MATHEMATICS OF COMPUTATION
Volume 78, Number 268, October 2009, Pages 1951–1973
S 0025-5718(09)02241-8
Article electronically published on February 11, 2009

A NEW APPROACH TO RICHARDSON EXTRAPOLATION IN THE FINITE ELEMENT METHOD FOR SECOND ORDER ELLIPTIC PROBLEMS

M. ASADZADEH, A. H. SCHATZ, AND W. WENDLAND

Abstract. This paper presents a nonstandard local approach to Richardson extrapolation, when it is used to increase the accuracy of the standard finite element approximation of solutions of second order elliptic boundary value problems in $\mathbb{R}^N$, $N \geq 2$. The main feature of the approach is that it does not rely on a traditional asymptotic error expansion, but rather depends on a more easily proved weaker a priori estimate, derived in [19], called an asymptotic error expansion inequality. In order to use this inequality to verify that the Richardson procedure works at a point, we require a local condition which links the different subspaces used for extrapolation. Roughly speaking, this condition says that the subspaces are similar about a point, i.e., any one of them can be made to locally coincide with another by a simple scaling of the independent variable about that point. Examples of finite element subspaces that occur in practice and satisfy this condition are given.

1. Introduction, preliminaries and statement of main results

The purpose of this paper is to present a nonstandard local approach to proving the validity of the Richardson extrapolation procedure for increasing the accuracy of approximations. Here, this will be done in the context of improving the accuracy of finite element approximations of solutions of second order elliptic boundary value problems in $\mathbb{R}^N$, $N \geq 2$.

An outline of this paper is as follows: Section 1.1 contains some notation and preliminaries, and a discussion of what we shall call the "traditional" approach to verifying Richardson extrapolation. This approach relies on the establishment of an asymptotic error expansion for the problem which is difficult to obtain and have been derived only for special classes of finite elements and boundary value problems in $\mathbb{R}^N$, $N \geq 2$. A discussion of some previous work will be given there and in other sections. Section 1.2 contains some basic technical tools that will be needed for our approach. Here we shall state two results, both referred to as "asymptotic error expansion inequalities". Section 1.3 contains statements of the first results using the non-standard approach. There we shall first introduce a so-called similarity property of the subspaces. Roughly speaking, the similarity condition allows us to relate the finite element solutions for different mesh sizes to each other. In effect we shall reduce the problem to a new problem on a single

Received by the editor November 21, 2007 and, in revised form, October 11, 2008.
2000 *Mathematics Subject Classification.* Primary 65N15, 65N30, 35J25.
Key words and phrases. Richardson extrapolation, local estimates, asymptotic error expansion inequalities, similarity of subspaces, scalings, finite element method, elliptic equations.

grid to which the asymptotic error expansion inequality may be applied to obtain a higher order of convergence. The main aim of this paper is to present a method for establishing the validity of the Richardson extrapolation procedure without relying on exact asymptotic error expansions, but rather on the weaker, and more easily proved asymptotic error expansion inequalities just mentioned. For this we shall need the similarity condition. In this section, we shall treat problems for principal part second order differential operators with constant coefficients, where the main ideas are transparent and proofs are very simple. In particular, Theorems 1.1 and 1.2 are concerned with Richardson extrapolation using two subspaces to increase the accuracy for $(u-u_h)(\widehat{x})$ and $\frac{\partial}{\partial x_i}(u-u_h)(\widehat{x})$, respectively, at similarity points $\widehat{x}$ of the grid. Theorem 1.3 contains a generalization to more than two subspaces. Section 1.4 deals with the case of variable coefficients. The results here are similar to Theorems 1.1 and 1.2. In Section 2 we give examples of subspaces of finite elements having the local similarity property about a point. Section 3 contains applications of our main results to a variety of boundary value problems. Section 4 is devoted to the proof of Theorem 1.4. Finally, in our conclusion, in Section 5 (Appendix I), we recall the usual finite element assumptions used throughout the paper. Below we denote by C a general constant independent of the parameters involved in the estimates unless otherwise explicitly stated or clear from the context.

1.1. **Some preliminaries and a discussion of the "traditional" approach.** Let Ω be a bounded domain in $\mathbb{R}^N$, $N \geq 2$. For $m \geq 0$ an integer, $1 \leq p \leq \infty$ and $G \subseteq \Omega$, $W_p^m(G)$ denotes the usual Sobolev space of functions with distributional derivatives of order $\leq m$ which are in $L_p(G)$. Define the seminorms

$$|u|_{W_p^j(G)} = \begin{cases} \Big(\sum\limits_{|\alpha|=j} \|D^\alpha u\|_{L_p(G)}^p \Big)^{1/p} & \text{if} \quad 1 \leq p < \infty, \\ \sum\limits_{|\alpha|=j} \|D^\alpha u\|_{L_\infty(G)} & \text{if} \quad p = \infty, \end{cases}$$

and the norms

$$\|u\|_{W_p^m(G)} = \begin{cases} \Big(\sum\limits_{j=1}^m |u|_{W_p^j(G)}^p \Big)^{1/p} & \text{if} \quad 1 \leq p < \infty, \\ \sum\limits_{j=1}^m |u|_{W_\infty^j(G)} & \text{if} \quad p = \infty. \end{cases}$$

If $m \geq 0$, $W_p^{-m}(G)$ is the completion of $C_0^\infty(G)$ under the norm

$$\|u\|_{W_p^{-m}(G)} = \sup_{\substack{\psi \in C_0^\infty(G) \\ \|\psi\|_{W_q^m(G)}=1}} \int_G u\psi dx, \qquad \frac{1}{p} + \frac{1}{q} = 1.$$

Let $\Omega_d \subset\subset \Omega$ and consider the second order differential equation

$$Lu = -\sum_{i,j=1}^N \frac{\partial}{\partial x_j}\Big(a_{ij}(x)\frac{\partial u}{\partial x_i}\Big) + \sum_{i=1}^N b_i(x)\frac{\partial u}{\partial x_i} + c(x)u = f(x) \text{ in } \Omega_d. \tag{1.1}$$

Concerning L, it will be assumed that the coefficients are smooth and that L is uniformly elliptic on Ω_d, i.e., there exists a constant $m > 0$ such that for all $x \in \Omega_d$,

$$m|\zeta|^2 \leq \sum_{i,j=1}^{N} a_{ij}(x)\zeta_i\zeta_j \quad \text{for all} \quad \zeta \in \mathbb{R}^n. \tag{1.2}$$

Let $u \in W_2^1(\Omega_d)$ be, locally, a weak solution of (1.1), i.e., satisfy the interior equations

$$A(u,v) = \int_{\Omega_d} fv dx \quad \text{for all} \quad v \in \overset{\circ}{W}{}_2^1(\Omega_d). \tag{1.3}$$

Here

$$A(u,v) = \int_{\Omega_d} \Big(\sum_{i,j=1}^{N} a_{ij}(x)\frac{\partial u}{\partial x_i}\,\frac{\partial v}{\partial x_j} + \sum_{i=1}^{N} b_i(x)\frac{\partial u}{\partial x_i}v + c(x)uv\Big)dx. \tag{1.4}$$

We shall be interested in error estimates at points $\widehat{x}$ of $\Omega_0 \subset\subset \Omega_d \subset\subset \Omega$ when u is approximated by the finite element method, and begin by giving a rough description of the finite element subspaces. Let $0 < h < 1$ be a parameter, $r \geq 2$ an integer, and let $S_r^h(\Omega_d) \subset W_\infty^1(\Omega_d)$ denote a family of finite element spaces defined on Ω_d. The precise assumptions A.1–A.4 on these spaces will be given in the Appendix I. However, for the purposes of this introduction, the reader may think of $S_r^h(\Omega_d)$ as any one of a large variety of commonly used spaces of continuous functions, defined on a disjoint quasi-uniform partition of roughly size h (of a set which covers Ω_d), whose restriction to each set τ_j^h of the partition are polynomials of degree $\leq r-1$. $\mathring{S}_r^h(\Omega_d)$ will denote the subspace of $S_r^h(\Omega_d)$ of functions with support in Ω_d. Roughly speaking, functions $u \in W_\infty^r(\Omega_d)$ can be approximated to order h^r in $L_\infty(\Omega_0)$, and to order h^{r-1} in $W_\infty^1(\Omega_0)$, by elements of $S_r^h(\Omega_d)$. We first wish to briefly describe the so-called "traditional" approach to extrapolation using the subspaces $S_r^h(\Omega_d)$. To this end, suppose we are given two mesh sizes $\lambda_i h$, $i = 0, 1$, where, say, $\lambda_0 = 1$ and $\lambda_1 > 1$, and let $u_{\lambda_i h} \in S_r^{\lambda_i h}(\Omega_d)$ be approximations of u satisfying the local Galerkin equations

$$A(u - u_{\lambda_i h}, \varphi) = 0 \quad \text{for all} \quad \varphi \in \mathring{S}_r^{\lambda_i h}(\Omega_d),\ i = 0, 1. \tag{1.5}$$

In the "traditional" approach, the proof that Richardson extrapolation works at a point $\widehat{x}$ relies on the establishment of an exact asymptotic error expansion at $\widehat{x}$, which for some h_0 is valid for $0 < h \leq h_0$, of the form

$$e_h(\widehat{x}) \equiv u(\widehat{x}) - u_h(\widehat{x}) = h^r E_1(\widehat{x}, u) + R_h(\widehat{x}, u) = h^r E_1(\widehat{x}, u) + \mathcal{O}(h^{r+1}). \tag{1.6}$$

Here it is important that the error term $E_1(\widehat{x}, u)$ is independent of h. It then follows that

$$e_{\lambda_1 h}(\widehat{x}) \equiv u(\widehat{x}) - u_{\lambda_1 h}(\widehat{x}) = (\lambda_1)^r h^r E_1(\widehat{x}, u) + R_{\lambda_1 h}(\widehat{x}, u). \tag{1.7}$$

Multiplying (1.6) by γ_1 and (1.7) by γ_2 respectively, where γ_1 and γ_2 are the Richardson weights given by

$$\gamma_1 = \frac{\lambda_1^r}{\lambda_1^r - 1}, \quad \gamma_2 = -\frac{1}{\lambda_1^r - 1}, \tag{1.8}$$

and then adding the two equations yields

$$u(\widehat{x}) - (\gamma_1 u_h(\widehat{x}) + \gamma_2 u_{\lambda_1 h}(\widehat{x})) = \mathcal{O}(h^{r+1}). \tag{1.9}$$

This argument shows that, if the expansion (1.6) holds, Richardson extrapolation gains at least one order of accuracy when extrapolating with two approximate solutions from different subspaces. The difficulty with this approach occurs in establishing the expansion (1.6), which is not easy to obtain. Expansions of this type have been known for some time for some finite difference methods (c.f., e.g. Wasow [32] and Böhmer [5]). The main contributions to the finite element literature are due to Q. Lin and coworkers. In the case of finite elements with $r = 2$ (in particular, piecewise linear elements), such expansions were first derived at points x which are the vertices of a uniform triangulation of the plane in Lin and Wang [15], Lin and Lü [14] and Lin and Zhu [18]. Improvements and extensions of these expansions were then given in the paper by Blum, Lin, and Rannacher [4] which contains an excellent presentation of the derivation of the exact asymptotic expansion. Further results, some of which will be discussed in more detail in Sections 1.3 and 1.4, can be found in Lin and Wang [16], Lin and Xie [17], Blum [3], Rannacher [19], Chen and Lin [8], Xie [33], Wang [31], Lin [13], Ding and Lin [10], and Chen and Rannacher [9], where other references can be found. We recommend the survey articles by Rannacher [20] and Blum [2].

1.2. **Asymptotic error expansion inequalities.** Here we shall introduce some technical results which, together with an additional idea to be presented in Section 1.3 (a local similarity condition on the subspaces), will form the basis of our approach to Richardson extrapolation. We begin with the so-called "asymptotic error expansion inequalities". They are weaker than the exact asymptotic error expansion (1.6) discussed previously, but have been proven to hold for rather general classes of subspaces defined on irregular grids in $\mathbb{R}^N$, $N \geq 2$. One way of viewing them, which will not be totally apparent from the way they will be stated below, is that they contain terms that are of the form of the expansions given in (1.6) except that the error term E_1 is now a bounded function of $\widehat{x}$, u and h. It is this dependence on h which does not allow us to use them in the traditional approach. For given $x \in \Omega_d$ and $d > 0$, let

$$B_d(x) = \{y : |y - x| < d\},$$

where throughout this paper it will be assumed that $B_d(x) \subset \Omega_d$. Now let $u \in W_2^1(B_d(x))$ and $u_h \in S_r^h(B_d(x))$ satisfy

$$A(u - u_h, \varphi) = F(\varphi) \quad \text{for all} \quad \varphi \in \mathring{S}_r^h(B_d(x)). \tag{1.10}$$

Here $F(v)$ is a bounded linear functional on $\mathring{W}_1^1(B_d(x))$. In the error estimates for the variable coefficient case we shall need the following norms:

$$|||\varphi|||_1 := h^{-1}||\varphi||_{L_1(B_d(x_0))} + \left\| \frac{|x - x_0| + h}{h} \nabla\varphi \right\|_{L_1(B_d(x_0))}, \tag{1.11}$$

and for $\ell > k$, arbitrary but fixed integer, $k = 1, 2$:

$$|||\varphi|||_2 := \sum_{j=0}^{\ell} \sum_{|\alpha|=j} \left(\ln \frac{1}{h} \right)^{\overline{j}} \int_{B_d(x_0)} (|x - x_0| + h)^{j-k} |D^\alpha \varphi| \, dx, \tag{1.12}$$

where $\overline{j} = 1$ if $N = 2$, $k = 2$, and $\overline{j} = 0$ otherwise.

The following estimates can be found in Schatz [24].

Lemma 1.1. *Suppose that $r \geq 3$ and Assumptions A.1–A.4 (given in Appendix I) are satisfied. Let t be a nonnegative integer, $1 \leq p \leq \infty$, and s an integer, $r \leq s \leq 2r-2$. Let $x \in \Omega_0$ and $d \geq kh$ for some k sufficiently large. Let $u \in W^s_\infty(B_d(x))$ and $u_h \in S^h_r(B_d(x))$ satisfy (1.10), then*

$$\begin{aligned}\|u-u_h\|_{L_\infty(B_h(x))} \leq C\Big(\ln\frac{d}{h}\Big)^{\overline{s}}\Big[h^r\sum_{|\alpha|=r}|D^\alpha u(x)|+\cdots\\ +h^{s-1}\sum_{|\alpha|=s-1}|D^\alpha u(x)|+h^s\|u\|_{W^s_\infty(B_d(x))}\Big]\\ +C\Big[d^{-t-N/p}\|u-u_h\|_{W_p^{-t}(B_d(x))}\\ +\Big(\ln\frac{d}{h}\Big)^{\overline{s}}h|||F|||_{-1,B_d(x)}+\Big(\ln\frac{d}{h}\Big)|||F|||_{-2,B_d(x)}\Big].\end{aligned} \tag{1.13}$$

Here, $\overline{s}=1$ if $s=2r-2$ and $\overline{s}=0$ if $r\leq s<2r-2$. Furthermore, for $j=1,2$,

$$|||F|||_{-j,B_d(x)} = \sup_{\substack{\psi\in C_0^\infty(B_d(x))\\ |||\psi|||_j=1}} F(\psi). \tag{1.14}$$

Before discussing the meaning of this result, let us first state the corresponding result for first derivatives which follows:

Lemma 1.2. *Suppose that $r \geq 2$ and Assumptions A.1–A.4 (given in Appendix I) are satisfied. Let t be a nonnegative integer, $1 \leq p \leq \infty$, and s an integer, $r \leq s \leq 2r-2$. Let $x \in \Omega_0$ and $d \geq kh$ for some k sufficiently large, and $u \in W^s_\infty(B_d(x))$ and $u_h \in S^h_r(B_d(x))$ satisfy (1.10). Then*

$$\begin{aligned}\|u-u_h\|_{W^1_\infty(B_h(x))} \leq C\Big(\ln\frac{d}{h}\Big)^{\overline{\overline{s}}}\Big(h^{r-1}\sum_{|\alpha|=r}|D^\alpha u(x)|+\cdots\\ +h^{s-1}\sum_{|\alpha|=s}|D^\alpha u(x)|+h^s\|u\|_{W^{s+1}_\infty(B_d(x))}\Big)\\ +C\Big(d^{-1-t-N/p}\|u-u_h\|_{W_p^{-t}(B_d(x))}+(\ln\frac{d}{h})|||F|||_{-1,B_d(x)}\Big).\end{aligned} \tag{1.15}$$

Here $\overline{\overline{s}}=1$ if $s=2r-2$, and $\overline{\overline{s}}=0$ if $r\leq s<2r-2$.

Let us first briefly discuss the role of each of the terms in the estimates (1.13) and (1.15), and then contrast them with the asymptotic expansion (1.6). The first terms on the right of (1.13) and (1.15) are expansions in terms of powers of h, up to a possible order of h^{2r-2}. Notice that each term, except for the last, is multiplied by appropriate derivatives of u which, most importantly, are evaluated only at x. It is this expansion on which the proof of the extrapolation procedure will be based. The term involving the negative norm $\|u-u_h\|_{W_p^{-t}(B_d(x))}$ measures the influence of the solution from outside the domain $B_d(x)$, and will be referred to, as is common, as a "pollution" term. Estimates for terms of this type will be given for specific boundary value problems in the examples given in Section 3. In our theorems we will assume, a priori, that these have a certain rate of convergence. Finally, the terms involving F simply measure the effect of the linear functional F. Terms of this type will be useful in Section 3 when we consider perturbations of the problem (1.10), in particular, problems with lower order terms and with variable coefficients. We remark that (1.10) and (1.13) are local, and that global versions for a smooth

Neumann problem are given in Schatz [25]. We shall now turn to the use of these inequalities instead of exact asymptotic expansions in Richardson extrapolation. Before doing so we mention that (1.13) and (1.15) have been used to obtain new results for problems in a posteriori estimates [11] and superconvergence.

1.3. **Similarity of subspaces at a point and a nonstandard approach to Richardson extrapolation in a special case.** The method we will use to validate the extrapolation procedure is based on the observation that, assuming the pollution error is small enough, it follows from either (1.13) or (1.15) (in the case that $F = 0$) that higher order accuracy for the errors occurs if

$$D^{\alpha}u(x) = 0 \quad \text{for all} \quad |\alpha| = r.$$

We, of course, have no right to expect that this condition holds for the solution u. However, we will construct a function v for which it does hold and which is equal to u at a point $\widehat{x}$. From this we will be able to deduce that Richardson extrapolation attains higher order accuracy at the point $\widehat{x}$. In order to accomplish this, we shall require a condition which links the subspaces with different values of h that are being used for extrapolation. This condition will allow us to combine the approximate solutions for two different subspaces so that an appropriate linear combination of finite element solutions is now the finite element solution v_h of a new relevant problem on a single subspace. This is the key, namely it is this problem on a single mesh to which we may apply the asymptotic error expansion inequalities (1.13) and (1.15). The following assumption says that two subspaces are similar in a neighborhood of a point $\widehat{x}$, if they coincide under a simple scaling. More precisely,

A.5. Let $h < h_1 = \lambda_1 h$, with $\lambda_1 > 1$, and let $\widehat{x}$ be a fixed point in Ω with $d > 0$, so that $B_{\lambda_1 d}(\widehat{x}) \subset \Omega$. The subspace $S_r^{\lambda_1 h}(B_{\lambda_1 d}(\widehat{x}))$ is said to be similar to the subspace $S_r^h(B_d(\widehat{x}))$ (under scaling about $\widehat{x}$), if the mapping

$$(T\varphi)(x) = \varphi(\widehat{x} + \lambda_1(x - \widehat{x})) \tag{1.16}$$

is a one-to-one mapping of $S_r^{\lambda_1 h}(B_{\lambda_1 d}(\widehat{x}))$ onto $S_r^h(B_d(\widehat{x}))$.

Examples of subspaces satisfying **A.5** will be given in Section 2. We now turn to our first results.

In this section we consider the special case where $A(\cdot,\cdot)$ is of the form

$$A(u,v) = \int_{\Omega} \Big(\sum_{i,j=1}^{N} a_{ij} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} \Big) dx, \quad a_{ij} = \text{constants}, \quad i,j = 1,\dots,N. \tag{1.17}$$

We begin with extrapolation using two subspaces to obtain a higher order accurate approximation for $u(x)$ at similarity points $\widehat{x}$.

Theorem 1.1. *Suppose that*

(a) *$r \geq 3$, and that the conditions of Lemma 1.1 hold with $A(u,v)$ of the form (1.17).*
(b) *Given $\widehat{x}$, there exists $d > 0$ and $1 = \lambda_0 < \lambda_1$ such that $B_{\lambda_1 d}(\widehat{x}) \subset \Omega_d$.*
(c) *The similarity condition* **A.5** *holds for the pair $S_r^{\lambda_1 h}(B_{\lambda_1 d}(\widehat{x}))$ and $S_r^h(B_d(\widehat{x}))$.*
(d) *u_h and $u_{\lambda_1 h}$ satisfy*

$$A(u - u_{\lambda_j h}, \varphi) = 0 \quad \text{for all} \quad \varphi \in \mathring{S}_r^{\lambda_j h}(B_{\lambda_j d}(\widehat{x})),\ j = 0, 1. \tag{1.18}$$

Set

$$v_h(x) = \gamma_1 u_h(x) + \gamma_2 u_{\lambda_1 h}(\widehat{x} + \lambda_1(x - \widehat{x})), \tag{1.19}$$

where

$$\gamma_1 = \frac{\lambda_1^r}{\lambda_1^r - 1}, \quad \gamma_2 = -\frac{1}{\lambda_1^r - 1}.$$

i) *If* $u \in W_\infty^{r+1}(B_{\lambda_1 d}(\widehat{x}))$, *then*

$$\begin{aligned} |u(\widehat{x}) - v_h(\widehat{x})| \le{}& Ch^{r+1}\Big(\ln\frac{d}{h}\Big)^{\overline{r}} \|u\|_{W_\infty^{r+1}(B_{\lambda_1 d}(\widehat{x}))} \\ &+ Cd^{-t-N/p} \sum_{j=0}^{1} \|u - u_{\lambda_j h}\|_{W_p^{-t}(B_{\lambda_j h}(\widehat{x}))}. \end{aligned} \tag{1.20}$$

Here $\overline{r} = 1$ *if* $r = 3$ *and* $\overline{r} = 0$ *otherwise.* C *is independent of* u, u_h, $u_{\lambda_1 h}$, h, d *and* $\widehat{x}$.

ii) *If, in addition, for some* $t \ge 0$, $1 \le p \le \infty$, *and* $\sigma > 0$,

$$d^{-t-N/p}\|u - u_{\lambda_j h}\|_{W_p^{-t}(B_{\lambda_j d}(x))} \le Ch^{r+\sigma}, \quad j = 0, 1, \tag{1.21}$$

then

$$|u(\widehat{x}) - v_h(\widehat{x})| \le C \max\Big(h^{r+1}\Big(\ln\frac{1}{h}\Big)^{\overline{r}},\ h^{r+\sigma}\Big) \tag{1.22}$$

where C *is independent of* h, *and* $\widehat{x}$.

Let us remark that the estimate (1.22) says that the Richardson extrapolation procedure results in higher order accuracy at a similarity point $\widehat{x}$ provided the "pollution" error (1.21) also has an order of convergence greater than r. We also remark that the distance d associated with the region of similarity of the subspaces can be small with h. As some of the examples in Section 3 will show, there are cases where $d = Ch^\delta$ for some $0 < \delta < 1$ and extrapolation yields, modulo possible logarithmic factors, a rate of convergence of h^{r+1} . It is important to remark that Theorem 1.1 is restricted to $r \ge 3$, which excludes the piecewise linear case ($r = 2$). One factor that contributes to this is that the similarity condition is only assumed to hold locally in a ball about a point $\widehat{x}$ and there are no other mesh restrictions outside of $B_d(\widehat{x})$ other than quasi-uniformity. In general, then, the best rate of convergence, in the piecewise linear case, in any negative norm that can be expected under our assumptions is $O(h^2)$, which is essentially attained without extrapolation. Let us briefly discuss some results obtained using exact asymptotic expansions which are generally proved under more stringent conditions. As previously mentioned, the paper by Blum, Lin, and Rannacher [4] gives an excellent derivation of such an expansion that is an extension of the paper by Lin and Wang [15]. Among other things, they treat a Dirichlet problem with homogeneous boundary data for Poisson's equation on a smooth two-dimensional domain using piecewise linear finite elements. The mesh is uniform except for a layer near the boundary and they prove an $O(h^3)$ rate of convergence at interior vertices for Richardson extrapolation. In Lin and Xie [17], an exact asymptotic expansion is established at interior vertices of a grid which is a piecewise smooth mapping of a piecewise regular grid in the plane. An $O(h^4)$ rate of convergence for Richardson extrapolation is established. The analyses of the methods are rather delicate and it seems hard to generalize to higher order elements and more space dimensions. Before proving Theorem 1.1, we

remark that the case $r = 2$ is included in Theorem 1.2 where extrapolation is used to achieve higher accuracy for derivatives.

Proof of Theorem 1.1. For convenience we take $\widehat{x}$ to be the origin. As previously remarked, the first important observation is that we can use the similarity assumption **A.5** to link the two finite element solutions via scaling, and reduce the problem to one on a single grid. In fact, we shall first show that for arbitrary scalars γ_1 and γ_2, the function

$$v_h(x) = \gamma_1 u_h(x) + \gamma_2 u_{\lambda_1 h}(\lambda_1 x) \in S_r^h(B_d(\widehat{x})) \tag{1.23}$$

is, locally, a finite element approximation of the function

$$v(x) = \gamma_1 u(x) + \gamma_2 u(\lambda_1 x), \tag{1.24}$$

on the "finer" grid. In fact, we will show that for any choice of scalars γ_1 and γ_2,

$$A(v - v_h, \varphi) = 0 \quad \text{for all} \quad \varphi \in \mathring{S}_r^h(B_d(x)). \tag{1.25}$$

To prove (1.25) first note that, by our similarity assumption, $u_{\lambda_1 h}(\lambda_1 x) \in S_r^h(B_d(\widehat{x}))$. We now show that

$$A(u(\lambda_1 x) - u_{\lambda_1 h}(\lambda_1 x), \varphi) = 0 \quad \text{for all} \quad \varphi \in \mathring{S}_r^h(B_d(\widehat{x})).$$

In fact, by the trivial change of variables $y = \lambda_1 x$ and (1.18)

$$\begin{aligned} &A(u(\lambda_1 x) - u_{\lambda_1 h}(\lambda_1 x), \varphi) \\ &\quad = \int_{B_d(\widehat{x})} \sum_{i,j=1}^{N} a_{ij} \frac{\partial (u(\lambda_1 x) - u_{\lambda_1 h}(\lambda_1 x))}{\partial x_i} \, \frac{\partial \varphi(x)}{\partial x_j} dx \\ &\quad = \lambda_1^{2-N} \int_{B_{\lambda_1 d}(\widehat{x})} \sum_{i,j=1}^{N} a_{ij} \frac{\partial (u(y) - u_{\lambda_1 h}(y))}{\partial y_i} \, \frac{\partial \varphi(\frac{y}{\lambda_1})}{\partial y_j} dy = 0. \end{aligned}$$

Here we have used the fact that $\varphi\big(\frac{y}{\lambda_1}\big) \in S_r^h(B_{\lambda_1 d}(\widehat{x}))$. It follows from (1.16) that (1.25) is satisfied. Hence by linearity, v_h and v given by (1.23) and (1.24) also satisfy (1.25). By this procedure, we have connected the two approximate solutions to a new problem on a single grid where we can apply the asymptotic error expansion (1.13) to the function $v(x) - v_h(x)$. From Lemma 1.1 we obtain

$$\begin{aligned} &|v(\widehat{x}) - v_h(\widehat{x})| \\ &\quad \le C\Big(\ln \frac{1}{h}\Big)^{\overline{r}} \Big(h^r \sum_{|\alpha|=r} |D^\alpha v(\widehat{x})| + h^{r+1} \|v\|_{W_\infty^{r+1}(B_d(\widehat{x}))}\Big) \\ &\qquad + C d^{-N/p-t} \|v - v_h\|_{W_p^{-t}(B_d(\widehat{x}))}. \end{aligned} \tag{1.26}$$

Note that $v(\widehat{x}) = u(\widehat{x})$ with the choice

$$\gamma_1 + \gamma_2 = 1. \tag{1.27}$$

As observed previously, higher order accuracy $\mathcal{O}(h^{r+1})$ in the expansion term for the local interpolation error in (1.26) will occur if $(D^\alpha v)(\widehat{x}) = 0$, for all multi-indices α, with $|\alpha| = r$. A simple computation now yields

$$(D^\alpha v)(\widehat{x}) = (\gamma_1 + \lambda_1^r \gamma_2)(D^\alpha u)(\widehat{x}) = 0, \tag{1.28}$$

for all $|\alpha| = r$, when γ_1 and γ_2 are chosen to satisfy

$$\gamma_1 + (\lambda_1)^r \gamma_2 = 0. \tag{1.29}$$

The unique solution of (1.27) and (1.29) is given by (1.8), which are the Richardson weights. The inequality (1.20) follows immediately from (1.26) by using the triangle inequality on the last term. This completes the proof of Theorem 1.1. □

We now turn to the problem of extrapolating for higher order accuracy for first derivatives. If $\widehat{x}$ is a point of the mesh where $\frac{\partial v_h}{\partial x_i}$ is discontinuous, then we define

$$\frac{\partial \widetilde{v}_h}{\partial x_i}(\widehat{x},\beta) = \lim_{s\to 0} \frac{\partial v_h}{\partial x_i}(\widehat{x}+s\beta), \tag{1.30}$$

where $\beta = (\beta_1,\dots,\beta_N)$ is any unit vector chosen so that for s sufficiently small, say, $0 < s \le s_0$, $\dfrac{\partial v_h}{\partial x_i}$ exists and has a limit as $s \to 0$. There may be many possible choices of $\dfrac{\partial \widetilde{v}_h}{\partial x_i}(\widehat{x},\beta)$. Obviously, $\dfrac{\partial \widetilde{v}_h}{\partial x_i}(\widehat{x},\beta) = \dfrac{\partial v_h}{\partial x_i}(\widehat{x})$ at points $\widehat{x}$ where $\dfrac{\partial v_h}{\partial x_i}$ is continuous. The analogue of Theorem 1.1 for first derivatives is as follows:

Theorem 1.2. *Suppose that the conditions of Theorem 1.1 hold, except that now Lemma 1.2 holds with $r \ge 2$, and γ_1 and γ_2 are given by*

$$\gamma_1 = \frac{\lambda_1^{r-1}}{\lambda_1^{r-1}-1}, \quad \gamma_2 = -\frac{1}{\lambda_1(\lambda_1^{r-1}-1)}. \tag{1.31}$$

i) *Then for any $i = 1,\dots,N$,*

$$\begin{aligned}\Big|\frac{\partial u(\widehat{x})}{\partial x_i} - \frac{\partial \widetilde{v}_h(\widehat{x},\beta)}{\partial x_i}\Big| &\le C\Big(\ln\frac{d}{h}\Big)^{\overline{r}} h^r \|u\|_{W_\infty^{r-1}(B_{\lambda_1 d}(\widehat{x}))} \\ &\quad + Cd^{-1-t-N/p}\sum_{j=0}^{1}\|u-u_{\lambda_j h}\|_{W_p^{-t}(B_{\lambda_j d}(\widehat{x}))},\end{aligned} \tag{1.32}$$

where $\overline{r} = 1$ if $r = 2$ and $\overline{r} = 0$ otherwise. C is independent of u, u_h, $u_{\lambda_1 h}$, d, h, and $\widehat{x}$.

ii) *If, in addition, for some $t \ge 0$, $1 \le p \le \infty$, and $\sigma > 0$,*

$$d^{-1-t-N/p}\|u-u_{\lambda_j h}\|_{W_p^{-t}(B_{\lambda_j d}(\widehat{x}))} \le Ch^{r-1+\sigma},\ j = 0,1, \tag{1.33}$$

then

$$\Big|\frac{\partial u(\widehat{x})}{\partial x_i} - \frac{\partial \widetilde{v}_h(\widehat{x},\beta)}{\partial x_i}\Big| \le C(u)\max\Big(h^r\Big(\ln\frac{1}{h}\Big)^{\overline{r}}, h^{r-1+\sigma}\Big). \tag{1.34}$$

Proof. Again, for simplicity, we take $\widehat{x}$ to be the origin. We have from the proof of Theorem 1.1 that (1.23) holds, and therefore Lemma 1.2 may be applied. It follows that for any direction β along which $\lim_{s\to 0}\frac{\partial v_h(\widehat{x}+s\beta)}{\partial x_i}$ exists,

$$\begin{aligned}&\Big|\frac{\partial v(\widehat{x})}{\partial x_i} - \frac{\partial \widetilde{v}_h(\widehat{x},\beta)}{\partial x_i}\Big| \\ &\quad\le C\Big(\ln\frac{d}{h}\Big)^{\overline{r}}\Big(h^{r-1}\sum_{|\alpha|=r}|D^\alpha v(\widehat{x})| + h^r\|v\|_{W_\infty^{r+1}(B_d(\widehat{x}))}\Big) \\ &\qquad + Cd^{-1-t-N/p}\|v-v_h\|_{W_p^{-t}(B_d(\widehat{x}))}.\end{aligned} \tag{1.35}$$

From (1.28) we have $D^\alpha v(\widehat{x}) = 0$ for all $|\alpha| = r$ when γ_1 and γ_2 are chosen to satisfy

$$\gamma_1 + (\lambda_1)^r\gamma_2 = 0. \tag{1.36}$$

Furthermore, we want $\left(\frac{\partial v}{\partial x_i}\right)(\widehat{x}) = \left(\frac{\partial u}{\partial x_i}\right)(\widehat{x})$. For any $i = 1, \ldots, m$ we have

$$\frac{\partial v}{\partial x_i}(\widehat{x}) = (\gamma_1 + \lambda_1\gamma_2)\frac{\partial u}{\partial x_i}(\widehat{x}),$$

and therefore we choose

$$\gamma_1 + \lambda_1\gamma_2 = 1. \tag{1.37}$$

The equations (1.36) and (1.37) lead to (1.32), which proves part i). The inequality (1.34) follows trivially from (1.32), and (1.33), which completes the proof of Theorem 1.2. □

The reason that we have first treated the case when $A(\cdot,\cdot)$ is of the form (1.17) is that our results are better than those for the general variable coefficient case, which will be treated in Theorem 1.4. It is specifically the present case that we can generalize to obtain higher order accuracy, up to order h^{2r-2}, by extrapolating with an appropriate number of subspaces. We shall now state this generalization. Let $1 = \lambda_0 < \lambda_1 < \cdots < \lambda_m$ for m an integer, $1 \le m \le r-2$. Suppose that, at the point $\widehat{x}$, each pair of subspaces $S_r^{\lambda_j h}(B_{\lambda_j d}(\widehat{x}))$ and $S_r^h(B_d(\widehat{x}))$, $j = 1, \ldots, m$ satisfy **A.5** for some $d \ge kh$ (k sufficiently large). Suppose further that for A of the form (1.17) the $u_{\lambda_j h} \in S_r^{\lambda_j h}(B_{\lambda_j d}(\widehat{x}))$ satisfies

$$A(u - u_{\lambda_j h}, \varphi) = 0 \quad \text{for all} \quad \varphi \in S_r^{\lambda_j h}(\widehat{x}),\ j = 0, \ldots, m. \tag{1.38}$$

We seek an approximation to u of the form

$$v_h(x) = \sum_{j=0}^{m} \gamma_j u_{\lambda_j h}(\widehat{x} + \lambda_j(x - \widehat{x})) \in S_r^h(B_d(\widehat{x})). \tag{1.39}$$

Theorem 1.3. i) *Suppose the conditions of Lemma 1.1 hold and that, with λ_j and $S_r^{\lambda_j h}(B_{\lambda_j d}(\widehat{x}))$ as above, the functions $u - u_{\lambda_j h}$ satisfy the equations (1.38). Let γ_j, $j = 0, \ldots, m$ be the unique solution of the system of equations*

$$\sum_{j=0}^{m} \gamma_j = 1, \tag{1.40}$$

$$\sum_{j=0}^{m} \lambda_j^{r+\ell}\gamma_j = 0 \quad \textit{for} \quad \ell = 0, \ldots, m-1. \tag{1.41}$$

Furthermore, let $u \in W_\infty^{r+m}(B_{\lambda_m d}(\widehat{x}))$, and suppose there exists a $\sigma > 0$ such that for $j = 0, \ldots, m$,

$$d^{-t-N/p}\|u - u_{\lambda_j h}\|_{W_p^{-t}(B_{\lambda_j d}(\widehat{x}))} \le Ch^{r+\sigma}. \tag{1.42}$$

Then

$$|(u - v_h)(\widehat{x})| \le C \max\left(h^{r+m}\Big(\ln\frac{1}{h}\Big)^{\overline{m}}, h^{r+\sigma}\right). \tag{1.43}$$

Here $\overline{m} = 1$ if $m = r-2$ and $\overline{m} = 0$ otherwise and C is independent of h and $\widehat{x}$.

ii) *Assume the conditions of i) hold, except now Lemma 1.2 holds, and the* γ_j, $j = 0, \dots, m$ *satisfy*

$$\sum_{j=0}^{m} \lambda_j \gamma_j = 1, \tag{1.44}$$

$$\sum_{j=0}^{m} \lambda_j^{r+\ell} \gamma_j = 0, \quad \ell = 0, \dots, m-1. \tag{1.45}$$

Furthermore, suppose that for $j = 0, \dots, m$ *there is* $\sigma > 0$ *such that*

$$d^{-1-t-N/p} \|u - u_{\lambda_j h}\|_{W_p^{-t}(B_{\lambda_j d}(\widehat{x}))} \le C h^{r-1+\sigma}. \tag{1.46}$$

Then

$$\Big| \frac{\partial u(\widehat{x})}{\partial x_i} - \frac{\partial \widetilde{v}_h(\widehat{x}, \beta)}{\partial x_i} \Big| \le C \max \Big(h^{r-1+m} \Big(\ln \frac{1}{h} \Big)^{\overline{m}}, h^{r-1+\sigma} \Big), \tag{1.47}$$

where $\overline{m} = 1$ *if* $m = r-1$, *and* $\overline{m} = 0$ *otherwise.*

The proof is an easy generalization of the proofs of Theorems 1.1 and 1.2 and will be left to the reader.

1.4. **Equations with variable coefficients.** Here we shall consider the case when the differential operators have variable coefficients, and where we are extrapolating with only two subspaces. Here we will be concerned with bilinear forms A of the form

$$A(u,v) = \int_{\Omega_d} \Big(\sum_{i,j=1}^{N} a_{ij}(x) \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} + \sum_{i=1}^{N} b_i(x) \frac{\partial u}{\partial x_i} v + c(x) u v \Big) dx, \tag{1.48}$$

where the $a_{ij}(x)$, $b_i(x)$ and $c(x)$ are smooth functions. We first have the following extension of Theorems 1.1 and 1.2.

Theorem 1.4. i) *Let* $r \ge 3$, *and suppose that the conditions of Theorem 1.1 hold except that* $A(\cdot,\cdot)$ *is of the form* (1.48). *Then, with* $v_h(x)$ *defined by* (1.19),

$$|u(\widehat{x}) - v_h(\widehat{x})| \le C(u) \Big(\ln \frac{1}{h} \Big)^2 \max(h^{r+1}, h^{r+\sigma}). \tag{1.49}$$

ii) *Suppose that the conditions of Theorem 1.2 hold except that now* $A(\cdot,\cdot)$ *is of the form* (1.48). *Then, with* $\frac{\partial \widetilde{v}_h(\widehat{x},\beta)}{\partial x_i}$ *defined by* (1.30), *and* γ_1 *and* γ_2 *satisfying* (1.31),

$$\Big| \frac{\partial u(\widehat{x})}{\partial x_i} - \frac{\partial \widetilde{v}_h(\widehat{x}, \beta)}{\partial x_i} \Big| \le C(u) \Big(\ln \frac{1}{h} \Big) \max(h^r, h^{r-1+\sigma}), \ i = 1, \dots, N. \tag{1.50}$$

The proof will be given in Section 4.

2. Examples of subspaces which are similar under scaling about a point

In this section we shall give some simple examples of subspaces which satisfy the local similarity condition **A.5**. Although it will not be stated explicitly each time, we wish to emphasize that this condition is local, i.e. for some $d > 0$ and $\lambda > 0$, **A.5** is required to only hold in some balls $B_d(\widehat{x})$ and $B_{\lambda d}(\widehat{x})$ respectively of radius d and λd about the point $\widehat{x}$. As is customary, we shall define our examples of finite element spaces in terms of

1) a disjoint partition of a set in R^N, and

2) a space of functions defined relative to the partition.

Specific examples will be given in a moment, but in general, for a given $0 < h < 1$, let $\{\tau_h^k\}$ denote a quasi-uniform partition of set which contains a ball $B_d(\widehat{x})$, for some $d > 0$. Here it will be assumed that the sets $\tau_h^j,\ j = 1, ..., m(h)$, are disjoint and roughly of size h. Our examples of finite element spaces $S_r^h(B_d(\widehat{x}))$, for $r \geq 2$, will be taken to be subspaces of the space of functions φ that are in $C^l(B_d(\widehat{x}))$, for some integer $0 \leq l$, and which on each set τ_h^j are of the form

$$\varphi = \sum_{\alpha \in I} c_\alpha x^\alpha, \quad c_\alpha \text{ constants.} \tag{2.1}$$

Here, $\alpha = (\alpha_1, ..., \alpha_N)$ is a multi-index, and I is some fixed index set, of which specific examples will be given below. In order to show that, for a given $\lambda > 0$, two subspaces $S_r^h(B_d(\widehat{x}))$ and $S_r^{\lambda h}(B_{\lambda d}(\widehat{x}))$ satisfy **A.5**, it is sufficient to show that they possess the following two properties:

Property i). Under the scaling $x \to \widehat{x} + \lambda(x - \widehat{x})$, the partition $\{\tau_h^j \bigcap B_d(\widehat{x})\}$ of $B_d(\widehat{x})$, coincides with the partition $\{\tau_{\lambda h}^j \bigcap B_{\lambda d}(\widehat{x})\}$ of $B_{\lambda d}(\widehat{x})$.

Property ii). For a given index set I, the set of functions φ of the form (2.1) is invariant in form under the same scaling as in **i)**.

It is easy to verify that the following two examples of classes of functions of the form (2.1) satisfy **Property ii)**.

$\mathcal{P}_{r-1}$: The set of functions of the form (2.1), with $\alpha \in I$ if and only if $|\alpha| = \sum_{i=1}^N \alpha_i \leq r - 1$. Thus, on each τ_h^k, the functions of this class are the set of all polynomials of degree $\leq r - 1$. For $r = 2$ these are just the linear functions and for $r = 3$ the quadratic functions, etc.

$\mathcal{Q}_{r-1}$: The set of functions of the form (2.1), with $\alpha \in I$ if and only if $\alpha_i \leq r-1$, $i = 1, ..., N$. Examples of such functions are, for $r = 2$, the piecewise bilinear functions (in two dimensions) and the trilinear functions (in three dimensions), etc., and for $r = 3$ the bi-quadratics functions (in two dimensions), etc.

Remark. One general way of constructing two subspaces in R^N satisfying **A.5** for any N is as follows: Simply take any quasi-uniform mesh together with any set of functions of the form (2.1) defined on it, and scale them by a factor λ_1 about any fixed point $\widehat{x}$ of the mesh. Then **Properties i) and ii)** above are automatically satisfied. We shall now give some more specific related examples of such subspaces which arise in practice, and which are defined in the neighborhood of certain points, when so-called "nested subspaces" are constructed for multigrid methods. We mention, however, that the nested spaces that will be constructed below have other properties at symmetry points which could be exploited to obtain other extrapolation results.

2.1. **A. Nested subspaces in two dimensions with triangular elements.** Consider a quasi-uniform triangulation $\{\tau_{h_0}^k\}$ of size h_0 (henceforth referred to as the coarse mesh size), of a polygonal domain Ω . Let $h_j = 2^{-j}h_0, j = 0, \ldots, J$ and construct the sequence of grids $\mathcal{M}_j$ in the usual way by, starting with the finite element space defined on the coarse mesh $\mathcal{M}_0$, subdividing each triangle of the previous mesh into four new triangles by introducing new vertices at the midpoints of the sides. This is done until $h_J = h$ for some integer J. We are now in a position

to give our first example of finite element subspaces and then identify the points at which they satisfy **A.5**.

Example 2.1. Let $S_r^{h_j}(\Omega)$ be a space of continuous functions on Ω whose restrictions to each triangle in $\mathcal{M}_j$ are elements of $\mathcal{P}_{r-1}$, the set of polynomials of degree $\leq r-1$ in two variables. As remarked above these may be taken to be the usual Lagrange or Hermite finite elements relative to this triangulation (cf., e.g. Brenner and Scott [6]).

Let $\mathcal{V}_{h_j}$ denote the set of vertices and $\mathcal{E}_{h_j}$ the set of edges of the triangulation $\mathcal{M}_j$. We shall separate the similarity points of these subspaces into two categories.

SP.1. Similarity points which are interior points of triangles of the coarse mesh. Consider a coarse mesh triangle $\tau_{h_0}^k$, and let $\widehat{x}$ be an interior point such that $\widehat{x} \in \mathcal{V}_h$, and $\widehat{x} \in \mathcal{V}_{2h}$, (i.e., $\widehat{x}$ is a vertex of triangles in both $\mathcal{M}_J$ and $\mathcal{M}_{J-1}$), and such that $dist(\widehat{x}, \mathcal{E}_0) = 2d$, where $h < d$. Then $S_r^h(B_d(\widehat{x}))$ is similar to $S_r^{2h}(B_{2d}(\widehat{x}))$. If $dist(\widehat{x}, \mathcal{E}_0) = 4d$, then $S_r^h(B_d(\widehat{x}))$ is similar to $S_r^{4h}(B_{4d}(\widehat{x}))$, etc. We remark that any interior point of the triangles of the coarse mesh, which is also the vertex of a finer mesh, eventually becomes a similarity point of the subspace provided enough further subdivisions are performed. We also remark that if we had created finer meshes by successively subdividing edges into three parts instead of two and creating 9 new triangles, then the interior points $\widehat{x}$ with $\widehat{x} \in \mathcal{V}_h$, $\widehat{x} \in \mathcal{V}_{3h}$, and $dist(\widehat{x}, \mathcal{E}_0) = 3d$ where $h < d$ are similarity points, then $S_r^h(B_d(\widehat{x}))$ is similar to $S_r^{3h}(B_{3d}(\widehat{x}))$. This is also true if $\widehat{x}$ is the midpoint of an edge of two successive grids with $dist(\widehat{x}, \mathcal{E}_1) \geq 3d$ and $h < d$.

SP.2. Similarity points on the boundary of triangles of the coarse triangulation. Let $\widehat{x} \in \mathcal{V}_h \bigcap \mathcal{E}_{h_0}, \widehat{x} \in \mathcal{V}_{2h} \bigcap \mathcal{E}_{h_0}, dist(\widehat{x}, \mathcal{V}_{h_0}) \geq 2d$, where $d \geq h$, then it is easily seen that $S_r^h(B_d(\widehat{x}))$ and $S_r^{2h}(B_{2d}(\widehat{x}))$ satisfy **A.5**. On the other hand, if $\widehat{x} \in \mathcal{V}_{h_0}$, i.e., is a vertex of the coarse triangulation, then $\widehat{x} \in \mathcal{V}_j$ for all, $j = 0, ..., J$. This means that $\widehat{x}$ is a vertex for each mesh, and is always a similarity point with $S_r^h(B_d(\widehat{x}))$ similar to $S_r^{2h}(B_{2d}(\widehat{x}))$.

Example 2.2. As mentioned above, the similarity condition **A.5** is local and is satisfied by some nested subspaces defined on refined grids. We illustrate this with an example of a nested family of geometrically refined grids that generalize Example 2.1, and are often used to resolve corner problems. For simplicity we assume that Ω contains a sectorial neighborhood of the origin given in polar coordinates $\{(r, \theta) : 0 \leq r \leq 1 \text{ and } \theta \leq \alpha\}$ for some $0 < \alpha \leq 2\pi$. For some fixed K we decompose this region into the union of the annuli

$$\Omega_k = \{(r, \theta) : 2^{-(k+1)} < r \leq 2^{-k}, 0 \leq \theta \leq \alpha\} \quad k = 0, \ldots, K,$$

and the ball

$$\Omega_I = \{(r, \theta) : r \leq 2^{-(k+1)}, 0 \leq \theta \leq \alpha\}.$$

We construct a nested sequence of meshes $\mathcal{M}_j$, for $j = 0, \ldots, J$, which for each j is a geometrically refined grid defined as follows: On the annulus Ω_k the mesh is a mesh of the type given in Example 2.1, this time with a mesh size $h_j d_k$, where $h_j = 2^{-j} h_0$ and $d_k = 2^{-(k+1)}$. Furthermore, on the ball Ω_I, the mesh size is $h_j 2^{-(k+1)}$. We leave the identification of similarity points of these meshes to the reader since they are locally the same as those given in Example 2.1.

2.2. **B. Subspaces in two dimensions with rectangular elements.** Finally, for the two-dimensional case, we consider a partition consisting of rectangles of roughly size h_0, which covers a domain Ω, and which we again denote by $\mathcal{M}_0$. Similar to the triangular case, the partitions $\mathcal{M}_j, j = 0, ..., J$ are obtained by successively introducing new nodes at the midpoints of the edges and subdividing each rectangle into four equal rectangles. We shall now give our third example of finite element subspaces and then briefly identify the similarity points at which they satisfy **A.5**.

Example 2.3. Let $S_r^h(\Omega)$ be the space of functions which are at least continuous on Ω, and whose restrictions to each solid are elements of $\mathcal{Q}_{r-1}$ (as defined above). The similarity points $\widehat{x}$, given in **SP.1** and **SP.2** above for triangular elements, remain similarity points of the rectangular grids if we replace the word triangle with rectangle there and in the definitions of $\mathcal{V}_{h_j}$ and $\mathcal{E}_{h_j}$. The centers of the rectangles are also similarity points if each edge is equally subdivided into 3 parts, thus leading to a rectangle with 9 equal rectangles.

Example 2.4. Serendipity elements. We mention in passing that a subspace of the subspaces given in Example 2.3 above are the Serendipity elements. As an example of these we consider the piecewise quadratic case. These are then the continuous functions, which on each rectangle are in $\mathcal{Q}_2$, whose value at the center of the rectangle is a fixed linear combination of the values at the vertices and midpoints of the edges. We again refer the reader to [6] for an exposition, and leave the verification that the similarity points are the same as those in Example 2.2 to the reader.

2.3. **C. Subspaces in three dimensions.** The subspaces defined for two-dimensional rectangular elements in Examples 2.2 and 2.3 can be easily generalized to the three- (or higher-) dimensional case without any difficulties. For example, the sequence of meshes $\mathcal{M}_j, j = 0, ..., J$ can be obtained by successively introducing new nodes at the midpoints of the edges and the centers of faces of rectangular solids. This results in 8 new rectangular solids. Unfortunately, a similar construction for tetrahedra does not lead to meshes which satisfies **Property i)** above. We shall now briefly discuss the rectangular case.

Example 2.5. We obtain an obvious generalization of Example 2.3 by taking $S_r^h(\Omega)$ to be the space of functions which are at least continuous on Ω and whose restrictions to each solid are elements of $\mathcal{Q}_{r-1}$. If we had equally subdivided the edges into 3 parts (instead of 2), and the faces into 9 equal parts (instead of 4), then in addition to the vertices, the midpoints of the edges and the centers of the faces are similarity points. There are other three-dimensional elements with which similar spaces can be constructed. One such class is the so-called prismatic elements which are sort of hybrid being triangular in one direction and rectangular in another.

3. Applications of the main results to specific boundary value problems

In this section we shall apply the main results of this paper to specific boundary value problems. For simplicity we shall treat the same problems as treated in Schatz, Sloan and Wahlbin [27] (Section 3) and Schatz [26]. In all of our examples, $\widehat{x}$ will

be an interior point of Ω which is a similarity point of the subspaces $S_r^h(B_d(\widehat{x}))$ and $S_r^{\lambda_1 h}B_{\lambda_1 d}((\widehat{x}))$, where $B_{\lambda_1 d}((\widehat{x})) \subset \Omega$.

Example 3.1. A homogeneous Neumann problem on a smooth domain in $\mathbb{R}^N$, $N \geq 2$. Let Ω be a smooth domain in $\mathbb{R}^N$, $N \geq 2$, with boundary $\partial\Omega$ and consider the boundary value problem

$$\begin{cases} Lu & = -\sum_{i,j=1}^N \frac{\partial}{\partial x_j}\Big(a_{ij}\ \frac{\partial u}{\partial x_i}\Big) + \sum_{i=1}^N b_i\ \frac{\partial u}{\partial x_i} + Cu = f \quad \text{in } \Omega, \\ \frac{\partial u}{\partial n_L} & = 0 \qquad \text{on } \partial\Omega, \end{cases}$$

where $\frac{\partial u}{\partial n_L}$ denotes the exterior conormal derivative. Suppose that the corresponding form (1.48) is coercive over $W_2^1(\Omega)$ and that the elements are taken to fit the boundary exactly. Let $u_h \in S_r^h(\Omega)$ satisfy the global equations

$$A(u - u_h, \varphi) = 0 \quad \text{for all} \quad \varphi \in S_r^h(\Omega).$$

It was shown in Rannacher and Scott [21] in 2 dimensions, and for arbitrary dimensions in Schatz and Wahlbin [28], that

$$\|e\|_{W_\infty^1(\Omega)} \leq Ch^{r-1}.$$

Using a duality argument (see [6]) we have that

$$\|u - u_h\|_{W_\infty^{2-r}(\Omega)} \leq C\Big(\ln\frac{1}{h}\Big)^2 h^{2r-2}. \tag{3.1}$$

If $r \geq 3$ and we extrapolate using two subspaces, then with variable coefficients, Theorem 1.4 yields, after disregarding logarithmic factors,

$$|(u - v_h)(\widehat{x})| \simeq C(h^{r+1} + d^{2-r}h^{2r-2}). \tag{3.2}$$

In Table 1 we have listed the rate of convergence given by (3.2) for different values of r, and also the minimal value of the radius d of the ball in which the subspaces have to be similar in order to guarantee a rate of convergence of h^{r+1}. Notice that the higher order subspaces are more local in that they require smaller domains of similarity,

TABLE 1

r	$\lvert u(\widehat{x}) - v_h(\widehat{x})\rvert \simeq$	h^{r+1}	minimal radius for similarity $d = h^{\frac{r-3}{r-2}}$
3		h^4	h^0
4		h^5	$h^{1/2}$
5		h^6	$h^{2/3}$

Table 2 contains results from Theorem 1.4 for $\frac{\partial(u(\widehat{x}) - \widetilde{v}_h(\widehat{x},\beta))}{\partial x_i}$, again disregarding logarithmic factors, for equations with variable coefficients.

Example 3.2. Consider Dirichlet's problem on a smooth domain in $\mathbb{R}^N$, $N \geq 2$, with the homogeneous boundary conditions $u = 0$ on $\partial\Omega$. Near the boundary, the finite element space consists of isoperimetrical elements, which approximate the boundary to order h^r, with imposed boundary conditions. It was proved in Schatz and Wahlbin [28] that if u is sufficiently smooth, then

$$\|u - u_h\|_{L_\infty(\Omega)} \leq Ch^r\Big(\ln\frac{1}{h}\Big)^{\overline{r}}. \tag{3.3}$$

TABLE 2

r	$\frac{\partial(u(\widehat{x})-\widetilde{v}_h(\widehat{x},\beta))}{\partial x_i} \simeq$	h^r	minimal radius for similarity $d = h^{\frac{r-2}{r-1}}$
2		h^2	h^0
3		h^3	$h^{1/2}$
4		h^4	$h^{2/3}$

Because of our approximation hypothesis on the boundary, we cannot deduce a higher rate of convergence than h^r in negative norms. This is not as good as the rate of convergence achieved in (3.2) which could be achieved if super-parametric elements were used to approximate the boundary. If we only have (3.3), then the theory does not predict any better rate of convergence for $u(\widehat{x}) - v_h(\widehat{x})$ using extrapolation than for $u(x)-u_h(x)$. However, extrapolation does improve the accuracy of first derivatives. Table 3 gives values from Theorem 1.4 for $\frac{\partial(u(\widehat{x})-\widetilde{v}_h(\widehat{x},\beta))}{\partial x_i}$, neglecting logarithmic factors.

TABLE 3

r	$\frac{\partial(u(\widehat{x})-\widetilde{v}_h(\widehat{x},\beta))}{\partial x_i} \simeq$	h^r	minimal radius for similarity
2		h^2	h^0
3		h^3	h^0
4		h^4	h^0

Example 3.3. Dirichlet's problem on a smooth plane domain. Scott [30] proposed a special way of treating Dirichlet's problems on smooth domains in $\mathbb{R}^2$ which was shown to satisfy

$$\|u - u_h\|_{W_2^{2-r}(\Omega)} \le Ch^{2r-2}. \tag{3.4}$$

Table 4 contains results from Theorem 1.4 for $u(\widehat{x}) - v_h(\widehat{x})$ and Table 5 contains results from Theorem 1.4 for $\frac{\partial u(\widehat{x})}{\partial x_i} - \frac{\partial \widetilde{v}_h(\widehat{x},\beta)}{\partial x_i}$. We emphasize here that $N = 2$ as compared to $N \ge 2$ in Example 3.1 and that the estimate (3.4) is, modulo logarithms, not as strong as (3.1).

TABLE 4

r	$\lvert u(\widehat{x}) - v_h(\widehat{x})\rvert \simeq$	h^{r+1}	minimal radius for similarity $d = h^{\frac{r-3}{r-1}}$
3		h^4	h^0
4		h^5	$h^{1/3}$
5		h^6	$h^{1/2}$

TABLE 5

r	$\left\lvert\frac{\partial(u(\widehat{x})-\tilde{v}_h(\widehat{x},\beta))}{\partial x_i}\right\rvert \simeq$	h^r	minimal radius for similarity $d = h^{\frac{r-2}{r}}$
2		h^2	h^0
3		h^3	$h^{1/3}$
4		h^4	$h^{1/2}$

Example 3.4. Dirichlet's problem on a plane polygonal domain Ω. We consider the problem

$$-\Delta u = f \text{ in } \Omega,\ u = 0 \text{ on } \partial\Omega.$$

It is well known that by using suitable mesh refinements (see Babuška and Rosenzweig [1]), that

$$\min_{\chi\in S_r^h(\Omega)} \|u-\chi\|_{W_2^1(\Omega)} \le Ch^{r-1}\|f\|_{W_2^{r-2}(\Omega)}.$$

A standard duality argument then gives

$$\|u-u_h\|_{W_2^{2-r}(\Omega)} \le C(u)h^{2r-2}.$$

Let $B_{\lambda_j d}(\widehat{x}) \subset \Omega_0, j = 0, 1$, where Ω_0 is an interior domain in which u is smooth and no refinement is done. Then in view of (3.4) and Theorem 1.4, we have that Table 4 and 5 hold. We finally remark that, using the results of Schatz and Wahlbin [28], the local results of this paper could be applied to do Richardson extrapolation to increase accuracy in regions where refined grids, like those given in Examples 2.1 and 2.2 of Section 2, are constructed.

4. Proof of Theorem 1.4

Here we have that $A(\cdot,\cdot)$ is of the form (1.48). As in the proof of Theorem 1.1 we take $\widehat{x}$ for convenience to be the origin then using a change of variables $y = \lambda_1 x$ it is easy to see that for all $\varphi \in \mathring{S}_r^h(B_d(\widehat{x}))$,

$$\int_{B_d(\widehat{x})} \sum_{i,j=1}^N a_{ij}(\lambda_1 x)\frac{\partial(u(\lambda_1 x)-u_{\lambda_1 h}(\lambda_1 x))}{\partial x_i}\frac{\partial\varphi(x)}{\partial x_j}\,dx$$
$$= \lambda_1^{2-N}\int_{B_{\lambda_1 d}(\widehat{x})} \sum_{i,j=1}^N a_{ij}(y)\frac{\partial(u(y)-u_{\lambda_1 h}(y))}{\partial y_i}\ \frac{\partial\varphi(y/\lambda_1))}{\partial y_j}\,dy.$$

Similarly,

$$\int_{B_d(\widehat{x})} \sum_{i=1}^N b_i(\lambda_1 x)\frac{\partial(u(\lambda_1 x)-u_{\lambda_1 h}(\lambda_1 x))}{\partial x_i}\varphi(x)\,dx$$
$$= \lambda_1^{1-N}\int_{B_{\lambda_1 d}(\widehat{x})} \sum_{i=1}^N b_i(y)\frac{\partial(u(y)-u_{\lambda_1 h}(y))}{\partial y_i}\ \varphi(y/\lambda_1)\,dy$$

and

$$\int_{B_d(\widehat{x})} c(\lambda_1 x)(u(\lambda_1 x) - u_{\lambda_1 h}(\lambda_1 x))\varphi(x)\,dx$$
$$= \lambda_1^{-N} \int_{B_{\lambda_1 d}(\widehat{x})} c(y)(u(y) - u_{\lambda_1 h}(y))\ \varphi(y/\lambda_1)\,dy.$$

Thus by change of variables $y = \lambda_1 x$, $u(\lambda_1 x) - u_{\lambda_1 h}(\lambda_1 x) = e(\lambda_1 x)$ satisfies

$$\int_{B_d(\widehat{x})} \Big(\sum_{i,j=1}^{N} a_{ij}(\lambda_1 x)\frac{\partial e(\lambda_1 x)}{\partial x_i}\ \frac{\partial \varphi(x)}{\partial x_j} + \sum_{i=1}^{N} \lambda_1 b_i(\lambda_1 x)\frac{\partial e(\lambda_1 x)}{\partial x_i}\varphi(x)$$
$$+ \lambda_1^2 c(\lambda_1 x)e(\lambda_1 x)\varphi(x)\Big)dx = 0 \text{ for all } \varphi \in \mathring{S}_r^h(B_d(\widehat{x})).$$

Hence for $v(x)$ and $v_h(x)$ defined by (1.24), and (1.25), it follows that

$$A(v - v_h, \varphi) = F_1(\varphi) + F_2(\varphi) + F_3(\varphi) \equiv F(\varphi) \quad \text{for all} \quad \varphi \in \mathring{S}_r^h(B_d(\widehat{x})), \tag{4.1}$$

where

$$F_1(\varphi) = \gamma_2 \int_{B_d(\widehat{x})} \sum_{i,j=1}^{N} \Big(a_{ij}(x) - a_{ij}(\lambda_1 x)\Big)\frac{\partial e(\lambda_1 x)}{\partial x_i}\ \frac{\partial \varphi}{\partial x_j}\ dx, \tag{4.2}$$

$$F_2(\varphi) = \gamma_2 \int_{B_d(\widehat{x})} \sum_{i=1}^{N} \Big(b_i(x) - \lambda_1 b_i(\lambda_1 x)\Big)\frac{\partial e(\lambda_1 x)}{\partial x_i}\ \varphi(x)dx, \tag{4.3}$$

and

$$F_3(\varphi) = \gamma_2 \int_{B_d(\widehat{x})} \Big(c(x) - \lambda_1^2 c(\lambda_1 x)\Big)e(\lambda_1 x)\varphi(x)dx. \tag{4.4}$$

Below we shall set $E := e(\lambda_1 x)$, and estimate each $F_i,\ i = 1, 2, 3$ term separately in the norms $|||\cdot|||_j,\ j = 1, 2$. Starting with F_1 for $\psi \in \mathring{W}_1^1(B_d(\widehat{x}))$, and $C = C(\lambda_1)$,

$$\begin{aligned} |F_1(\psi)| &\le Ch \sum_{i,j=1}^{N} \Big(\int_{B_d(\widehat{x})} \frac{|x| + h}{h} |\frac{\partial \psi}{\partial x_j}|\,dx\Big) \Big\| \frac{\partial E}{\partial x_i} \Big\|_{L_\infty(B_d)} \\ &\le Ch \sum_{i=1}^{N} \Big\| \frac{\partial E}{\partial x_i} \Big\|_{L_\infty(B_d)} |||\psi|||_1. \end{aligned} \tag{4.5}$$

As for the $|||\cdot|||_2$-norm, it follows on integrating by parts that, for all $\psi \in \mathring{W}_1^1(B_d(\widehat{x}))$,

$$\begin{aligned} F_1(\psi) = &- \gamma_2 \int_{B_d(\widehat{x})} \sum_{i,j=1}^{N} \Big(\frac{\partial}{\partial x_i}(a_{ij}(x) - a_{ij}(\lambda_1 x))\Big)e(\lambda_1 x)\frac{\partial \psi}{\partial x_j}\ dx \\ &- \gamma_2 \int_{B_d(\widehat{x})} \sum_{i,j=1}^{N} (a_{ij}(x) - a_{ij}(\lambda_1 x))e(\lambda_1 x)\frac{\partial^2 \psi}{\partial x_i \partial x_j}\,dx \\ &\equiv F_{1a}(\psi) + F_{1b}(\psi), \end{aligned} \tag{4.6}$$

$$\begin{aligned} |F_{1a}(\psi)| &\le C \sum_{i,j=1}^{N} \Big(\Big| \frac{\partial}{\partial x_i}\Big(a_{ij}(x) - a_{ij}(\lambda_1 x)\Big)\frac{\partial \psi}{\partial x_j}\Big|_{W_1^1(B_d)}\Big) \|E\|_{W_\infty^{-1}(B_d)} \\ &\le C\,\|E\|_{W_\infty^{-1}(B_d)}\,|||\psi|||_2, \end{aligned} \tag{4.7}$$

where $C = C(\lambda_1)$. Continuing with a similar argument for $F_{1b}(\psi)$ we have that

$$
\begin{aligned}
|F_{1b}(\psi)| &\le C \sum_{i,j=1}^{N} \Big(\Big| \Big(a_{ij}(x) - a_{ij}(\lambda_1 x) \Big) \frac{\partial^2 \psi}{\partial x_i \partial x_j} \Big|_{W_1^1(B_d)} \Big) \|E\|_{W_\infty^{-1}(B_d)} \\
&\le C \|E\|_{W_\infty^{-1}(B_d)} \sum_{i,j=1}^{N} \Big(\Big| \sqrt{x^2+h^2} \frac{\partial^2 \psi}{\partial x_i \partial x_j} \Big|_{W_1^1(B_d)} \Big) \\
&\le C \|E\|_{W_\infty^{-1}(B_d)} \Big[\Big| \frac{\partial}{\partial x_i} (\sqrt{x^2+h^2}) \frac{\partial^2 \psi}{\partial x_i \partial x_j} \Big|_{L_1(B_d)} \\
&\qquad + \Big| \sqrt{x^2+h^2} \frac{\partial^3 \psi}{\partial x_i^2 \partial x_j} \Big|_{L_1(B_d)} \Big] \le C \|E\|_{W_\infty^{-1}(B_d)} |||\psi|||_2.
\end{aligned} \tag{4.8}
$$

We proceed to $F_2(\psi)$, using partial integration we get

$$
\begin{aligned}
F_2(\psi) = &-\gamma_2 \int_{B_d(\widehat{x})} \sum_{i=1}^{N} \frac{\partial}{\partial x_i} \Big(b_i(x) - \lambda_1 b_i(\lambda_1 x) \Big) e(\lambda_1 x) \psi(x) dx \\
&- \gamma_2 \int_{B_d(\widehat{x})} \sum_{i=1}^{N} \Big(b_i(x) - \lambda_1 b_i(\lambda_1 x) \Big) e(\lambda_1 x) \frac{\partial \psi}{\partial x_i} \, dx.
\end{aligned} \tag{4.9}
$$

Hence, by the same procedure as the estimate of $F_1(\psi)$ we have that

$$
\begin{aligned}
|F_2(\psi)| &\le Ch \|E\|_{L_\infty(B_d(\widehat{x}))} \Big(\sum_{i=1}^{N} \Big\| \frac{\partial}{\partial x_i} [b_i(x) - \lambda_1 b_i(\lambda_1 x)] \Big\|_{L_\infty(B_d(\widehat{x}))} \Big) \\
&\quad \times h^{-1} \|\psi\|_{L_1(B_d(\widehat{x}))} \\
&\quad + \|E\|_{L_\infty(B_d(\widehat{x}))} \Big(\sum_{i=1}^{N} \|[b_i(x) - \lambda_1 b_i(\lambda_1 x)] \psi\|_{L_1(B_d(\widehat{x}))} \Big) \\
&\le C \|E\|_{L_\infty(B_d(\widehat{x}))} |||\psi|||_1,
\end{aligned} \tag{4.10}
$$

and as for the $|||\cdot|||_2$-norm

$$
\begin{aligned}
|F_2(\psi)| &\le C \|E\|_{W_\infty^{-1}(B_d)} \Big(\sum_{i=1}^{N} \Big| \Big(\frac{1}{|x|+h} \Big) \psi \Big|_{W_1^1(B_d(\widehat{x}))} + |\psi|_{W_1^2(B_d(\widehat{x}))} \Big) \\
&\le C \|E\|_{W_\infty^{-1}(B_d)} |||\psi|||_2.
\end{aligned} \tag{4.11}
$$

Finally, for the remaining c or $F_3(\psi)$-term we have analogously

$$
\begin{aligned}
|F_3(\varphi)| &\le \gamma_2 \int_{B_d(\widehat{x}))} |c(x) - \lambda_1^2 c(\lambda_1 x)| e(\lambda_1 x) \varphi(x) dx| \\
&\le Ch \|E\|_{L_\infty(B_d(\widehat{x}))} h^{-1} \|\psi\|_{L_1(B_d(\widehat{x})} \le Ch \|E\|_{L_\infty(B_d(\widehat{x}))} |||\psi|||_1.
\end{aligned} \tag{4.12}
$$

Similarly,

$$
|F_3(\varphi)| \le Ch \|E\|_{W_\infty^{-1}(B_d)} \Big| (|x|+h) \psi \Big|_{W_1^1(B_d(\widehat{x}))} \le Ch \|E\|_{W_\infty^{-1}(B_d)} |||\psi|||_2. \tag{4.13}
$$

Summing up we have, recalling the definition of $|||F|||_j$, $j = 1, 2$ norms (1.11) and (1.12), that

$$
\begin{aligned}
|||F|||_{-1} &\le C \Big(h \textstyle\sum_{i=1}^{N} \Big\| \frac{\partial E}{\partial x_i} \Big\|_{L_\infty(B_d(\widehat{x}))} + (1+h) \|E\|_{L_\infty(B_d(\widehat{x}))} \Big) \\
|||F|||_{-2} &\le C(1+h) \|E\|_{W_\infty^{-1}(B_d(\widehat{x}))} .
\end{aligned} \tag{4.14}
$$

Now we claim that

$$|(u-v_h)(\widehat{x})| \le C(u)\Big(\ln\frac{1}{h}\Big)^2\Big(h^{r+1}+d^{-N/p-t}\sum_{i=0}^{1}\|u-u_{\lambda_i h}\|_{W_p^{-t}(B_{\lambda_i d(\widehat{x})})}\Big). \tag{4.15}$$

To prove (4.15) we need to estimate the norms of the error E appearing in (4.14). To this end it follows from Lemmas 1.1 and 1.2, for $F\equiv 0$, with $r\ge 3$, that

$$\begin{aligned}\|E\|_{L_\infty(B_d(\widehat{x}))} &= \|u-u_{\lambda_1 h}\|_{L_\infty(B_{\lambda_1 d}(\widehat{x}))}\\ &\le C\big(h^r\|u\|_{W_\infty^r(B_{2\lambda_1 d}(\widehat{x}))}+d^{-t-N/p}\|u-u_{\lambda_1 h}\|_{W_p^{-t}(B_{2\lambda_1 d}(\widehat{x}))}\big),\end{aligned} \tag{4.16}$$

and similarly,

$$\begin{aligned}&\|u-u_{\lambda_1 h}\|_{W_\infty^1(B_{\lambda_1 d}(\widehat{x}))}\\ &\quad\le C\big(h^{r-1}\|u\|_{W_\infty^r(B_{2\lambda_1 d}(\widehat{x}))}+d^{-1-t-N/p}\|u-u_{\lambda_1 h}\|_{W_p^{-t}(B_{2\lambda_1 d}(\widehat{x}))}\big).\end{aligned} \tag{4.17}$$

Using (4.16) and (4.17) and the fact that $h/d\le 1$, the first estimate in (4.14) can be written as

$$|||F|||_{-1}\le C\big(h^r\|u\|_{W_\infty^r(B_{2\lambda_1 d}(\widehat{x}))}+d^{-t-N/p}\|u-u_{\lambda_1 h}\|_{W_p^{-t}(B_{2\lambda_1 d}(\widehat{x}))}\big). \tag{4.18}$$

Furthermore, using Lemma 1.1 and a local duality argument, (see Cameron [7] or Schatz and Nitsche [12]), it is not hard to prove that if $r\ge 3$, and $t\ge 1$, then

$$\begin{aligned}&\|u-u_{\lambda_1 h}\|_{W_\infty^{-1}(B_{\lambda_1 d}(\widehat{x}))}\\ &\quad\le C\Big(\ln\frac{d}{h}\Big)\Big(h^{r+1}\|u\|_{W_\infty^r(B_{2\lambda_1 d}(\widehat{x}))}+d^{1-t-N/p}\|u-u_{\lambda_1 h}\|_{W_p^{-t}(B_{2\lambda_2 d}(\widehat{x}))}\Big).\end{aligned} \tag{4.19}$$

Thus in view of the second estimate in (4.14), (4.18) and (4.19) we get

$$\begin{aligned}&\Big(\ln\frac{1}{h}\Big)\big(h|||F|||_{-1}+|||F|||_{-2}\big)\\ &\le C\Big(\ln\frac{1}{h}\Big)^2\Big(h^{r+1}\|u\|_{W_\infty^r(B_{2\lambda_1 d}(\widehat{x}))}+d^{1-t-N/p}\|u-u_{\lambda_1 h}\|_{W_p^{-t}(B_{2\lambda_1 d}(\widehat{x}))}\Big).\end{aligned} \tag{4.20}$$

Using (4.20), (1.27), and (1.28), in (1.13), we have proved the claim (4.15). Now (4.15) together with (1.21) leads to (1.48). This completes the proof of part i) of Theorem 1.4. The inequality (1.49) follows in exactly the same way, except this time we use Lemma 1.2 instead of 1.1. □

5. Appendix I. Properties of the finite element subspaces

Here we shall state our assumptions on the finite element subspaces used in this paper. They are basically the same as those given in Schatz and Wahlbin [28] and [29]. The precise statements here are taken from [29].

For $0<h<1$ a parameter and $r\ge 2$ an integer, $S_r^h(\Omega)$ will denote a family of finite dimensional subspaces of $W_\infty^1(\Omega)$. If $D\subseteq\Omega$, then $S_r^h(D)$ will denote the restriction of functions in $S_r^h(\Omega)$ to D and $\mathring{S}_r^h(D)$ is the subspace of $S_r^h(D)$ consisting of functions whose support is contained in D. In what follows $D_0\subset\subset D_1\subset\subset D_2$, etc. denote concentric balls which are contained in Ω. Assume that there exists a constant k such that if $\operatorname{dist}(D_0,\partial D_1)\ge kh$ and $\operatorname{dist}(D_1,\partial D_2)\ge kh$, then the following hold:

A.1 (Approximation). If $t = 0, 1$, $t \le \ell \le r$, $1 \le p \le \infty$, then for each $v \in W_p^\ell(D_2)$ there exists a $\chi \in S_r^h(D_1)$ such that

$$\|v - \chi\|_{W_p^t(D_1)} \le Ch^{\ell - t}|v|_{W_p^\ell(D_2)}.$$

Here

$$|v|_{W_p^\ell} = \begin{cases} \Big(\sum\limits_{|\alpha|=\ell} \|D^\alpha v\|_{L_p}^p \Big)^{1/p} & \text{if} \quad 1 \le p < \infty, \\ \sum\limits_{|\alpha|=\ell} \|D^\alpha v\|_{L_\infty} & \text{if} \quad p = \infty. \end{cases}$$

Furthermore, if $v \in \mathring{W}_p^\ell(D_0)$, then $\chi \in \mathring{S}_r^h(D_2)$. Here C is independent of h, v, χ and D_i, $i = 0, 1, 2$.

A.2 (Inverse Properties). If $\chi \in S_r^h(D_2)$, then for $t = 0, 1$,

$$\|\chi\|_{W_\infty^t(D_1)} \le Ch^{-N/2-t}\|\chi\|_{L_2(D_2)},$$

and for $\ell = 0, 1$,

$$\|\chi\|_{W_2^t(D_1)} \le Ch^{\ell - t}\|\chi\|_{W_2^{-\ell}(D_2)}.$$

Here C is independent of h, χ, D_1 and D_2.

A.3 (Super-approximation). Let $\omega \in C_0^\infty(D_1)$, then for each $\chi \in S_r^h(D_2)$ there exists an $\eta \in \mathring{S}_r^h(D_2)$ such that for some integer $\gamma > 0$,

$$\|\omega\chi - \eta\|_{W_2^1(D_2)} \le Ch\|\omega\|_{W_\infty^\gamma(D_1)}\|\chi\|_{W_2^1(D_3)}.$$

Furthermore, if $\omega \equiv 1$ on D_0, and $D_{-1} \subset\subset D_0$ with $\operatorname{dist}(D_{-1}, \partial D_0) \ge k$, then $\eta = \chi$ on D_{-1}, and

$$\|\omega\chi - \eta\|_{W_2^1(D_2)} \le Ch\|\omega\|_{W_\infty^\gamma(D_1)}\|\chi\|_{W_2^1(D_2 \setminus D_{-1})}.$$

Here C is independent of ω, χ, η, h, D_i, $i = -1, 0, 1, 2$.

A.4 (Scaling). Let $x_0 \in \overline{\Omega}$ and $d \ge kh$. The linear transformation $y = x_0 + (x - x_0)/d$ takes $B_d(x_0) = \{x : |x - x_0| < d\} \cap \Omega$ into a new domain $\widehat{B}_1(x_0)$ and $S_r^h(B_d(x_0))$ into a new function space $\widehat{S}_r^{h/d}(\widehat{B}_1(x_0))$. Then $\widehat{S}_r^{h/d}(\widehat{B}_1(x_0))$ satisfies A.1, A.2 and A.3 with h replaced by h/d. The constants occurring in A.1, A.2 and A.3 remain unchanged, in particular, independent of d.

Acknowledgments

Parts of this work were carried out while the first author was visiting Cornell University. He wishes to thank the math department at Cornell for their hospitality and Professors A. H. Schatz and L. B. Wahlbin for many fruitful discussions and their generosity. The research of M. Asadzadeh was supported by the Swedish Foundation for Strategic Research (SSF) in Gothenburg Mathematical Modelling Centre (GMMC). The research of A. H. Schatz was supported by NSF Grant 06-12599.

References

[1] I. Babuška, M. B. Rosenzweig and B. Michael, *A finite element scheme for domains with corners*, Numer. Math. 20 (1972/73), 1–21. MR0323129 (48:1487)

[2] H. Blum, *Numerical treatment of corner and crack singularities,* in Finite Element and Boundary Element Techniques from a Mathematical and Engineering Point of View, CISM Courses and Lectures. vol. 301 (1988), pp. 171–212. MR1002579 (91b:65122)

[3] H. Blum, *On Richardson extrapolation for linear finite elements on domains with reentrant corners,* Z. Angew. Math. Mech. 29 (1987), pp. 351–353. MR907630 (88h:65201)

[4] H. Blum, Q. Lin and R. Rannacher, *Asymptotic error expansions and Richardson extrapolation for linear finite elements,* Numer. Math. 49 (1986), pp. 11–37. MR847015 (87m:65172)

[5] K. Böhmer, *Asymptotic expansions for the discretization error in linear elliptic boundary value problems for general region,* Math. Z. 177 (1982), pp. 235–255. MR612877 (82d:65064)

[6] S. Brenner and R. Scott, The mathematical theory of finite element methods, Second Edition. Texts in Applied Mathematics, 15, Springer, 2002. MR1894376 (2003a:65103)

[7] A. Cameron, *Weighted L_∞-based negative norm estimates for the finite methods for second order elliptic problems.* In preparation.

[8] C. M. Chen, Q. Lin, *Extrapolation of finite element approximations in a rectangular domain,* J. Comput. Math. 7 (1989), pp. 227–233. MR1017183 (90i:65190)

[9] H. Chen and R. Rannacher, *Local error expansions and Richardson extrapolation for the streamline diffusion finite element method,* East-West J. Numer. Math. 1 (1993), pp. 253–265. MR1318805 (95k:65104)

[10] Y.H. Ding and Q. Lin, *Finite element expansion for variable coefficient elliptic problems,* Systems Sci. Math. Sci. 2 (1989), pp. 54–69. MR1110121 (92d:65190)

[11] W. Hoffmann, A. H. Schatz, L. B. Wahlbin and G. Wittum, *Asymptotically exact a posteriori estimators for the pointwise gradient error on each element in irregular meshes. I. A smooth problem and globally quasi-uniform meshes,* Math. Comp. 70 (2001), no. 235, 897–909 (electronic). MR1826572 (2002a:65178)

[12] J. A. Nitsche and A. H. Schatz, *Interior estimates for Ritz-Galerkin methods.* Math. Comp. 28 (1974), 937–958. MR0373325 (51:9525)

[13] Q. Lin, *Fourth order eigenvalue approximation by extrapolation on domains with reentrant corners,* Numer. Math. 58 (1991), pp. 631–640. MR1083525 (92d:65198)

[14] Q. Lin and T. Lü, *Asymptotic expansions for the finite element approximation of elliptic problems on polygonal domains,* Computing methods in applied sciences and engineering, VI (Versailles, 1983), pp. 317–321, North-Holland, Amsterdam, 1984. MR806787

[15] Q. Lin and J.P. Wang, *Some expansions for finite element approximation,* Shuli Kexue, Mathematical Sciences, Research report IMS 15, Academia Sinica, Inst. of Math. Sciences, Chendu (1984), pp. 1–11. MR777686 (86d:65148)

[16] Q. Lin and J.P. Wang, *Asymptotic expansions and extrapolation for the finite element method* (Chinese), J. Systems Sci. Math. Sci. 5 (1985), pp. 114–120. MR841406 (87f:65137)

[17] Q. Lin and R.F. Xie, *Error expansion for FEM and superconvergence under natural assumption,* J. Comput. Math. 7 (1989) pp. 402–411. MR1149709

[18] Q. Lin and Q. Zhu, *Asymptotic expansion for the derivative of finite elements,* J. Computing Math. 2 (1982), pp. 361–363. MR869509 (87m:65185)

[19] R. Rannacher, *Richardson extrapolation for a mixed finite element approximation of a plate bending problem,* Z. Angew. Math. Mech. 67 (1987), pp. 381–383. MR907636 (88g:65118)

[20] R. Rannacher, *Extrapolation techniques in the finite element method (survey),* of Mathematics, Report C7, Feb. 1988, pp. 80–113.

[21] R. Rannacher and R. Scott, *Some optimal error estimates for piecewise linear finite element approximations,* Math. Comp. 38 (1982), no. 158, 437–445. MR645661 (83e:65180)

[22] U. Rüde, *The hierarchical basis extrapolation method,* SIAM J. Sci. Statist. Comput., 13 (1992), pp. 307–318. MR1145188 (92h:65181)

[23] U. Rüde, *Extrapolation and related techniques for solving elliptic equation,* Bericht I-9135, Institut fur Informatik der TU Munchen, (1991).

[24] A. H. Schatz, *Perturbations of forms and error estimates for the finite element method at a point, with an application to improved superconvergence error estimates for subspaces that are symmetric with respect to a point*, SIAM J. Numer. Anal. 42 (2005), no. 6, 2342–2366. MR2139396 (2006g:65194)

[25] A. H. Schatz, *Pointwise error estimates and asymptotic error expansion inequalities for the finite element method on irregular grids.* I. Global estimates, Math. Comp. 67 (1998), no. 223, 877–899. MR1464148 (98j:65082)

[26] A. H. Schatz, *Pointwise error estimates and asymptotic error expansion inequalities for the finite element method on irregular grids.* II. Interior estimates, SIAM J. Numer. Anal. 38(2000), 1269–1293. MR1786140 (2001i:65119)

[27] A. H. Schatz, I. Sloan and L. B. Wahlbin, *Superconvergence in finite element method and meshes that are locally symmetric with respect to a point,* SIAM J. Numer. Anal. 33 (1996), pp. 505–521. MR1388486 (98f:65112)

[28] A. H. Schatz, and L. B. Wahlbin, *Maximum norm estimates in the finite element method on plane polygonal domains.* I, Math. Comp. 32 (1978), no. 141, 73–109. MR0502065 (58:19233a)

[29] A. H. Schatz, and L. B. Wahlbin, *Maximum norm estimates in the finite element method on plane polygonal domains.* II, Refinements. Math. Comp. 33 (1979), no. 146, 465–492. MR0502067 (58:19233b)

[30] R. Scott, *Optimal L^∞ estimates for the finite element method on irregular meshes*, Math. Comp. 30 (1976), no. 136, 681–697. MR0436617 (55:9560)

[31] J. P. Wang, *Asymptotic expansions and L^∞- error estimates for mixed finite element methods for second order elliptic problems,* Numer. Math. 55 (1989), pp. 401–430. MR997230 (90e:65166)

[32] W. Wasow, *Discrete approximation to elliptic differential equations,* Z. Angnew. Math. Phys. 6 (1955), pp. 81–97. MR0080369 (18:236e)

[33] R.F. Xie, *Pointwise estimates for finite element approximations to Green functions on a concave polygonal domain, and finite element extrapolation* (Chinese), Math. Numer. Sinica, 10 (1988), pp. 232–241. MR985474 (90f:65211)

Department of Mathematics, Chalmers University of Technology, SE-412 96 Goteborg, Sweden

E-mail address: mohammad@chalmers.se

Current address: Department of Mathematics, 310 Malott Hall, Cornell University, Ithaca, New York 14853

E-mail address: asadzadeh@math.cornell.edu

Department of Mathematics, 310 Malott Hall, Cornell University, Ithaca, New York 14853

E-mail address: schatz@math.cornell.edu

Institute for Applied Analysis and Numerical Simulations, University of Stuttgart, Pfaffenwaldring 57, D-750550, Germany

E-mail address: wendland@mathematik.uni-stuttgart.de