PROBABILISTIC ANALYSIS OF AN *A POSTERIORI* ERROR ESTIMATOR FOR FINITE ELEMENTS

PEDRO DÍEZ* and JUAN JOSÉ EGOZCUE

Departament de Matemàtica Aplicada III,
E.T.S. Ingenieros de Caminos Universitat Politècnica de Catalunya,
Modulo C-2 Campus Nord, E-08034 Barcelona, Spain

A residual type *a posteriori* error estimator for finite elements is analyzed using a new technique. In this case, the error estimate is the result of two consecutive projections of the exact error on two finite-dimensional subspaces. The analysis introduced in this paper is based on a probabilistic approach, that is, the idea is to assess the average value of the effectivity index (the ratio estimated error over exact error) by assuming the randomness of the exact error. The average value characterizes the mean behavior of the estimator and it is found to be related with some geometric properties of the subspaces. These geometric properties are obtained from the standard matrices of the linear systems arising in the formulation of the finite element method.

Keywords: Error estimation, probabilistic analysis, finite elements, adaptivity

1. Introduction

A posteriori error estimators are needed to perform practical finite element (FE) computations and to control the quality of the numerical solution. They are also required to drive adaptive procedures leading to optimal meshes.[5]

The analysis of *a posteriori* error estimators for finite elements is usually performed in terms of finding lower and upper bounds of the effectivity index[1,7] which is the ratio of the estimated error and the exact error. This may be seen, in fact, as an *a priori* analysis of the *a posteriori* error estimator. This kind of analysis is used to ensure a good behavior of the estimator in the asymptotic range. Nevertheless, the practical application of an error estimator to an FE computation is usually far from this asymptotic range and, consequently, this analysis does not furnish any clue about the actual behavior of the estimator in a current case.

---

*E-mail: pedro.diez@upc.es

This work introduces a new approach to the analysis of error estimators using a probabilistic viewpoint. Instead of finding pessimistic bounds of the behavior of the studied error estimator under certain assumptions, the exact error is assumed to be a random function. This analysis is applied to a specific residual type error estimator based on the approximation of an h-refined reference solution.[3] In the analyzed error estimator the error is undervalued, i.e. the (measure of the) estimated error is lower than the (measure of the) exact error and, consequently, the effectivity index is lower than one. The effectivity index is therefore a random variable that ranges from zero to one. If the expected value of this random variable is close to one, the estimator shows a good average behavior.

Let us introduce some notation. The finite element method (FEM) is used to solve a PDE with an unknown solution u, which belongs to a functional space V. The FEM provides an approximate solution u_h, lying in a finite-dimensional interpolation space V_h. The interpolation space V_h is generated by a mesh of finite elements with characteristic size h. A reference mesh of characteristic size $\tilde{h}$ ($\tilde{h}$ much smaller than h) is introduced: the associated reference solution $u_{\tilde{h}} \in V_{\tilde{h}}$, is much more accurate than u_h and the reference error, $e_{\tilde{h}} := u_{\tilde{h}} - u_h$, fairly approximates the exact error, $e := u - u_h$. The approximate solution, u_h, is the projection of the exact solution, u, on V_h following the energetic scalar product that appears in the weak form of the problem. In the following, this scalar product is denoted by $\langle \cdot, \cdot \rangle$ and the induced energy norm is denoted by $\| \cdot \|$.

In order to assess the efficiency of any error estimator, the obtained estimate, e_L, should be compared with the exact error. This is usually done in terms of the effectivity index, ν, introduced by Zienkiewicz and Zhu[10] and defined as

$$\nu := \frac{\text{estimated error}}{\text{exact error}} = \frac{\|e_L\|}{\|e\|} \approx \frac{\|e_L\|}{\|e_{\tilde{h}}\|} = \frac{\text{estimated error}}{\text{reference error}} . \tag{1.1}$$

The exact error e is unknown and the definition given in the left-hand side terms of the previous equation is often replaced by the approximation given by the right-hand side terms.

Thus, the estimator can be seen as a function from $V_{\tilde{h}}$ to $V_{\tilde{h}}$, mapping the reference error function $e_{\tilde{h}}$ into the estimate e_L. In the following, the value of the effectivity index defined in Eq. (1.1) is denoted by $\nu(e_{\tilde{h}})$, $e_{\tilde{h}}$ is assumed to be a random vector in $V_{\tilde{h}}$. This is developed in detail in Sec. 2. The idea is to assess the average value of $\nu(e_{\tilde{h}})$, which is understood as a random variable defined over $V_{\tilde{h}}$, for a given choice of the probability distribution of $e_{\tilde{h}}$. This idea has already been introduced,[4] however the goal of this paper is to provide a theoretical framework for this probabilistic analysis.

The error estimator introduced by Díez *et al.*[3,8] is based on the idea of approximating a reference error $e_{\tilde{h}}$ by local low-cost computations. A reference mesh is built up assembling a set of elementary submeshes discretizing each of the elements of the computational mesh. Thus, two different subspaces of $V_{\tilde{h}}$ are defined: $V^{\mathrm{I}} \subset V_{\tilde{h}}$ associated with the interior of the elements and $V^{\mathrm{P}} \subset V_{\tilde{h}}$ associated with

subdomains covering the element edges denoted patches. The projection of $e_{\tilde{h}}$ on V^{I} is the interior estimate ε and the projection of $e_{\tilde{h}}$ on V^{P} is the patch estimate η. Moreover, the patch estimate, η, is forced to be orthogonal to the interior estimate, ε. The complete estimate is $e_L := \varepsilon + \eta$. Thus, the complete estimate is computed through two projections, the first is free (interior estimate, projection on V^{I}), the latter is restricted with an orthogonality constraint (η is the projection on V^{P} orthogonal to ε). This process is denoted as multiprojection strategy.

For analysis purposes, the error estimation process is split into three phases.

(1) From e to $e_{\tilde{h}}$ (from an infinite-dimensional problem to a finite-dimensional one).
(2) From $e_{\tilde{h}}$ to the projection of $e_{\tilde{h}}$ on $V^{\mathrm{I}} + V^{\mathrm{P}}$, denoted by $e_{\tilde{h}}^*$. This phase accounts for the effect of the remaining hidden points.
(3) From the latter to e_L (pure multiprojection).

None of these phases correspond to the practical computation of the estimate. However, all of them are projections and the undervaluation introduced in each may be analyzed separately. In fact, the analysis of each of these phases is carried out using different techniques.

(1) The first phase can be studied using *a priori* error estimates and Richardson extrapolation to obtain an explicit expression:

$$\|u_{\tilde{h}} - u_h\| = \|e_{\tilde{h}}\| \simeq \left[1 - \left(\frac{\tilde{h}}{h}\right)^{2p}\right]^{1/2} \|e\|, \tag{1.2}$$

where p stands for the degree of the interpolating polynomial. That is, if $\tilde{h}$ is one fourth of h and p is one, the reference error, $e_{\tilde{h}}$, is 97% of the exact error e.
(2) The second phase is, in fact, a cancellation of a small amount of degrees of freedom. It accounts for the effect of the points where the error estimate is forced to be zero. This can be seen as a single projection (from $V_{\tilde{h}}$ onto $V^{\mathrm{I}} + V^{\mathrm{P}}$) and the undervaluation introduced in this phase is assessed in Sec. 3.1.
(3) The third phase contains the essence of the multiprojection strategy: the assessment of the efficiency of this phase is carried out in Sec. 3.2.

2. Probabilistic Analysis

In the following the randomness of the error is characterized in the simplest way. Then, without any previous consideration, the probability distribution of the error is assumed to be uniform. This is the main assumption used in the probabilistic assessment of the error estimator. Of course, this may be improved in further works by assuming more realistic probability distributions for the error (related with superconvergence properties). Thus, in the remainder of the paper $e_{\tilde{h}}$ is a uniformly distributed random vector in a finite-dimensional space.

Notice that, for any real number β ($\beta \neq 0$), $\nu(\beta e_{\tilde{h}}) = \nu(e_{\tilde{h}})$, i.e. the value of the effectivity index does not depend on the "size" of the error but only on the

"direction" of $e_{\tilde{h}}$ (if $e_{\tilde{h}}$ is seen as a vector). Then, in order to study the mean behavior of the effectivity index, the norm of vector $e_{\tilde{h}}$ is taken constant, i.e. $e_{\tilde{h}}$ is assumed to yield in a hypersphere $S_R(V_{\tilde{h}}) := \{e_{\tilde{h}} \in V_{\tilde{h}} \text{ such that } \|e_{\tilde{h}}\| = R\}$ of radius R. Thus, the mean value of $\nu(e_{\tilde{h}})$ is defined as

$$\phi := \frac{1}{\text{meas}(S_R(V_{\tilde{h}}))} \int_{S_R(V_{\tilde{h}})} \nu(e_{\tilde{h}}) dS \,, \qquad (2.3)$$

where $\text{meas}(S_R(V_{\tilde{h}}))$ stands for the measure of $S_R(V_{\tilde{h}})$ and the density of probability of $e_{\tilde{h}}$ is assumed to be uniform over $S_R(V_{\tilde{h}})$. The measure of $S_R(V_{\tilde{h}})$ depends on the dimension of the ambient space $V_{\tilde{h}}$, n, and may be analytically computed[2]:

$$\text{meas}(S_R) = \begin{cases} R^{n-1} \dfrac{(2\pi)^{n/2}}{2 \cdot 4 \cdots (n-2)} & \text{for even } n \\[3mm] R^{n-1} \dfrac{2(2\pi)^{(n-1)/2}}{3 \cdot 5 \cdots (n-2)} & \text{for odd } n \,. \end{cases} \qquad (2.4)$$

Since the error estimator is a combination of error projections, the obtained estimate undervaluates the exact error and, consequently, ϕ is lower than one. In fact, if $\phi = 1$, the projection strategy is said to be optimal because for every $e_{\tilde{h}}$, $\|e_L\| = \|e_{\tilde{h}}\|$. Usually the squared norms of the error magnitudes are easier to handle. Consequently, it is convenient to introduce the mean value of $[\nu(e_{\tilde{h}})]^2$,

$$\psi := \frac{1}{\text{meas}(S_R(V_{\tilde{h}}))} \int_{S_R(V_{\tilde{h}})} [\nu(e_{\tilde{h}})]^2 dS \,. \qquad (2.5)$$

The definition of ψ is introduced because the expressions for ψ are much simpler than the expressions for ϕ (see Sec. 3). In fact, both ϕ and ψ can be used to evaluate the average underestimation and the optimality of the projection strategy is also equivalent to $\psi = 1$.

3. Analysis of the Average Behavior of the Estimator

This section deals with the assessment of the main undervaluation introduced in the transformation from $e_{\tilde{h}}$ to e_L. Attending to the three steps of the analysis introduced in Sec. 1, this transformation is split into two partial transformations: from $e_{\tilde{h}}$ to $e_{\tilde{h}}^*$ and from $e_{\tilde{h}}^*$ to $e_L := \varepsilon + \eta$. The first one is a single projection from the space $V_{\tilde{h}}$ to the space $V^I + V^P$ (recall that $V^I + V^P \subset V_{\tilde{h}}$) and, therefore, the average efficiency of the projection depends only on the dimensions of $V_{\tilde{h}}$ and $V^I + V^P$. The second one is more complex because ε is a single projection of $e_{\tilde{h}}^*$ on V^I but η is obtained via a restricted projection on V^P.

3.1. *Analysis of the efficiency of a single projection*

Let us denote by n the dimension on $V_{\tilde{h}}$ and by $n - m$ the dimension of $V^I + V^P$. Thus, m dimensions are lost in the transformation from $e_{\tilde{h}}$ to $e_{\tilde{h}}^*$. The average efficiency of this transformation is obviously a function of n and m. The expression

of ϕ as a function of n and m is explicitly found using a set of generalized hyperspherical coordinates.[2,4] It is worth noting that the expression for ϕ depends on the parity of n and m and that it is cumbersome.[4]

On the contrary, the expression for ψ (recall Eq. (2.5)) is much simpler and does not depend on the parity on n and m:

$$\psi = 1 - \frac{m}{n}\,. \tag{3.6}$$

Recall that the dimension of $V_{\bar{h}}$, n, is directly related with the number of nodes in the reference mesh generating $V_{\bar{h}}$. On the other hand, m is related with the number of points where the estimate e_L is forced to vanish. Thus, the values of n and m are determined by simple node (i.e. degrees of freedom) counting. Once n and m are known, ϕ and ψ are obtained in the expressions presented in Ref. 4 and Eq. (3.6), respectively. Using simple numerical experimentation it can be found that, for large values of n ($n \geq 100$), ϕ is very similar to $\sqrt{\psi}$ (less than 1% error), that is, the variance of ν, given by $\psi - \phi^2$, is small.

Thus, two main conclusions may be extracted from the study the mean effectivity of the single projection. First, it is worth noting that analytical expressions are available for the expected value of the efficiency of the single projection. Second, the resulting analytical expression for ψ is much simpler than the expression for ϕ. As it is shown in Sec. 3.2, this stands also for the evaluation of the efficiency of the multiprojection strategy and, consequently, ψ is preferred to measure the mean undervaluation of the complete multiprojection.

3.2. *Analysis of the efficiency of the multiprojection*

The main goal of this section is to describe the expected behavior of the multiprojection strategy and to relate it with some properties of the subspaces V^{I} and V^{P}. These properties are measured by a set of magnitudes that may be interpreted under a geometric viewpoint and that can be evaluated for given subspaces. Then, once the values of these magnitudes are known, they are used to compute the expected value of the (squared) effectivity index ψ and to assess the performance of the error estimator.

Let us study first a very particular case of multiprojection. For this purpose we set $\dim V^{\mathrm{I}} = 2$, $\dim V^{\mathrm{P}} = 2$ and $\dim V^{\mathrm{I}} + V^{\mathrm{P}} = 3$, i.e. V^{I} and V^{P} may be seen as two planes in a three-dimensional space. Then, the multiprojection may be seen as a way to approximate a vector $e_{\bar{h}}^*$ in $\mathbb{R}^3$ by the sum of two vectors, $e_{\bar{h}}^* \approx e_L := \varepsilon + \eta$. The first, ε, is the projection of $e_{\bar{h}}^*$ on the plane V^{I} and the second, η, is the projection of $e_{\bar{h}}^*$ of the straight line in V^{P} orthogonal to ε, see Fig. 1. The aim is to characterize the main value, ϕ, of $\|e_L\|/\|e_{\bar{h}}^*\|$ or, alternatively, the main value, ψ, of $\|e_L\|^2/\|e_{\bar{h}}^*\|^2$ for $e_{\bar{h}}^*$ ranging in the unit sphere.

It is worth noting that, in this case, the relative position of the two planes V^{I} and V^{P} is characterized by the diedric angle α and, consequently, the values of ψ and ϕ are functions of α. In the limit case $\alpha = 0$, V^{I} and V^{P} are the same plane
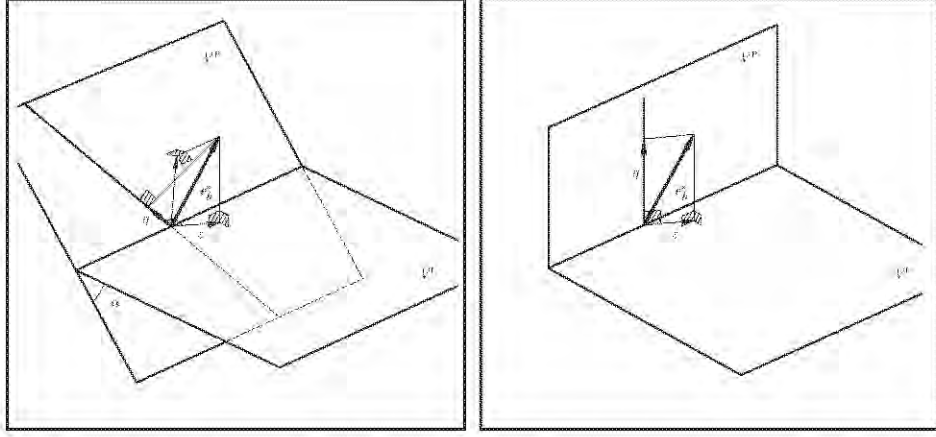
Fig. 1. Geometric illustration of the multiprojection strategy in a 3-D simple case (left) and optimal behavior for perpendicular subspaces (right).
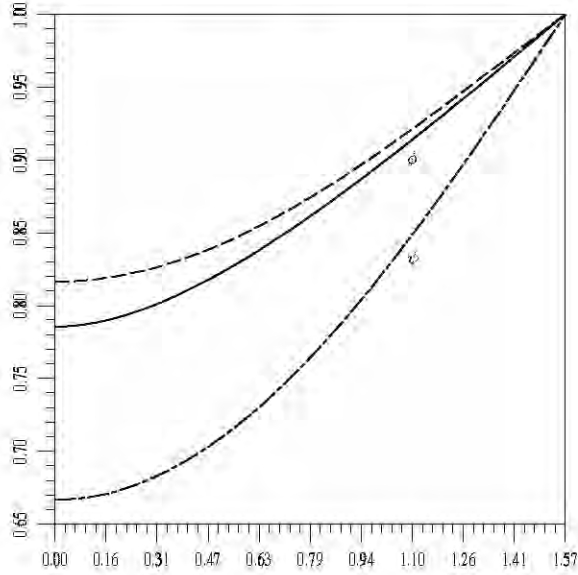

Fig. 2. Representation of ϕ and ψ vs α. The upper dashed line is the representation of $\sqrt{\psi}$.

and the second projection is useless. Then, for $\alpha = 0$, we are in the case of Sec. 3.1 with $n = 3$ and $m = 1$, consequently $\phi = \pi/4 \approx 0.785$ and $\psi = 2/3 \approx 0.667$. Moreover, for the particular case of perpendicular planes ($\alpha = \pi/2$), it can be easily seen, see Fig. 1, that the multiprojection is optimal and therefore $\psi = \phi = 1$. In this simple case both ϕ and ψ may be easily computed for every intermediate value of α between 0 and π. The curves representing ϕ and ψ versus α are plotted in Fig. 2.

The values of ϕ plotted in Fig. 2 have been numerically computed using a standard quadrature over the sphere because we did not find any analytical expression for ϕ as a function of α. On the contrary, it can be easily found as a particular case of the developments showed in the following, that ψ has a very simple expression:

$$\psi = 1 - \frac{1}{3}\cos\alpha . \tag{3.7}$$

Of course, for $\alpha = 0$ the value given by Eq. (3.7) coincides with Eq. (3.6) for $n = 3$ and $m = 1$. The variance of ν (recall Eq. (1.1)) is $\psi - \phi^2$ and must be positive, consequently $\sqrt{\psi} \geq \phi \geq \psi$. This is valid for all the cases and not only for this particular case with geometric interpretation. However, in this case the relative position of the subspaces V^{I} and V^{P} is controlled by only one parameter, the angle α, and the relationship $\sqrt{\psi} \geq \phi \geq \psi$ may be easily illustrated in Fig. 2.

From this example we can conclude that, in this particular case, the average behavior of the multiprojection process (i.e. the values of ϕ and ψ) is directly related with the angle α that describes the relative position of V^{I} and V^{P}. Moreover, while ϕ requires to be computed using a numerical quadrature, ψ has a very simple analytical expression.

In the remainder of this section this result is generalized for any dimensions of the subspaces V^{I} and V^{P}. First, an extended definition of perpendicularity is given that ensures optimality of the multiprojection strategy. Second, the general multiprojection is analyzed in a proper basis in order to obtain an algebraic expression for ψ as simple as possible. Third, this expressions are related to geometric magnitudes, analogous to the diedric angle α, that characterize the relative position of the two subspaces.

Let us denote by n_c the dimension of $V^{\mathrm{I}} \cap V^{\mathrm{P}}$ and introduce the following notation:

$$n_c := \dim(V^{\mathrm{I}} \cap V^{\mathrm{P}}), \quad n_i := \dim V^{\mathrm{I}} - n_c \quad \text{and} \quad n_p := \dim V^{\mathrm{P}} - n_c . \tag{3.8}$$

Thus, $\dim(V^{\mathrm{I}} + V^{\mathrm{P}}) = n_i + n_c + n_p$. Let us build up a particular basis of $V^{\mathrm{I}} + V^{\mathrm{P}}$, $\mathcal{B}$ that simplifies the algebraic expressions in the following. We first select an orthonormal basis, $\mathcal{B}_c$ of $V^{\mathrm{I}} \cap V^{\mathrm{P}}$. Then, two families of vectors $\mathcal{B}_i$ and $\mathcal{B}_p$ are given such that $\mathcal{B}_i \cup \mathcal{B}_c$ is an orthonormal basis of V^{I} ($\mathcal{B}_i$ completes an orthonormal basis of V^{I}) and $\mathcal{B}_c \cup \mathcal{B}_p$ is an orthonormal basis of V^{P} ($\mathcal{B}_p$ completes an orthonormal basis of V^{P}). Then, $\mathcal{B} := \mathcal{B}_i \cup \mathcal{B}_c \cup \mathcal{B}_p$ is a basis of $V^{\mathrm{I}} + V^{\mathrm{P}}$ such that the matrix of the scalar product, $\langle \cdot, \cdot \rangle$, in $\mathcal{B}$ has quite a simple shape:

$$[\langle \cdot, \cdot \rangle]_{\mathcal{B}} = \begin{pmatrix} I_{n_i} & 0_{n_i \times n_c} & A \\ 0_{n_c \times n_i} & I_{n_c} & 0_{n_c \times n_p} \\ A^{\mathrm{T}} & 0_{n_p \times n_c} & I_{n_p} \end{pmatrix} , \tag{3.9}$$

where for any value of n_1 and n_2, I_{n_1} is the $n_1 \times n_1$ identity matrix, $0_{n_1 \times n_2}$ is the $n_1 \times n_2$ null matrix, and A is a rectangular $n_i \times n_p$ which contains the cross scalar products of the elements of the bases $\mathcal{B}_i$ and $\mathcal{B}_p$. For the particular simple case with

geometrical interpretation discussed above, we have $n_i = n_c = n_p = 1$ and A is a scalar that coincides with $\cos \alpha$, such that the expression of the matrix of Eq. (3.9) particularizes in

$$[\langle \cdot, \cdot \rangle]_{\mathcal{B}} = \begin{pmatrix} 1 & 0 & \cos \alpha \\ 0 & 1 & 0 \\ \cos \alpha & 0 & 1 \end{pmatrix}. \tag{3.10}$$

Using the basis $\mathcal{B}$, the error e^*_h to be approximated with the multiprojection is expressed in vectorial form:

$$[e^*_h]_{\mathcal{B}} = \begin{pmatrix} e_i \\ e_c \\ e_p \end{pmatrix}. \tag{3.11}$$

Then, the projections ε and η may be expressed in terms of $[\langle \cdot, \cdot \rangle]_{\mathcal{B}}$ and $[e^*_h]_{\mathcal{B}}$. After some algebra, explicit expressions for the components of the multiprojection are found[2]:

$$[\varepsilon]_{\mathcal{B}} = \begin{pmatrix} e_i + A e_p \\ e_c \\ 0 \end{pmatrix} \tag{3.12}$$

and

$$[\eta]_{\mathcal{B}} = \begin{pmatrix} 0 \\ \eta_c \\ \eta_p \end{pmatrix}, \tag{3.13}$$

where

$$\begin{pmatrix} \eta_c \\ \eta_p \end{pmatrix} = \left\{ I_{n_c + n_p} - \frac{1}{\mathbf{a}^{\mathrm{T}} \mathbf{a}} \mathbf{a} \mathbf{a}^{\mathrm{T}} \right\} \begin{pmatrix} e_c \\ A^{\mathrm{T}} e_i + e_p \end{pmatrix} \tag{3.14}$$

and $\mathbf{a}$ is a $n_c + n_p$ vector defined by:

$$\mathbf{a} := \begin{pmatrix} e_c \\ A^{\mathrm{T}} e_i + A^{\mathrm{T}} A e_p \end{pmatrix}. \tag{3.15}$$

As shown in Eqs. (3.12) and (3.14) the result of the multiprojection strategy depends on (and only on) the rectangular matrix A. In fact, all the information regarding the relative position of the two subspaces V^{I} and V^{P} is contained in A. Hence, the expected values ϕ and ψ are also functions of A. The matrix A may be decomposed as

$$A = B \Sigma C^{\mathrm{T}}, \tag{3.16}$$

using the singular value decomposition (SVD),[9] where B and C are $n_i \times n_i$ and $n_p \times n_p$ unit matrices and Σ is an $n_i \times n_p$ diagonal matrix with diagonal entries σ_i, $i = 1, \ldots, \min(n_i, n_p)$ such that

$$\Sigma^{\mathrm{T}}\Sigma = \begin{pmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_{n_p}^2 \end{pmatrix} \quad \text{and} \quad \Sigma\Sigma^{\mathrm{T}} = \begin{pmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_{n_i}^2 \end{pmatrix}, \qquad (3.17)$$

where the scalars σ_i, $i = 1, \ldots, \max(n_i, n_p)$, are the singular values of A. Denoting by r the rank of A, $\sigma_i = 0$ for $i > r$. Moreover, due to the definition of the basis $\mathcal{B}$ ($\mathcal{B}_i$ and $\mathcal{B}_p$ are both orthonormal) the singular values σ_i are lower than 1. The set of singular values, σ_i, $i = 1, \ldots, r$, may be interpreted as the cosinus of a set of angles describing the relative position of V^{I} and V^{P}. Indeed, for $n_i = n_c = n_p = 1$ we are in the simple 3-D case previously introduced and, for $\alpha \neq 0$, we have $r = 1$ and $\sigma_1 = \cos \alpha$. If all the singular values are close to 0, the spaces V^{I} and V^{P} are nearly "perpendicular". The notion of perpendicularity between two vectorial subspaces must be here understood in the sense of the following definition:

Definition 3.1. Two vectorial subspaces V^{I} and V^{P} are said to be perpendicular if the orthogonal to the first, $(V^{\mathrm{I}})^{\perp}$, is included in the latter $((V^{\mathrm{I}})^{\perp} \subset V^{\mathrm{P}})$.

Remark 3.1. The orthogonal space is defined in the ambient space $V^{\mathrm{I}} + V^{\mathrm{P}}$, i.e.

$$(V^{\mathrm{I}})^{\perp} = \{\mathbf{x} \in V^{\mathrm{I}} + V^{\mathrm{P}} \text{ such that } \forall \, \mathbf{y} \in V^{\mathrm{P}}, \langle \mathbf{x}, \mathbf{y} \rangle = 0\}. \qquad (3.18)$$

Remark 3.2. The definition of perpendicularity is symmetric because

$$(V^{\mathrm{I}})^{\perp} \subset V^{\mathrm{P}} \Leftrightarrow (V^{\mathrm{P}})^{\perp} \subset V^{\mathrm{I}}. \qquad (3.19)$$

As it has already been stated for the simple case with geometrical interpretation, perpendicularity implies optimality also in the general case.

Theorem 3.1. *Let V^{I} and V^{P} be two perpendicular subspaces, then the multiprojection strategy is optimal, i.e. for every $e_h^* \in V^{\mathrm{I}} + V^{\mathrm{P}}$, $e_h^* \neq 0$, $\nu(e_h^*) = 1$.*

Proof. ε is the projection of e_h^* in V^{I}. Then, $\varepsilon^{\perp} := e_h^* - \varepsilon$ is in $(V^{\mathrm{I}})^{\perp}$. Using the perpendicularity condition, $(V^{\mathrm{I}})^{\perp} \subset V^{\mathrm{P}}$, it is found that $\varepsilon^{\perp} \in V^{\mathrm{P}}$ and, consequently, $\eta = \varepsilon^{\perp}$. Then, $e_L = \varepsilon + \eta = e_h^*$ and $\nu(e_h^*) = \|e_L\|/\|e_h^*\| = 1$. $\qquad \square$

As previously mentioned, the singular values of A are related to the relative position of V^{I} and V^{P} and represent the cosinus of a set of generalized angles between V^{I} and V^{P}. In particular, if all the singular values are 0 (A is a null matrix), the spaces V^{I} and V^{P} are perpendicular. If all the singular values are close to one (recall they must be lower than one), the angles between the subspaces are small,

the spaces V^{I} and V^{P} are almost coincident and, consequently, the contribution of the second part of the projection is small. In this case, the projection strategy behaves almost as a single projection on the first space.

Thus, the expected value of the squared effectivity index, ψ is found to be[2]

$$\psi = 1 - \frac{1}{n}\sum_{i=1}^{r}\sigma_i^2 - \frac{1}{\mathrm{meas}(S_R)}\int_{e_{\tilde{h}}^* \in S_R}\frac{[e_i^{\mathrm{T}}\Sigma(I_{n_p} - \Sigma^{\mathrm{T}}\Sigma)^{1/2}e_p]^2}{e_i^{\mathrm{T}}\Sigma\Sigma^{\mathrm{T}}e_i + e_c^{\mathrm{T}}e_c}\,, \qquad (3.20)$$

where $n := n_i + n_c + n_p$ is the dimension of $V^{\mathrm{I}} + V^{\mathrm{P}}$ and the variable $e_{\tilde{h}}^*$ of the integral ranges in the hypersphere S_R, recall Eq. (3.11). The integral on the right-hand side term of Eq. (3.20) is a function of A and, in particular, of its singular values σ_i. However, this term is difficult to handle and we did not find any analytical expression for it. In the simple 3-D case, setting $\sigma_1 = \cos\alpha$ and using the proper spherical coordinates, the expression of Eq. (3.7) is obtained from Eq. (3.20). Indeed, in this case, the integral in Eq. (3.20) is such that

$$\frac{1}{\mathrm{meas}(S_R)}\int_{e_{\tilde{h}}^* \in S_R}\frac{[e_i^{\mathrm{T}}\Sigma(I_{n_p} - \Sigma^{\mathrm{T}}\Sigma)^{1/2}e_p]^2}{e_i^{\mathrm{T}}\Sigma\Sigma^{\mathrm{T}}e_i + e_c^{\mathrm{T}}e_c}$$

$$= \frac{1}{4\pi R^2}\int_{e_{\tilde{h}}^* \in S_R}\frac{[e_i\cos\alpha(1 - \cos^2\alpha)^{1/2}e_p]^2}{\cos^2\alpha\, e_i^2 + e_c^2}$$

$$= \frac{1}{3}\cos\alpha - \frac{1}{3}\cos^2\alpha\,, \qquad (3.21)$$

recall that, in this case, $n_i = n_c = n_p = r = 1$ and $n = 3$.

Thus, in the general case, the right-hand side term of Eq. (3.20) is unknown. Nevertheless we can use this equation to find an approximate expression for ψ. Such approximate expression of ψ must be valid for the limit cases that have already been mentioned. For instance, if $r = 0$ (all $\sigma_i = 0$), $\psi = 1$ (perpendicular subspaces). Moreover, if $r = n_p$ and $\sigma_i = 1$, $i = 1,\ldots,n_p$ ($V^{\mathrm{P}} \subset V^{\mathrm{I}}$), ψ must coincide with the expression of Eq. (3.6) with $n = n_i + n_c + n_p$ and $m = n_p$.

All these requirements are fulfilled if the following approximation is assumed:

$$\psi \approx 1 - \frac{1}{n}\sum_{i=1}^{r}\sigma_i\,. \qquad (3.22)$$

It must be remarked that the expression of Eq. (3.22) is not exact, i.e. the integral of Eq. (3.20) has not been explicitly computed as a (simple) function of the singular values of A. However, the very simple expression of Eq. (3.22) gives quite a good answer to evaluate a number that must be greater than $1 - \frac{r}{n}$ and lower than $1 - \frac{1}{n}\sum_{i=1}^{r}\sigma_i^2$. Especially if it is noticed that, for the simple 3-D case with analytical expression, the exact answer is a particular case of this general approximate expression.

Moreover, next section shows an example demonstrating that the approximate evaluation fits in a general and realistic situation the exact value.

4. Numerical Example

In this section the technique introduced above is used to *a priori* evaluate the expected value of the effectivity index of the studied estimator in a concrete problem. This can be used to predict and, hence, to correct the estimate. The problem is a standard example used in the validation of error estimators. The Poisson equation is solved in an L-shaped domain discretized with the mesh shown in Ref. 3. The proper source term and Dirichlet boundary conditions are imposed to obtain a given exact solution (in this case $u(x, y) = x^2 + y^2$). Thus, the exact error is evaluated and the behavior of the error estimate may be studied and compared with the exact error. This example has been used to demonstrate the robustness of the studied error estimator.[3] The global effectivity index is found to be 91.9% and the distribution of the local effectivity index (element by element) is quite uniform (ranging from 82% to 95%).

As previously mentioned the analysis of the average undervaluation introduced in the error estimation is split into three phases that are studied independently.

First, the effect of approximating the exact error by a reference error belonging to a finite dimensional (even if fine) space is accounted for using Eq. (1.2). In this case bilinear elements are used ($p = 1$) and the refinement factor is 4 ($\tilde{h}/h = 0.25$). Thus,

$$\frac{\|e_{\tilde{h}}\|}{\|e\|} \approx [1 - 0.25^2]^{1/2} = 0.968 \,. \tag{4.23}$$

Second, the effect of the points where the estimate is forced to vanish is accounted for. The number of free nodes in reference mesh is 1,425. The number of points where the estimate is forced to vanish (center points of the interior edges of the elements of the computational mesh) is 148. Then, the expected value of the undervaluation introduced in this second phase is described either by $\phi_{148}^{1425} = 0.946628$ or $\psi_{148}^{1425} = 0.89614$. Note that, as previously mentioned, the value of $\sqrt{\psi} = 0.946647$ is very close to ϕ. Up to the required accuracy (three significant digits) $\sqrt{\psi}$ and ϕ may be considered equal. Thus, the undervaluation associated with this phase is taken to be 0.947. That is,

$$\frac{\|e_{\tilde{h}}^*\|}{\|e_{\tilde{h}}\|} \approx 0.947 \,. \tag{4.24}$$

Third, the undervaluation introduced in the multiprojection strategy is assessed.

In the considered case the dimension of V^{I} is $n_i + n_c = 765$, the dimension of V^{P} is $n_p + n_c = 802$ and the dimension of the intersection, $V^{\mathrm{I}} \cup V^{\mathrm{P}}$, is $n_c = 340$. Thus, the A matrix is a 425×462 matrix. The study of the singular value decomposition of this matrix and the corresponding application of Eq. (3.22) to obtain an approximate value for the average underestimation is cumbersome and computationally expensive. However, both V^{I} and V^{P} are generated as a sum of vectorial spaces associated with (almost) identical meshes. Denoting by M the

number of elements in the original computational mesh and by M' the number of patches (which coincides with the number of free nodes)

$$V^I = \bigoplus_{k=1}^{M} V_k^I \quad \text{and} \quad V^P = \bigoplus_{l=1}^{M'} V_l^P, \tag{4.25}$$

where V_k^I is the interpolation space associated with the submesh discretizing the kth element and V_l^P is the interpolation space associated with the submesh discretizing the lth patch. The set of element spaces $\{V_k^I\}_{k=1,...,M}$ is orthogonal as well as the set of the set patch spaces $\{V_l^P\}_{l=1,...,M'}$. Moreover, if the kth element and the lth patch are disjoint, V_k^I and V_l^P are also orthogonal. On the other hand, the value of ψ depends only on the "relative position" of V^I and V^P. The singular values of A are interpreted as the cosinus of a set of angles describing precisely this relative position. Due to the decomposition described in Eq. (4.25), the singular values are different from zero, i.e. the angles different to the right angle, must correspond to local spaces V_k^I and V_l^P associated with an element and a patch which are non-disjoint. These angles are essentially a function of the topology of the intersection between the element and the patch and therefore will be the same for every couple of local spaces V_k^I and V_l^P. In the following it is assumed that the singular values of A are represented by the singular values arising in a much simpler problem related with the sample spaces V_1 and V_2 generated by the mesh shown in Fig. 3. Note that if the singular values of A are simply a repetition of the singular values of the sample (and simple) problem, the result obtained applying Eq. (3.22) is the same for the original problem and the simplified one.

Thus, in the following, the average undervaluation introduced in the multiprojection strategy is analyzed replacing V^I and V^P by V_1 and V_2, respectively. The
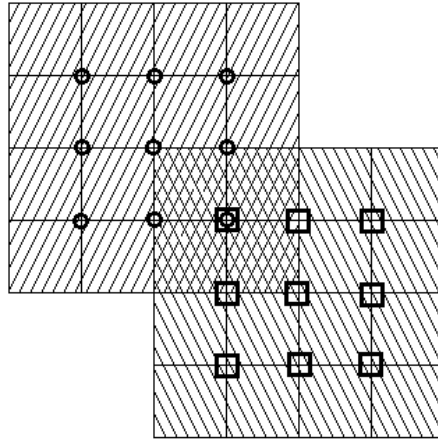

Fig. 3. Sample meshes generating V_1 (nodes marked by circles) and V_2 (nodes marked by squares). Measuring the "relative position" between V_1 and V_2 suffices to characterize the relative position between V^I and V^P and, hence, the average behavior of the multiprojection strategy.

dimension of V_1 is $n_i + n_c = 8 + 1 = 9$, the dimension of V_2 is $n_p + n_c = 8 + 1 = 9$. Consequently, in this case, A is an 8×8 matrix. The analysis of the matrix shows that the range is $r = 3$ and the corresponding singular values are

$$\sigma_1 = 0.23932, \quad \sigma_2 = 0.11452 \quad \text{and} \quad \sigma_3 = 0.01276.$$

Recall that small singular values of A are associated with nearly perpendicular spaces and a multiprojection strategy with a good behavior. The approximate expression of Eq. (3.22) yields

$$\psi \simeq 0.978435.$$

In order to verify that the approximation introduced in Eq. (3.22) is acceptable, at least in this simple case, 100,000 vectors are randomly generated on the unit sphere of the 17 dimensions space $V_1 + V_2$. The effectivity index of the projection strategy, ν, is computed for each of these vectors and the mean value of ν, ϕ, and ν^2, ψ, are evaluated. The obtained values are

$$\phi = 0.990121 \quad \text{and} \quad \psi = 0.980736.$$

Consequently, the variance of ν is $0.019924 \simeq 0.02$. It is worth noting that the predicted value of ψ (0.978435) and the computed value (0.980736) are equal up to the second significant digit, i.e. the error introduced in the approximation of Eq. (3.22) is much less than the variance of ν.

Thus, the predicted undervaluation in the transformation from $e_{\tilde{h}}^*$ to e_L is taken as the square root of the predicted value for ψ, i.e.

$$\frac{\|e_L\|}{\|e_{\tilde{h}}^*\|} \simeq 0.989. \tag{4.26}$$

Resuming the results of Eqs. (4.23), (4.24) and (4.26), an evaluation of the expected value of the effectivity index is found:

$$\frac{\|e_L\|}{\|e\|} = \frac{\|e_{\tilde{h}}\|}{\|e\|} \frac{\|e_{\tilde{h}}^*\|}{\|e_{\tilde{h}}\|} \frac{\|e_L\|}{\|e_{\tilde{h}}^*\|} \simeq 0.968 \times 0.947 \times 0.989 = 0.907. \tag{4.27}$$

Recall that, in this problem the obtained effectivity index is 0.919. If the prediction on the behavior of the effectivity index is used to improve the error estimate, the corrected estimate would be $e_L/0.907$. The new value for the effectivity index is $0.919/0.907 = 1.013$. That results on improving the quality of the error estimate. The original estimate has an error of 8.1% in the evaluation of the error. The "error in the error" for the corrected estimate is 1.3%.

References

▶1. I. Babuška, R. Durán and R. Rodríguez, *Analysis of the efficiency of an a posteriori error estimator for linear triangular finite elements*, SIAM J. Numer. Anal. **29** (1992) 947–964.

2. P. Díez, *Un nuevo estimador de error para el método de los elementos finitos*, Doctoral Thesis, Universitat Politècnica de Catalunya, Barcelona (1996).

▶3. P. Díez, J. J. Egozcue and A. Huerta, *A posteriori error estimation for standard finite element analysis, Computer Method Appl. Mech. Engrg.* **163** (1998) 141–157.

4. P. Díez, J. J. Egozcue and A. Huerta, *Analysis of the average efficiency of an error estimator,* in **Finite Element Methods: Superconvergence, Post-processing and a Posteriori Estimates,** eds. M. Křížek *et al.* (Marcel Dekker, 1998), pp. 113–126.

▶5. P. Díez and A. Huerta, *A unified approach to remeshing strategies for finite element h-adaptivity, Computer Methods Appl. Mech. Engrg.* **176** (1999) 215–229.

▶6. P. Díez, M. Arroyo and A. Huerta, *Adaptivity based on error estimation for viscoplastic softening materials, Mech. Cohesive-Frictional Materials* **5** (2000) 87–112.

▶7. R. Durán and R. Rodríguez, *On the asymptotic exactness of Bank–Weiser's estimator, Numer. Math.* **62** (1992) 297–303.

▶8. A. Huerta and P. Díez, *Error estimation including pollution assessment for nonlinear finite element analysis, Computer Methods Appl. Mech. Engrg.* **181** (2000) 21–41.

9. D. W. Lewis, **Matrix Theory** (World Scientific, 1991)

▶10. O. C. Zienkiewicz and J. Z. Zhu, *A simple error estimator and adaptive procedure for practical engineering analysis, Internat. J. Numer. Methods Engrg.* **24** (1987) 337–357.