COMPUTATIONAL METHODS IN APPLIED MATHEMATICS, Vol. 1 (2001), No. 1, pp. 125–131

ON THE SATURATION ASSUMPTION AND HIERARCHICAL A POSTERIORI ERROR ESTIMATOR

A. AGOUZAL

Maply. U.M.R 5558

Bat. 101, Univresité Lyon1, 43 Bd 18 Novembre 1918. 69 622, villeurbanne cedex. France

E-mail: agouzal@maply.univ-lyon1.fr

Abstract — The saturation assumption is widely used in a posteriori error analysis of finite element methods. It asserts, in its simplest form, that the solution can be approximated asymptotically with quadratic finite elements than with linear ones. In this paper, we show that a simple modification of this "hypothesis" is valid, and the proof of a posteriori error estimators by several authors becomes rigoruous with this simple modification.

2000 Mathematics Subject Classification: 65N30.

Keywords: saturation assumption, hierarchical spaces, a posteriori error estimators.

1. Introduction

The classical proof of equivalence of some a posteriori error estimators with the energy error requires the *saturation assumption* [2, 3]: this means, in its simplest form, that the solution can be approximated asymptotically with quadratic finite elements than with linear ones. The saturation assumption was shown to be superfluous by Nochetto in [5] for Bank-Weiser estimators, howeover, removing this assumption requires comparison with residual estimators. More recently W. Dorfler and R.H. Nochetto [4] have given a sufficient condition for the validity of the saturation assumption, more precisely, they have proved that "*small data oscillation implies the saturation assumption*"; the technique used in the last paper was taken from the a posteriori error analysis and comparison with residual estimators. Recently the author of [1], have given a direct proof of the reliability and efficiency of the hierarchical estimator without comparison with the residual estimator or the saturation assumption for conforming and nonconforming approximations.

In this paper, we show that a simple modification of this "assumption" is valid, and the proof of a posteriori error estimates of several authors is rigorous with this simple modification. As an example, we give "elementary" and direct proof of the efficiency and reliability of the hierarchical a posteriori error estimators.

2. Modified saturation "assumption"

We consider the following model problem:

$$\begin{cases} u \in H_0^1(\Omega), \\ -\Delta u = f \text{ in } \Omega, \end{cases}$$

where Ω is a bounded polygonal domain in R^2 and $f \in L^2(\Omega)$. Let T_h be a conforming and shape regular triangulation by triangular elements, we set

$$V_1 = \{v_h \in H_0^1(\Omega); \quad \forall T \in T_h, v_{h|T} \in P_1(T)\},$$
$$V_2 = \{v_h \in H_0^1(\Omega); \quad \forall T \in T_h, v_{h|T} \in P_2(T) + span\{b_T\}\},$$

where b_T is the cubic bubble function, and consider the following problems $(k = 1, 2)$:

$$\begin{cases} \text{Find } u_k \in V_k \\ \forall v_k \in V_k; \quad \displaystyle\int_\Omega \nabla u_k \cdot \nabla v_k dx = \int_\Omega f v_k dx. \end{cases}$$

In the sequel, we denote by C the positive generic constant independent of the mesh size which can change from one line to another.

Theorem 2.1. *There exists a constant $\alpha \in [0, 1[$ depending on the minimum angles of T_h so that*

$$|u - u_2|_{1,\Omega} \leqslant \alpha |u - u_1|_{1,\Omega} + C \left(\sum_{T \in T_h} h_T^2 \|f - f_T\|_{0,T}^2 \right)^{1/2},$$

where $f_T = \dfrac{1}{meas(T)} \displaystyle\int_T f dx.$

Proof. First, we introduce the operator Π^0 defined from $H^1(\Omega)$ onto V_2 by : for all $v \in H^1(\Omega)$, $\Pi^0 v$ is the unique element of V_2 such that

$$\text{for all nodes } a \text{ of } T_h, \quad \Pi^0 v(a) = 0,$$

$$\text{for all edges } e, \int_e \left(\Pi^0 v - v\right) d\sigma = 0 \ ; \quad \text{for all } T \in T_h, \int_T \left(\Pi^0 v - v\right) dx = 0.$$

Using the Green formula, we have

$$\begin{cases} \forall T \in T_h, \quad \forall w_h \in P_1(T), \quad \forall v \in H^1(T), \\ \displaystyle\int_T \nabla w_h \nabla \left(v - \Pi^0 v\right) dx = \int_{\partial T} \frac{\partial w_h}{\partial n_T} \left(v - \Pi^0 v\right) d\sigma = 0, \end{cases}$$

and by the scaling arguments and the trace theorem it is easy to prove that

$$\forall v \in H^1(\Omega), \quad \forall T \in T_h, \quad \left|\Pi^0 v\right|_{1,T} \leqslant C \left(h_T^{-1} \|v\|_{0,T} + |v|_{1,T}\right).$$

We now set $e_k = u_k - u$, $k = 1, 2$. We have for all $e_h \in V_1$:

$$
\begin{aligned}
|u-u_1|^2_{1,\Omega} &= \int_\Omega \nabla u_1 \nabla e_1 dx - \int_\Omega f e_1 dx \\
&= \int_\Omega \nabla u_1 \nabla (e_1 - e_h) dx - \int_\Omega f(e_1 - e_h) dx \\
&= \int_\Omega \nabla u_1 \nabla \Pi^0 (e_1 - e_h) dx - \int_\Omega f(e_1 - e_h) dx \\
&= \int_\Omega \nabla \left(u_1 - u_2\right) \nabla \Pi^0 \left(e_1 - e_h\right) dx + \int_\Omega f \left(\Pi^0 \left(e_1 - e_h\right) - e_1 + e_h\right) dx \\
&= \int_\Omega \nabla \left(u_1 - u_2\right) \nabla \Pi^0 \left(e_1 - e_h\right) dx + \int_\Omega \left(f - f_T\right) \left(\Pi^0 \left(e_1 - e_h\right) - e_1 + e_h\right) dx \\
&\leqslant |u_1 - u_2|_{1,\Omega} \left|\Pi^0 (e_1 - e_h)\right|_{1,\Omega} \\
&\qquad + C \left(\sum_{T \in T_h} h_T^2 \|f - f_T\|^2_{0,T} \right)^{1/2} \left(\left|\Pi^0 (e_1 - e_h)\right|_{1,\Omega} + |e_1 - e_h|_{1,\Omega} \right) \\
&\leqslant C \left(|u_1 - u_2|_{1,\Omega} + \left(\sum_{T \in T_h} h_T^2 \|f - f_T\|^2_{0,T} \right)^{1/2} \right) \\
&\qquad \times \left(\sum_{T \in T_h} h_T^{-2} \|e_1 - e_h\|^2_{0,T} + |e_1 - e_h|^2_{1,T} \right)^{1/2} .
\end{aligned}
$$

Then

$$
\begin{aligned}
|u-u_1|^2_{1,\Omega} &\leqslant C \left(|u_1 - u_2|_{1,\Omega} + \left(\sum_{T \in T_h} h_T^2 \|f - f_T\|^2_{0,T} \right)^{1/2} \right) \\
&\qquad \times \inf_{v_h \in V_1} \left(\sum_{T \in T_h} h_T^{-2} \|e_1 - v_h\|^2_{0,T} + |e_1 - v_h|^2_{1,T} \right)^{1/2} \\
&\leqslant C \left(|u_1 - u_2|_{1,\Omega} + \left(\sum_{T \in T_h} h_T^2 \|f - f_T\|^2_{0,T} \right)^{1/2} \right) |u - u_1|_{1,\Omega} ,
\end{aligned}
$$

we deduce that

$$
|u - u_1|_{1,\Omega} \leqslant C |u_1 - u_2|_{1,\Omega} + C \left(\sum_{T \in T_h} h_T^2 \|f - f_T\|^2_{0,T} \right)^{1/2} .
$$

Now, using the equality

$$
|u - u_1|^2_{1,\Omega} = |u - u_2|^2_{1,\Omega} + |u_1 - u_2|^2_{1,\Omega},
$$

and since we can choose $C \geqslant 1$, we obtain

$$|u-u_2|^2_{1,\Omega} \leqslant (1-C^{-1})|u-u_1|^2_{1,\Omega} + C\left(\sum_{T\in T_h} h_T^2 \|f-f_T\|^2_{0,T}\right)^{1/2}$$
$$= \alpha^2 |u-u_1|^2_{1,\Omega} + C\left(\sum_{T\in T_h} h_T^2 \|f-f_T\|^2_{0,T}\right)^{1/2}.$$

with $\alpha \in [0,1[$. □

3. Hierarchical a posteriori error estimators

In this section, we are interested in the hierarchical a posteriori error estimator. From the previous results we have

$$|u_1-u_2|_{1,\Omega} \leqslant |u-u_1|_{1,\Omega} \leqslant C\left\{|u_1-u_2|_{1,\Omega} + \left(\sum_{T\in T_h} h_T^2 \|f-f_T\|^2_{1,T}\right)^{1/2}\right\}.$$

Since u_2 is expensive to compute, following Bank et al [2, 3], we introduce " hierarchical" space and solve a discrete problem with smaller dimensions than V_2. For this purpose we set

$$V_2^0 = \{v_h \in V_2, \text{ so that for all nodes } a \text{ of } T_h, v_h(a) = 0\},$$

and we introduce the discrete problem

$$\begin{cases} \text{Find} \quad u_2^0 \in V_2^0, \\ \forall v_2^0 \in V_2^0; \quad \displaystyle\int_\Omega \nabla u_2^0 \nabla v_2^0 dx = \int_\Omega \nabla u_1 \nabla v_2^0 dx - \int_\Omega f v_2^0 dx. \end{cases}$$

We have the following theorem.

Theorem 3.1. *There exists a constant C_0 depending only on the minimum angle of T_h such that*

$$|u_2^0|_{1,\Omega} \leqslant |u-u_1|_{1,\Omega} \leqslant C_0|u_2^0|_{1,\Omega} + C\left(\sum_{T\in T_h} h_T^2 \|f-f_T\|^2_{1,T}\right)^{1/2}.$$

Proof. The first inequality is trivial. Let us prove the second one. Arguing as in the proof of Theorem 2.1, we can assume that

$$f = f_T \quad \text{on } T, \quad \forall T \in \mathcal{T}_h,$$

Using the same notations as in the proof of the last theorem, we have for all $e_h \in V_1$:

$$\begin{aligned}
|u_1 - u_2|^2_{1,\Omega} &= \int_\Omega \nabla u_1 \nabla (u_1 - u_2) - \int_\Omega f\,(u_1 - u_2)\,dx \\
&= \int_\Omega \nabla u_1 \nabla \Pi^0 (u_1 - u_2 - e_h) - \int_\Omega f\,(u_1 - u_2 - e_h)\,dx \\
&= \int_\Omega \nabla u_2^0 \nabla \Pi^0 (u_1 - u_2 - e_h) - \int_\Omega f\left(u_1 - u_2 - e_h - \Pi^0(u_1 - u_2 - e_h)\right) dx \\
&= \int_\Omega \nabla u_2^0 \nabla \Pi^0 (u_1 - u_2 - e_h) \leqslant |u_2^0|_{1,\Omega} \left|\Pi^0(u_1 - u_2 - e_h)\right|_{1,\Omega},
\end{aligned}$$

arguing as above, we obtain

$$|u_1 - u_2|_{1,\Omega} \leqslant C|u_2^0|_{1,\Omega}.$$

Combining this inequality with

$$|u - u_1|_{1,\Omega} \leqslant C|u_1 - u_2|_{1,\Omega},$$

we obtain

$$|u - u_1|_{1,\Omega} \leqslant C|u_2^0|_{1,\Omega}.$$

□

Now, we can introduce a local a posteriori error estimator. To do this, we begin with some notations. We denote by E_I the set of interior edges for all $e \in E_I$, by ϕ_e – the canonical continuous piecewise quadratic basis function corresponding to the midpoint of the edge e, and, finally, by b_T the bubble function corresponding to the triangle T of T_h is denoted. We set

$$\forall e \in E_I, \quad a_e = \frac{\displaystyle\int_\Omega \nabla u_1 \nabla \phi_e dx - \int_\Omega f\phi_e dx}{|\phi_e|^2_{1,\Omega}},$$

$$\forall T \in T_h, \quad c_T = \frac{\displaystyle\int_\Omega \nabla u_1 \nabla b_T dx - \int_\Omega f b_T dx}{|b_T|^2_{1,\Omega}} := -\frac{1}{|b_T|^2_{1,\Omega}} \int_T f b_T dx$$

and

$$\epsilon = \sum_{e \in E_I} a_e \phi_e + \sum_{T \in T_h} c_T b_T \in V_2^0.$$

We introduce the bilinear form $d(.,.)$ defined on $(V_2^0)^2$ by

$$\forall (v,w) \in (V_2^0)^2; \quad v = \sum_{e \in E_I} v_e \phi_e + \sum_{T \in T_h} v_T b_T, \quad w = \sum_{e \in E_I} w_e \phi_e + \sum_{T \in T_h} w_T b_T,$$

$$d(v,w) := \sum_{e \in E_I} v_e w_e |\phi_e|^2_{1,\Omega} + \sum_{T \in T_h} v_T w_T |b_T|^2_{1,\Omega}.$$

It is clear that

$$\forall v_2^0 \in V_2^0; \quad d(\epsilon, v_2^0) = \int_\Omega \nabla u_1 \nabla v_2^0 dx - \int_\Omega f v_2^0 dx = \int_\Omega \nabla u_2^0 \nabla v_2^0 dx.$$

We have the following theorem.

Theorem 3.2. *There exist constants C_0 and C_1 depending only on the minimum angle of T_h so that*

$$|u - u_1|_{1,\Omega} \leqslant C_0 |\epsilon|_{1,\Omega} + C \left(\sum_{T \in T_h} h_T^2 \|f - f_T\|_{1,T}^2 \right)^{1/2},$$

and

$$\forall T \in T_h, \quad |\epsilon|_{1,T} \leqslant C_1 |u - u_1|_{1,\Delta(T)},$$

where $\Delta(T)$ is the set of elements sharing an edge with T.

Proof. First, using the definition and scaling arguments we have

$$|u_2^0|_{1,\Omega}^2 = \int_\Omega \nabla u_1 \nabla u_2^0 dx - \int_\Omega f u_2^0 dx = d(\epsilon, u_2^0) \leqslant C |u_2^0|_{1,\Omega} |\epsilon|_{1,\Omega}.$$

Therefore, using Theorem 3.1, we have

$$|u - u_1|_{1,\Omega} \leqslant C_0 |\epsilon|_{1,\Omega} + C \left(\sum_{T \in T_h} h_T^2 \|f - f_T\|_{1,T}^2 \right)^{1/2}.$$

As for the lower bound of the error, let $T \in T_h$, for each edge $e \in E_I$ of T, since $|\phi_e|_{1,\Omega} \geqslant C$ and $supp(\phi_e) \subset \Delta(T)$, we have

$$|a_e| \leqslant C |u - u_1|_{1,\Delta(T)}.$$

Likewise, using the same arguments, we also have

$$|c_T| \leqslant C |u - u_1|_{1,\Delta(T)}.$$

Using the last inequalities and the regularity of the mesh, we obtain

$$|\epsilon|_{1,T} \leqslant \sum_{e \in \partial T} |a_e| |\phi_e|_{1,\Omega} + |c_T| |b_T|_{1,\Omega} \leqslant C |u - u_1|_{1,\Delta(T)}.$$

□

References

[1] A. Agouzal and J. Olaiz, *A posteriori error estimator on stars for nonconforming finite elements approximations*, submitted to Numer. Math.

[2] R. E. Bank and A. Weiser, *Some a posteriori error estimators for elliptic partial differential equations*, Math. Comp., **44** (1985), pp. 283–301.

[3] R. E. Bank and B. D. Welfert, *A posteriori error estimates for the Stokes problem*, SIAM J. Numer. Anal., **28** (1991), pp. 591–623.

[4] W. Dorfler and R. H. Nochetto, *Small data oscillation implies the saturation assumption*, to appear in Numer. Math.

[5] R. H. Nochetto, *Removing the saturation assumption in a posteriori error analysis*, Rend., Sci. Mat. Appl. A, **127** (1993), pp. 67–82.

Received 29 Oct. 2001
Revised 15 Jul. 2002