $l(\phi)$ is potential energy of loads and $I(\phi)$ is the total potential energy of the system described by (1).

Theorem 4.1. *The weak form* $B(\phi_h, v) = l(v)$ *resulting from GM/WF for self adjoint differential operator A in* $A\phi - f = 0$ *in which* $B(\cdot,\cdot)$ *is symmetric is variationally consistent.*

Proof. Variational consistency of the weak form $B(\phi_h, v) = l(v)$ requires that there exist a functional $I(\phi_h)$ such that $\delta I(\phi_h) = 0$ gives the weak form, the Euler's equation resulting from $\delta I(\phi_h) = 0$ is the BVP, and $\delta^2 I(\phi_h)$ yields unique extremum principle. Following Section 4.1 the existence of the functional $I(\phi_h)$ is by construction (Equation (5))

$$I(\phi_h) = \frac{1}{2}B(\phi_h, \phi_h) - l(\phi_h)$$

If $I(\phi_h)$ is differentiable in ϕ_h, then $\delta I(\phi_h) = 0$ is a necessary condition for an extremum of $I(\phi_h)$. Using $\delta\phi_h = v$ (due to GM/WF),

$$\delta I(\phi_h) = \frac{1}{2}B(v, \phi_h) + \frac{1}{2}B(\phi_h, v) - l(v) = 0$$

Since $B(\cdot,\cdot)$ is symmetric, we obtain

$$\delta I(\phi_h) = B(\phi_h, v) - l(v) = 0$$

or

$$B(\phi_h, v) = l(v), \text{ the weak form}$$

The unique extremum principle (or sufficient condition) is given by

$$\delta^2 I(\phi_h) = \delta\left(B(\phi_h, v) - l(v)\right) = B(v, v) > 0, \ \forall v \in V_h \subset H$$

Hence, a unique extremum principle.

To show that the Euler's equation resulting from the weak form is in fact the BVP, we just have to transfer differentiation back to ϕ (or ϕ_h) from v in the weak form using integration by parts. This is rather straightforward. Thus, the weak form $B(\phi_h, v)$ resulting from the GM/WF is variationally consistent. $\delta^2 I(\phi_h) = B(v, v) > 0$ implies that a ϕ_h from the weak form minimizes $I(v), \forall v \in V_h$,

$$I(\phi_h) \le I(v), \ \forall v \in V_h$$ □

Theorem 4.2. *Let* $A\phi - f = 0$ *be a BVP in which A is self adjoint and let* $B(\phi_h, v) = l(v)$ *be weak form resulting from GM/WF in which* $B(\phi_h, v) = B(v, \phi_h)$ *and* $\phi_h, v \in V_h \subset H$, *then* ϕ_h *has best approximation property in* $B(\cdot,\cdot)$*-norm. That is, if* $e = \phi - \phi_h$, $\phi \in H$ *being theoretical solution, then*

(a) $B(e, v) = 0, \ \forall v \in V$

(b) $B(e, e) \le B(\phi - w, \phi - w), \ \forall w \in V$

Proof.

a)

$$B(\phi_h, v) = l(v)$$

$$B(\phi, v_1) = l(v_1), v_1 \in H$$

Choosing $v_1 = v \in V \subset H$

$$B(\phi, v) = l(v)$$

Hence,

$$B(\phi - \phi_h, v) = 0$$

or

$$B(e,v)=0$$

This implies that no element of V is a better approximation of ϕ than ϕ_h, the solution for the weak form when measured in $B(\cdot,\cdot)$ as e is $B(\cdot,\cdot)$-orthogonal to every element v of V. This is called the best approximation property of GM/WF for self adjoint operators.

b) For any $v \in V_h$

$$B(e+v,e+v)=B(e,e)+2B(e,v)+B(v,v)$$

But $B(e,v)=0$, hence

$$B(e+v,e+v)=B(e,e)+B(v,v)$$

Since $B(v,v)>0$ we have

$$B(e,e)\le B(e+v,e+v)$$

$$e+v=\phi-\phi_h+v=\phi-(\phi_h-v)$$

$$\phi_h, v \in V; \quad \text{hence } \phi_h - v = w \in V_h$$

Thus,

$$B(e,e)\le B(\phi-w,\phi-w),\ \forall w \in V_h \subset H$$

or

$$B(\phi-\phi_h,\phi-\phi_h)\le B(\phi-w,\phi-w)$$

or

$$\|\phi-\phi_h\|_B \le \|\phi-w\|_B$$

That is, error in ϕ_h in B-norm is the lowest compared to any other solution w. This completes the proofs of a) and b). □

4.2. GM/WF for Non-Self Adjoint and Non-Linear Operators

Theorem 4.3. *Let $A\phi-f=0$ in Ω be a BVP in which A is a non-self adjoint differential operator. Let $B(\phi,v)-l(v)=0$ be all possible weak forms. Then all such integral forms are variationally inconsistent.*

Proof. Let there exist a functional $I(\phi)$ such that $\delta I(\phi)=0$ yield the weak form $B(\phi,v)-l(v)=0$. Since A is non-self adjoint, $B(\phi,v)$ is bilinear but not symmetric (*i.e.* $B(\phi,v)\neq B(v,\phi)$), hence

$$\delta^2 I(\phi)=\delta\left(B(\phi,v)-l(v)\right)=B(\delta\phi,v)$$

$$=B(v,v)\begin{cases}>0\\=0,\ \forall v\in V\\<0\end{cases}$$

is not possible because $B(\cdot,\cdot)$ is not symmetric. Therefore, $\delta^2 I(\phi)$ is not a unique extremum principle. Thus, the integral form $B(\phi,v)-l(v)=0$ with $v=\delta\phi$ is VIC when the differential operator is non-self adjoint. □

Theorem 4.4. *Let $A\phi-f=0$ in Ω be a BVP in which A is a non-linear differential operator and let $B(\phi,v)-l(v)=0$ be all possible weak forms of $A\phi-f=0$ in Ω. Then, all such integral forms or weak forms are variationally inconsistent.*

Proof. Let there exist a functional $I(\phi)$ such that $\delta I(\phi)$ yields the integral form $B(\phi,v)-l(v)=0$. Since the differential operator A is non-linear, $B(\phi,v)$ is linear in v but not linear in ϕ and $l(v)$ is linear in v. Therefore, the second variation of I

$$\delta^2 I(\phi)=\delta\left(B(\phi,v)-l(v)\right)=B(\delta\phi,v)$$

is a function of ϕ due to the fact that $B(\phi,v)$ is a non-linear function of ϕ. Thus, $\delta^2 I(\phi)$ does not represent

a unique extremum principle and, hence, the integral form or weak form $B(\phi, v) - l(v) = 0$ is VIC. □

4.3. Least-Squares Method Based on Residual Functional: Self-Adjoint and Non-Self-Adjoint Operators

Theorem 4.5. *The integral form in least-squares method based on residual functional is variationally consistent when the BVP is described by self adjoint differential operator.*

Proof. Consider the BVP

$$A\phi - f = 0, \quad \forall x \in \Omega$$

in which A is self adjoint. Let $\phi_h \in V_h \subset H$ be an approximation of ϕ over discretization $\bar{\Omega}^T = \bigcup_e \bar{\Omega}^e$ of $\bar{\Omega}$. Let

$$E = A\phi_h - f, \quad \forall x \in \bar{\Omega}^T$$

be residual function. We define residual functional

$$I(\phi_h) = (E, E)$$

If $I(\phi_h)$ is differentiable in ϕ_h then the necessary condition is given by $\delta I(\phi_h) = 0$.

$$\delta I(\phi_h) = 2(E, \delta E) = 2(A\phi_h - f, Av) = 0, \quad v = \delta\phi_h \in V_h$$

or

$$(A\phi_h, Av) = (f, Av)$$

or

$$B(\phi_h, v) = l(v)$$

$B(\phi_h, v)$ is bilinear and symmetric and $l(v)$ is linear.

$$\delta^2 I(\phi_h) = (\delta E, \delta E) = (Av, Av) > 0, \quad \forall v \in V_h$$

Hence, the integral form resulting from $\delta I(\phi_h) = 0$ is variationally consistent. □

Theorem 4.6. *The integral form in least-squares method based on residual functional is variationally consistent when the BVP is described by non-self adjoint operator.*

Proof. Since non-self adjoint operators are linear the proof of this theorem is same as that for self adjoint operators (Theorem 4.5) which are also linear. □

4.4. Least-Squares Method Based on Residual Functional for Non-Linear Operators

Theorem 7 *Let $A\phi - f = 0$ in Ω be a boundary value problem in which A is a non-linear differential operator. Let ϕ_h be approximation of ϕ in $\bar{\Omega}^T = \bigcup_e \bar{\Omega}^e$, discretization of $\bar{\Omega}$ and let $A\phi_h - f = E$ be the residual function in $\bar{\Omega}$. Then the integral form resulting from the first variation of the residual functional $I(\phi_h) = (E, E)$ set to zero is VC provided $\delta^2 I(\phi_h) \cong (\delta E, \delta E)$ and the system of non-linear algebraic equations resulting from $\delta I(\phi_h) = 0$ are solved using Newton-Raphson or Newton's linear method.*

Proof. Since A is non-linear, E is a non-linear function of ϕ_h, hence δE is a function of ϕ_h.

$$I(\phi_h) = (E, E) = (A\phi_h - f, A\phi_h - f); \quad \text{existence of } I(\phi_h)$$

If $I(\phi_h)$ is differentiable in ϕ_h, then

$$\delta I(\phi_h) = 2(E, \delta E) = 2g(\phi_h) = 0$$

Hence, $g(\phi_h) = 0$ is a necessary condition.

$$\text{Since} \quad \delta E = \delta(A\phi_h - f) = \delta A(\phi_h) + Av,$$

$$g(\phi_h) = (A\phi_h - f, \delta A(\phi_h) + Av) = 0$$

or

$$(A\phi_h, Av + \delta A(\phi_h)) = (f, \delta A(\phi_h) + Av)$$

or

$$B(\phi_h, v) = l(v)$$

Also

$$\delta^2 I(\phi_h) = 2(\delta E, \delta E) + 2(E, \delta^2 E) \begin{cases} > 0 \\ = 0, \ \forall \phi_h, v \in V_h \subset H \\ < 0 \end{cases}$$

is not possible. Hence, we do not have a unique extremum principle. At this stage, the least-squares process is VIC. We rectify the situation in the following.

We note that based on the necessary condition $g(\phi_h) = 0$ must hold. Since $g(\phi_h)$ is a non-linear function of ϕ_h, we must find a ϕ_h iteratively that satisfies $g(\phi_h) = 0$. Let ϕ_h^0 be an initial (or assumed) solution, then

$$g(\phi_h^0) \neq 0$$

Let $\Delta\phi_h$ be a change in ϕ_h^0 such that

$$g(\phi_h^0 + \Delta\phi_h) = 0$$

Expanding $g(\phi_h^0 + \Delta\phi_h)$ in Taylor series about ϕ_h^0 and retaining only up to linear terms in $\Delta\phi_h$ (Newton-Raphson or Newton's linear method)

$$g(\phi_h^0 + \Delta\phi_h) \cong g(\phi_h^0) + \left.\frac{\partial g(\phi_h)}{\partial \phi_h}\right|_{\phi_h^0} \Delta\phi_h = 0$$

Therefore

$$\Delta\phi_h = -\left[\left.\frac{\partial g(\phi_h)}{\partial \phi_h}\right|_{\phi_h^0}\right]^{-1} g(\phi_h^0)$$

But $\dfrac{\partial g(\phi_h)}{\partial \phi_h} = \dfrac{1}{2}\delta(\delta I(\phi_h)) = \dfrac{1}{2}\delta^2 I(\phi_h)$. Hence

$$\Delta\phi_h = -\frac{1}{2}\left[\delta^2 I(\phi_h)\right]_{\phi_h^0}^{-1} g(\phi_h^0)$$

Thus, in order for the coefficient matrix $\left[\delta^2 I(\phi_h)\right]$ to be positive-definite,

$$\delta^2 I(\phi_h) \cong 2(\delta E, \delta E) > 0$$

This gives a unique extremum principle. The improved value of ϕ_h is given by

$$\phi_h = \phi_h^0 + \alpha^* \Delta\phi_h$$

We choose α^* such that $I(\phi_h) \leq I(\phi_h^0)$. This is referred to as line search. With this approximation of $\delta^2 I(\phi_h)$, the integral form $\delta I(\phi_h) = (A\phi_h - f, Av + \delta A(\phi_h))$ is variationally consistent. □

Remarks

1) Justification for approximating $\delta^2 I(\phi_h)$ is important to discuss.

2) We note that

$$g(\phi_h) = (E, \delta E)$$

Justification of $\delta^2 I(\phi_h) \cong 2(\delta E, \delta E)$ is only necessary in the asymptotic range of convergence as the a priori error estimation only holds in this range, thus establishing best approximation property of LSM method in some norm is also only required in this range.

$$\delta^2 I(\phi_h) = (\delta E, \delta E) + (E, \delta^2 E)$$

In the asymptotic range $E \to 0$ in the pointwise sense if the approximation spaces are minimally conforming to ensure that all integrals over $\bar{\Omega}^T$ are Riemann. When $E \to 0 \quad \forall x \in \bar{\Omega}^T$ then $(E, \delta^2 E) \to 0$, hence $\delta^2 I(\phi_h) \cong (\delta E, \delta E)$ is valid. Further discussion on the validity of this approximation can be found in reference [35].

Theorem 4.8. *The integral form resulting from the least-squares method based on residual functional has best approximation property in L_2-norm of E.*

Proof. From Section 4.3, we have

$$(E, \delta E) = 0$$

or

$$(A\phi_h - f, Av) = 0$$

For theoretical or exact solution ϕ, we have

$$A\phi - f = 0 \Rightarrow f = A\phi$$

Hence,

$$(A(\phi_h - \phi), Av) = (Ae, Av) = 0, \quad e = \phi_h - \phi$$

Thus, $A(\phi_h - \phi)$ or Ae is orthogonal to $Av \in {}^A V_h$ (dual of V_h). We note that

$$\|A(\phi_h - \phi)\|_{L_2} = \|Ae\|_{L_2} = \|E\|_{L_2}$$

That is L_2-norm of E obtained using ϕ_h is lowest out of all $v \in V_h$. Hence, LSP has best approximation property in L_2-norm of E or $\|E\|_{L_2}$. □

Theorem 4.9. *A variationally consistent integral form has a best approximation property in some associated norm. Conversely, if an integral form has a best approximation property in some norm, then it is variationally consistent.*

Proof. Proof of this theorem follows due to the fact that VC integral form in GM/WF has best approximation property in B-norm because $B(\cdot,\cdot)$ is bilinear and symmetric. The integral form in the LSP is also VC but LSP has best approximation property in L_2-norm of E. Both GM/WF and LSP are VC but have best approximation property in different norms. In both cases, VC integral form is not possible without best approximation property and the best approximation property is not possible without VC integral form. This is obviously due to the fact that they both require the functional $B(\cdot,\cdot)$ to be bilinear and symmetric. As long as this holds, how $B(\cdot,\cdot)$ is derived is not important. □

We note that

1) Since the integral forms for non-self adjoint and non-linear differential operators are VIC in GM/WF, the approximation ϕ_h from GM/WF does not have best approximation property in B-norm (Theorem 4.9).

2) Lack of best approximation property and lack of VC of the integral form resulting from GM/WF for non-self adjoint and non-linear differential operators are both obviously due to the fact that the functional $B(\cdot,\cdot)$ in the weak forms is not symmetric.

3) In LSP for all classes of differential operator $I(\phi_h) = (E, E)_{\bar{\Omega}^T}$ is minimized, therefore ϕ_h has best approximation property in E-norm

$$\left(\|E\|_{L_2} = (E, E)^{\frac{1}{2}} \right).$$

4) We note that variational consistency of the integral form holds for all choices of h, p, and k whereas the best approximation property only holds in the asymptotic range.

4.5. Integral Forms Based on Other Methods of Approximation

The integral forms used in finite element method based on Petrov-Galerkin method, Galerkin method, and weighted residual method are not considered as these always yield integral forms that are variationally inconsistent. Hence, when using these integral forms computations may not even be possible.

4.6. General Remarks

1) We have established that GM/WF yields VC integral form only for self adjoint operators when the functional $B(\cdot,\cdot)$ in the integral form is symmetric and this method has best approximation property in *B*-norm.

2) LSP based on residual functional yields VC integral forms for self adjoint, non-self adjoint, and non-linear (in the asymptotic range) differential operators and has best approximation property in *E*-norm.

3) VC integral form implies best approximation property in some norm and vice versa.

4) Best approximation property is necessary in a priori error estimation (in the asymptotic range), as shown in subsequent sections.

5) In general, when using GM, PGM, WRM, etc. error estimation is not possible as in these methods the approximation ϕ_h of ϕ does not have best approximation property in any norm.

5. A Priori Error Estimates: GM/WF and LSP

We consider simple model problems to demonstrate the best approximation properties of GM/WF for self adjoint operators and LSP for linear operators and present derivations of the a priori error estimates and convergence rates when $\phi_h \in V_h \subset H^{k,p}\left(\bar{\Omega}^e\right)$. These estimates are derived using model problems (as illustrations) and are then generalized for all BVPs.

5.1. Model Problem 1: GM/WF

Consider the following BVP:

$$-\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} = f(x) \text{ or } -\phi'' = f(x),\ \forall x \in \Omega_x = (0,L) \tag{6}$$

$$\text{BCs:}\quad \phi(0) = \phi(L) = 0 \tag{7}$$

GM/WF for (6) with BCs (7) gives

$$(\phi', v') = (f, v),\ \forall v \in V \subset H \tag{8}$$

Let $\phi_h \in V_h \subset H$ be the finite element approximation of ϕ, then we have

$$(\phi_h', v') = (f, v),\ \forall v \in V_h \tag{9}$$

Using (8) and (9) and since $V_h \subset V$, v in (9) is also in V and we have

$$(\phi' - \phi_h', v') = 0,\ \forall v \in V_h \tag{10}$$

Theorem 5.1. *For any* $v \in V_h$ *we have*

$$\|\phi' - \phi_h'\| \le \|\phi' - v'\|,\ \forall v \in V_h$$

Proof.

$$\|\phi' - \phi_h'\|^2 = (\phi' - \phi_h', \phi' - \phi_h')$$

Since

$$(\phi' - \phi_h', w') = 0,\ \forall w \in V_h$$

we can choose $w = \phi_h - v$ as both $\phi_h, v \in V_h$, then

$$(\phi' - \phi_h', \phi_h' - v') = 0,\ \forall v \in V_h$$

Hence,

$$\begin{aligned}\left\|\phi'-\phi_h'\right\|_{L_2}^2 &= \left(\phi'-\phi_h', \phi'-\phi_h'\right)+\left(\phi'-\phi_h', \phi_h'-v'\right)\\ &= \left(\phi'-\phi_h', \left(\phi'-\phi_h'\right)+\left(\phi_h'-v'\right)\right)\\ &= \left(\phi'-\phi_h', \phi'-v'\right)\end{aligned}$$

Using Cauchy-Schwarz inequality [35]

$$\left\|\phi'-\phi_h'\right\|_{L_2}^2 \le \left\|\phi'-\phi_h'\right\|_{L_2}\left\|\phi'-v'\right\|_{L_2}$$

or

$$\left\|\phi'-\phi_h'\right\|_{L_2} \le \left\|\phi'-v'\right\|_{L_2},\ \ \forall v \in V$$

or

$$\left\|\phi-\phi_h\right\|_B \le \left\|\phi-v\right\|_B$$

That is, in this case for the model problem (6) - (7) the derivative of ϕ_h has the best approximation property in L_2-norm. Alternatively, $\phi-\phi_h$ has best approximation property in B-norm. This completes the proof. □

5.2. Model Problem 2: LSP

Consider the following BVP described by non-self adjoint differential operator.

$$\phi' = f,\ \ \forall x \in (0,1) = \Omega \subset \mathbb{R}^1 \tag{11}$$

$$\text{BC:}\quad \phi(0) = 0 \tag{12}$$

LSP based on residual functional gives (for $f = 0$)

$$\left(\phi_h', v'\right) = 0,\ \ v = \delta\phi_h,\ \ \phi_h, v \in V_h \subset H \tag{13}$$

ϕ_h is approximation of ϕ over $\bar{\Omega}$. This integral form is VC. Also for theoretical solution

$$\left(\phi', v_t'\right) = 0,\ \ v_t = \delta\phi \tag{14}$$

Setting $v_t = v$ in (14)

$$\left(\phi', v'\right) = 0 \tag{15}$$

Subtracting (13) from (15)

$$\left(\phi'-\phi_h', v'\right) = \left(e', v'\right) = 0,\ \ v \in V_h \tag{16}$$

Using interpolant ϕ_I of ϕ (interpolant matches ϕ at end nodes); $\phi_I \in V_h$, let $e = \phi-\phi_I+\phi_I-\phi_h$, then we have

$$\begin{aligned}\left\|e'\right\|_{L_2}^2 &= \left(e', e'\right) = \left(e', \left(\phi'-\phi_I'\right)+\left(\phi_I'-\phi_h'\right)\right)\\ &= \left(e', \phi'-\phi_I'\right)+\left(e', \phi_I'-\phi_h'\right)\end{aligned} \tag{17}$$

We note that $\phi_I-\phi_h = w \in V_h$, hence

$$\left(e', \phi_I'-\phi_h'\right) = \left(e', w'\right) = 0 \quad \text{(due to (5.11))} \tag{18}$$

Thus, (17) reduces to

$$\left\|e'\right\|_{L_2}^2 = \left(e', \phi'-\phi_I'\right) \tag{19}$$

Using Cauchy-Schwarz inequality

$$\left\|e'\right\|_{L_2}^2 \le \left\|e'\right\|_{L_2}\left\|\phi'-\phi_I'\right\|_{L_2} \tag{20}$$

Thus,

$$\left\|e'\right\|_{L_2} \le \left\|\phi' - \phi_I'\right\|_{L_2} \tag{21}$$

That is L_2-norm of the derivative of error $\phi - \phi_h$ is bounded by the finite element interpolant. Using proposition 5.1 (shown subsequently) and (21), we can write

$$\left\|e'\right\|_{L_2} \le h\left|\phi\right|_2 \tag{22}$$

L_2-norm of e; that is, $\left\|e\right\|_{L_2}$ for LSP is derived using Aubin-Nitsche trick (Oden and Carey [33] and Reddy [34]). We consider details in the following.

Consider the same BVP (for $f = 0$),

$$\begin{aligned} &\phi' = f, \quad \forall x \in (0,1) = \Omega \\ &\phi(0) = 0 \end{aligned} \tag{23}$$

Let $e = \phi - \phi_h$. Assume that w is the solution of the second order differential equation

$$\left.\begin{aligned} &-w'' = e, \ \forall x \in (0,1) = \Omega \\ &w(0) = 0 \\ &w'(1) = 0 \end{aligned}\right\} \tag{24}$$

The finite element interpolant w_I ($w_I(0) = w_I(1) = 0$) satisfies

$$\left\|w' - w_I'\right\|_{L_2} \le h\left|w\right|_2 = h\left\|w''\right\|_{L_2} \tag{25}$$

$$\le h\left\|e\right\|_{L_2} \ \text{(using (5.19))} \tag{26}$$

Consider

$$(e,e) = -(e,w'') \tag{27}$$

Using integration by parts and the fact that $e = 0$ at $x = 0$ and $w' = 0$ at $x = 1$ and $(e', w_I') = 0$ (orthogonal property)

$$(e,e) = -(e,w'') = (e',w') = (e',w' - w_I') \tag{28}$$

Hence, (using Cauchy-Schwarz inequality)

$$\left\|e\right\|_{L_2}^2 \le \left\|e'\right\|_{L_2} \left\|w' - w_I'\right\|_{L_2} \le \left(h\left\|\phi''\right\|_{L_2}\right)\left(h\left\|e\right\|_{L_2}\right) \tag{29}$$

Dividing by $\left\|e\right\|_{L_2}$

$$\left\|e\right\|_{L_2} \le h^2\left\|\phi''\right\| = h^2\left|\phi\right|_2 \tag{30}$$

We make the following remarks.

1) For a first order BVP, the rate of convergence of the L_2-norm of the error in the finite element solution is proportional to h^2 and the rate of convergence of the L_2-norm of the derivative of the error is proportional to h.

2) These estimates are same as those for a second order BVP when using GM/WF in which the integral form is variationally consistent.

General Remarks

1) The error estimates have been derived for a second order BVP using GM/WF in which the integral form is VC and the local approximation is linear ($p = 1$) over an element. In case of LSP the BVP is first order ODE, the integral form is VC, and $p = 1$ for local approximation.

2) We note that the integral forms in both cases are VC and contain only up to first order derivatives, hence the reason for same convergence rates of $\left\|\phi - \phi_h\right\|_{L_2}$ and $\left\|\phi' - \phi_h'\right\|_{L_2}$ even though in case of GM/WF the BVP is a second order ODE and in case of LSP it is only a first order ODE. This is rather significant to note that VC of the integral form and the highest order of the derivative in the integral form control the rates of convergence.

3) We need to extend these estimates for higher degree local approximation (*i.e.* p-level of "p").

4) The order of approximation space k needs to be incorporated in the error estimates.

5.3. Proposition and Proof

Proposition 1 *Let the theoretical solution* ϕ *of* (6) - (7) *be at least of class* $C^2[0,L]$ *and let* ϕ_h *be approximation of* ϕ *over the discretization* $\bar{\Omega}^T = \bigcup_e \bar{\Omega}^e$ *of* $\bar{\Omega} = [0,1]$ *in which* $\bar{\Omega}^e = [x_i, x_{i+1}]$ *is an element e. Let h be the characteristic length of* $\bar{\Omega}^T$ *such that* $h = \max_e h_e$. *Let* $\phi_I = \bigcup_e \phi_I^e$ *be interpolant of* ϕ *that agrees with* ϕ *at the nodes* [*i.e.* $\phi(x_i) = \phi_I(x_i)$, $i = 0,1,\cdots$]. *Then*

a)

$$\left|\tilde{E}(x)\right| = \left|\phi(x) - \phi_I(x)\right| \le h^2 \max_{[0,L]} \left|\phi''\right| \tag{31}$$

b)

$$\left|\tilde{E}'(x)\right| = \left|\phi'(x) - \phi_I'(x)\right| \le h \max_{[0,L]} \left|\phi''\right| \tag{32}$$

c) When (31) and (32) hold, the following hold

$$\left\|\tilde{E}'(x)\right\|_{L_2} = \left\|\phi'(x) - \phi_I'(x)\right\|_{L_2} = \left|\phi(x) - \phi_I(x)\right|_{H^1} = \left|\phi(x) - \phi_I(x)\right|_1 \le h\left\|\phi''\right\|_{L_2} = h\left|\phi\right|_2 \tag{33}$$

$$\left\|\tilde{E}(x)\right\|_{L_2} = \left\|\phi(x) - \phi_I(x)\right\|_{L_2} = \left|\phi(x) - \phi_I(x)\right|_{H^0} = \left|\phi(x) - \phi_I(x)\right|_0 \le h^2\left\|\phi''\right\|_{L_2} = h^2\left|\phi\right|_2 \tag{34}$$

$$\left\|\tilde{E}(x)\right\|_{H^1} = \left\|\tilde{E}(x)\right\|_1 = \left\|\phi(x) - \phi_I(x)\right\|_{H^1} = \left\|\phi(x) - \phi_I(x)\right\|_1 \le h\left\|\phi''\right\|_{L_2} = Ch\left|\phi\right|_2 \tag{35}$$

in which

$$\left\|\phi''\right\|_{L_2} = \left|\phi\right|_{H^2} = \left|\phi\right|_2 = \left(\int_0^1 \left(\phi''(x)\right)^2 \mathrm{d}x\right)^{\frac{1}{2}} \tag{36}$$

Proof. Consider linear $\phi_h^e(x)$ and $\phi_I^e(x)$ (*i.e.* $p = 1$).

For an element *e* let $\tilde{E}^e(x) = \phi(x) - \phi_I^e(x), \forall x \in [x_i, x_{i+1}]$ be the interpolation error between ϕ and interpolant $\phi_I^e(x)$. Since $\tilde{E}^e(x)$ vanishes at x_i and x_{i+1} of an element *e*, by virtue of Rolle's theorem there exists at least one point β between x_i and x_{i+1} at which $\left(\tilde{E}^e(x)\right)' = 0$. Then for any *x*

$$\left(\tilde{E}^e(x)\right)' = \int_\beta^x \left(\tilde{E}^e(x)\right)'' \mathrm{d}x, \quad \left|\left(\tilde{E}^e(x)\right)'\right| \le \int_\beta^x \left|\left(\tilde{E}^e(x)\right)''\right| \mathrm{d}x \tag{37}$$

Since $\phi_I^e(x)$ is linear, $\tilde{E}^e(x) = \phi - \phi_I^e$ implies that

$$\left(\tilde{E}^e(x)\right)'' = \phi''(x) - \left(\phi_I^e(x)\right)'' = \phi''(x) \tag{38}$$

Applying Cauchy-Schwarz inequality to (37) and using (38)

$$\left|\left(\tilde{E}^e(x)\right)'\right| \le \left(\int_\beta^x (1)^2 \mathrm{d}x\right)^{\frac{1}{2}} \left(\int_\beta^x \left|\phi''(x)\right|^2 \mathrm{d}x\right)^{\frac{1}{2}} \tag{39}$$

$$\le (h_e)^{\frac{1}{2}} \left(\int_\beta^x \left|\phi''(x)\right|^2 \mathrm{d}x\right)^{\frac{1}{2}} \tag{40}$$

$$\le (h_e)^{\frac{1}{2}} \left(\int_{x_i}^{x_{i+1}} \left(\max_{\bar{\Omega}^e} \left|\phi''(x)\right|\right)^2 \mathrm{d}x\right)^{\frac{1}{2}} \tag{41}$$

$$\le (h_e)^{\frac{1}{2}} \left(h_e \left(\max_{\bar{\Omega}^e} \left|\phi''(x)\right|\right)^2\right)^{\frac{1}{2}} \tag{42}$$

or

$$\left|\left(\tilde{E}^e(x)\right)'\right| \le h_e \max_{\bar{\Omega}^e}\left|\phi''(x)\right| \tag{43}$$

Let

$$\left.\begin{aligned} h &= \max_e h_e \\ \max_{\bar{\Omega}^e}\left|\phi''(x)\right| &\le \max_{[0,1]}\left|\phi''(x)\right| \end{aligned}\right\} \tag{44}$$

Hence for $\bar{\Omega}^T$, we can write

$$\left|\left(\tilde{E}^e(x)\right)'\right| = \left|\phi'(x) - \phi_I'(x)\right| \le h \max_{[0,1]}\left|\phi''(x)\right| \tag{45}$$

This proves (32).
Likewise (since $\tilde{E}^e(x_i) = 0$),

$$\tilde{E}^e(x) = \int_{x_i}^{x}\left(\tilde{E}^e(x)\right)' \mathrm{d}x, \quad \left|\tilde{E}^e(x)\right| \le \int_{x_i}^{x}\left|\left(\tilde{E}^e(x)\right)'\right|\mathrm{d}x \tag{46}$$

Applying Cauchy-Schwarz inequality

$$\left|\tilde{E}^e(x)\right| \le \left(\int_{x_i}^{x_{i+1}} (1)^2\,\mathrm{d}x\right)^{\frac{1}{2}}\left(\int_{x_i}^{x}\left|\left(\tilde{E}^e(x)\right)'\right|^2\mathrm{d}x\right)^{\frac{1}{2}} \tag{47}$$

Substituting from (43) into (47)

$$\begin{aligned} \left|\tilde{E}^e(x)\right| &\le (h_e)^{\frac{1}{2}}\left(\int_{x_i}^{x}\left(h_e\int_{x_i}^{x_{i+1}}\left(\max_{\bar{\Omega}^e}\left|\phi''(x)\right|\right)^2\mathrm{d}x\right)\mathrm{d}x\right)^{\frac{1}{2}} \\ &\le (h_e)^{\frac{1}{2}}\left(h_e h_e\int_{x_i}^{x_{i+1}}\left(\max_{\bar{\Omega}^e}\left|\phi''(x)\right|\right)^2\mathrm{d}x\right)^{\frac{1}{2}} \\ &= (h_e)^{\frac{3}{2}}\left(\int_{x_i}^{x_{i+1}}\left(\max_{\bar{\Omega}^e}\left|\phi''(x)\right|\right)^2\mathrm{d}x\right)^{\frac{1}{2}} \end{aligned} \tag{48}$$

Hence,

$$\left|\tilde{E}^e(x)\right| \le h_e^2 \max_{\bar{\Omega}^e}\left|\phi''(x)\right| \tag{49}$$

Using (44), (49) reduces to

$$\left|\tilde{E}(x)\right| \le h^2 \max_{[0,1]}\left|\phi''(x)\right| \tag{50}$$

This proves (31):

$$\left\|\tilde{E}'(x)\right\|_{L_2}^2 \le \sum_e \int_{x_i}^{x_{i+1}}\left|\left(\tilde{E}^e(x)\right)'\right|^2\mathrm{d}x \tag{51}$$

Substituting $\left|\left(\tilde{E}^e(x)\right)'\right|$ from (40) into (51)

$$\left\|\tilde{E}'(x)\right\|_{L_2}^2 \le \sum_e \int_{x_i}^{x_{i+1}}\left(h_e\int_{\beta}^{x}\left|\phi''(x)\right|^2\mathrm{d}x\right)\mathrm{d}x \tag{52}$$

$$\leq h\sum_{e}\int_{x_i}^{x_{i+1}}\left(\int_{x_i}^{x_{i+1}}\left|\phi''(x)\right|^2\,\mathrm{d}x\right)\mathrm{d}x \tag{53}$$

$$\leq h\sum_{e}\int_{x_i}^{x_{i+1}}\left(\left\|\phi''(x)\right\|_{L_2}^2\right)_{\bar{\Omega}^e}\mathrm{d}x \tag{54}$$

$$\leq h^2\sum_{e}\left(\left\|\phi''(x)\right\|_{L_2}^2\right)_{\bar{\Omega}^e} \tag{55}$$

Thus,

$$\left\|\tilde{E}'(x)\right\|_{L_2}^2 \leq h^2\left\|\phi''(x)\right\|_{L_2}^2 \tag{56}$$

Hence,

$$\left\|\tilde{E}'(x)\right\|_{L_2} = \left|\tilde{E}'(x)\right|_{H^1} = \left|\tilde{E}'(x)\right|_1 \leq h\left\|\phi''(x)\right\|_{L_2} = h\left|\phi\right|_{H^2} = h\left|\phi\right|_2 \tag{57}$$

This proves (33).
Consider

$$\tilde{E}^e(x) = \int_{x_i}^{x}\left(\tilde{E}^e(x)\right)'\mathrm{d}x \tag{58}$$

or

$$\left|\tilde{E}^e(x)\right| \leq \int_{x_i}^{x}\left|\left(\tilde{E}^e(x)\right)'\right|\mathrm{d}x \tag{59}$$

Using Cauchy-Schwarz inequality

$$\left|\tilde{E}^e(x)\right| \leq \left(\int_{x_i}^{x_{i+1}}(1)^2\,\mathrm{d}x\right)^{\frac{1}{2}}\left(\int_{x_i}^{x}\left|\left(\tilde{E}^e(x)\right)'\right|^2\mathrm{d}x\right)^{\frac{1}{2}} \tag{60}$$

$$\leq (h_e)^{\frac{1}{2}}\left(\int_{x_i}^{x}\left|\left(\tilde{E}^e(x)\right)'\right|^2\mathrm{d}x\right)^{\frac{1}{2}} \tag{61}$$

Substituting from (40)

$$\left|\tilde{E}^e(x)\right| \leq (h_e)^{\frac{1}{2}}\left(\int_{x_i}^{x}\left(h_e\int_{\beta}^{x}\left|\phi''(x)\right|^2\mathrm{d}x\right)\mathrm{d}x\right)^{\frac{1}{2}} \tag{62}$$

$$\left\|\tilde{E}(x)\right\|_{L_2}^2 = \sum_{e}\int_{x_i}^{x_{i+1}}\left|\tilde{E}^e(x)\right|^2\mathrm{d}x \tag{63}$$

$$\leq \sum_{e}\int_{x_i}^{x_{i+1}}h_e\left(\int_{x_i}^{x}\left(h_e\int_{\beta}^{x}\left|\phi''(x)\right|^2\mathrm{d}x\right)\mathrm{d}x\right)\mathrm{d}x \tag{64}$$

$$\leq \sum_{e}\int_{x_i}^{x_{i+1}}h_e\left(\int_{x_i}^{x}\left(h_e\int_{x_i}^{x_{i+1}}\left|\phi''(x)\right|^2\mathrm{d}x\right)\mathrm{d}x\right)\mathrm{d}x \tag{65}$$

$$\leq h^2\sum_{e}\int_{x_i}^{x_{i+1}}\left(\int_{x_i}^{x_{i+1}}\left(\left\|\phi''(x)\right\|_{L_2}^2\right)_{\bar{\Omega}^e}\mathrm{d}x\right)\mathrm{d}x \tag{66}$$

$$\le h^4 \sum_e \left(\left\| \phi''(x) \right\|_{L_2}^2 \right)_{\bar{\Omega}^e} = h^4 \left\| \phi''(x) \right\|_{L_2}^2 \tag{67}$$

Hence

$$\left\| \tilde{E}(x) \right\|_{L_2} = \left\| \phi(x) - \phi_I \right\|_{L_2} = \left| \phi - \phi_I \right|_{H^0} = \left| \phi - \phi_I \right|_0 \le h^2 \left\| \phi'' \right\|_{L_2} = h^2 \left| \phi \right|_{H^2} = h^2 \left| \phi \right|_2 \tag{68}$$

Now

$$\left\| \phi - \phi_I \right\|_{H^1}^2 = \left\| \phi - \phi_I \right\|_1^2 = \left\| \phi - \phi_I \right\|_{L_2}^2 + \left\| \phi' - \phi_I' \right\|_{L_2}^2 \tag{69}$$

Using (56) and (68), we have

$$\left\| \phi - \phi_I \right\|_{H^1}^2 \le \left(h^2 \left| \phi \right|_2^2 + h^4 \left| \phi \right|_2^2 \right) \tag{70}$$

$$\le \left| \phi \right|_2^2 \left(h^2 + h^4 \right) \tag{71}$$

$$< C^2 h^2 \left| \phi \right|_2^2 ; \left(h^4 \le h^2 \right) \tag{72}$$

Hence

$$\left\| \phi - \phi_I \right\|_{H^1} = \left\| \phi - \phi_I \right\|_1 \le Ch \left| \phi \right|_2 \tag{73}$$

This proves (35).

Remarks

From theorem 5.1, we have

$$\left\| \phi' - \phi_h' \right\|_{L_2} \le \left\| \phi' - \phi_I' \right\|_{L_2} = \left\| E'(x) \right\|_{L_2} \tag{74}$$

and

$$\left\| \phi - \phi_h \right\|_{L_2} \le \left\| \phi - \phi_I \right\|_{L_2} = \left\| E(x) \right\|_{L_2} \tag{75}$$

Hence using (74), (75), (33), and (34), we finally have

$$\left\| \phi' - \phi_h' \right\|_{L_2} \le h \left| \phi \right|_2 = h \left\| \phi'' \right\|_{L_2} = h \left| \phi \right|_{H^2} \tag{76}$$

and

$$\left\| \phi - \phi_h \right\|_{L_2} \le h^2 \left| \phi \right|_2 = h^2 \left\| \phi'' \right\|_{L_2} = h^2 \left| \phi \right|_{H^2} \tag{77}$$

and likewise

$$\left\| \phi - \phi_h \right\|_{H^1} \le Ch \left| \phi \right|_2 = Ch \left\| \phi'' \right\|_{L_2} = Ch \left| \phi \right|_{H^2} \tag{78}$$

5.4. Proposition and Proof

Proposition 5.2. The derivation of the error estimates in proposition 1 are presented for model problem 1 using GM/WF in which the operator is self adjoint, hence the weak form is VC. In model problem 2 (Section 5.2) the differential operator is non-self adjoint and the error estimates are derived for LSP in which the integral form is also VC. In this section we consider a more general approach of deriving a priori error estimates for arbitrary degree of approximation p only based on the assumption that the integral form is VC.

If the integral form resulting from a method of approximation is VC, then the following hold.

$$\begin{aligned} &\left\| \phi - \phi_h \right\|_{L_2} \le C_1 h^{p+1} \left| \phi \right|_{p+1} \\ &\left| \phi - \phi_h \right|_{H^q} \le C_2 h^{p+1-q} \left| \phi \right|_{p+1}, \text{ seminorm of order } q \\ &\left\| \phi - \phi_h \right\|_{H^q} \le C_3 h^{p+1-q} \left| \phi \right|_{p+1} \end{aligned} \tag{79}$$

And if

$$E = A\phi - f, \text{ residual function} \tag{80}$$

then

$$\|E\|_{L_2} \le C_4 h^{p+1-2m} |\phi|_{p+1} \tag{81}$$

In (81), $2m$ is the highest order of the derivative in the differential operator A. The constants C_1, C_2, C_3, and C_4 do not depend upon h and p.

Proof. Consider one dimensional BVP:

$$A\phi - f = 0, \ \forall x \in (0, L) = \Omega \tag{82}$$

Let $\bar{\Omega}^T = \bigcup_e \bar{\Omega}^e$ be discretization of $\bar{\Omega}$ in which $\bar{\Omega}^e = [x_i, x_{i+1}]$ is an element e. Let ϕ_h be finite element approximation of ϕ over $\bar{\Omega}^T$ such that $\phi_h = \bigcup_e \phi_h^e$ in which ϕ_h^e is local approximation of ϕ over $\bar{\Omega}^e$. Let ϕ_I and ϕ_I^e be interpolants of ϕ of class C^0 over $\bar{\Omega}^T$ and $\bar{\Omega}^e$ such that at the nodes ϕ_I agrees with the theoretical solution ϕ. Thus, error estimation reduces to estimating error between ϕ and ϕ_I over an element $\bar{\Omega}^e$ of length h_e. When $\phi(x)$ is analytic, it can be expanded in Taylor series in h_e over $\bar{\Omega}^e$ about some point j.

$$\phi(x) = \phi(h_e) = \phi_j + h_e \frac{\partial \phi_j}{\partial x} + \frac{h_e^2}{2!}\frac{\partial^2 \phi_j}{\partial x^2} + \cdots + \frac{h_e^p}{p!}\frac{\partial^p \phi_j}{\partial x^p} + \frac{h_e^{p+1}}{(p+1)!}\frac{\partial^{p+1} \phi_j}{\partial x^{p+1}} + \cdots \tag{83}$$

Consider a ϕ_h^e over $\bar{\Omega}^e$ of degree p resulting from a VC integral form (hence, ensuring well-behaved solution), then the local approximation ϕ_h^e at the same point j can also be written as (assuming ϕ_h^e agrees with $\phi(x)$ up to degree of p),

$$\phi_h^e(x) = \phi_h(h_e) = \phi_j + h_e \frac{\partial \phi_j}{\partial x} + \frac{h_e^2}{2!}\frac{\partial^2 \phi_j}{\partial x^2} + \cdots + \frac{h_e^p}{p!}\frac{\partial^p \phi_j}{\partial x^p} \tag{84}$$

Subtracting (84) from (83), we obtain

$$\left|\phi(x) - \phi_h^e(x)\right| \le O\left(h_e^{p+1}\right)\left|\frac{\partial^{p+1}\phi}{\partial x^{p+1}}\right| \tag{85}$$

$$\left\|\phi(x) - \phi_h^e(x)\right\|_{L_2}^2 \le \int_{x_i}^{x_{i+1}} C_1^2 \left(h_e^{p+1}\right)^2 \left|\frac{\partial^{p+1}\phi}{\partial x^{p+1}}\right|^2 \mathrm{d}x \tag{86}$$

$$\left\|\phi(x) - \phi_h^e(x)\right\|_{L_2}^2 \le \sum_e \left(\int_{x_i}^{x_{i+1}} C_1^2 \left(h_e^{p+1}\right)^2 \left|\frac{\partial^{p+1}\phi}{\partial x^{p+1}}\right|^2 \mathrm{d}x \right) \tag{87}$$

Let

$$h = \max_e h_e \tag{88}$$

Then

$$\left\|\phi(x) - \phi_h(x)\right\|_{L_2}^2 \le C_1^2 \left(h^{p+1}\right)^2 \sum_e \left(\int_{x_i}^{x_{i+1}} \left|\frac{\partial^{p+1}\phi}{\partial x^{p+1}}\right|^2 \mathrm{d}x \right) \le C_1^2 \left(h^{p+1}\right)^2 |\phi|_{p+1}^2 \tag{89}$$

Therefore

$$\left\|\phi(x) - \phi_h(x)\right\|_{L_2} \le C_1 h^{p+1} |\phi|_{p+1} \tag{90}$$

Using (83)-(90), it is rather straightforward to establish

$$\left\|\phi'(x) - \phi_h'(x)\right\|_{L_2} \le C_2 h^p |\phi|_{p+1} \tag{91}$$

and by induction

$$\left\|\phi^q(x)-\phi_h^q(x)\right\|_{L_2}=\left|\phi(x)-\phi_h(x)\right|_{H^q}\le C_2h^{p+1-q}\left|\phi\right|_{p+1} \tag{92}$$

Using (90) and (92), we can establish that

$$\left\|\phi(x)-\phi_h(x)\right\|_{H^q}\le C_3h^{p+1-q}\left|\phi\right|_{p+1},\ \left(q=0\text{ implies }L_2\text{-norm}\right) \tag{93}$$

□

Remarks

1) The estimates in (92) and (93) apply to VC integral forms regardless of the method of approximation. Thus, these estimates hold for GM/WF for self adjoint operators and also hold for LSP for all three classes of differential operators.

2) The local approximations used are always of class C^0.

3) The constants C_1, C_2, and C_3 do not depend on h and p.

4) The estimates (92) and (93) apply to all finite element processes in which the integral form is variationally consistent.

5) From (92) and (93), we note that progressively increasing order of derivatives of the finite element solution converge progressively slower. That is

$$\left\|\phi(x)-\phi_h(x)\right\|_{L_2}\propto h^{p+1} \tag{94}$$

$$\left\|\phi'(x)-\phi_h'(x)\right\|_{L_2}\propto h^{p} \tag{95}$$

and so on. Likewise

$$\left\|\phi(x)-\phi_h(x)\right\|_{H^0}\propto h^{p+1} \tag{96}$$

$$\left\|\phi(x)-\phi_h(x)\right\|_{H^1}\propto h^{p} \tag{97}$$

and so on. From (95) and (97), we note that convergence rate in H^1-norm is controlled by the convergence rate of the seminorm $\left|\cdot\right|_{H^1}$ (*i.e.* highest order derivative in $\left\|\cdot\right\|_{H^1}$). This property holds universally for all operators and integral forms as long as they are variationally consistent.

6) When examining $\left\|E\right\|_{L_2}$, if the highest order of derivative in E is $2m$, then we have

$$\left\|E\right\|_{L_2}\simeq\left\|\phi-\phi_h\right\|_{H^{2m}}\le C_4h^{p+1-2m}\left|\phi\right|_{p+1} \tag{98}$$

5.5. Convergence Rates

In this section, we present details of the convergence rates of various error norms for finite element solutions obtained using GM/WF for self adjoint operators when $B(\cdot,\cdot)$ is symmetric and LSP for all three classes of differential operators. We recall that when the integral form has best approximation property in some norm, hence is variationally consistent, we have the following a priori error estimate (derived for 1D BVP, Equation (93)) in the asymptotic range:

$$\left\|e\right\|_{H^q}=\left\|\phi(x)-\phi_h(x)\right\|_{H^q}\le\left(C_3\left|\phi\right|_{p+1}\right)h^{p+1-q} \tag{99}$$

Taking log of both sides

$$\log\left\|e\right\|_{H^q}\le\log\left(C_3\left|\phi\right|_{p+1}\right)+\left(p+1-q\right)\log h \tag{100}$$

or

$$y\le C+mx \tag{101}$$

in which

$$\begin{aligned} y &= \log\left(\|e\|_{H^q}\right) \\ C &= \log\left(C_3 |\phi|_{p+1}\right) \\ m &= p+1-q \\ x &= \log h \end{aligned} \tag{102}$$

We note that (101) is the equation of a straight line (when we use equality) in *xy*-space in which *m* is the slope and *C* is the *y*-intercept. That is, if we plot $\log h$ versus $\log\left(\|e\|_{H^q}\right)$ on an *xy*-plot, then we obtain a straight line whose slope is $(p+1-q)$ and intercept is $\log\left(C_3 |\phi|_{p+1}\right)$. Slope $(p+1-q)$ is called the rate of convergence of $\|e\|_{H^q}$. Higher values of $(p+1-q)$ imply faster convergence of ϕ_h to ϕ measured in $\|e\|_{H^q}$. Equation (101) can be expressed in terms of total degrees of freedom which is perhaps more appealing in applications as dofs are more easily accessible than characteristic length or size "*h*" of the discretization $\bar{\Omega}^T$. As the discretization $\bar{\Omega}^T$ is refined, the characteristic length *h* reduces and the total dofs increase, thus dofs are inversely proportional to *h*,

$$h \propto \frac{1}{dofs}, \quad h = O\left(\frac{1}{dofs}\right) \tag{103}$$

Using $h = \dfrac{1}{dofs}$ in (100) and since $\log(1) = 0$ we obtain

$$\log\|e\|_{H^q} \le \log\left(C_3 |\phi|_{p+1}\right) - (p+1-q)\log(dofs) \tag{104}$$

We keep in mind that dofs in (104) are purely due to uniform mesh refinement. Thus, in order to determine convergence rate of $\|e\|_{H^q}$ for finite element processes with VC integral forms we need to plot $\log\|e\|_{H^q}$ versus $\log(dofs)$ and determine the slope of this curve $(p+1-q)$, which is the convergence rate in the asymptotic range. For a sequence of fixed discretizations, as *p* increases convergence rate increases linearly.

Remarks

I) We note that $\|e\|_{H^q}$ requires knowledge of theoretical solution ϕ, which may not be possible to determine for a practical application.

II) When the approximation space $V_h \subset H^{k,p}\left(\bar{\Omega}^e\right)$ is minimally conforming or of higher order (*i.e.* $k \ge 2m+1$ for integrals over $\bar{\Omega}^T$ to be Riemann or $k = 2m$ if the Lebesgue integrals over $\bar{\Omega}^T$ are acceptable), then $\sqrt{I} = \sqrt{(E,E)_{\bar{\Omega}^T}} = \|E\|_{L_2}$ in which the residual function can be computed using $E = A\phi_h - f$ over $\bar{\Omega}^T$ and $E^e = A\phi_h^e - f$ over $\bar{\Omega}^e$.

$$\|E\|_{L_2} \le C_4 h^{p+1-2m} |\phi|_{p+1} \tag{105}$$

using $h = \dfrac{1}{dofs}$ and taking log of both sides

$$\log\|E\|_{L_2} \le \log\left(C_4 |\phi|_{p+1}\right) - (p+1-2m)\log(dofs) \tag{106}$$

The dofs in (106) are also due to uniform *h*-refinement. Since $\|E\|_{L_2}$ does not require theoretical solution ϕ, it can be computed using $\phi_h = \bigcup_e \phi_h^e$. Equation (106) can be used for any application without the knowledge of theoretical solution as long as the approximation space is minimally conforming or of higher order than minimally conforming.

5.6. Proposition and Proof

Proposition 5.3. When local approximation ϕ_h^e is of progressively higher order global differentiability, that is, in $V_h \subset H^{k,p}\left(\bar{\Omega}^e\right)$ scalar product spaces for progressively increasing *k*, the accuracy of the finite element solution progressively improves. In this proposition we answer two important questions:

1) Dependence of the a priori error estimates derived so far for local approximations of class C^0 on the order of the space k; that is, if the local approximations are in $V_h \subset H^{k,p}\left(\bar{\Omega}^e\right)$ space how do the a priori estimates change and the influence of k on convergence rate.

2) The influence of the order k of the approximation space on the accuracy of the finite element computation.

Of course (1) and (2) are interdependent because when we have determined (1), the assessment of accuracy may be inferred from it.

The following a priori error estimate derived for 1D BVPs using C^0 p-version local approximation can be extended when the local approximations are in $V_h \subset H^{k,p}\left(\bar{\Omega}^e\right)$ spaces using the following two important considerations or properties of local approximations in $V_h \subset H^{k,p}\left(\bar{\Omega}^e\right)$ spaces:

$$\left\|\phi(x)-\phi_h(x)\right\|_{H^q} = \left\|\phi(x)-\phi_h(x)\right\|_q \le C_3 h^{p+1-q}\left|\phi\right|_{p+1} \tag{107}$$

Property I

We consider a simple illustration of a 1D discretization using three node p-version hierarchical local approximation finite elements in $V_h \subset H^{k,p}\left(\bar{\Omega}^e\right)$ space. Let m be the number of elements in the discretization, then the total degrees of freedom (dofs) are given by

$$dofs = (m+1)k + m(p-2k+1) \tag{108}$$

p is the degree of local approximation (assumed same for all elements of the discretization). Let us choose a p-level, say nine (9) and a one hundred (100) element discretization, then using (108) we can determine total degrees of freedom for $k = 1, 2, \cdots, 5$ corresponding to the local approximations of class C^0, C^1, ···, C^4.

From **Table 1**, we observe that as k increases (*i.e.* progressively higher order local approximations) the total degrees of freedom are progressively reduced. This is a significant property of the higher order local approximations. From **Table 1**, we note that for C^0, 901 dofs are reduced to 802 in the case of C^1 without much effect on accuracy of the solution. The same holds for progressively higher order local approximations C^2, C^3, and so on; that is, the dofs continue to reduce with progressively increasing order of space without much effect on the accuracy. This behavior of the solution accuracy (say in $\|E\|_{L_2}$) holds regardless of the type of differential operator and regardless of the method of approximation used to oconstruct the integral form as long as the integral form is variationally consistent. **Figure 2** shows typical plots of $\|E\|_{L_2}$ versus dofs at $p = 5$ for solutions of classes C^0, C^1, and C^2. Typical points *A*, *B*, *C* correspond to solutions of classes C^0, C^1, and C^2 for the same discretization and p-level (*i.e.* fixed h and p), with almost same value of $\|E\|_{L_2}$ but progressively reducing degrees of freedom. In view of the a priori error estimate (107) we can conclude that if h and p are fixed, then the dependence of the a priori estimate on k lies in C_3, C_4 [*i.e.* $C_3 = C_3(k)$ in (107) and $C_4 = C_4(k)$ in (98)].

Property II

If we choose $\phi_h \in V_h \subset H^{k,p}\left(\bar{\Omega}^e\right)$ and if ϕ_I is the interpolant that agrees with ϕ at the inter-element nodes of the discretization; that is, $\phi(x_i), \dfrac{\partial^j \phi(x_i)}{\partial x^j}, j = 1, 2, \cdots$ agree with $\phi_I(x_i), \dfrac{\partial^j \phi_I(x_i)}{\partial x^j}; j = 1, 2, \cdots$ corresponding to local approximations of classes $C^1, C^2, \cdots$ respectively, then in the consideration of the a priori error estimates we only need to consider (x_i, x_{i+1}) (*i.e.* interior of the element). This suggests that in a priori estimate in (98) (for example) only C_4 depends on k. That is, (98) holds when $\phi_h \in V_h \subset H^{k,p}$ except that $C_4 = C_4(k)$.

Table 1. Total dofs for a 100 element discretization at p = 9 for different values of the order of space k.

Type of local approximation	dofs
$C^0;\ k = 1$	901
$C^1;\ k = 2$	802
$C^2;\ k = 3$	703
$C^3;\ k = 4$	604
$C^4;\ k = 5$	505

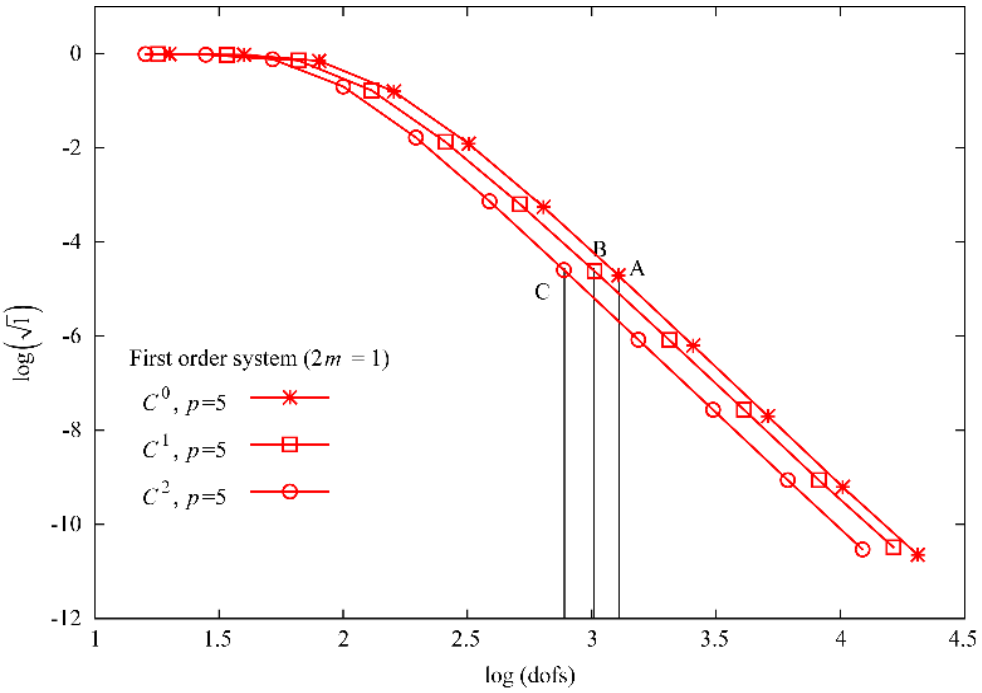

Figure 2. Typical $\|E\|_{L_2}$ versus dofs behavior for k = 1, 2, and 3.

Remarks

1) From properties I and II it is clear that when $\phi_h \in V_h \subset H^{k,p}\left(\bar{\Omega}^e\right)$, in the error estimate (98) the terms $h^{p+1-2m}|\phi|_{p+1}$ and likewise the terms $h^{p+1-q}|\phi|_{p+1}$ in (107) remain unaffected. Only the coefficients C_4 and C_3 show mild dependence on k.

2) In view or properties I and II, we conclude that if $C^0, C^1, \cdots$ solutions of a BVP are to be computed for a fixed number of degrees of freedom, then progressively more degrees of freedom can be added to solutions of class $C^1, C^2, \cdots$ so that the total dofs in all classes of solutions are the same. We recall that with same values of h and p in the solutions of class $C^0, C^1, C^2, \cdots, \|E\|_{L_2}$ (**Figure 2**) remains virtually the same for all classes, however the total dofs are progressively reduced.

The consequence of adding more dofs (through h-refinement) with progressively increasing order of space so that in each case the dofs match with C^0 solutions is clearly improved accuracy of ϕ_h reflected by progressively reducing $\|E\|_{L_2}$. Clearly, in doing so the convergence rate $(p+1-q)$ or $(p+1-2m)$ is not affected. Thus, $\log\|E\|_{L_2}$ versus log(dofs) graphs for solutions of classes $C^0, C^1, \cdots$ in the asymptotic range are parallel to each other but with progressively lower values of $\|E\|_{L_2}$ as shown in **Figure 2**. That is graph for C^1 is below C^0 and that of C^2 is below C^1 and so on, but they are all parallel.

5.7. General Remarks

1) The a priori error estimates are presented for one dimensional boundary value problems. Their extensions to 2D and 3D require more elaborate derivations (see references) and new definitions of h and $|\phi|_{p+1}$, but the convergence rates remain the same as $(p+1-q)$ or $(p+1-2m)$ derived for 1D BVPs.

2) We remark again that the rates only hold in the asymptotic range.

3) The integral forms must be VC so that the best approximation property of ϕ_h holds in some norm in order for these estimates to remain valid. The estimates derived here hold for: (a) GM/WF for self adjoint operators when the bilinear functional $B(\cdot,\cdot)$ is symmetric and (b) for LSP based on residual functional for all three classes of differential operator.

4) In case of GM/WF for non-self adjoint and non-linear operators, the a priori estimates derived here do not hold. In case of such operators the functional $B(\cdot,\cdot)$ generally consists of a symmetric part and a non-symmetric part. With sufficient mesh refinement if we can ensure that the behavior is dominated by the symmetric part,

then the estimates derived here hold in the range of calculations when asymptotic range is realized. We illustrate this aspect through model problems presented in a later section.

6. Computations of a Priori Error Estimates and Convergence Rates

In this section, we present numerical studies related to the computation of a priori error estimates and convergence rates for BVPs described by self adjoint, non-self adjoint, and non-linear differential operators in which VC integral forms are constructed using GM/WF for BVP described by self adjoint differential operators and using LSP for BVPs described by all three classes of differential operators.

6.1. Model Problem 1: Self-Adjoint Operator, 1D Diffusion Equation

We consider the 1D steady-state diffusion equation.

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(a\frac{\mathrm{d}\phi}{\mathrm{d}x}\right)=q(x),\quad \forall x\in(0,L)=\Omega\subset\mathbb{R}^1 \tag{109}$$

$$BC:\quad \phi(0)=0,\quad a\frac{\mathrm{d}\phi}{\mathrm{d}x}\bigg|_{x=L}=0 \tag{110}$$

If we choose $a=1$, $L=1$, $q(x)=x^n$, $n=6$, then the theoretical solution ϕ or ϕ_t is given by

$$\phi_t(x)=\phi(x)=\left(-\frac{1}{a(x+1)(x+2)}\right)x^{n+2}+x\left(\frac{1}{a(x+1)}\right)L^{n+1} \tag{111}$$

a) GM/WF: The differential operator $A=-\frac{\mathrm{d}}{\mathrm{d}x}\left(a\frac{\mathrm{d}}{\mathrm{d}x}\right)$ is linear and $A^*=A$. The integral form using GM/WF is given by (over $\bar{\Omega}=[0,1]$)

$$\left(\frac{\mathrm{d}\phi}{\mathrm{d}x},\frac{\mathrm{d}v}{\mathrm{d}x}\right)_{\bar{\Omega}}=\left(q(x),v\right)_{\bar{\Omega}},\quad v=\delta\phi,\quad \forall v\in V_h\subset H^{k,p} \tag{112}$$

or

$$B(\phi,v)=l(v) \tag{113}$$

$B(\cdot,\cdot)$ is bilinear and symmetric and $l(\cdot)$ is linear. The integral form (weak form) is VC due to the fact that $\delta(B(\phi,v))=\left(\frac{\mathrm{d}v}{\mathrm{d}x},\frac{\mathrm{d}v}{\mathrm{d}x}\right)_{\bar{\Omega}}>0,\ \forall v\in V_h$ hence a solution ϕ from (113) minimizes $I(\phi)=\frac{1}{2}B(\phi,\phi)-l(\phi)$.

b) LSP based on residual functional

I) LSP using higher order system (without auxiliary equation)

Using (109), referred to as the higher order differential equation or system, if we let ϕ_h be approximation of ϕ over $\bar{\Omega}^T$ then

$$I(\phi_h)=(E,E)_{\bar{\Omega}^T};\quad E=A\phi_h-f=\frac{\mathrm{d}^2\phi_h}{\mathrm{d}x^2}+q(x),\quad \forall x\in[0,1] \tag{114}$$

$$\delta I(\phi_h)=2(E,\delta E)=0\Rightarrow(A\phi_h,Av)_{\bar{\Omega}^T}+\left(q(x),Av\right)_{\bar{\Omega}^T}=0 \tag{115}$$

or

$$\left(\frac{\mathrm{d}^2\phi_h}{\mathrm{d}x^2},\frac{\mathrm{d}^2v}{\mathrm{d}x^2}\right)_{\bar{\Omega}^T}+\left(q(x),\frac{\mathrm{d}^2v}{\mathrm{d}x^2}\right)_{\bar{\Omega}^T}=0 \tag{116}$$

or

$$B(\phi_h,v)-l(v)=0 \tag{117}$$

$$\delta^2 I(\phi_h) = B(v,v) = \left(\frac{\mathrm{d}^2 v}{\mathrm{d}x^2}, \frac{\mathrm{d}^2 v}{\mathrm{d}x^2}\right)_{\bar{\Omega}^T} > 0, \quad \forall v \in V_h \subset H^{k,p} \tag{118}$$

Hence, the integral form (117) is variationally consistent.

II) LSP using first order system

Let $\tau = \dfrac{\mathrm{d}\phi}{\mathrm{d}x}$, hence (109) can be written as a system of two first order equations.

$$\begin{aligned} &\frac{\mathrm{d}\tau}{\mathrm{d}x} + q(x) = 0 \\ &\tau - \frac{\mathrm{d}\phi}{\mathrm{d}x} = 0 \end{aligned} \tag{119}$$

LSP for (119) follows standard procedure. Let ϕ_h and $\tau_h \in V_h \subset H^{k,p}$ be approximations of ϕ and τ, then

$$I = \sum_{i=1}^{2} (E_i, E_i)_{\bar{\Omega}^T}, \quad E_1 = \frac{\mathrm{d}\tau_h}{\mathrm{d}x} + q(x), \quad E_2 = \tau_h - \frac{\mathrm{d}\phi_h}{\mathrm{d}x} = 0 \tag{120}$$

$$\delta I = 2\sum_{i=1}^{2} (E_i, \delta E_i)_{\bar{\Omega}^T} = 0 \tag{121}$$

$$\delta^2 I = \sum_{i=1}^{2} (\delta E_i, \delta E_i) > 0 \tag{122}$$

Hence the integral form (121) resulting from LSP is variationally consistent.

Remarks

I) All other methods of approximation yield VIC integral forms, hence are not considered as in such cases the a priori error estimates and the convergence rates are not valid.

II) In the numerical studies, we consider GM/WF and LSP for higher order as well as first order system of differential equations describing BVPs.

6.1.1. GM/WF

In this section, we present numerical studies for the integral form (112) for $\bar{\Omega}^T = \bigcup_e \bar{\Omega}^e$, discretization of $\bar{\Omega} = [0,1]$. We consider uniform discretizations employing three node *p*-version hierarchical 1D elements with local approximations in scalar product space $V_h \subset H^{k,p}(\bar{\Omega})$. We begin with two element uniform discretization and perform uniform mesh refinement containing 4, 8, 16, ...elements. Since in this model problem the theoretical solution ϕ is known, various error norms can be computed. We note from the description of the BVP (109) that in this case $2m = 2$ (highest order of the derivative in the BVP) and the integral form resulting from GM/WF contains only up to first order derivatives of the dependent variable and the test function. We consider computations using solutions of class C^0, C^1, and C^2 at different *p*-levels with uniform mesh refinements. Computed results for solution of class C^0 are shown in **Figure 3**. The integral form is VC and ϕ_h, the computed solution, has best approximation property in $B(\cdot,\cdot)$-norm.

In this case, the following a priori error estimates hold (Proposition 5.2):

$$\left.\begin{aligned} &\|\phi - \phi_h\|_{H^q} \le C_3 h^{p+1-q} |\phi|_{p+1} \\ &|\phi - \phi_h|_{H^q} \le C_2 h^{p+1-q} |\phi|_{p+1} \\ &\|A\phi - f\|_{L_2} = \|E\|_{L_2} \le C_4 h^{p+1-2m} \end{aligned}\right\} \tag{123}$$

For this BVP, $2m = 2$ and q depends on the type of norm. **Figure 3** also shows the theoretical values of the convergence rates of various error norms for solutions of class C^0 at *p*-levels of 2 and 5. Graphs of the log of error norms versus log of dofs for these solutions are shown in **Figure 3**. We note that due to smoothness of the theoretical solution even the two element discretization yields the error norms in the asymptotic range; that is,

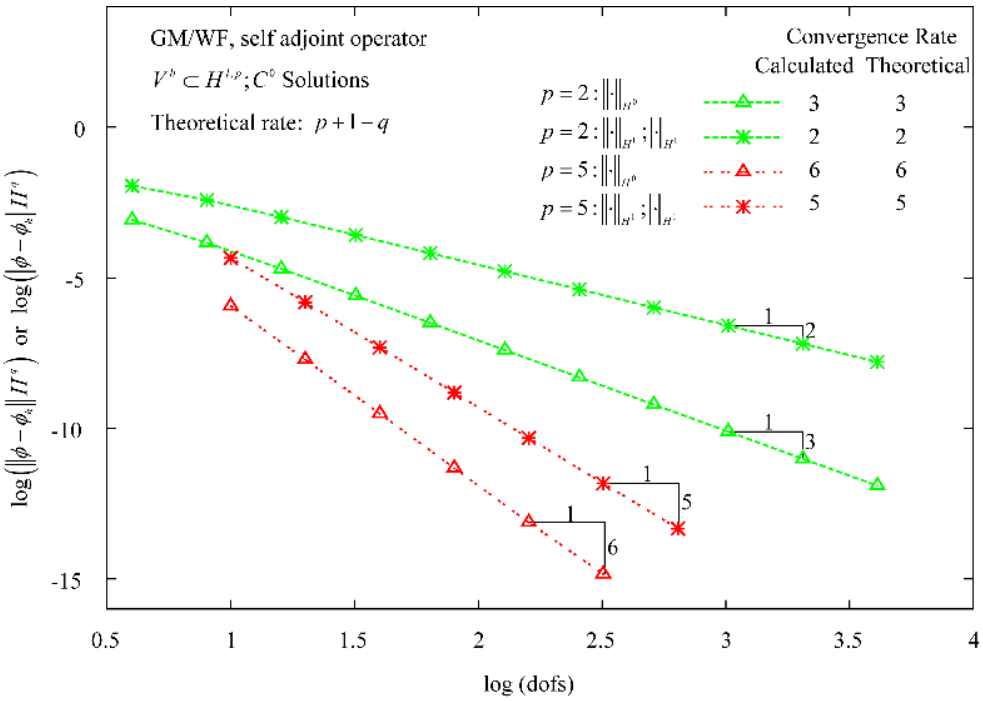

Figure 3. $\log\|\cdot\|_{H^q}$ versus log(dofs) for solutions of class C^0 (GM/WF, model problem 1, $p=2$ and 5).

pre-asymptotic and onset of asymptotic ranges in these solutions do not appear in **Figure 3**. All computations are in the asymptotic range, hence onset of post-asymptotic and post-asymptotic ranges are also absent. Calculated convergence rates are in perfect agreement with the theoretical convergence rates calculated using (123). We note that in $\|\cdot\|_{H^1}$ and $|\cdot|_{H^1}$ error norms the integrals over $\bar{\Omega}^T$ are Lebesgue, but the norms are well-behaved due to smoothness of ϕ.

Figure 4 shows plots of log of various norms and seminorms versus log of degrees of freedom at $p=3$ and 5 for C^1 solutions. The computed convergence rates of various error norms and comparison with the theoretical convergence rates obtained using (123) are shown in **Figure 4**. The agreement is perfect. Here we note that in computing $\|\cdot\|_{H^2}$ and $|\cdot|_{H^2}$, the integrals over $\bar{\Omega}^T$ are Lebesgue but error norms are well-behaved due to smoothness of ϕ. Also, nearly all computations shown in **Figure 4** are in the asymptotic range, except for the last point for $\|\cdot\|_{H^0}$ at $p=3$ and $\|\cdot\|_{H^1}$ at $p=5$.

Log of various error norms and seminorms versus log of degrees of freedom for solutions of class C^2 at $p=5$ and 7 are shown in **Figure 5**. Since $k=3$ for $V_h \subset H^{k,p}\left(\bar{\Omega}^e\right)$, all integrals in all error norms are Riemann over the discretization $\bar{\Omega}^T$. Computed error norms using (123) and comparison with the computed convergence rates of error norm are also shown in **Figure 5**. We observe perfect match between the theoretical values and the computed values. Except for the last point shown in **Figure 5** for $\|\cdot\|_{H^1}$ at $p=5$, all other computed results are in the asymptotic range due to smoothness of ϕ.

Figure 6 shows plots of $\log\|\cdot\|_{H^0}$ (or $\|\cdot\|_{L_2}$) versus log of dof for solutions of class C^0, C^1, and C^2 ($k=1,2,3$) at $p=5$. All three graphs of $\log\|\cdot\|_{H^0}$ versus log of dofs for $k=1,2,3$ are parallel, confirming that the convergence rate of $\|\cdot\|_{H^0}$ is independent of the order k of the approximation space. We note that graph for C^1 appears below C^0 and the graph for C^2 is below C^1 confirming that for given dofs, as the order k of space is increased, the error in the computed solution ϕ_h (measured in H^0-norm) decreases without affecting the convergence rate.

The BVP in this model problem is described by a second-order differential operator ($2m=2$); hence, $k=3$ corresponds to minimally conforming space $V_h \subset H^{k,p}\left(\bar{\Omega}^e\right)$ for which the integrals are always Riemann. However, due to smoothness of ϕ, when $k=2$ (solutions of class C^1) in which case the integrals over $\bar{\Omega}^T$ are Lebesgue, the solution ϕ_h is expected to converge weakly to class C^2. Next we consider solutions of class

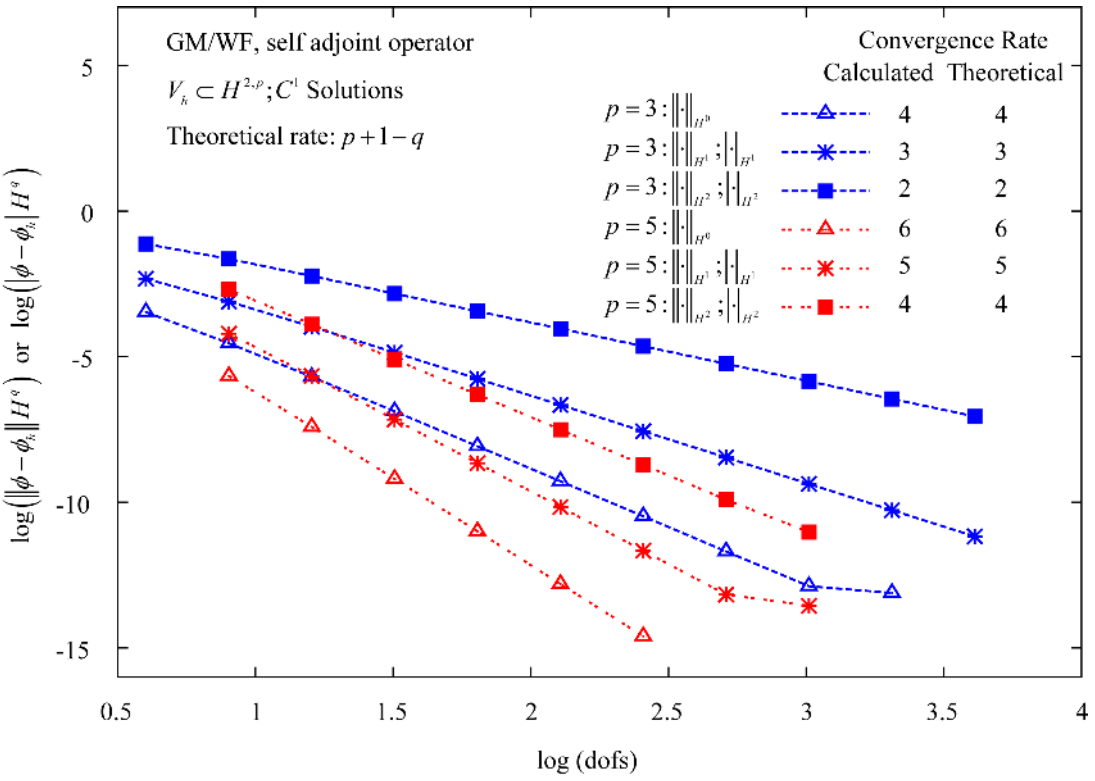

Figure 4. $\log\|\cdot\|_{H^q}$ versus log(dofs) for solutions of class C^1 (GM/WF, model problem 1, $p=3$ and 5).

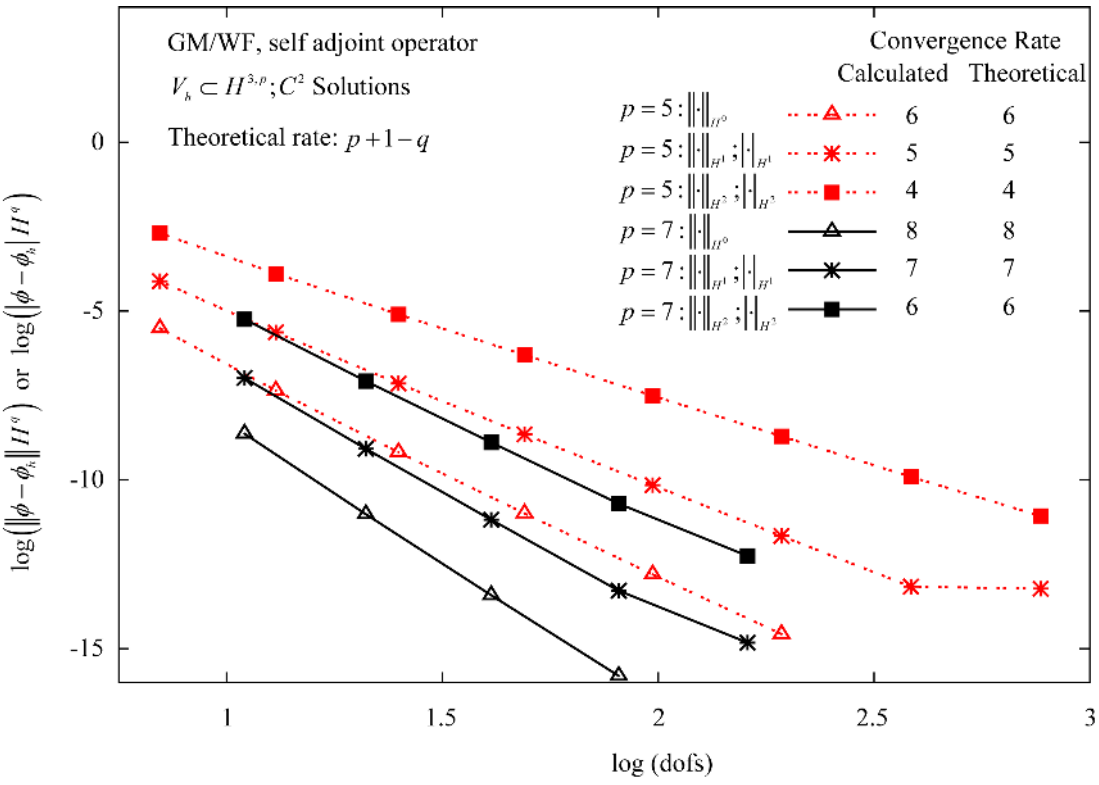

Figure 5. $\log\|\cdot\|_{H^q}$ versus log(dofs) for solutions of class C^2 (GM/WF, model problem 1, $p=5$ and 7).

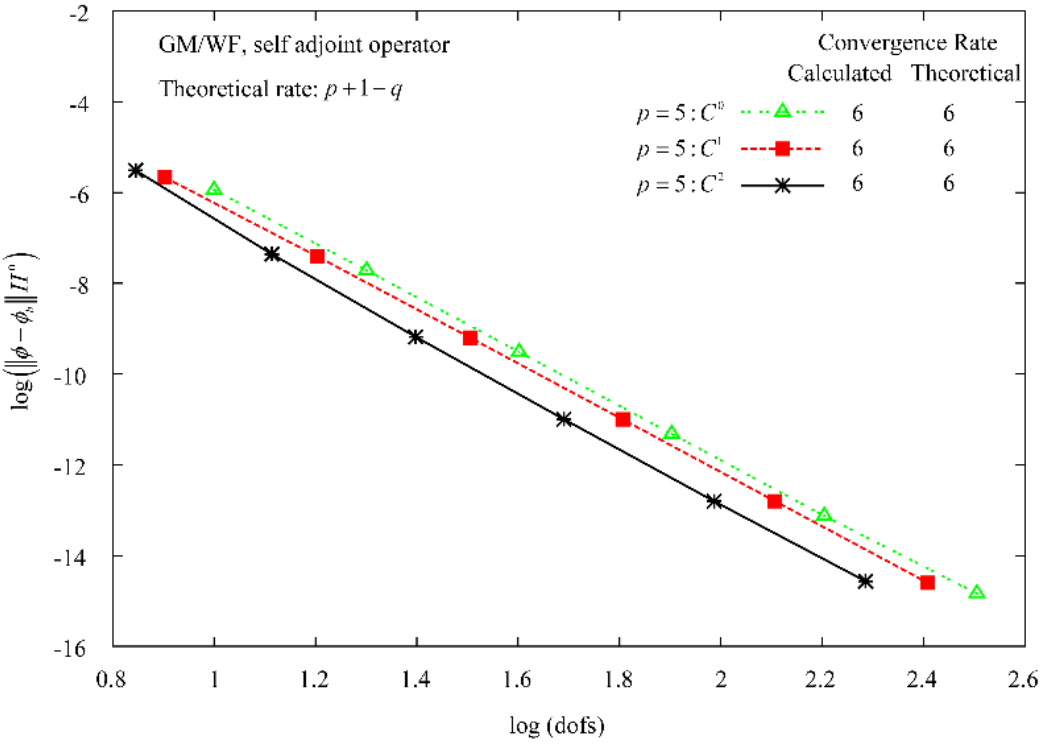

Figure 6. $\log\|\cdot\|_{H^0}$ versus log(dofs) for solutions of classes C^0, C^1, and C^2 at $p=5$ (GM/WF, model problem 1).

C^2 with $p=5$ (minimum). Numerical solutions are computed for uniform mesh refinements beginning with a two-element uniform discretization. For each discretization we calculate $\|\cdot\|_{H^2}$ and $\|E\|_{L_2}=\|A\phi_h-f\|_{L_2}$. Since the rate of convergence of $\|\cdot\|_{H^2}$ is controlled by $|\cdot|_{H^2}$, we expect the convergence rates of $\|\cdot\|_{H^2}$ and $\|E\|_{L_2}$ to be nearly same. We clearly see this in **Figure 7**. Graphs for C^1 and C^2 are parallel, confirming the same convergence rates of $\|\cdot\|_{H^2}$ (or $\|E\|_{L_2}$) for solutions of class C^1 ($k=2$) and class C^2 ($k=3$). The convergence rate in the case of $\|\cdot\|_{H^2}$ is $p+1-q=5+1-2=4$, whereas in the case of $\|E\|_{L_2}$ is $p+1-2m=5+1-2=4$. Plots in **Figure 7** confirm that rate of convergence of error norms is independent of the order of the approximation space.

6.1.2. LSP, Higher-Order System (No Auxiliary Equation)

In this study, we consider finite element formulation of model problem (109) using least-squares process based on residual functional. We consider solutions of class C^1 as well as C^2. In case of C^1 solutions integrals over $\bar{\Omega}^T$ are Lebesgue whereas for solutions of class C^2 the integrals are Riemann. **Figure 8** shows plots of $\log\|\cdot\|_{H^q}$ and $\log|\cdot|_{H^q}$ versus log of dofs for p-levels of 3 and 5 calculated using uniform mesh refinement. Calculated convergence rates of various error norms are also shown in **Figure 8**. The theoretical convergence rates of various error norms and a comparison with calculated convergence rates is also shown in **Figure 8**.

Agreement between theoretical and calculated values is excellent. Here also we observe absence of pre-asymptotic and onset of asymptotic ranges due to smoothness of the theoretical solution. Some graphs for significant refinement show appearance of post-asymptotic (or onset of post-asymptotic) range.

Similar studies for solutions of class C^2 are shown in **Figure 9** for $\|\cdot\|_{H^0}$; $\|\cdot\|_{H^1}$, $|\cdot|_{H^1}$; and $\|\cdot\|_{H^2}$, $|\cdot|_{H^1}$ norms at p-levels of 5 and 7. The computed convergence rates of the error norms are in perfect agreement with theoretical rates calculated using ($p+1-q$), shown in **Figure 9**.

Graphs of $\log\|\cdot\|_{H^2}$ and $\log\|E\|_{L_2}$ (or $\sqrt{I}$) versus log of dofs for solutions of class C^1 and C^2 obtained using uniform mesh refinement are shown in **Figure 10**. Calculated convergence rates are also shown in **Figure 10**. For $\|\cdot\|_{H^2}$ error norm the theoretical rate is ($p+1-q$) whereas for $\|E\|_{L_2}$ it is ($p+1-2m$). The theoretical convergence rates are in perfect agreement with those calculated using graphs in **Figure 10**. We note that

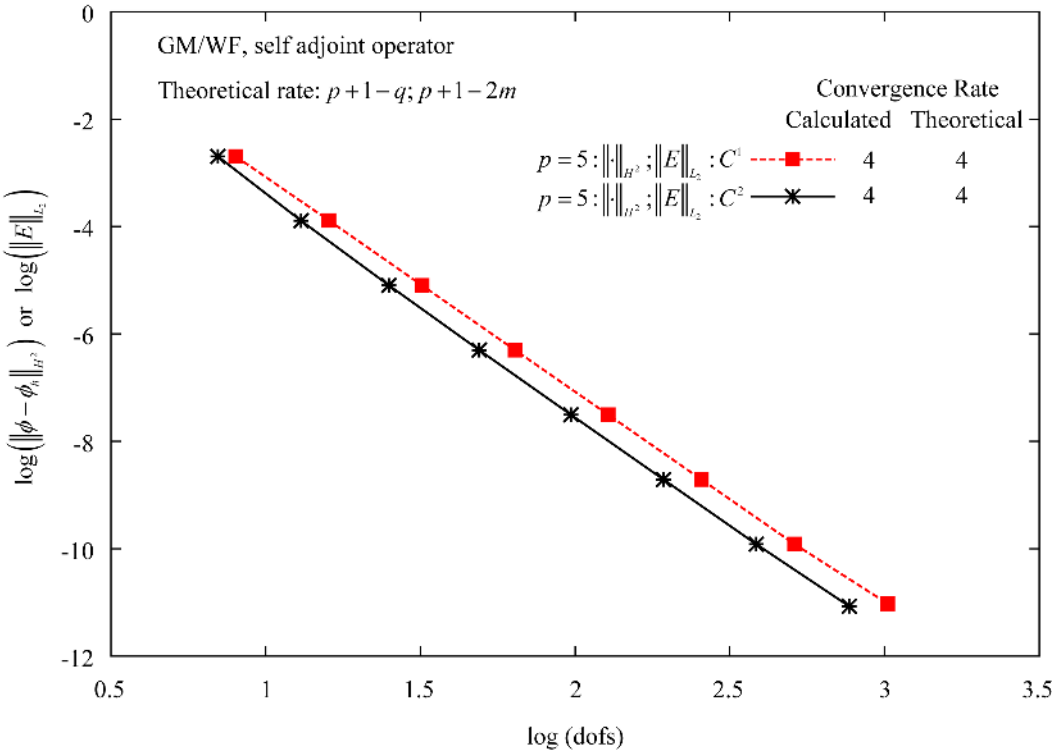

Figure 7. $\log\|\cdot\|_{H^2}$ or $\log\|E\|_{L_2}$ versus log(dofs) for solutions of classes C^1 and C at $p=5$ (GM/WF, model problem 1).

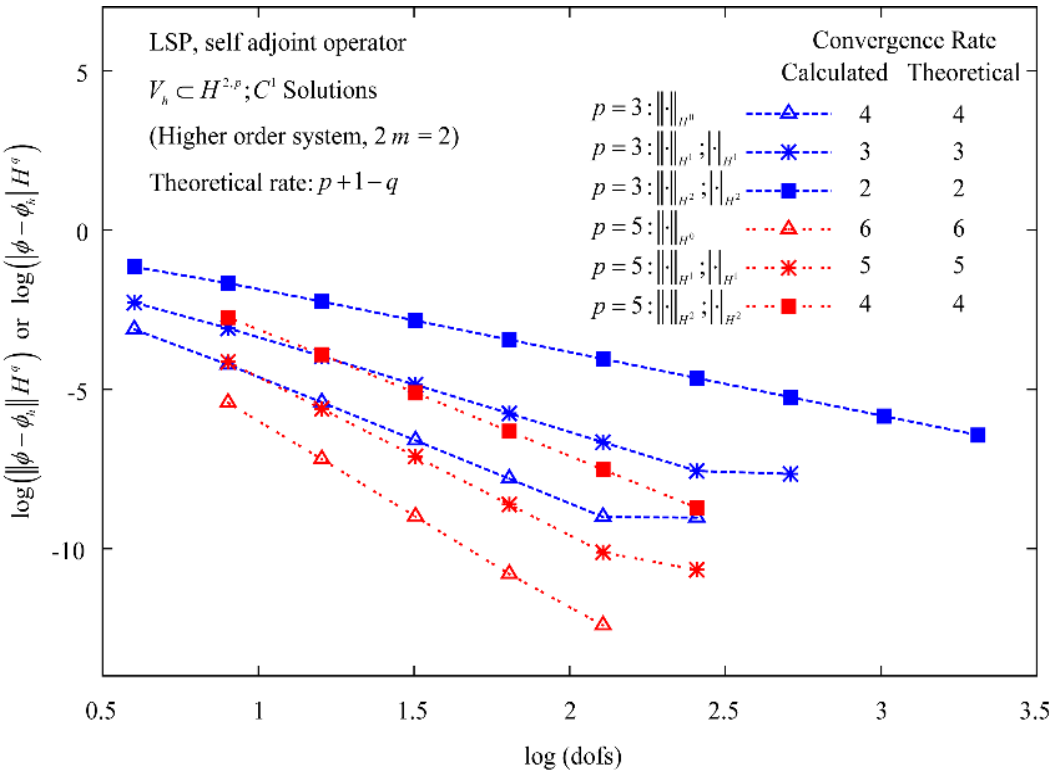

Figure 8. $\log\|\cdot\|_{H^q}$ versus log(dofs) for solutions of class C^1 (LSP, model problem 1, $p=3$ and 5).

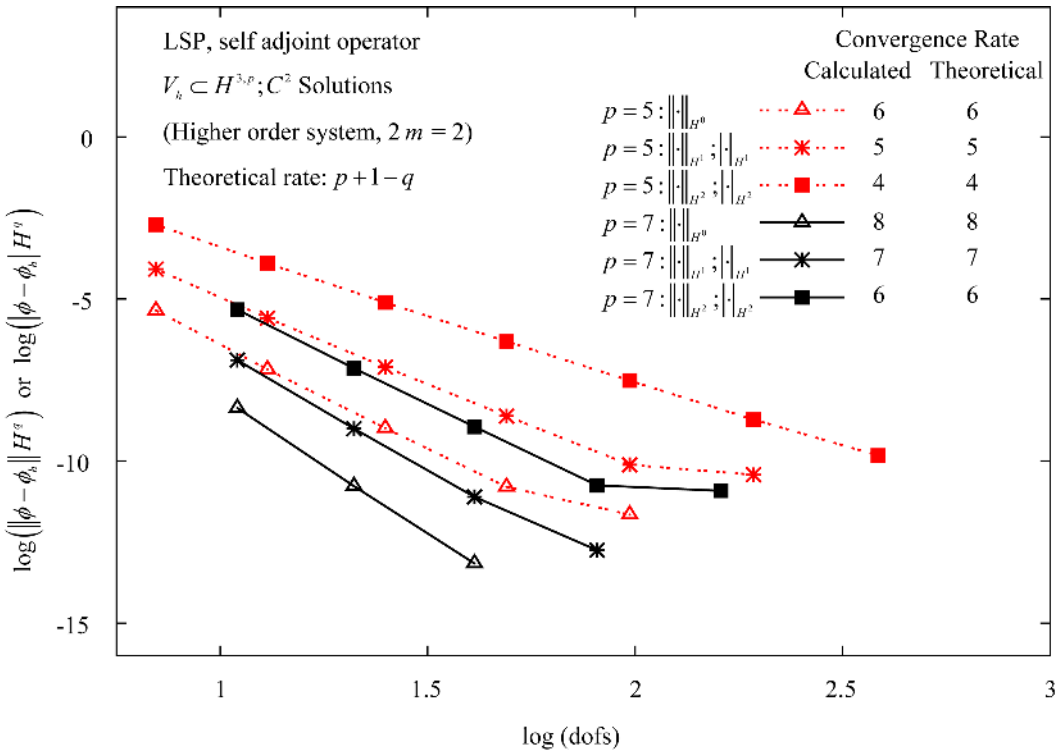

Figure 9. $\log\|\cdot\|_{H^q}$ versus log(dofs) for solutions of class C (LSP, model problem 1, p = 5 and 7).

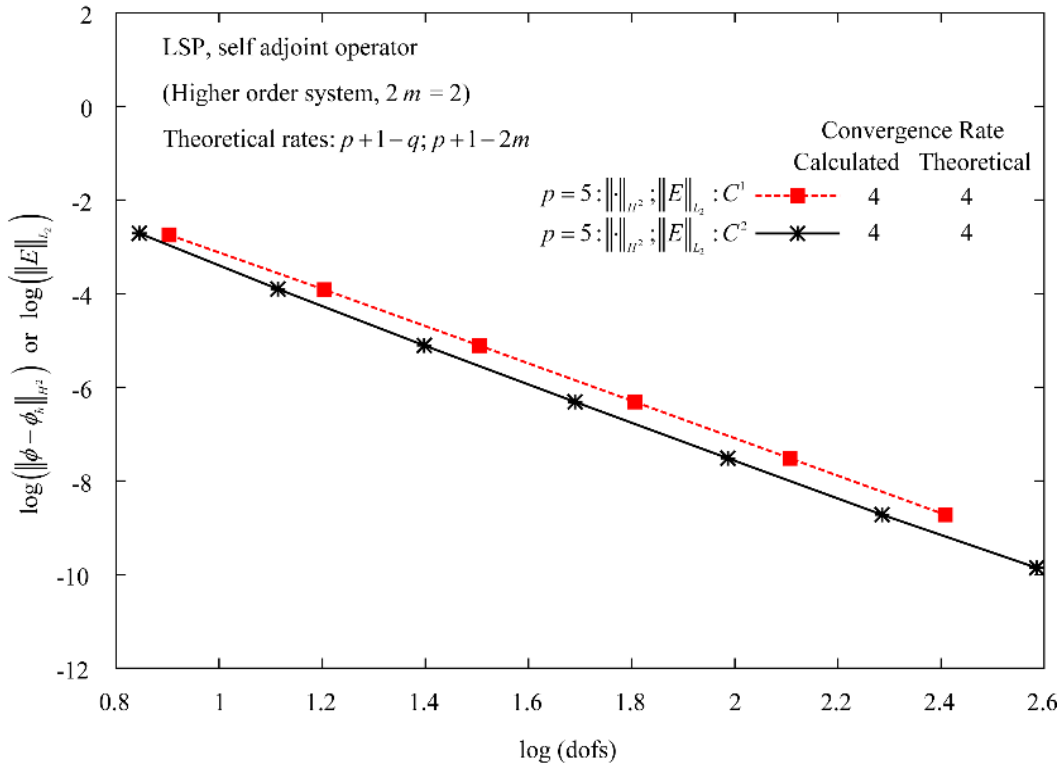

Figure 10. $\log\|\cdot\|_{H^2}$ or $\log\|E\|_{L_2}$ versus log(dofs) for solutions of classes C^1 and C^2 at p = 5 (LSP, model problem 1).

C^1 and C^2 graphs for same *p*-level (5) are parallel to each other and C^2 graph is below C^1, confirming that the convergence rates for $k=2$ and $k=3$ are same (*i.e.* independent of *k*), the order of space, but for $k=3$ the solution has better accuracy compared to $k=2$.

Remarks. Numerical studies for LSP using auxiliary equation (*i.e.* a first-order system) are not presented for this model problem but will be presented for the next model problem, 1D convection-diffusion equation.

6.2. Model Problem 2: Non-Self-Adjoint Operator, 1D Convection-Diffusion Equation

We consider 1D convection-diffusion equation described by non-self adjoint operator for computing a priori error estimates and convergence rates and compare them with their theoretical values,

$$\frac{\mathrm{d}\phi}{\mathrm{d}x}-\frac{1}{Pe}\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}=f(x);\quad \forall x\in(0,1)=\Omega \tag{124}$$

$$\text{BCs:}\quad \phi(0)=1,\quad \phi(1)=0 \tag{125}$$

We consider $f(x)=0$. Theoretical solution of (124)-(125), finite element solution using GM/WF and LSP using higher order system (no auxiliary variables) and using first order system (using auxiliary variables) is given in [35]. For this BVP, the operator $A=\frac{\mathrm{d}}{\mathrm{d}x}-\frac{1}{Pe}\frac{\mathrm{d}^2}{\mathrm{d}x^2}$ is linear but $A^*=-\frac{\mathrm{d}}{\mathrm{d}x}-\frac{1}{Pe}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\neq A$, hence the integral form from GM/WF is VIC, but LSP for higher order as well as first order system of differential equations is VC.

a) GM/WF: The integral form of (124)-(125) is given by (for $f(x)=0$)

$$\left(\frac{\mathrm{d}\phi}{\mathrm{d}x},v\right)_{\bar{\Omega}}+\frac{1}{Pe}\left(\frac{\mathrm{d}\phi}{\mathrm{d}x},\frac{\mathrm{d}v}{\mathrm{d}x}\right)_{\bar{\Omega}}=0,\quad v=\delta\phi,\ \forall v\in V\subset H^{k,p}(\bar{\Omega}) \tag{126}$$

$$B(\phi,v)=l(v);\quad l(v)=0 \tag{127}$$

$B(\phi,v)$ is bilinear but not symmetric and

$$\delta B(\phi,v)=\left(\frac{\mathrm{d}v}{\mathrm{d}x},v\right)_{\bar{\Omega}}+\frac{1}{Pe}\left(\frac{\mathrm{d}v}{\mathrm{d}x},\frac{\mathrm{d}v}{\mathrm{d}x}\right)_{\bar{\Omega}} \tag{128}$$

does not yield a unique extremum principle. Hence, the integral form (127) is VIC.

b) LSP based on residual functional:

I) Higher order system (without auxiliary equation)

In this case we use (124) without introducing auxiliary equation, that is without reducing (124) into a first order system of equations. Let ϕ_h be approximation of ϕ over $\bar{\Omega}^T$, then

$$I(\phi_h)=(E,E)_{\bar{\Omega}^T},\ E=A\phi_h-f=\frac{\mathrm{d}\phi_h}{\mathrm{d}x}-\frac{1}{Pe}\frac{\mathrm{d}^2\phi_h}{\mathrm{d}x^2} \tag{129}$$

$$\delta I(\phi_h)=2(E,\delta E)=0\Rightarrow(A\phi_h,Av)=0 \tag{130}$$

or

$$\left(\frac{\mathrm{d}\phi_h}{\mathrm{d}x}-\frac{1}{Pe}\frac{\mathrm{d}^2\phi_h}{\mathrm{d}x^2},\ \frac{\mathrm{d}v}{\mathrm{d}x}-\frac{1}{Pe}\frac{\mathrm{d}^2v}{\mathrm{d}x^2}\right)_{\bar{\Omega}^T}=0 \tag{131}$$

or

$$B(\phi_h,v)=0 \tag{132}$$

$$\begin{aligned}\delta^2 I(\phi_h)&=B(v,v)=(Av,Av)_{\bar{\Omega}^T}\\&=\left(\frac{\mathrm{d}v}{\mathrm{d}x}-\frac{1}{Pe}\frac{\mathrm{d}^2v}{\mathrm{d}x^2},\frac{\mathrm{d}v}{\mathrm{d}x}-\frac{1}{Pe}\frac{\mathrm{d}^2v}{\mathrm{d}x^2}\right)_{\bar{\Omega}^T}>0,\ \forall v\in V_h\subset H^{k,p}(\bar{\Omega}^T)\end{aligned} \tag{133}$$

Hence, the integral form (132) is VC.

II) First order system

Let $\tau = \dfrac{\mathrm{d}\phi}{\mathrm{d}x}$, hence (124) can be written as

$$\begin{aligned} &\frac{\mathrm{d}\phi}{\mathrm{d}x} - \frac{1}{Pe}\frac{\mathrm{d}\tau}{\mathrm{d}x} = 0 \\ &\tau - \frac{\mathrm{d}\phi}{\mathrm{d}x} = 0 \end{aligned} \tag{134}$$

LSP for (134) follows standard procedure (parallel to Equations (119)-(122)). Details are straightforward. See [35] for many model problems of similar type.

Remarks

1) Since GM/WF yields VIC integral form and does not have best approximation property as the operator A is not self adjoint, hence the a priori error estimates derived in earlier sections using best approximation property in *B*-norm do not hold in this case. Nonetheless we present numerical studies for GM/WF for this model problem to illustrate some important aspects of error norms in a later section.

2) Integral form derived using LSP is VC and has best approximation property in *E*-norm or $\sqrt{I}$, *I* being residual functional, hence the same a priori error estimates derived for LSP for self adjoint operators hold here as well.

6.2.1. LSP: First Order System

Domain $\bar{\Omega} = [0,1]$ is discretized using 3-node *p*-version 1D elements of higher order global differentiability into 2, 4, 6, ...element uniform meshes. The solutions are computed using finite element formulation based on LSP for first order system of equations. Solutions of classes C^0, C^1, and C^2 are considered at different *p*-levels. For this problem the a priori estimates (123) hold as well with $2m = 1$ due to the fact that it is a first order system of equations. LSP has best approximation property in E-norm and the integral form is VC,

$$\left.\begin{aligned} &\left\|\phi - \phi_h\right\|_{H^q} \le C_3 h^{p+1-q} \left|\phi\right|_{p+1} \\ &\left|\phi - \phi_h\right|_{H^q} \le C_2 h^{p+1-q} \left|\phi\right|_{p+1} \\ &\left\|A\phi - f\right\|_{L_2} = \left\|E\right\|_{L_2} \le C_4 h^{p+1-2m} \end{aligned}\right\} \tag{135}$$

First, we consider solutions of class C^0 at $p = 2$ and 5 and with $Pe = 100$. Due to C^0 local approximation and the first order system, integrals over $\bar{\Omega}^T$ are Lebesgue but due to smoothness of ϕ weak convergence of computed ϕ_h to C^1 class is expected. **Figure 11** shows plots of log of various error norms versus log of the dofs at $p = 2$ and 5. Details of the studies are also given in **Figure 11**. Theoretical convergence rates are in perfect agreement with the calculated rates shown in **Figure 11**. As p-level is increased from 2 to 5 convergence rates also show increase by 3 at $p = 5$ compared to those at $p = 2$. We clearly observe pre-asymptotic, onset of asymptotic, and asymptotic ranges in all cases. For $p = 5$ also observe onset of post-asymptotic and post-asymptotic ranges. We note that even though LSP does not have best approximation property in *B*-norm but due to the fact that the integral form is VC, the convergence rate of LSP (135) is same as those of GM/ WF for self adjoint operators (123).

As *p*-level is increased convergence rate increases proportionately. Derivatives converge more slowly than functions, hence convergence rate of $\left\|\cdot\right\|_{H^1}$ is one order lower than that of $\left\|\cdot\right\|_{H^0}$ or $\left\|\cdot\right\|_{L_2}$. Since the convergence rate of $\left\|\cdot\right\|_{H^1}$ is dominated by the first derivative, $\left\|\cdot\right\|_{H^1}$ and $\left|\cdot\right|_{H^1}$ have same convergence rates (also clear from (135)).

Solutions of class C^1 at $p = 3$ and 5 are considered here. Results obtained using uniform mesh refinement are given in **Figure 12**. Plots of log of various error norms versus log of dofs and calculated convergence rates are shown in **Figure 12** and are compared with theoretical convergence rates. Calculated and theoretical convergence rates are in perfect agreement. We note that when $p = 5$ the convergence rates of error norms are independent of *k* (*i.e.* at $p = 5$), solutions of class C^0 and C^1 have same convergence rates for the same norm, confirming that convergence rates of the error norms are not a function of *k*, the order of the approximation space. Pre-asymptotic, onset of asymptotic, and asymptotic ranges are clearly observed in **Figure 12**.

Solutions of class C^2 at $p = 5$ and 7 are considered next. Results obtained using uniform mesh refinement are shown in **Figure 13** and are compared with the theoretical convergence rates obtained using (135). Once

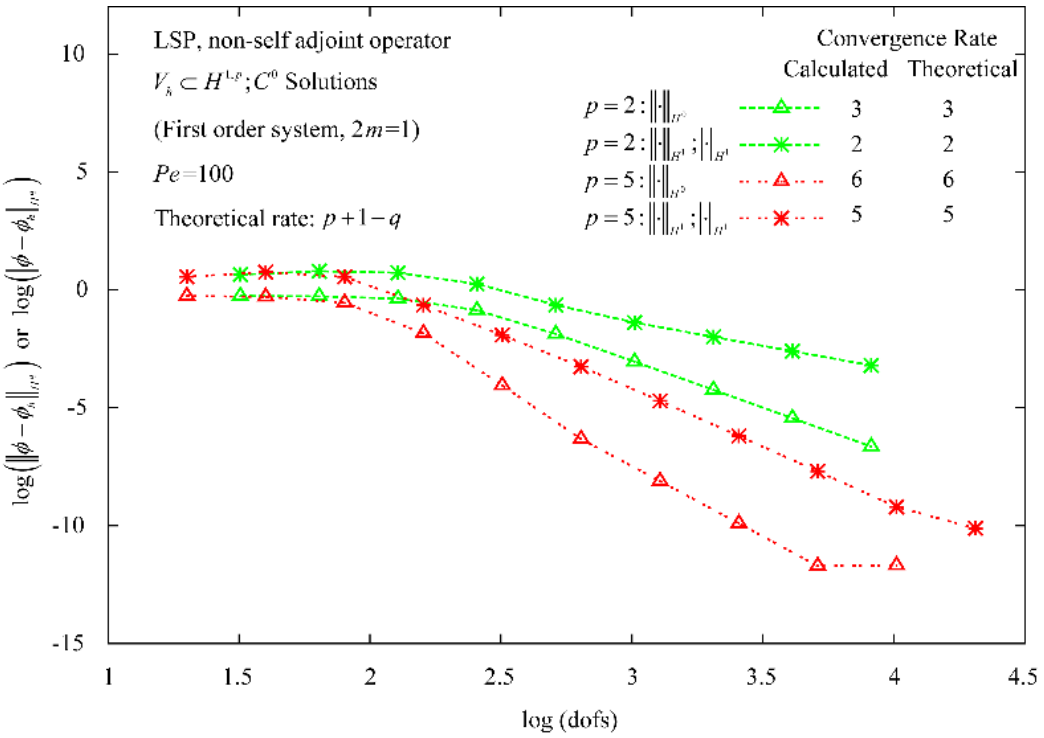

Figure 11. $\log\|\cdot\|_{H^q}$ versus log(dofs) for solutions of class C^0 (LSP, model problem 2, first order system, $p=2$ and 5, $Pe=100$).

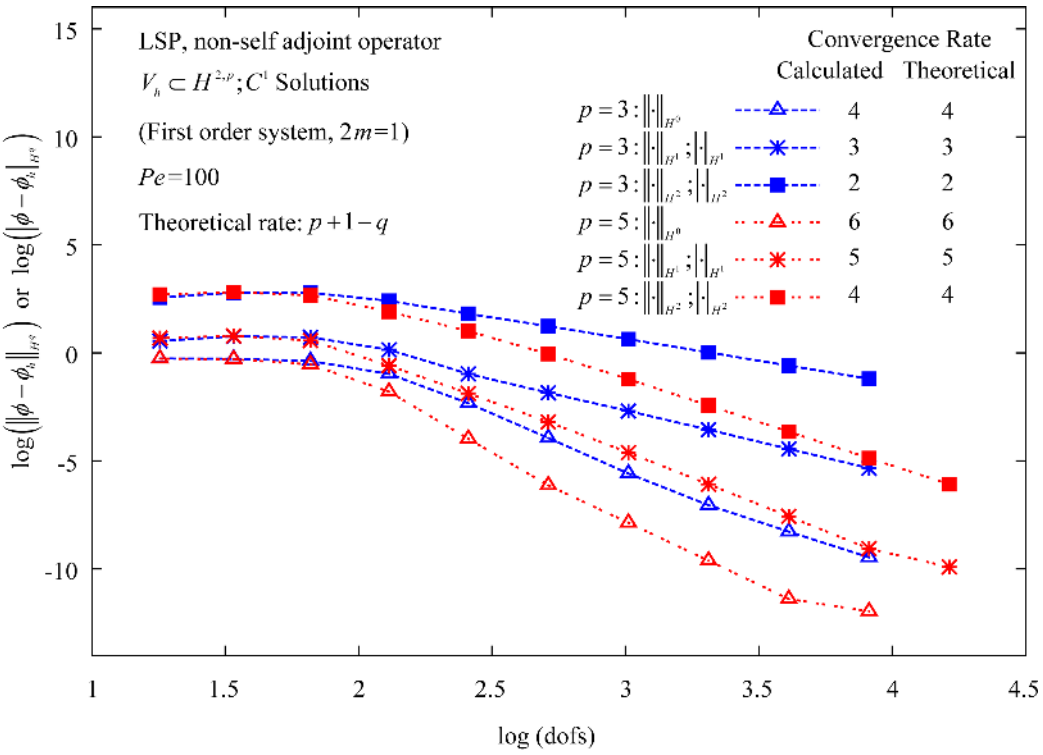

Figure 12. $\log\|\cdot\|_{H^q}$ versus log(dofs) for solutions of class C^1 (LSP, model problem 2, first order system, $p=3$ and 5, $Pe=100$).

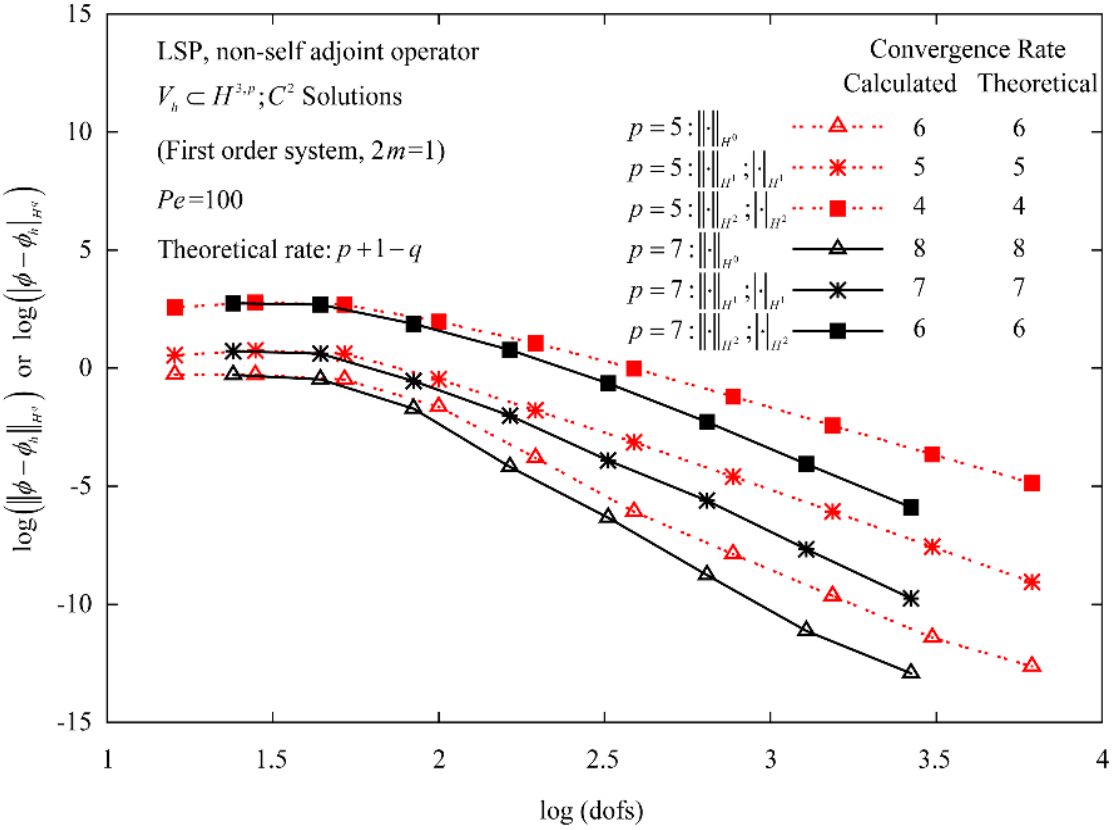

Figure 13. $\log\|\cdot\|_{H^q}$ versus log(dofs) for solutions of class C^2 (LSP, model problem 2, first order system, $p=5$ and 7, $Pe=100$).

again the agreement is perfect. Again, we note from **Figure 12** and **Figure 13** that at $p=5$ the convergence rates are independent of k. In this case, the integrals in the computations of the error norms are always Riemann.

Figure 14 shows plots of $\log\|\cdot\|_{H^2}$ and $\log\sqrt{I}$ versus log of dof for solutions of class C^1 and C^2 at $p=5$. Since the differential operator has the highest derivative of order 2, the convergence rate of $\sqrt{I}$ is expected to be same as that of $\|\cdot\|_{H^2}$ or $|\cdot|_{H^2}$ for both classes of solutions. This is confirmed in **Figure 14**. Convergence rate in case of C^1 and C^2 solutions are same (4 in this case), but C^2 solutions have better accuracy for a given dofs, confirming again that convergence rates of error norms or residual functional are not a function of the order *k* of the approximation space. Calculated rates are in perfect agreement with the theoretical rates. **Figure 15** shows plots of log of $\|\cdot\|_{H^0}=\|\cdot\|_{L_2}$ versus log of dofs for solution of classes C^0, C^1, and C^2 at $p=5$. We observe same convergence rates for $k=1$, 2, and 3 but better accuracy of the solution with progressively increasing *k*. These rates for LSP match perfectly with GM/WF for self adjoint operators due to the fact that in both cases the integral forms are variationally consistent. This proves again that the best approximation property in *B*-norm is not a requirement for establishing convergence rate. It is the variational consistency of the integral form that matters. Clearly the LSP does not have best approximation property in *B*-norm, yet has same convergence rates as GM/WF for self adjoint operators due to the fact that in both cases the integral forms are variationally consistent.

6.2.2. GM/WF

Since the differential operator is non-self adjoint the GM/WF will yield VIC integral form in which $B(\cdot,\cdot)$ is nonsymmetric and we lose the best approximation property in *B*-norm. Nonetheless we conduct some numerical experiments to monitor convergence rates of various error norms. First, we note that GM/WF in this model problem will yield the following element equations for an element *e* (when $p=1$ and $f(x)\neq 0$). See reference [35].

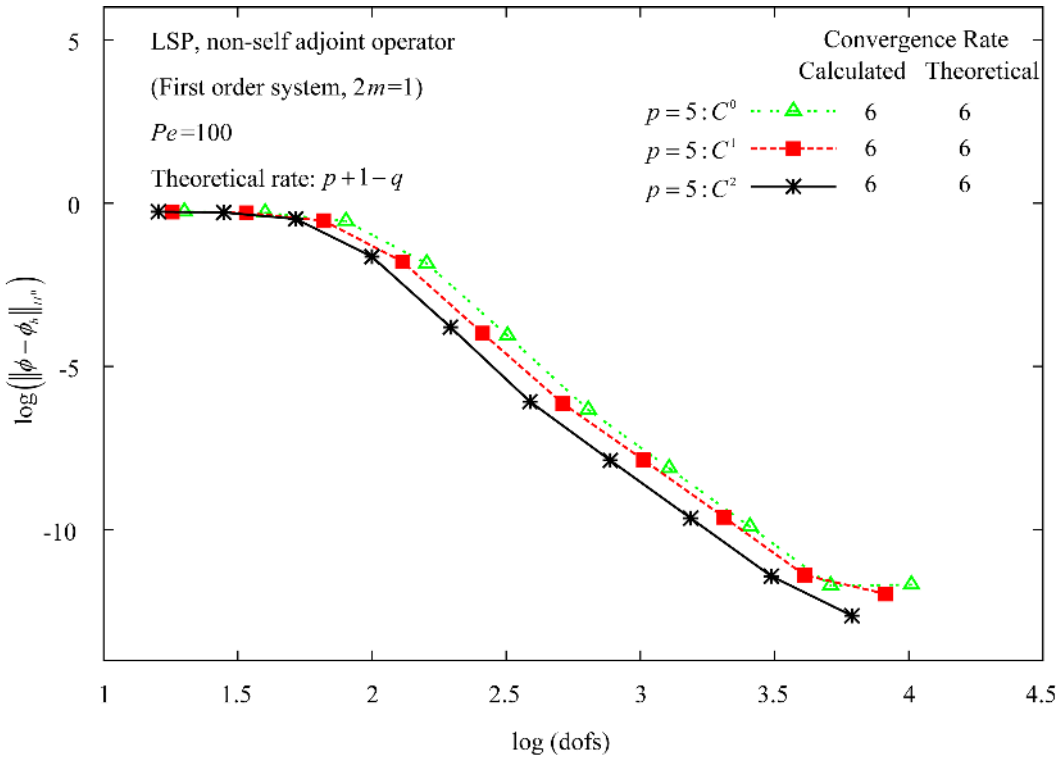

Figure 14. $\log\|\cdot\|_{H^0}$ versus log(dofs) for solutions of classes C^0, C^1, and C^2 at $p=5$ (LSP, first order system, model problem 2).

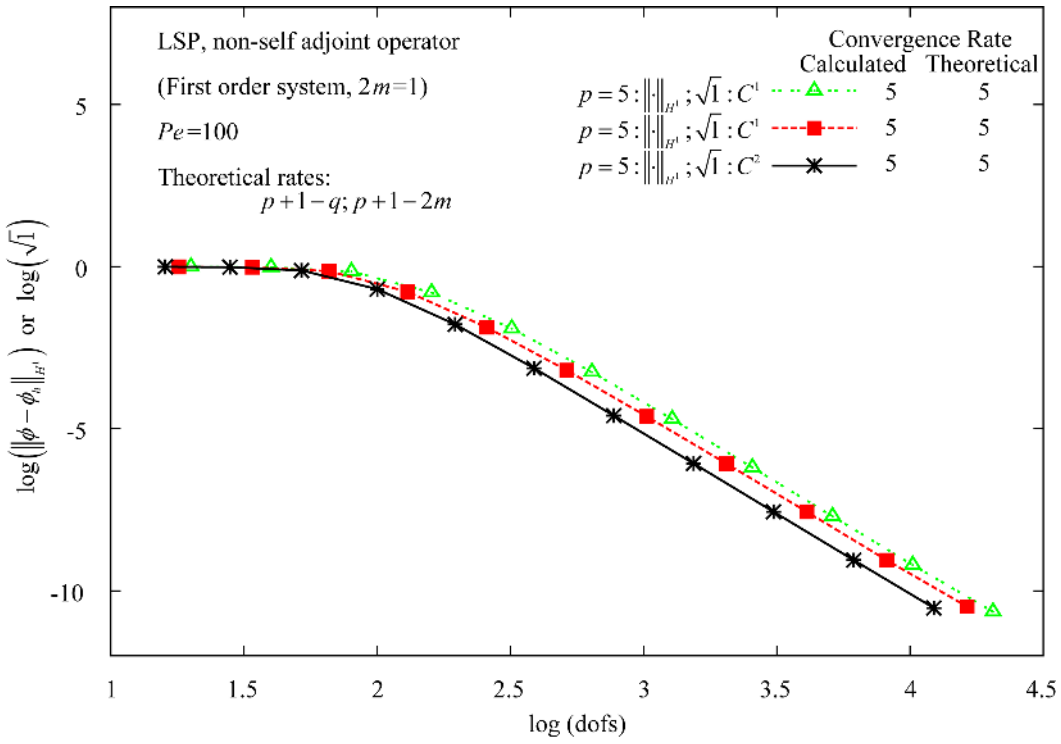

Figure 15. $\log\|\cdot\|_{H^2}$ or $\log\|E\|_{L_2}$ versus log(dofs) for solutions of classes C^1 and C^2 at $p=5$ (LSP, first order system, model problem 2).

$$\left[\left[K_1^e\right]+\frac{1}{Peh_e}\left[K_2^e\right]\right]\left\{\delta^e\right\}=\left\{P^e\right\}+\left\{f^e\right\} \tag{136}$$

and the assembled equations for discretization $\bar{\Omega}^T=\bigcup_e \bar{\Omega}^e$ are

$$\left[K\right]\left\{\delta\right\}=\left[\left[K_1\right]+\frac{1}{Peh_e}\left[K_2\right]\right]\left\{\delta\right\}=\left\{P\right\}+\left\{f\right\};\quad h=h_e \tag{137}$$

in which $\left\{\delta\right\}=\bigcup_e\left\{\delta^e\right\}$ and $\left[K_1\right]$, $\left[K_2\right]$ are due to assembly of $\left[K_1^e\right]$ and $\left[K_2^e\right]$ for $\bar{\Omega}^T$. As shown in reference [35], $\left[K_1\right]$ is due to convection term (*i.e.* $\frac{\mathrm{d}\phi}{\mathrm{d}x}$) and $\left[K_2\right]$ is due to diffusion (*i.e.* $\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}$); $\left[K_1\right]$ is nonsymmetric with zeros on the diagonals after $\phi(0)=1$ and $\phi(1)=0$ BCs are imposed, thus if Peh is large, the contribution of $\left[K_2\right]$ to $\left[K\right]$ is almost insignificant compared to the contribution of $\left[K_1\right]$ and the computations using (137) will fail. On the other hand if the discretization $\bar{\Omega}^T$ is sufficiently refined, the contribution of $\left[K_2\right]$ to $\left[K\right]$ overshadows that of $\left[K_1\right]$ and the behavior will be dominated by $\left[K_2\right]$ (*i.e.* $\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}$ term in the differential operator). When this happens the integral form from GM/WF will behave like a VC integral form as it is primarily due to $\frac{1}{Pe}\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}$ term in the differential operator which is self adjoint, hence the convergence rates of various error norms will be similar to GM/WF for self adjoint operator.

For numerical experiments, we consider $Pe=100$ and $Pe=1000$. For $Pe=1000$ the solution gradients are more isolated near $x=1$ and are higher in magnitude compared to $Pe=100$. We consider solutions of class C^1 at $p=3$ for both Peclet numbers. Progressively refined uniform discretizations are used for computing solutions and error norms. **Figure 16** shows error norms versus dof plots for solutions of class C^1, $p=3$ for $Pe=100$. We note that due to smoothness of the solutions, the asymptotic range in which $\left[K_2\right]$ dominates is quickly achieved, and the computations succeed for meshes of 16 elements or more. In this range calculated convergence rates match perfectly with the theoretical rates for self adjoint operator. In this range the BVP reduces to $\frac{1}{Pe}\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}=0$ as the contribution of $\frac{\mathrm{d}\phi}{\mathrm{d}x}$ term in this range is insignificant. For meshes with 16 elements or fewer the calculated solution from (137) does not satisfy (137) when substituted in them, implying lack of equilibrium due to spuriousness of the computed solution. **Figure 17** shows similar graphs for $Pe=1000$. The computations fail for discretizations resulting in $\log(dofs)\le 2.5$ (meshes coarser than 256 elements) where equilibrium is not achieved, that is, calculated solution from (137) does not satisfy (137) when substituted into the equations. This is due to VIC nature of the integral form resulting from GM/WF. Correspondingly, the values of the error norms for the failed discretizations grow out of control. When $\log(dofs)\ge 2.5$ (discretization contains 256 elements or more), $\left[K_1\right]$ contribution becomes insignificant and the BVP behaves like $\frac{1}{Pe}\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}=0$, hence the asymptotic range is observed with calculated convergence rates of the indicated error norms of 3.7, 2.9, 2 are achieved compared to their theoretical values of 4, 3, 2 for self adjoint operators, rather amazingly good performance for VIC integral form.

When performing the error computations for Pe higher than 1000 with uniform mesh refinement of 2, 4, ...elements failure of computations occurs when $\left[K_1\right]$ dominates the total $\left[K\right]$ as expected.

6.2.3. LSP: Higher Order System (Without Auxiliary Equation)

In this study, we consider 1D convection-diffusion equation (124) without converting it to a system of first order equations through the use of auxiliary equation. In this case $V_h\subset H^{k,p}\left(\bar{\Omega}^e\right)$, $k=3$ is minimally conforming approximation space if the integrals over $\bar{\Omega}^T$ are to be Riemann. For $k=2$ the integrals over $\bar{\Omega}^T$ are Lebesgue and $k=1$ (solutions of class C^0) is not admissible.

Error norms are computed for progressively refined uniform discretizations for $k=2,3$ (solutions of classes C^1 and C^2) at p-levels of 3 and 5 for $k=2$ and $p=5$ and 7 for $k=3$. Plots of error norms versus dof for

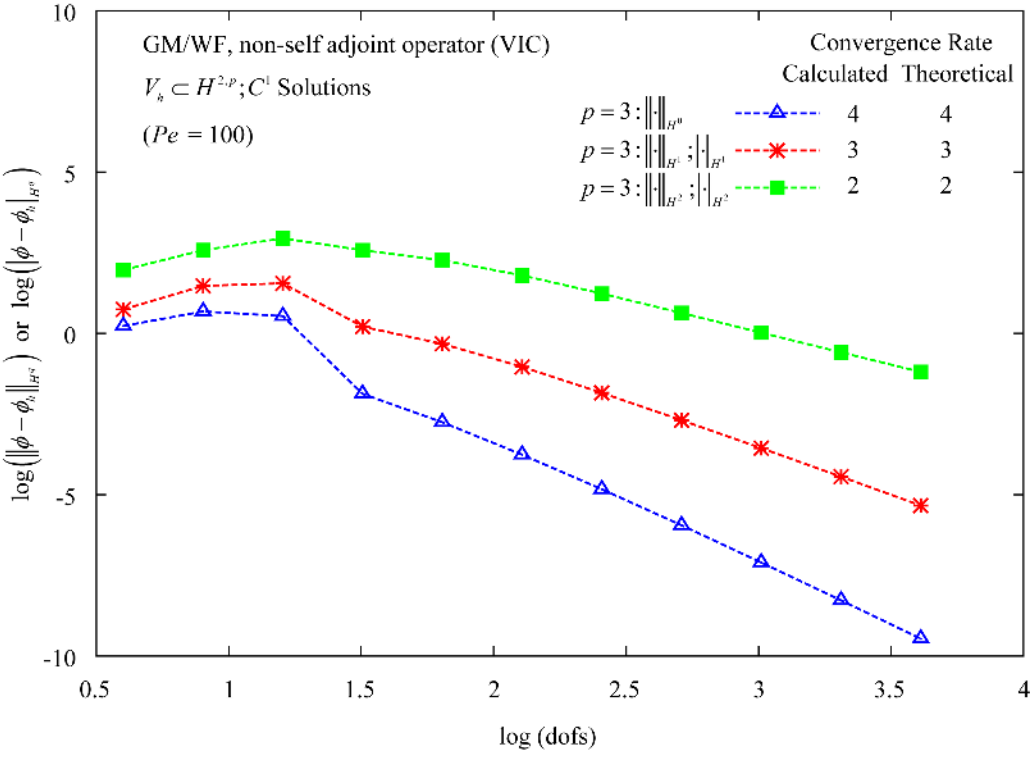

Figure 16. $\log\|\cdot\|_{H^q}$ versus log(*dofs*) for solutions of class C^1 (GM/WF, model problem 2, $p=3$, $Pe=1000$).

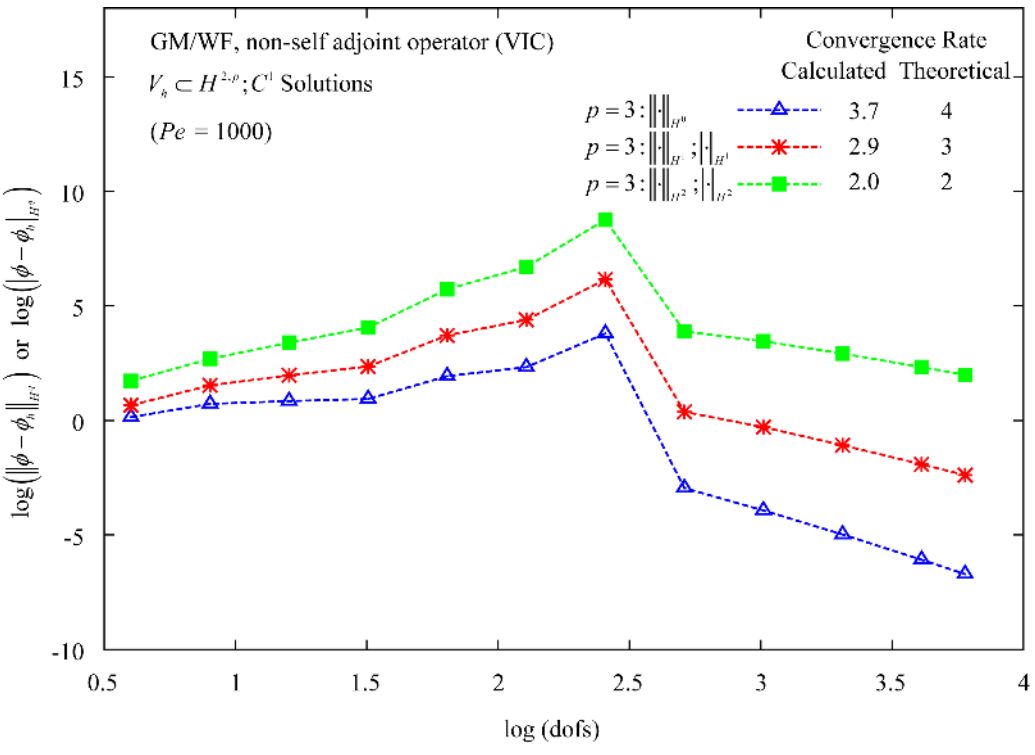

Figure 17. $\log\|\cdot\|_{H^q}$ versus log(*dofs*) for solutions of class C^1 (GM/WF, model problem 2, $p=3$, $Pe=1000$).

solutions of classes C^1 and C^2 and the calculated convergence rates and comparisons with the theoretical values are shown in **Figure 18** and **Figure 19**. We note that the highest order of the derivative in the mathematical model ($2m$) in this case is 2 as the convection-diffusion equation is not reduced to a first order system using auxiliary equations. Theoretical convergence rates are overall in good agreement with the calculated convergence rates confirming importance of the variational consistency of the integral form. In solutions of both classes, the convergence rate of $\|\cdot\|_{H^0}$ is higher than predicted for $p=5$. In this case $2m=2$ whereas in case of first order system derived using auxiliary equation $2m=1$, thus the first order system has higher convergence rate of $\sqrt{I}$ in the LSP.

Figure 20 shows plots of $\|\cdot\|_{H^2}$ versus dofs and $\sqrt{I}$ versus dofs for solutions of classes C^1 and C^2 at $p=5$. Since the highest order derivative is two in the differential operator, the convergence rate of $\|\cdot\|_{H^2}$ is same as that of $\sqrt{I}$. C^1 and C^2 solutions have same convergence rates but C^2 solutions have better accuracy for a given dofs, confirming that the convergence rates of error norm and residual functional are not a function of the order *k* of the approximation space. Thus, for higher order system we also observe that the rates for LSP match with GM/WF for self adjoint operators due to the fact that in both the integral forms are VC even though the two methods of approximation have best approximation property in different norms.

6.3. Model Problem 3: Non-Linear Operator, 1D Burgers Equation

We consider 1D Burgers equation described by a non-linear operator (see reference [35]) to compute a priori error estimates and convergence rates of various error norms and compare them with their theoretical values,

$$\phi\frac{\mathrm{d}\phi}{\mathrm{d}x}-\frac{1}{Re}\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}=0,\quad \forall x\in(0,1)=\Omega\subset\mathbb{R}^1 \tag{138}$$

$$\text{BCs:}\quad \phi(0)=1,\quad \phi(1)=0 \tag{139}$$

For the studies presented in the following sections, a value of $Re=100$ is used. Theoretical solution ϕ of

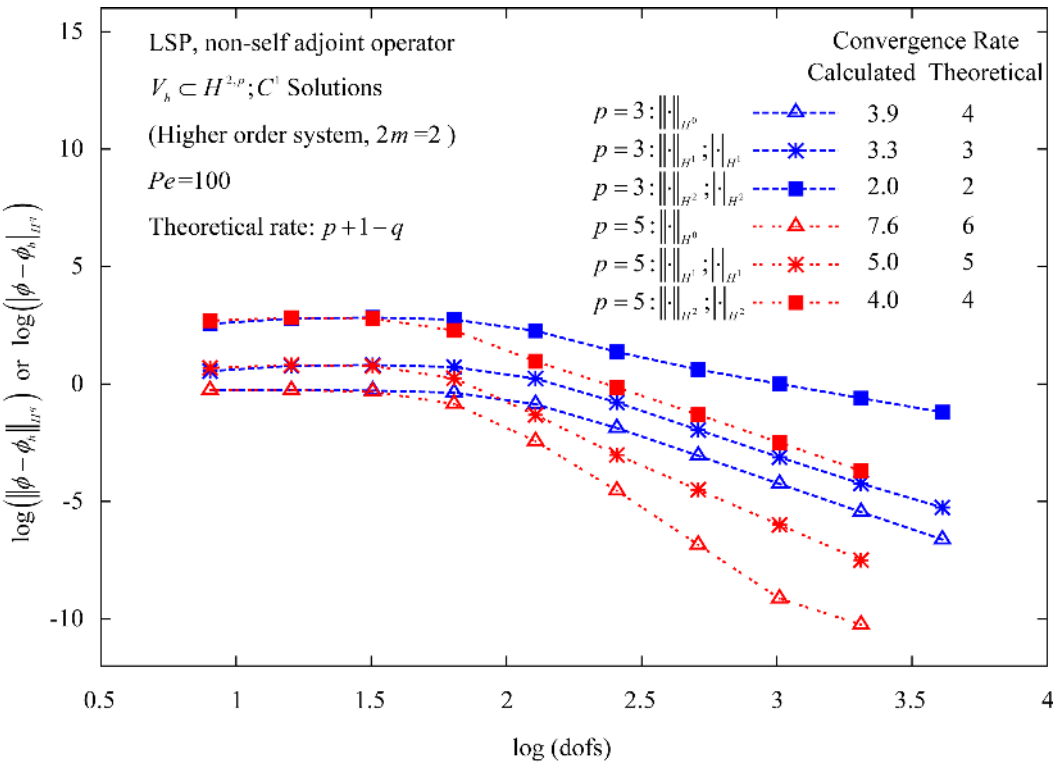

Figure 18. $\log\|\cdot\|_{H^q}$ versus log(dofs) for solutions of class C^1 (LSP, model problem 2, higher order system, $p=3$ and 5, $Pe=100$).

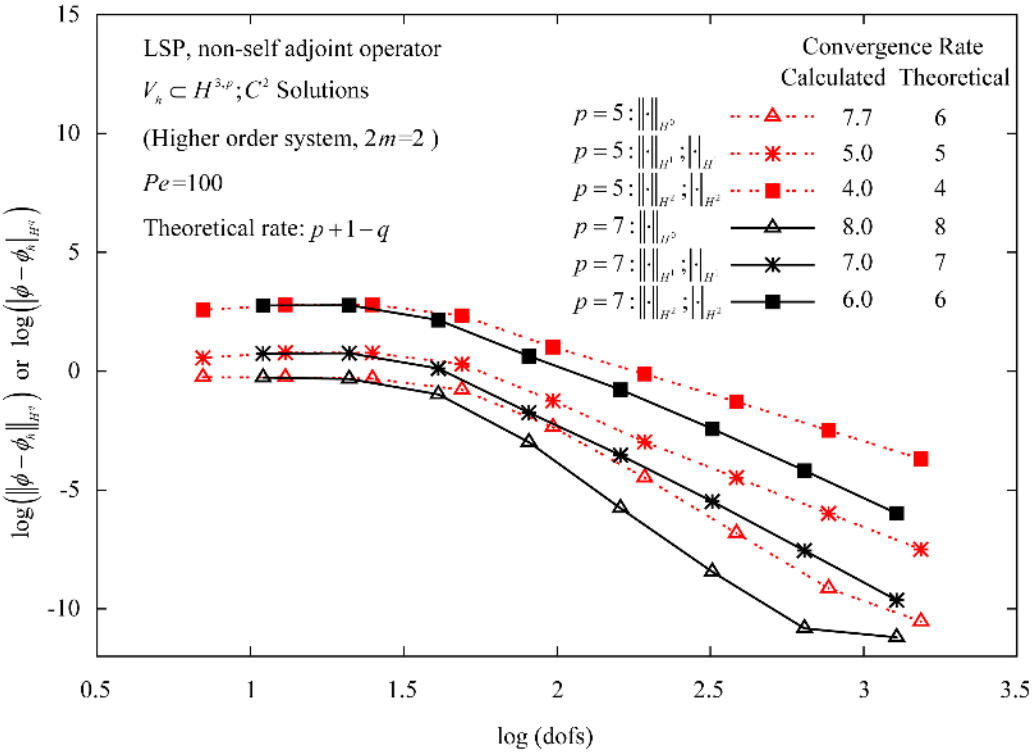

Figure 19. $\log\|\cdot\|_{H^q}$ versus log(dofs) for solutions of class C^2 (LSP, model problem 2, higher order system, $p=5$ and 7, $Pe=100$).

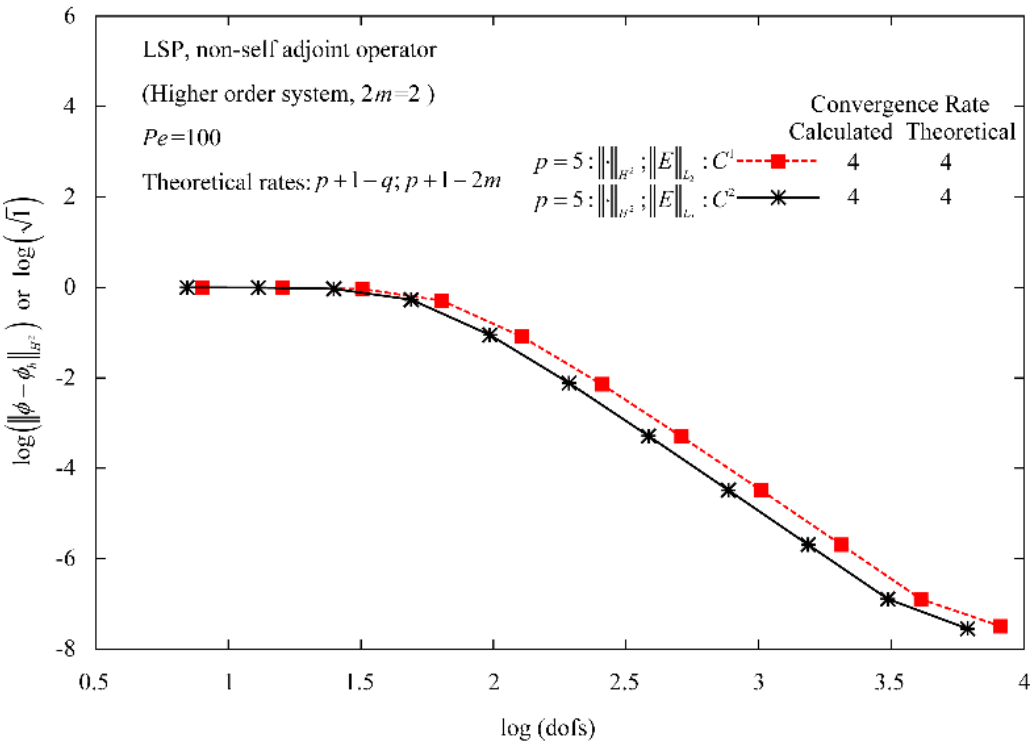

Figure 20. $\log\|\cdot\|_{H^2}$ or $\log\|E\|_{L_2}$ versus log(dofs) for solutions of classes C^1 and C^2 at $p=5$ (LSP, higher order system, model problem 2).

(138) and (139) and finite element solution ϕ_h using GM/WF and LSP (higher order and first order systems) are given in reference [35]. Some details are given in the following as a review. It is shown [35] that in this case $A = \phi \dfrac{\mathrm{d}}{\mathrm{d}x} - \dfrac{1}{Re}\dfrac{\mathrm{d}^2}{\mathrm{d}x^2}$ which is a function of ϕ, hence non-linear. The GM/WF yields VIC integral form. The integral form from the LSP is VC with minor adjustments (see theorem 7) of little consequence but immense benefit as they yield variational consistency of the integral form.

a) GM/WF: The integral form of (138) and (139) over $\bar{\Omega}$ is given by

$$\left(\phi\frac{\mathrm{d}\phi}{\mathrm{d}x}, v\right)_{\bar{\Omega}} + \frac{1}{Re}\left(\frac{\mathrm{d}\phi}{\mathrm{d}x}, \frac{\mathrm{d}v}{\mathrm{d}x}\right)_{\bar{\Omega}} = 0\,; \quad v = \delta\phi, \quad \forall v \in V \subset H^{k,p}\left(\bar{\Omega}\right) \tag{140}$$

or

$$B\left(\phi, v\right) = 0 \tag{141}$$

Functional $B\left(\cdot,\cdot\right)$ is linear in v but not linear in ϕ and is obviously not symmetric.

$$\delta B\left(\phi, v\right) = \left(v\frac{\mathrm{d}\phi}{\mathrm{d}x} + \phi\frac{\mathrm{d}v}{\mathrm{d}x}, v\right) + \frac{1}{Re}\left(\frac{\mathrm{d}v}{\mathrm{d}x}, \frac{\mathrm{d}v}{\mathrm{d}x}\right) \tag{142}$$

is obviously not >0, $=0$, or <0 $\forall v \in V \subset H^{k,p}\left(\bar{\Omega}\right)$, hence the integral form (141) is VIC.

b) LSP based on residual functional: These can be constructed in two alternate ways, as a higher order system (138) or by recasting (138) as a system of first order equations. [(I)]

I) Higher order system

$$E = A\phi_h - f = \phi_h\frac{\mathrm{d}\phi_h}{\mathrm{d}x} - \frac{1}{Re}\frac{\mathrm{d}^2\phi_h}{\mathrm{d}x^2}, \quad \forall x \in \bar{\Omega}^T = \bigcup_e \bar{\Omega}^e \tag{143}$$

and residual functional $I\left(\phi_h\right)$ is given by

$$I\left(\phi_h\right) = \left(E, E\right)_{\bar{\Omega}} \tag{144}$$

$$\delta I\left(\phi_h\right) = 2\left(E, \delta E\right) = 2g = 0$$

$$\delta E = v\frac{\mathrm{d}\phi_h}{\mathrm{d}x} + \phi_h\frac{\mathrm{d}v}{\mathrm{d}x} - \frac{1}{Re}\frac{\mathrm{d}^2 v}{\mathrm{d}x^2} \tag{145}$$

$$\delta^2 I\left(\phi_h\right) \simeq 2\left(\delta E, \delta E\right) \tag{146}$$

The necessary condition $g = 0$ is satisfied by calculating a solution using Newton's linear method. See reference [35] for full details. The integral form in this case is variationally consistent.

II) First order system

Let $\tau = \dfrac{\mathrm{d}\phi}{\mathrm{d}x}$, then (138) reduces to

$$\begin{aligned} &\phi\frac{\mathrm{d}\phi}{\mathrm{d}x} - \frac{1}{Re}\frac{\mathrm{d}\tau}{\mathrm{d}x} = 0 \\ &\tau - \frac{\mathrm{d}\phi}{\mathrm{d}x} = 0 \end{aligned} \tag{147}$$

LSP for (147) is described in detail in reference [35] and is omitted here. This integral form is also VC.

6.3.1. LSP: Higher-Order System (Without Auxiliary Equation)

For this model problem we only present studies related to convergence rates of various error norms using (138) (*i.e.* without recasting it as a system of first order equations). As in other problems $\bar{\Omega} = [0,1]$ is discretized using uniform meshes of 2, 4, 8, ...3-node p-version higher order global differentiability elements and the solutions are computed using finite element formulations based on GM/WF and LSP. In case of LSP, since the integral form is VC the same convergence rate estimates hold as in (135):

$$\left.\begin{aligned} &\left\|\phi-\phi_h\right\|_{H^q} \le C_3 h^{p+1-q}\left|\phi\right|_{p+1} \\ &\left|\phi-\phi_h\right|_{H^q} \le C_2 h^{p+1-q}\left|\phi\right|_{p+1} \\ &\left\|A\phi-f\right\|_{L_2} = \left\|E\right\|_{L_2} \le C_4 h^{p+1-2m} \end{aligned}\right\} \tag{148}$$

In this BVP, $2m=2$. Since the differential operator has derivative of ϕ up to second order, the minimally conforming space in this case is $k=3$ for the integrals over $\overline{\Omega}^T$ to be Riemann and the integrals are in Lebesgue sense when $k=2$. $k=1$ is not admissible. **Figure 21** and **Figure 22** show plots of various error norms versus dofs for solutions of class C^1 and C^2 as well as calculated and theoretical convergence rates.

First, we note from **Figure 21** and **Figure 22** large pre-asymptotic and onset of asymptotic ranges. The asymptotic range is rather limited, due to which accurate computation of convergence rates is difficult. Nonetheless we observe that for most error norms the theoretical and calculated convergence rates are in good agreement. Once again, we observe that due to VC integral form in LSP for nonlinear operators the convergence rate estimates for GM/WF for self adjoint operators and the same for LSP for linear operators hold here, again confirming the significance and importance of VC integral forms.

Figure 23 shows plots of $\left\|\cdot\right\|_{H^2}$ versus dof and $\sqrt{I}$ versus dof for solutions of class C^1 and C^2 at $p=5$. Since the differential operator is second order operator, the convergence rate of $\left\|\cdot\right\|_{H^2}$ is same as that of $\sqrt{I}$ for both C^1 and C^2 local approximations. However, C^2 solutions have better accuracy for a given dofs. We clearly observe that the convergence rate is not a function of *k*, the order of approximation space. Calculated convergence rates of $\left\|\cdot\right\|_{H^2}$ and $\sqrt{I}$ are the same and are in exact agreement with the theoretical convergence rates.

6.3.2. GM/WF

Since the differential operator is non-linear the integral form from GM/WF is VIC. $B(\cdot,\cdot)$ is not bilinear and is

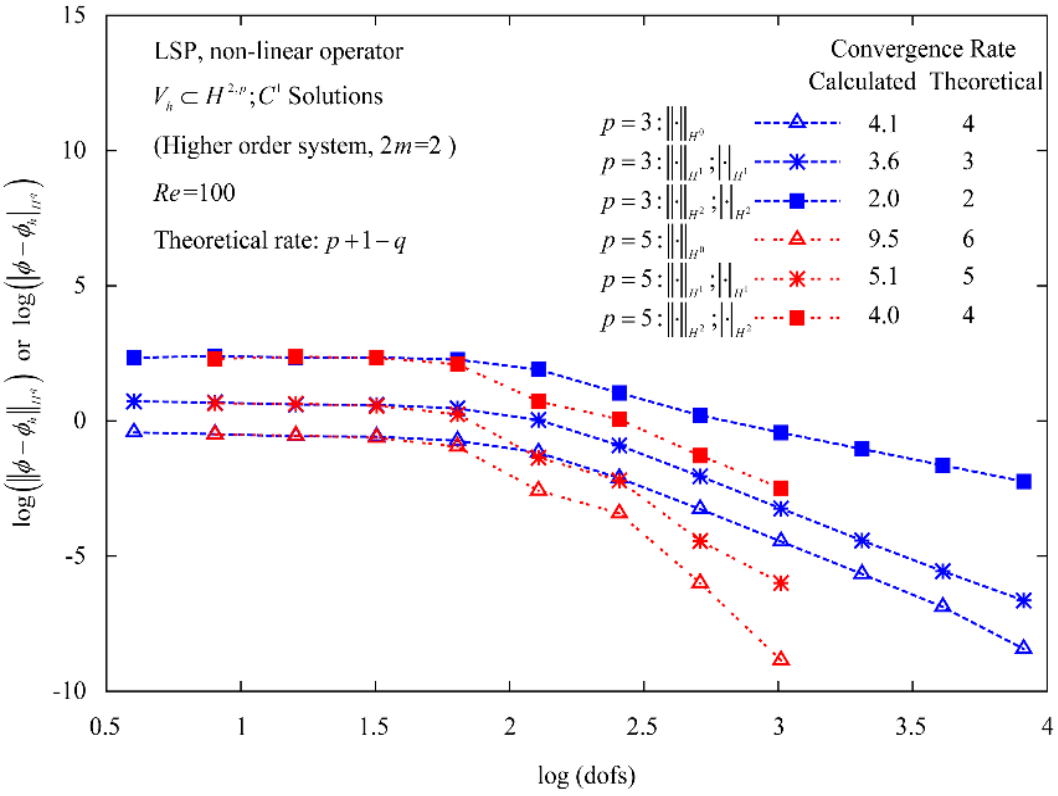

Figure 21. $\log\left\|\cdot\right\|_{H^q}$ versus log(dofs) for solutions of class C^1 (LSP, model problem 3, higher order system, $p=3$ and 5, $Re=100$).

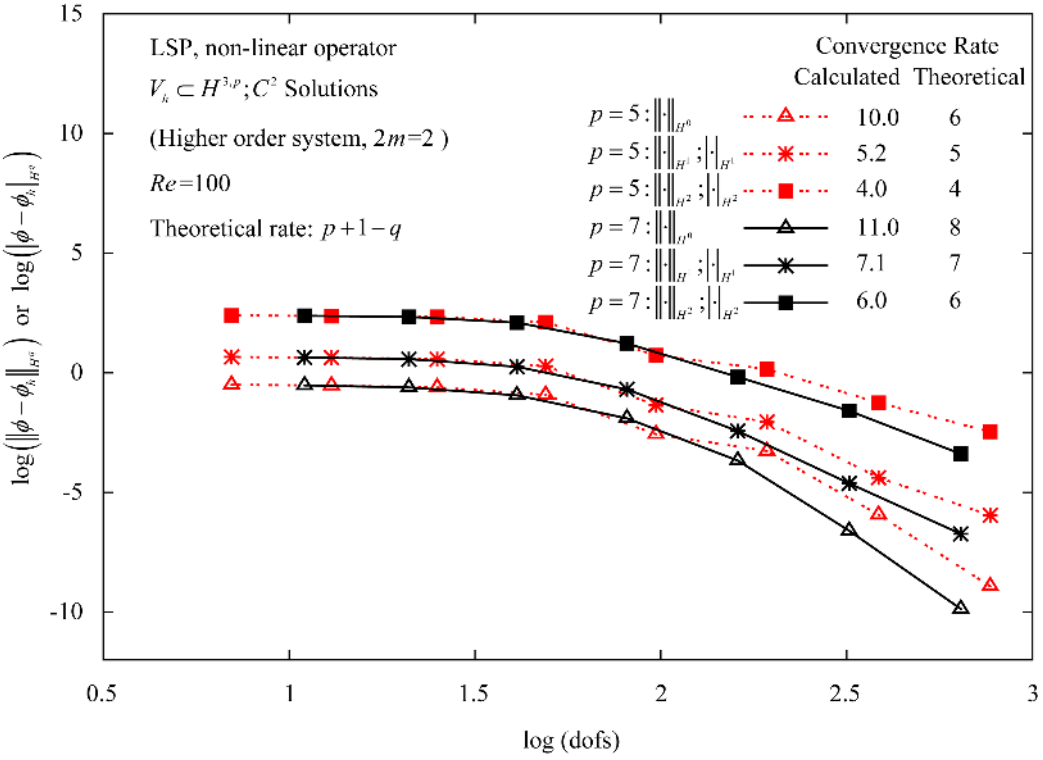

Figure 22. $\log\|\cdot\|_{H^q}$ versus log(dofs) for solutions of class C^2 (LSP, model problem 3, higher order system, $p=5$ and 7, $Re=100$).

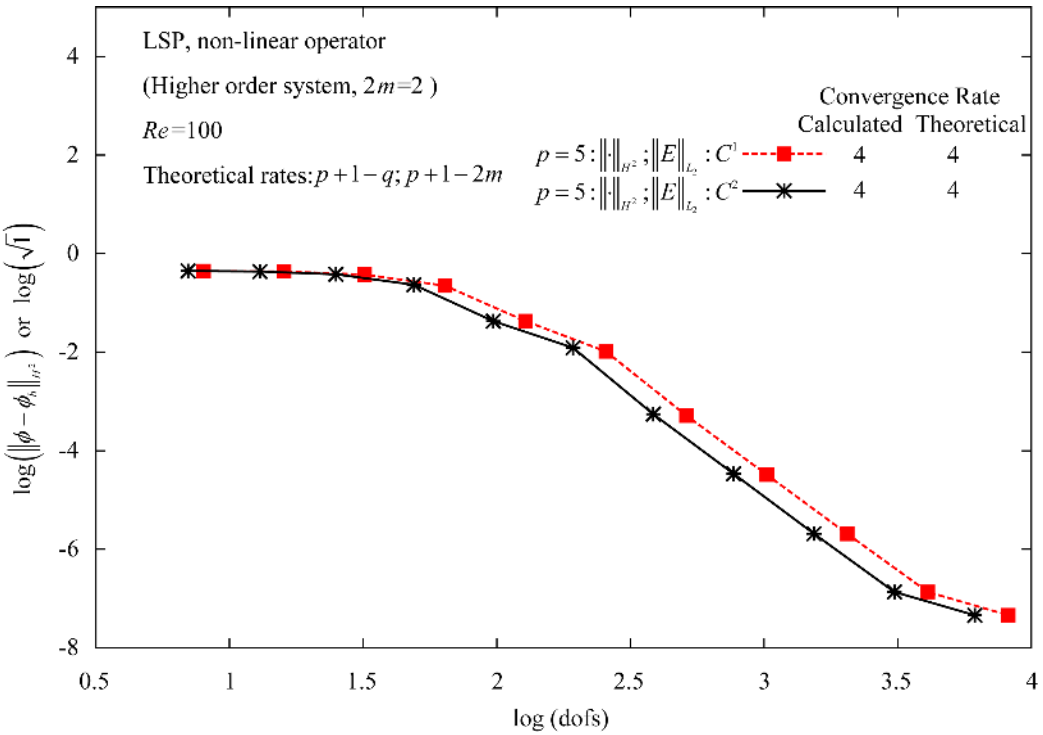

Figure 23. $\log\|\cdot\|_{H^2}$ or $\log\|E\|_{L_2}$ versus log(dofs) for solutions of classes C^1 and C^2 at $p=5$ (LSP, higher order system, model problem 3).

not symmetric, hence we lose best approximation property of the GM/WF in -norm. GM/WF will yield the following form of the assembled equations for $\bar{\Omega}^T$ (when $p=1$ and $Bf(x)\neq 0$) assuming uniform discretization ($h=h_e$):

$$[K]\{\delta\}=\left[[K_1]+\frac{1}{Reh}[K_2]\right]\{\delta\}=\{P\}+\{F\} \tag{149}$$

in which $[K_1]=\left[K_1(\{\delta\})\right]$ and $(K_1)_{ij}\neq(K_1)_{ji}$. $[K_2]$ is symmetric. $[K_1]$ is due to $\phi\frac{\mathrm{d}\phi}{\mathrm{d}x}$ and $[K_2]$ is due to $\frac{1}{Re}\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}$ term in the differential equation. Furthermore $[K_1]$ has zeros on the diagonal after $\phi(0)=1$ and $\phi(1)=0$ boundary conditions are imposed, thus if Reh is large, the contribution of $[K_2]$ to $[K]$ is almost insignificant and the computations using (149) will fail. On the other hand if the discretization $\bar{\Omega}^T$ is sufficiently refined then contribution of $[K_2]$ to $[K]$ overshadows that of $[K_1]$ and the solution behavior will be dominated by $[K_2]$ (*i.e.* $\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}$ term in the differential equation). When this happens the integral form from GM/WF will behave like a VC integral form and the convergence rates of various error norms will be same as those of GM/WF for self adjoint operator.

For numerical studies, we consider $Re=100$. Uniform mesh refinement is carried out for solutions of class C^1 at $p=3$. **Figure 24** shows plots of error norms versus dofs. We note that for discretizations coarser than 128 elements the error norms correspond to erroneous computed solutions in which equilibrium condition is violated for the assembled equations. For finer discretizations (128 elements or more) asymptotic range is observed. In this range discretization is sufficiently refined so that the integral form is dominated by the diffusion term. Calculated convergence rates (of $\|\cdot\|_{H^0}$; $\|\cdot\|_{H^1}$, $|\cdot|_{H^1}$; $\|\cdot\|_{H^2}$, $|\cdot|_{H^2}$) 3.7, 3, and 2 are in close agreement with the theoretical convergence rates 4, 3, 2. In this study for $Re=100$ computations failed for discretizations coarser than 128 elements where equilibrium was not achieved.

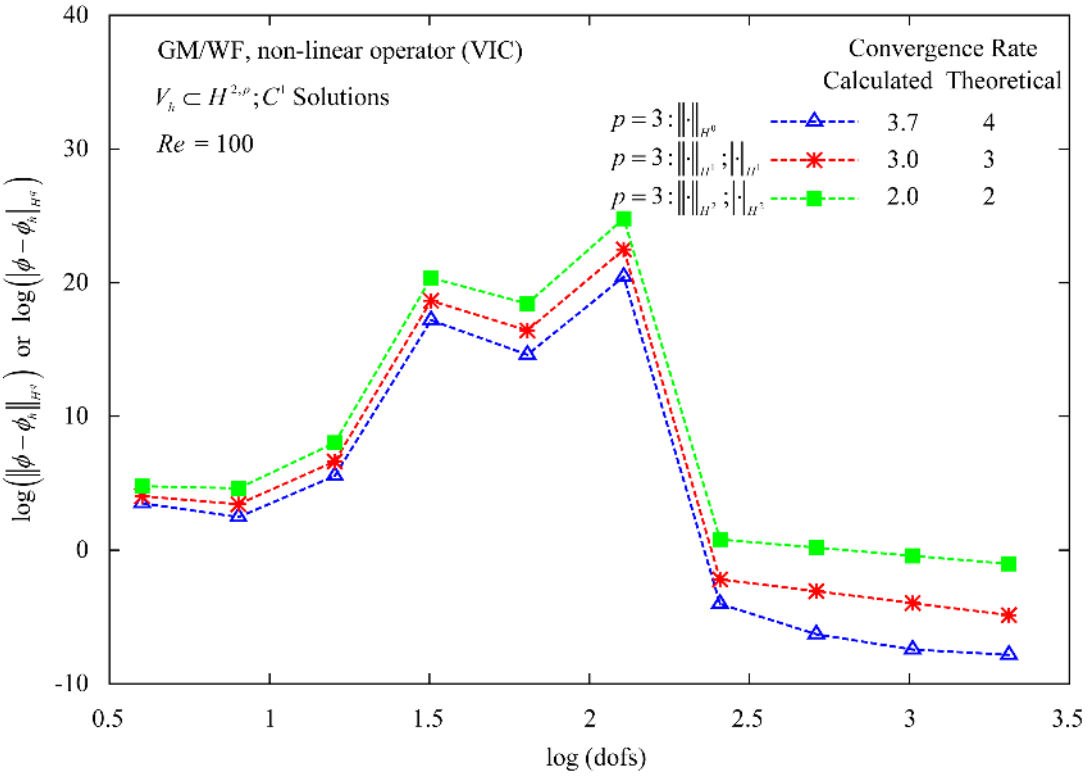

Figure 24. $\log\|\cdot\|_{H^q}$ versus log(dofs) for solutions of class C^1 (GM/WF, model problem 3, $p=3$, $Re=100$).

7. A Posteriori Error Estimation and Computation

7.1. A Posteriori Error Estimation

A posteriori error estimation refers to estimation of errors in the computed solution. The primary purpose is to be able to devise some element-wise measures as well as in the whole discretization that quantify the errors in the computed solution as well as provide some guidance on the portions of the domain where the computed solution needs to be improved. Based on these measures one could design mesh refinement, *p*-level change, etc. strategies that result in the desired accuracy of the computed solution. This process of changing *h*, *p*, and possibly *k* based on measures estimated using the computed solution is referred to as *adaptive process* (*i.e.* we adapt *h*, *p*, and *k* as dictated by the current state of the solution and a posteriori error estimators or indicators).

During the development of finite element technology and even now, solutions of class C^0 have been used predominantly. The local approximations of class C^0 result in interelement discontinuity of the derivatives normal to the interelement boundaries. When the solutions of the BVPs are smooth, these interelement jumps in the derivatives are reduced upon *h*, *p* refinements and we say C^0 solutions converge weakly to class C^1. The a posteriori error estimations largely exploit the interelement discontinuities of the derivatives inherent in C^0 local approximations. We note the following.

1) When the local approximations are considered in higher order spaces, the a posteriori error estimates used currently that are derived based on C^0 local approximations are meaningless as for higher order global differentiability local approximations the interelement jumps in the derivatives of the solutions used currently do not exist.

2) The C^0 local approximations can only be used in a system of first order differential equations to calculate the residuals and residual functionals over $\bar{\Omega}^e$ as well as over $\bar{\Omega}^T$, but only in Lebesgue sense. For higher order BVPs such computations are not possible with local approximations of class C^0. Even though the residual functional over $\bar{\Omega}^e$ and $\bar{\Omega}^T$ are true measures of how well the local approximation satisfies the BVP, the emphasis has been largely on a posteriori error estimation, primarily due to the insistence on the use of C^0 local approximations.

3) Our view is that in a finite element computational framework the physics of the BVP must be preserved and in such a framework, once a finite element solution has been calculated, the computational framework must permit a posteriori computations of any desired measures otherwise the computational framework is deficient.

7.2. A Posteriori Error Computation

As mentioned in Section 7.1, the computational framework must be designed such that it permits a posteriori computations of all desired measures that are necessary and meaningful in adaptivity. Minimally conforming spaces play a crucial role in accomplishing this. We present details in the following. Let

$$\boldsymbol{A}\boldsymbol{\phi} - \boldsymbol{f} = 0 \quad \text{over } \bar{\Omega} \tag{150}$$

be a boundary value problem in which the differential operator may be self adjoint, non-self adjoint, or nonlinear. Let $2m$ be the highest order of the derivative of ϕ in (150). Let ϕ_h^e and ϕ_h be approximations of ϕ over $\bar{\Omega}^e$ and $\bar{\Omega}^T$. The approximation ϕ_h is assumed to be computed from any of the methods of approximation in which the integral forms may be VC or VIC. Let

$$\phi_h^e \in V_h \subset H^{k,p}\left(\bar{\Omega}^e\right); \quad k \geq 2m+1 \tag{151}$$

$$\phi_h = \bigcup_e \phi_h^e \tag{152}$$

The approximation space V_h is minimally conforming ensuring that the integrals over $\bar{\Omega}^T$ are Riemann. Using (150) and (152), we can define residual functions E_i

$$E_i = \sum_j A_{ij}\left(\phi_h\right)_j - f_i; \quad i = 1, 2, \cdots, n \quad \text{over } \bar{\Omega}^T \tag{153}$$

where *n* is the number of differential equations in (150). Let

$$E_i^e = \sum_j A_{ij}\left(\phi_h^e\right)_j - f_i; \quad i = 1, 2, \cdots, n \quad \text{over } \bar{\Omega}^e \tag{154}$$

We define residual functionals I and I^e over $\bar{\Omega}^T$ and $\bar{\Omega}^e$ by

$$I = \sum_{i=1}^{n} \left(E_i, E_i\right)_{\bar{\Omega}^T} \tag{155}$$

$$I^e = \sum_{i=1}^{n} \left(E_i^e, E_i^e\right)_{\bar{\Omega}^e} \tag{156}$$

Since $\bar{\Omega}^T = \bigcup_e \bar{\Omega}^e$, we can write (using (155) and (156))

$$I = \sum_e I^e = \sum_e \sum_{i=1}^{n} \left(E_i^e, E_i^e\right)_{\bar{\Omega}^e} \tag{157}$$

If ϕ_h is the theoretical solution ϕ then

$$E_i = \sum_j A_{ij}\left(\phi_h\right)_j - f_i = 0, \ \ i = 1, 2, \cdots, n \tag{158}$$

and

$$I = 0, \ \ I^e = 0 \tag{159}$$

over $\bar{\Omega}^T$ and each $\bar{\Omega}^e$. Minimally conforming space V_h ensures that integrals over $\bar{\Omega}^T$ are Riemann, hence proximity of $I\left(\phi_h\right)$ to zero (theoretical value of functional I; that is, $I\left(\phi\right)$) is a measure of error in the solution ϕ_h over $\bar{\Omega}^T$. When $I\left(\phi_h\right) \to 0$, $\left(\boldsymbol{E}, \boldsymbol{E}\right) \to 0 \Rightarrow E_i \to 0; i = 1, 2, \cdots, n, \forall x \in \bar{\Omega}^T$, thus $E_i^e \to 0, \forall x \in \bar{\Omega}^e$ for each $\bar{\Omega}^e$ in $\bar{\Omega}^T$, implying that differential Equation (150) are satisfied in the pointwise sense. Thus, the main steps in a posteriori error computation can be summarized in the following.

1) Choose minimally conforming space $k \geq 2m+1$ thereby ensuring integrals over $\bar{\Omega}^T$ in Riemann sense.

2) Regardless of the method of approximation to construct integral form in the finite element process, the following steps are possible and help in quantifying solution error. Calculate finite element solution ϕ_h and hence ϕ_h^e.

3) Calculate E_i^e, $I^e = \sum_{i=1}^{n} \left(E_i^e, E_i^e\right)_{\bar{\Omega}^e}$ for each element e with domain $\bar{\Omega}^e$ of the discretization $\bar{\Omega}^T$.

4) Calculate $I = \sum_e I^e$ for $\bar{\Omega}^T$.

5) When $I \simeq 0$ ($O\left(10^{-8}\right)$ or lower), ϕ_h is reasonably converged to ϕ for the h, p, and k employed, hence no need for adaptive refinements.

6) When $I \neq 0$, we examine I^e values for individual elements of $\bar{\Omega}^T$ to determine which elements have I^e values larger than a certain threshold value I^e. These elements can be considered for adaptive refinement (h or p or both) depending on the strategy adopted. Some of these are presented in the next section.

7) In this approach, a posteriori error estimations derived and used presently (of little value in higher order spaces) are eliminated altogether.

8) Errors in the computed solution are quantified without the knowledge of theoretical solution and there is built-in adaptivity due to I^e for individual elements. The elements with I^e values larger than a threshold value I^e are candidates for refinement.

9) Adaptive processes based on I^e values for elements of descretization $\bar{\Omega}^T$ are presented in the next section.

8. Summary and Conclusions

In this paper, we have considered a priori and a posteriori error estimations, a posteriori error computation, and convergence rates of the finite element computations for BVPs described by self-adjoint, non-self-adjoint, and nonlinear differential operators. Concepts of h-, p-, and k-versions and h-, p-, and k-convergences in finite element processes are presented and discussed. It is shown that a desired measure of error norm or residual functional versus degrees of freedom behavior has distinct features that can be classified as pre-asymptotic range, onset of asymptotic range, asymptotic range, onset of post-asymptotic range, and post-asymptotic range. The

significance and importance of these ranges in finite element computations has been discussed and demonstrated through three model problems described by self adjoint, non-self adjoint, and non-linear differential operators.

The a priori estimates only hold in asymptotic range and their derivation in the currently published literature are only valid for self adjoint operators in GM/WF when functional $B(\cdot,\cdot)$ is symmetric, thus GM/WF has best approximation property in *B*-norm. New work presented in this paper establishes correspondence between best approximation property of an integral form in some norm and the variational consistency of the integral form and demonstrates that when one exists the other is ensured. Thus, for establishing a priori error estimates, variational consistency becomes an essential property of the integral form. Of course best approximation property in some norm if it exists is equally good as best approximation property and variational consistency of integral form can not exist without each other, *i.e*. they co-exist. In case of GM/WF, VC integral form is possible for self adjoint operator and in case of LSP VC integral form is possible for all three classes of differential operators, hence a priori estimates for GM/WF for self adjoint operators and a priori estimates for LSP for all three classes of operators can be derived. The derivation of a priori error estimates presented in proposition 5.2 applies to GM/WF for self adjoint operators and in case of LSP for all three classes of operators as well as any other integral form resulting from a chosen method of approximation as long as the integral form is VC. Numerical studies for the model problems containing the three classes of operators confirm that when the integral form is VC, same a priori estimates and convergence rates hold. Thus, for the first time we have a priori error estimates for non-self adjoint and non-linear differential operators. Extensive numerical studies are presented for various *p* and *k* values for uniform h-refinements demonstrating that the theoretically derived convergence rates in a priori estimates are always in agreement with calculated values when the integral forms are VC. The a priori error estimates derived here also hold for 2D and 3D BVPs as long as the integral forms in these BVPs are variationally consistent. This can be confirmed numerically and is in agreement with published literature for self adjoint operators.

A posteriori error estimation based on the work presented here is viewed unnecessary when the approximation spaces are minimally conforming or of orders higher than minimally conforming due to the fact that when using such spaces a posteriori error computations of any desired quantity (for example I^e and *I*) that can help guide adaptivity is possible. I^e residual values for elements of $\bar{\Omega}^T$ are shown to be a perfect choice for adaptivity.

In short, VC integral form permits derivation of a priori error estimates and determination of convergence rates for all three classes of differential operators and use of minimally conforming spaces make a posteriori error estimation unnecessary and permit determination of desired a posteriori measures (such as I^e and *I*) that can be used to quantify errors in the currently computed solution and to design adaptive processes (presented in a followup paper). The same estimates and convergence rates hold for 2D and 3D BVPs when the integral forms are VC. The details are somewhat involved and have been presented in published literature for self-adjoint operators.

Acknowledgments

The first and third authors are grateful for the support provided by their endowed professorships during the course of this research. The computational infrastructure provided by the Computational Mechanics Laboratory (CML) of the Mechanical Engineering department of the University of Kansas is gratefully acknowledged. The financial support provided to the second author by the Naval Air Warfare Center is greatly appreciated.

References

[1] Surana, K.S., Ahmadi, A.R. and Reddy, J.N. (2002) The k-Version of Finite Element Method for Self-Adjoint Operators in BVP. *International Journal of Computational Engineering Science*, **3**, 155-218. http://dx.doi.org/10.1142/S1465876302000605

[2] Surana, K.S., Ahmadi, A.R. and Reddy, J.N. (2003) The k-Version of Finite Element Method for Non-Self-Adjoint Operators in BVP. *International Journal of Computational Engineering Science*, **4**, 737-812. http://dx.doi.org/10.1142/S1465876303002179

[3] Surana, K.S., Ahmadi, A.R. and Reddy, J.N. (2004) The k-Version of Finite Element Method for Non-Linear Operators in BVP. *International Journal of Computational Engineering Science*, **5**, 133-207. http://dx.doi.org/10.1142/S1465876304002307

[4] Surana, K.S., Allu, S. and Reddy, J.N. (2007) The k-Version of Finite Element Method for Initial Value Problems:

Mathematical and Computational Framework. *International Journal of Computational Engineering Science*, **8**, 123-136. http://dx.doi.org/10.1080/15502280701252321

[5] Babuska, I. and Rheinboldt, W.C. (1978) A Posteriori Error Estimates for the Finite Element Method. *International Journal for Numerical Methods in Engineering*, **12**, 1597-1615. http://dx.doi.org/10.1002/nme.1620121010

[6] Babuska, I. and Rheinboldt, W.C. (1978) Error Estimates for Adaptive Finite Element Computations. *SIAM Journal on Numerical Analysis*, **18**, 736-754. http://dx.doi.org/10.1137/0715049

[7] Babuska, I. and Rheinboldt, W.C. (1979) Adaptive Approaches and Reliability Estimations in Finite Element Analysis. *Computer Methods in Applied Mechanics and Engineering*, **17**, 519-540. http://dx.doi.org/10.1016/0045-7825(79)90042-2

[8] Babuska, I. and Rheinboldt, W.C. (1981) A Posteriori Error Analysis of Finite Element Solutions for One Dimensional Problems. *SIAM Journal on Numerical Analysis*, **18**, 435-463. http://dx.doi.org/10.1137/0718036

[9] Ainsworth, M. and Oden, J.T. (2000) A Posteriori Error Estimation in Finite Element Analysis. Wiley-Interscience, Hoboken.

[10] Szabo, B.A. and Babuska, I. (1991) Finite Element Analysis. Wiley-Interscience, Hoboken.

[11] Schwab, Ch. (1998) p and hp Finite Element Methods. Clarendon Press, Oxford.

[12] Guo, G. and Babuska, I. (1986) The hp Version of the Finite Element Method. Part 1: The Basic Approximation Results. Part 2: General Results and Applications. *Computational Mechanics*, **1**, 21-41, 203-220.

[13] Gui, W. and Babuska, I. (1986) The h, p and hp Versions of the Finite Element Method in One Dimension. Part 1: The Error Analysis of the p-Version. Part 2: The Error Analysis of the h- and hp-Versions. Part 3: The Adptive hp-Versions. *Numerische Mathematik*, **49**, 577-683. http://dx.doi.org/10.1007/BF01389733

[14] Ainsworth, M. and Senior, B. (1997) An Adaptive Refinement Strategy for hp-Finite Element Computations. *Applied Numerical Mathematics*, **26**, 165-178. http://dx.doi.org/10.1016/S0168-9274(97)00083-4

[15] Oden, J.T., Patra, A. and Feng, Y. (1992) An hp Adaptive Strategy. In Noor, A.K., Ed., *Adaptive Multilevel and Hierarchical Computational Strategies*, ASME Publication, 23-46.

[16] Rachowicz, W. (1989) An hp Finite Element Method for One-Irregular Meshes, Error Estimation and Mesh Refinement Strategy. PhD Thesis, University of Texas at Austin, Austin.

[17] Demkowicz, L. (2007) Computing with hp-Adaptive Finite Elements. Chapman and Hall/CRC, Boca Raton.

[18] Babuska, I. and Strouboulis, T. (2001) The Finite Element Method and Its Reliability. Oxford University Press Inc., New York.

[19] Jiang, B. (1998) The Least-Squares Finite Element Method: Theory and Applications in Computational Fluid Dynamics and Electromagnetics. Springer, Berlin. http://dx.doi.org/10.1007/978-3-662-03740-9

[20] Strouboulis, T. and Haque, K.A. (1992) Recent Experiences with Error Estimation and Adaptivity. Part I: Review of Error Estimators for Scalar Elliptic Problems. *Computer Methods in Applied Mechanics and Engineering*, **97**, 399-436. http://dx.doi.org/10.1016/0045-7825(92)90053-M

[21] Strouboulis, T. and Haque, K.A. (1992) Recent Experiences with Error Estimation and Adaptivity. Part II: Error Estimation for h-adaptive Approximations on Grids of Triangles and Quadrilaterals. *Computer Methods in Applied Mechanics and Engineering*, **100**, 359-430. http://dx.doi.org/10.1016/0045-7825(92)90090-7

[22] Apel, T. (1999) Anisoptropic Finite Elements: Local Estimates and Applications. Teubner.

[23] Surana, K.S., Stone, T., Reddy, J.N. and Romkes, A. (2011) Adaptivity in *hpk* Finite Element Processes. Proceedings of the 11th US Congress on Computational Mechanics (USNCCM-11), Minneapolis, 25-28 July 2011.

[24] Surana, K.S., Stone, T., Romkes, A. and Reddy, J.N. (2009) Adaptivity in Finite Element Processes in *hpk* Mathematical and Computational Framework. Proceedings of the 10th US Congress on Computational Mechanics (USNCCM-10), Columbus, 15-19 July 2009.

[25] Romkes, A., Bryant, C.M. and Reddy, J.N. (2010) A Posteriori Error Estimation of *hpk* FE Solutions of Linear Boundary Value Problems in Terms of Quantities of Interest. Proceedings of the International Conference on Multiscale Modeling and Simulation (ICMMS-2010), Guangzhou, 17-19 December 2010.

[26] Surana, K.S., Stone, T., Romkes, A. and Reddy, J.N. (2009) Adaptivity in Finite Element Processes in *hpk* Mathematical and Computational Framework. Proceedings of the ICCMES, Hyderabad, 8-10 January 2009.

[27] Romkes, A., Surana, K.S., Reddy, J.N. and Stone, T. (2008) Error Estimation for the K-Version of the Finite Element Method. Proceedings of the International Conference on Multiscale Modeling and Simulation (ICMMS-2008), Bangalore, 2-4 January 2008.

[28] Romkes, A., Reddy, J.N., Stone, T. and Surana, K.S. (2007) A Priori Error Estimation in *hpk* FE Analysis. Proceedings of the 9th US Congress on Computational Mechanics (USNCCM-9), San Francisco, 22-26 July 2007.

[29] Reddy, J.N. (2006) An Introduction to the Finite Element Method. 3rd Edition, McGraw Hill Inc., New York.

[30] Claes J. (1994) Numerical Solutions of Partial Differential Equations. Cambridge University Press, New York.

[31] White, R.E. (1985) An Introduction to the Finite Element Method with Applications to Nonlinear Problems. John Wiley & Sons, New York.

[32] Carey, G.F. and Oden J.T. (1983) Finite Elements: A Second Course, Volume II. Prentice Hall, Upper Saddle River.

[33] Oden, J.T. and Carey, G.F. (1983) Finite Elements: Mathematical Aspects. Prentice Hall, Upper Saddle River.

[34] Reddy, J.N. (1986) Applied Functional Analysis and Variational Methods in Engineering. McGraw Hill Company, New York.

[35] Surana, K.S. and Reddy, J.N. (2016) The Finite Element Method for Boundary Value Problems: Mathematics and Computations. CRC/Taylor and Francis, London. (In Press)