A General Lagrangian Formulation for the computation of a-*posteriori* Finite Element Bounds

A.T. Patera and J. Peraire

Massachusetts Institute of Technology, Cambridge MA 02139, USA

Abstract. We present our general framework for implicit a-*posteriori* finite element procedures to compute upper and lower bounds for functional outputs of finite element solutions. The approach is based on a finite element domain decomposition technique and the construction of an augmented Lagrangian, in which the objective is a "quadratic" energy re-formulation of the desired output, and the constraints are the finite element equilibrium equations and inter-subdomain continuity requirements. Bounds for the output, on a suitably refined "truth" mesh, are then obtained by appealing to a dual min-max relaxation, evaluated for Lagrange multipliers computed on a coarse "working" mesh. For coercive problems, the computed bounds are uniformly valid regardless of the size of the underlying coarse discretization. An extension of the basic method to deal with non-coercive problems is also presented. In this case, the bounds computed can only be proven to be asymmmptotic although in practice, they are also found to be valid for very coarse meshes. Adaptive mesh procedures can be incorporated naturally into the bounding framework to produce optimized meshes that meet a target bound gap. The methods presented are illustrated with several application examples.

1 Introduction

Engineering applications require the prediction of certain quantities of interest, or outputs – such as compliance, stresses, flow rates and heat transfer. These outputs are functionals of field variables – such as displacements, velocity field, temperature – which, in turn, are solutions of partial differential equations that need to be approximated numerically. Engineering decisions are thus based on approximations to the desired outputs that are generated from computed approximations to the field variables. It is often difficult to determine a-*priori* the size of the discretization that would yield the outputs at a required level of accuracy. In fact, it is well understood that for a given field solution, different outputs can be predicted at varying levels of accuracy. The situation is further complicated in a real design, or optimization, setting where one may require multiple appeals to the PDE solver to evaluate solutions for different values of the design parameters. In this case, using a conservative fine discretization for every solution, in an attempt to guarantee a prescribed accuracy, may prove prohibitive.

In these notes, we present a summary of our work on a-posteriori error estimation. We propose a finite element a-posteriori method to compute inexpensive upper and lower bounds for engineering outputs of interest. The proposed method requires the solution of non-linear local Neumann sub-problems and is therefore referred to as an implicit method [2]. When compared to the simpler more inexpensive a-*posteriori* explicit methods which only require residual evaluations (e.g. [3]), implicit methods offer the potential for quantitative constant-free bounds. The original implicit methods [9], [1], [4] were developed, for linear self-adjoint problems, to provide bounds for the energy norm of the error. In reality however, it is not the error in the energy norm which is of interest, but the error in the quantities on which an engineering decision will be based, such as stresses, deflection etc. The first attempts at developing a-posteriori error estimation for functional outputs were carried out in the context of explicit error estimation [5]. It turns out that the error in the quantities of interest can be related to a weighted norm of the residual where the specific weights are determined by the solution of an adjoint problem. Unfortunately, the explicit nature of the procedure means that these estimates contain unknown constants, which severely limit its quantitative value.

The development of an implicit procedure yielding a-posteriori constant-free bounds for linear-functional outputs of partial differential equations was presented in [15], [16], [18]. The method is applicable to elliptic coercive problems including non-symmetric terms. The approach is based on a finite element domain decomposition technique and the construction of an augmented Lagrangian, in which the objective is a "quadratic" energy re-formulation of the desired output, and the constraints are the finite element equilibrium equations and inter-subdomain continuity requirements. Bounds for the output, on a suitably refined "truth" mesh, are then obtained by appealing to a dual min-max relaxation, evaluated for Lagrange multipliers computed on a coarse "working" mesh. The focus of these notes is on the computation of bounds, but the method presented provides naturally a local indicator that can be used to drive a mesh adaptive procedure [18].

The framework presented extends naturally to the computation of bounds for reduced basis methods. This extension is treated in detail in [20].

1.1 Motivation example

We consider an example that was introduced and analyzed in [13]. Figure 1 shows a stack of of heat generating units, such as, for example, computer chips. These units are supported by a material with low thermal conductivity and surrounded by air. A possible question a designer may want to ask is : for a given Q and a given geometry, is the average temperature $\bar{\theta}$ on Γ, within the "acceptable" range $[\bar{\theta}_{min}, \bar{\theta}_{max}]$?

An appropriate mathematical model to describe the relevant physical phenomena is given by the incompressible Navier-Stokes equations complemented

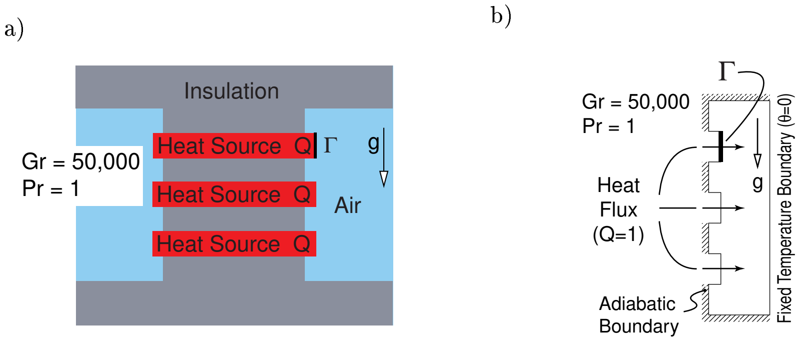


Fig. 1. Illustrative example showing three heat generating units supported by an insulating material and surrounded by air: (a) problem description, and (b) computational domain.

with the energy equation to account for density variations due to temperature changes via the Boussinesq approximation. The relevant design parameters for this problem are the Prandtl and Grashof numbers [10], and the geometry. An output of engineering interest is the average temperature $\bar{\theta}$ over the boundary of the top unit.

Given a suitable mathematical model we can now use a standard finite element solver to predict the flow and temperature distributions in an attempt to answer the original design question. One of the first decisions we must make is about the discretization to be employed. Figure 2 shows a coarse "working" triangulation $\mathcal{T}_H$, leading to an inexpensive but uncertain result $\bar{\theta}_H$, and a very fine "truth" triangulation $\mathcal{T}_h$, leading to an accurate but expensive prediction $\bar{\theta}_h$. The fine triangulation is sufficiently conservative so that we confidently assume that $\bar{\theta}_h \approx \bar{\theta}$.

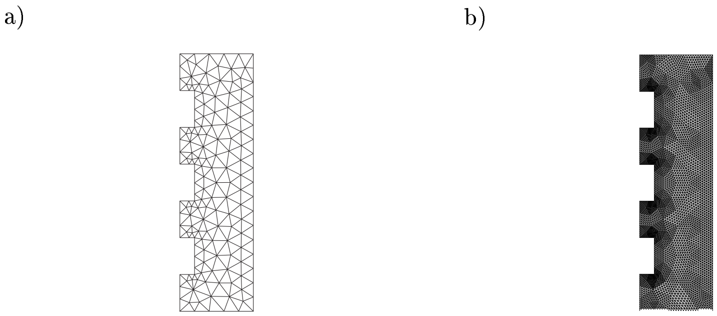


Fig. 2. Finite Element meshes: (a) "working" mesh, and (b) "truth" mesh.

The use of rigorous a-*posteriori* bounds can help determine a suitable discretization. The proposed procedure allows us to calculate upper and lower bounds for the output of interest computed on the fine mesh with a computational cost which is roughly proportional to that of the coarse mesh. Of course if the initial coarse mesh happens to be inadequate to resolve the features of interest the computed upper and lower bounds will most likely be very far apart. The finer the mesh used to compute the solution the smaller the bound gap we can expect.

In figure 3 we illustrate the results obtained by our a-*posteriori* error estimation approach. On the initial mesh, shown on the left, the bound procedure predicts that the solution $\bar{\theta}^{init.}$ lies inside the interval $[0.253, 0.297] = 0.275 \pm 0.022(8\%)$. If the bound gap is deemed to be too large, the error estimation procedure can be used to drive the a mesh adaption algorithm. The adapted mesh together with the solution computed is shown on the right, and for this mesh, the bound procedure predicts that the temperature $\bar{\theta}^{adapt.}$ is within the interval $[0.2745, 0.2795] = 0.277 \pm 0.0025(1\%)$.

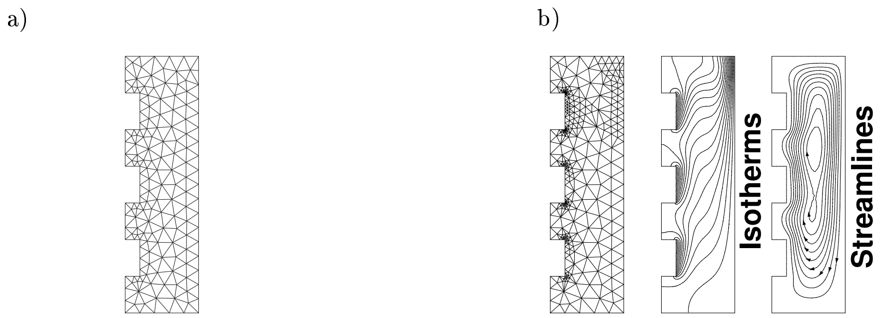


Fig. 3. (a) Initial mesh, and (b) adapted mesh and computed solution

In the following sections we develop the methods which are used to produce the results described above. First, we introduce the general formulation in a relatively abstract setting. We then present, in the second part, efficient relaxations to compute accurate bounds for finite element solutions of partial differential equations.

2 Model Problem

2.1 Continuous Problem

In order to facilitate the description of the methods to be presented we will introduce a simple model problem. We consider the following convection-diffusion problem

$$-\nu \nabla^2 u + U \cdot \nabla u = f^I, \quad \text{for all} \;\; x \in \Omega \tag{1}$$

and

$$u = 0 \quad \text{on} \quad \Gamma \; .$$

Here, u is the unknown function, U is a given velocity field satisfying $\nabla \cdot U = 0$, ν is a diffusivity constant and f^I is a known forcing term which can be a function of x. We also consider our output of interest to be of the form

$$s = \int_\Omega f^O u \, d\Omega,$$

where f^O is a given function defined over Ω.

In order to introduce the weak form of this problem we define $X(\Omega) \equiv H^1_0(\Omega)$ (or simply X) to be the space of solutions and variations. The weak form of the problem is then: find $u \in X$ such that, $\forall v \in X$,

$$\int_\Omega \nu \nabla u \cdot \nabla v \, d\Omega + \int_\Omega U \cdot \nabla u \, v \, d\Omega = \int_\Omega f^I v \, d\Omega, \tag{2}$$

or, in shorthand form: find $u \in X$ such that

$$a(u, v) = l(v), \qquad \forall v \in X,$$

where $a : X \times X \mapsto \mathbb{R}$ and $l : X \mapsto \mathbb{R}$ are given by

$$\begin{aligned} a(w, v) &= \int_\Omega \nu \nabla w \cdot \nabla v \, d\Omega + \int_\Omega U \cdot \nabla w \, v \, d\Omega \\ l(v) &= \int_\Omega f^I v \, d\Omega \end{aligned}$$

It is relatively straightforward to show that all solutions to the strong statement (1) will also be a solution of the weak form (2). To do that, we simply select an arbitrary function $v \in X$, multiply it by the differential equation, and integrate over Ω

$$\int_\Omega \{ -\nu \nabla^2 u + U \cdot \nabla u - f^I \} \, v \, d\Omega = 0.$$

Given that v belongs to X, we can always integrate by parts the term involving the second derivative to obtain the weak form.

To show that solution of the weak statement (2) is also a solution of the original strong form (1), requires reversing the steps followed above. This will only be possible when the solution to our weak problem is sufficiently regular so that the integration by parts can be performed. This will not be the case in general. In summary, we note that the weak statement places weaker regularity requirements on the solution than the original strong form.

2.2 Galerkin Approximation

Given a finite dimensional subspace $\tilde{X} \subset X$, $dim\{\tilde{X}\} = n$, the Galerkin approximation $\tilde{u} \in \tilde{X}$ satisfies

$$a(\tilde{u}, v) = l(v), \quad \forall v \in \tilde{X}.$$

Let the functions $\phi_i, i = 1, \ldots, n$ form a basis of $\tilde{X}$

$$\tilde{X} = span\{\phi_1, \ldots, \phi_n\}.$$

By writing

$$\tilde{u} = \sum_{i=1}^{n} \tilde{u}_i \phi_i(x), \qquad v = \sum_{i=1}^{n} v_i \phi_i(x),$$

we can explicitly carry out the spatial integrations and generate a linear set of equations for the coefficients of $\tilde{u}$, $\tilde{u}_i$. This can be expressed in vector from as

$$\underline{v}^T \tilde{A} \underline{\tilde{u}} = \underline{v}^T \underline{\tilde{F}}, \quad \forall \underline{v} \in \mathbb{R}^n,$$

or equivalently,

$$\tilde{A}\, \underline{\tilde{u}} = \underline{\tilde{F}},$$

where $\tilde{A} \in \mathbb{R}^{n \times n}$ and $\underline{\tilde{u}}, \underline{\tilde{F}} \in \mathbb{R}^n$.

Finally, an approximation to the output of interest, $\tilde{s}$, is obtained as

$$s \approx \tilde{s} = \int_{\Omega} F^O \tilde{u}\, d\Omega.$$

Given an approximation to s, one may ask how good is this approximation. The aim of our a-*posteriori* bounding procedures is to answer this question by computing approximations in the $\tilde{X}$ space which can yield guaranteed upper and lower bounds for s.

3 Function Minimization

In this section we present a brief review of some basic concepts in function minimization. The results are presented in the context of general finite or infinite dimensional spaces. For clarity purposes, we use the vector-like notation so that all spaces can be readily interpreted as $\mathbb{R}^n$. That is, for $w, v \in X$ with $dim\{X\}$ *finite* or *infinite* we write $a(w, v)$"="$v^T A w$ and $l(v)$"="$v^T F$. The term $v^T A w$ can be interpreted as a vector multiplying a matrix times a vector when the dimension of X is finite. But we will also assume that, when the dimension of X happens to be infinite in the sense of the usual distributional operator and associated duality pairing, it has the meaning of $a(w, v)$. Regarding the function minimization procedures described below we refer to references [22,6] for additional details.

3.1 Unconstrained Minimization

We consider the minimization of an objective function $f : Y \mapsto \mathbb{R}$ defined over a space Y (e.g. $Y \equiv \mathbb{R}^n$)

$$f^* = \min_{v \in Y} f(v)$$

Here, $f^* \equiv f(v^*)$ denotes the minimum which is attained by the minimizer v^*. A more precise statement would replace minimum (maximum) by infimum (supremum), since in general we can not guarantee that the infimum (supremum) will belong to Y. This subtlety will not be important for our discussions and in these notes we will simply use min and max.

Upper Bound for f^* An upper bound, f^+, for the minimum f^*, can be easily computed by setting $f^+ = f(\bar{v})$ for any $\bar{v} \in Y$,

$$f^* = \min_{v \in Y} f(v) \equiv f(v^*) \leq f(\bar{v}) \equiv f^+.$$

For f sufficiently regular, we expect a reasonable bound if $\bar{v} \approx v^*$.

For illustration purposes we consider the case in which f is a linear quadratic function of $v \in \mathbb{R}^n$ of the form

$$f(v) = \frac{1}{2} v^T A v - b^T v$$

with $A \in \mathbb{R}^n \times \mathbb{R}^n$ is symmetric positive definite, and $b \in \mathbb{R}^n$, as illustrated in figure 4. The stationarity condition implies that

$$Av^* = b,$$

and hence,

$$f^* = \min_{v \in Y} f(v) = f(v^*) = -\frac{1}{2} v^{*T} A v^*.$$

In order to show that the stationary point is a minimum we need to show that $f(v^*) < f(\bar{v}), \quad \forall \bar{v} \in Y$. To this end we consider $v = v^* + \epsilon$ for $\epsilon \in \mathbb{R}^n$ small (in some appropriate norm), and write

$$\begin{aligned}
f(\bar{v}) &= \frac{1}{2}(v^* + \epsilon)^T A (v^* + \epsilon) - (v^* + \epsilon)^T b \\
&= \frac{1}{2} v^{*T} A v^* - v^{*T} b \\
&+ \frac{1}{2}\epsilon^T A v^* + \frac{1}{2} v^{*T} A \epsilon - \epsilon^T b \\
&+ \frac{1}{2}\epsilon^T A \epsilon \\
&= f(v^*) + \underbrace{\epsilon^T (Av^* - b)}_{=0} + \frac{1}{2}\epsilon^T A \epsilon
\end{aligned}$$

or,

$$f^{+} = f(\bar{v}) = f(v^{*}) + \frac{1}{2}\epsilon^{T} A \epsilon,$$

which implies that the error in the objective function $f^{+} - f^{*}$ is of order $\mathcal{O}(\epsilon^{2})$.

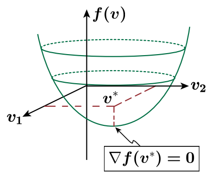


Fig. 4. Illustration of a linear-quadratic objective function in $\mathbb{R}^2$

We see that in the case of unconstrained minimization, upper bounds are trivially obtained. The computation of lower bounds is less straightforward and, for this purpose, we will need some additional tools which are considered in the following section.

3.2 Equality Constrained Minimization and Duality

We now consider the minimization of an objective function $f : Y \mapsto \mathbb{R}$ subject to equality constraints, which we write as $g(v) = 0$, where $g : Y \mapsto \Lambda$. In the finite dimensional case we will typically have $Y \equiv \mathbb{R}^{n}$, and $\Lambda \equiv \mathbb{R}^{m}$. We write

$$f^{*} = \min_{\substack{v \in Y \\ g(v)=0}} f(v).$$

We note that one obvious way of obtaining lower bounds in the constrained problem is to simply remove the constraint and perform the minimization over a larger space. Unfortunately, this approach is often too simplistic and does not yield useful bounds.

A convenient way to address these type of constrained minimization problems is to introduce a Lagrangian $\mathcal{L}(v, \lambda) : Y \times \Lambda \rightarrow \mathbb{R}$, as

$$\mathcal{L}(v, \lambda) = f(v) + \lambda^{T} g(v)$$

where λ are the Lagrange multipliers, and write

$$f^{*} = \min_{v \in Y} \max_{\lambda \in \Lambda} \mathcal{L}(v, \lambda). \tag{3}$$

The proof of the above statement is straightforward. Performing first the maximization with respect to lambda, we note that

$$\max_{\lambda \in \Lambda} \mathcal{L}(v, \lambda) = \begin{cases} f(v) \text{ if } g(v) = 0 \\ +\infty \text{ if } g(v) \neq 0 \end{cases}$$

since, by construction, $\mathcal{L}(v, \lambda)$ is a linear function of λ. Hence, (3) follows after taking the minimum over $v \in Y$.

By maximizing the Lagrangian with respect to λ, and minimizing with respect to v, we can write

$$\mathcal{L}(v, \lambda^*(v)) \equiv \max_{\lambda \in \Lambda} \mathcal{L}(v, \lambda),$$
$$\mathcal{L}(v^*(\lambda), \lambda) \equiv \min_{v \in Y} \mathcal{L}(v, \lambda),$$

and consequently,

$$\mathcal{L}(v, \lambda^*(v)) \geq \mathcal{L}(v, \lambda) \geq \mathcal{L}(v^*(\lambda), \lambda), \quad \forall v, \lambda.$$

Since this expression is applicable to all v and λ, we can now take the minimum with respect to v and the maximum with respect to λ to prove the following duality statement

$$\min_{v \in Y} \max_{\lambda \in \Lambda} \mathcal{L}(v, \lambda) \geq \max_{\lambda \in \Lambda} \min_{v \in Y} \mathcal{L}(v, \lambda). \tag{4}$$

We can define the duality gap as

$$\delta = \min_{v \in Y} \max_{\lambda \in \Lambda} \mathcal{L}(v, \lambda) - \max_{\lambda \in \Lambda} \min_{v \in Y} \mathcal{L}(v, \lambda).$$

For linear-quadratic objective functions and linear constraints [22], and more generally, for convex functions with convex constraints [6], the duality gap can be shown to be zero. For more general functions, $\delta \geq 0$.

Although duality can be invoked to calculate lower bounds for s, the computation of the $\max - \min$ appearing in the above expression is often just as complicated as solving the original $\min - \max$ problem. Below we consider some relaxations that can be used to compute good lower bounds at a much reduced cost.

Lower Bound for f^* Combining equation (3) with duality we can write,

$$f^* = \min_{v \in Y} \max_{\lambda \in \Lambda} \mathcal{L}(v, \lambda) \geq \max_{\lambda \in \Lambda} \min_{v \in Y} \mathcal{L}(v, \lambda) \geq \min_{v \in Y} \mathcal{L}(v, \bar{\lambda}), \quad \forall \bar{\lambda} \in \Lambda, \tag{5}$$

and we simply define $f^- \equiv \min_{v \in Y} \mathcal{L}(v, \bar{\lambda})$. In addition, we would expect accurate bounds when $\bar{\lambda} \approx \lambda^*$.

For illustration purposes, we consider a constrained minimization problem involving a linear-quadratic functional, with linear constraints of the form

$$\begin{aligned} f(v) &= \frac{1}{2} v^T A v - b^T v \\ g(v) &= Bv - c \end{aligned}$$

with $A \in \mathbb{R}^n \times \mathbb{R}^n$ is symmetric positive definite, $B \in \mathbb{R}^m \times \mathbb{R}^n$, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}^m$.

The Lagrangian takes the form

$$\mathcal{L}(v, \lambda) = v^T A v + b^T v + \lambda^T (Bv - c),$$

and the constrained minimization problem can be written as,

$$\begin{aligned} f^* &= \min_{v \in \mathbb{R}^n} \ \max_{\lambda \in \mathbb{R}^m} \ \mathcal{L}(v, \lambda) \\ &= \max_{\lambda \in \mathbb{R}^m} \ \min_{v \in \mathbb{R}^n} \mathcal{L}(v, \lambda) \ \equiv \mathcal{L}(v^*, \lambda^*). \end{aligned}$$

The last equality follows from the fact that the duality gap for this problem is zero [6].

The stationarity conditions can be written as

$$\begin{aligned} 2Av^* + b + B^T \lambda^* &= 0 \\ Bv^* - c &= 0, \end{aligned}$$

and it is not difficult to show that the point (v^*, λ^*) does indeed correspond to a saddle point.

The value of f^* can thus be expressed in terms of (v^*, λ^*) as

$$f^* = -v^{*T} A v^* - \lambda^{*T} c$$

If we now consider an approximate multiplier $\bar{\lambda} = \lambda^* + \epsilon$, for $\epsilon \in \mathbb{R}^m$ small (in some appropriate norm), and use (5), we have that,

$$f^* \geq \min_{v \in \mathbb{R}^n} \mathcal{L}(v, \bar{\lambda}) \equiv f^-$$

The solution $\hat{v}$ satisfying the above problem will be determined by solving

$$2A\hat{v} + b + B^T \bar{\lambda} = 0.$$

Since $\bar{\lambda} = \lambda^* + \epsilon$, we have that

$$\hat{v} = -\frac{1}{2} \ A^{-1} (b + B^T (\lambda^* + \epsilon)) = v^* - \frac{1}{2} A^{-1} B^T \epsilon.$$

Finally,

$$f^- = \hat{v}^T A \hat{v} + b^T \hat{v} + (\lambda^{*T} + \epsilon^T)(B\hat{v} - c),$$

or,

$$\begin{aligned} f^- &= -v^{*T}Av^* - \lambda^{*T}c \\ &\quad -\frac{1}{4}\,\epsilon^T BA^{-1}B^T\epsilon \\ &\quad +\epsilon^T(Bv^* - c) \\ &= f^* - \frac{1}{4}\,\epsilon^T BA^{-1}B^T\epsilon. \end{aligned}$$

This expression shows that the error in approximating the solution is quadratic in the error in approximating the Lagrange multiplier, ϵ. e.g. $f^* - f^- = \mathcal{O}(\epsilon^2)$.

4 Abstract Bound Formulations

4.1 "Energy" Output

In this section, we extend the bounding strategies presented above to outputs which are functionals of solutions of our abstract problem

$$A\,u = F, \tag{6}$$

where $A : Y \mapsto Y'$ is a symmetric positive definite operator, $u \in Y$, and $F \in Y'$. Here, Y' stands for the dual space of Y. We recall that for $\mathbb{R}^n$, $Y \equiv Y'$.

To simplify the presentation we start by considering a particular output s, which is given simply as

$$s = u^T F.$$

By introducing the linear-quadratic functional $J : Y \mapsto \mathbb{R}$,

$$J(v) = \frac{1}{2}v^T Av - v^T F,$$

problem (6) can be re-formulated as an unconstrained minimization problem

$$J(u) = \min_{v\in Y} J(v),$$

where the minimizer $v^* \equiv u$.

At the minimum, we will have that

$$J(u) = \frac{1}{2}u^T Au - u^T F = -\frac{1}{2}u^T Au = -\frac{1}{2}u^T F,$$

and therefore our "energy" output can be easily related to $J(u)$

$$s = u^T F = u^T Au = -2J(u). \tag{7}$$

In some situations, it will prove advantageous to work with the difference between an approximation to the solution, $\tilde{u}$, and the exact solution, u, rather than with u itself. Generally, the approximation $\tilde{u}$ will be a Galerkin solution computed on a subspace $\tilde{Y} \subset Y$. That is, $\tilde{u}$ will be the solution to a problem $\tilde{A}\,\tilde{u} = \tilde{F}$.

Given an approximation $\tilde{u} \in \tilde{Y} \subset Y$ to u, $\tilde{u} \approx u$, we define the error $e \in Y$ as

$$e = u - \tilde{u}.$$

We can also write down the problem satisfied by the error as

$$A\,e = F - A\tilde{u} \equiv R, \tag{8}$$

and express the output in terms of the approximate output and the error

$$s = \tilde{u}^T F + e^T F = \tilde{s} + e^T F.$$

Analogously, we can introduce the linear quadratic functional $J^{err} : Y \mapsto \mathbb{R}$,

$$J^{err}(v) = \frac{1}{2} v^T A v - v^T R,$$

and characterize the solution e of problem (8) as the minimizer, v^*, of $J^{err}(v)$. That is,

$$J^{err}(e) = \min_{v \in Y} J^{err}(v).$$

At the minimum

$$J^{err}(e) = \frac{1}{2} e^T A e - e^T R = -\frac{1}{2} e^T A e,$$

and since $e^T F = e^T R + e^T A\tilde{u} = e^T A e + R^T \tilde{u}$ we have that

$$s = \tilde{s} + \tilde{u}^T R + e^T A e = \tilde{s} + \tilde{u}^T R - 2J^{err}(e). \tag{9}$$

We note that, in such cases where the approximation $\tilde{u}$ is a Galerkin approximation carried out in a subspace $\tilde{Y} \subset Y$, the residual R will be orthogonal to all the elements of $\tilde{Y}$, and in particular $\tilde{u}^T R = 0$.

Lower Bounds Given that

$$J(\bar{v}) \ge \min_{v \in Y} J(v), \qquad \forall \bar{v} \in Y,$$

we can replace u by any $\bar{v} \in Y$ in (7) to obtain a lower bound for s,

$$s^- \equiv -2J(\bar{v}) \le -2J(u) \equiv s.$$

Alternatively, a lower bound can also be obtained by replacing e for any $\bar{v} \in Y$ in (9),

$$s^- \equiv \tilde{s} + \tilde{u}^T R - 2J^{err}(\bar{v}) \le \tilde{s} + \tilde{u}^T R - 2J^{err}(e) \equiv s \tag{10}$$

It is worth noting that, in the case when $\tilde{u}$ corresponds to a Galerkin approximation, the term $\tilde{u}^T R$ is zero, and the choice $\bar{v} = 0$ in 10 gives $s^- \equiv \tilde{s}$, thus showing that the value of the output computed using the Galerkin approximaion is indeed a lower bound for the true output.

Upper Bounds Since the output of interest is $-2J(u)$, upper bounds for the output of interest will be readily obtained if we can compute lower bounds for $J(u)$ (or $J^{err}(e)$).

We will consider here three possible approaches.

Approach I : Relax Y

If we can construct $\hat{Y} \supset Y$, we can write

$$J(\hat{u}) = \min_{\hat{v} \in \hat{Y}} J(\hat{v}) \leq J(u).$$

Although simple, this approach is unlikely to yield useful bounds in practice and often we will find that $J(\hat{u}) \to -\infty$. A more useful approach will often result from relaxing Y, and the same time introducing some additional constraints thus formulating the problem as a constrained minimization problem.

Let $B : \hat{Y} \mapsto \Lambda$ be such that

$$Y = \{\hat{v} \in \hat{Y} \mid B\hat{v} = 0\}. \tag{11}$$

For instance, when $\hat{v} \in \hat{Y} \equiv \mathbb{R}^{\hat{n}}$, we can think of $B : \mathbb{R}^{\hat{n}} \mapsto \mathbb{R}^m$ as a rectangular matrix where m, is the number of constraints.

We can write,

$$\begin{aligned} J(u) = \min_{v \in Y} \ J(v) = & \min_{\substack{\hat{v} \in \hat{Y} \\ B\hat{v} = 0}} \ J(\hat{v}) = \min_{\hat{v} \in \hat{Y}} \max_{\lambda \in \Lambda} \ J(\hat{v}) + \lambda^T B \hat{v} \\ & \geq \max_{\lambda \in \Lambda} \min_{\hat{v} \in \hat{Y}} \ \mathcal{L}(\hat{v}, \lambda) \geq \min_{\hat{v} \in \hat{Y}} \ \mathcal{L}(\hat{v}, \bar{\lambda}), \quad \forall \bar{\lambda} \in \Lambda \end{aligned}$$

where the Lagrangian is given (implicitly) by $\mathcal{L}(\hat{v}, \lambda) = J(\hat{v}) + \lambda^T B\hat{v}$.

Now, we can easily obtain an upper bound, s^+, for s by setting $s^+ \equiv -2\min_{\hat{v} \in \hat{Y}} \ \mathcal{L}(\hat{v}, \bar{\lambda})$ since

$$s \equiv -2J(u) \leq -2 \min_{\hat{v} \in \hat{Y}} \ \mathcal{L}(\hat{v}, \bar{\lambda}) \equiv s^+$$

Analogously, we can define for the error formulation,

$$\mathcal{L}^{err}(\hat{v}, \lambda) = J^{err}(\hat{v}) + \lambda^T B\hat{v},$$

and note that

$$J^{err}(e) \geq \min_{\hat{v} \in \hat{Y}} \ \mathcal{L}^{err}(\hat{v}, \bar{\lambda}), \quad \forall \bar{\lambda} \in \Lambda.$$

Finally, for any $\bar{\lambda}$),

$$\begin{aligned} s &= \tilde{s} + \tilde{u}^T R - 2J^{err}(e) \\ &\leq \tilde{s} + \tilde{u}^T R - 2 \min_{\hat{v} \in \hat{Y}} \mathcal{L}^{err}(\hat{v}, \bar{\lambda}) \equiv s^+ . \end{aligned}$$

As an example of a space and its relaxation, we consider $Y \equiv \mathbb{R}^n$ to the space of grid functions over a one dimensional grid with n nodes; $\hat{Y} \equiv \mathbb{R}^{2n-2}$ is the space of grid functions over the same grid, but now the functions are allowed to be multi-valued at the interior grid nodes as shown in figure 5. Clearly, $Y \subset \hat{Y}$.

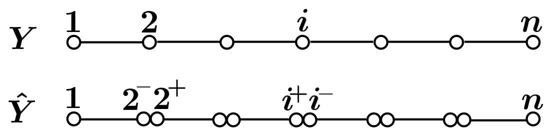


Fig. 5. A space Y consisting of all the grid functions over a one dimensional grid, of size n, and an enlarged space $\hat{Y}$, where for each grid node i, functions are allowed to take different values at i^+ and i^-.

In addition, one could define constraints that would to identify all the members of $\hat{Y}$ that belong to Y. In this case the constraints would be simply $\hat{v}_{i^+} - \hat{v}_{i^-} = 0$, for $i = 2, \ldots n-1$.

Approach II : Use a modified form A_{mod}

Let A_{mod} (SPD) be such that:

$$v^T A_{mod}\, v \leq v^T A v, \quad \forall v \in Y.$$

Then we can introduce a modified functional,

$$J_{mod}(v) = \frac{1}{2} v^T A_{mod}\, v - v^T F \leq J(v), \quad \forall v \in Y,$$

and,

$$\min_{\hat{v} \in \hat{Y}} J_{mod}(\hat{v}) = J_{mod}(u_{mod}) = -\frac{1}{2} u_{mod}^T A_{mod}\, u_{mod} .$$

Here, u_{mod}, satisfies the stationarity condition $A_{mod}\, u_{mod} = F$. In addition, since

$$J_{mod}(u_{mod}) \leq J_{mod}(u) \leq J(u),$$

we have that

$$s \equiv -2J(u) \leq -2J_{mod}(u_{mod}) \equiv s^+ .$$

A graphical "proof" is given in figure 6.

This approach is useful in the context of the reduced basis method [20]. As one may expect, good bounds rely on a good choice for A_{mod}, which is not

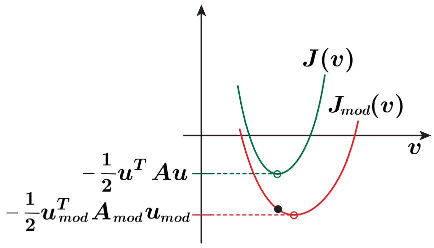


Fig. 6. One dimensional illustration of the modified functional $J_{mod}(v)$.

generally obvious. A much more effective approach, which is less sensitive to the choice of A_{mod}, is obtained if we work with the error rather than u. Here, we define

$$J^{err}_{mod}(v) = \frac{1}{2} v^T A_{mod}\; v - v^T R.$$

Then

$$J^{err}_{mod}(e_{mod}) \leq J^{err}(e),$$

where e_{mod} satisfies the stationarity condition $A_{mod}\, e_{mod} = R$ Finally, we have

$$s \equiv \tilde{s} + \tilde{u}^T R - 2J^{err}(e) \leq \tilde{s} + \tilde{u}^T R - 2J^{err}_{mod}(e_{mod}) \equiv s^+.$$

Approach III : Additive Decomposition
Here, we assume that A and F can be decomposed as

$$A = \sum_{k=1}^{K} A_k, \qquad F = \sum_{k=1}^{K} F_k,$$

where $A_k : Y \mapsto Y'$ and is symmetric positive semi-definite for all k. Using this decomposition we can define $J_k(v) = \frac{1}{2}(v^T A_k v - v^T F_k)$ and write

$$\begin{aligned} J(v) &= \frac{1}{2} v^T A v - v^T F = \frac{1}{2} v^T (\sum_{k=1}^{K} A_k) v - v^T (\sum_{k=1}^{K} F_k) \\ &= \frac{1}{2} \sum_{k=1}^{K} (v^T A_k v - v^T F_k) = \sum_{k=1}^{K} J_k(v). \end{aligned}$$

Now,

$$J(u) = \min_{v \in Y} J(v) = \min_{v \in Y} \sum_{k=1}^{K} J_k(v) \geq \sum_{k=1}^{K} \min_{v \in Y} J_k(v),$$

and thus

$$s = -2J(u) \leq -2 \sum_{k=1}^{K} \min_{v \in Y} J_k(v) \equiv s^+.$$

Similarly,

$$s \leq \tilde{s} + \tilde{u}^T R - 2 \sum_{k=1}^{K} \min_{v \in Y} J_k^{err}(v) \equiv s^+.$$

4.2 More general outputs

The bounds developed in the previous section were applicable to symmetric positive-definite operators and to the particular case of the so called "energy" (or "compliance") output. We consider now the more general case where A may be non-symmetric, and the output is any (bounded) linear functional. However, we will still require that its symmetric part of A, A^S, be positive definite.

We consider the problem

$$A\,u = F,$$

and attempt to develop bounds for the scalar output

$$s = u^T L^O.$$

here $A : Y \mapsto Y'$, $A^S = \frac{1}{2}(A + A^T)$ is symmetric positive definite, $u \in Y$, and $F, L^O \in Y'$.

We introduce the following functionals

$$\mathcal{E}(v) \equiv v^T A^S v - v^T F,$$

and

$$\mathcal{E}^{err}(v) \equiv v^T A^S v - v^T R,$$

and re-write our output s as

$$\pm s = \pm u^T L^O + \kappa \mathcal{E}(u), \quad \forall \kappa \in \mathbb{R}^+$$

The parameter κ is any positive number and, at this stage, it is kept fixed. We shall see below how we can optimize our computed bounds with respect to κ, thus obtaining sharper bounds.

We now write,

$$\pm s = \min_{\substack{v \in Y \\ Av = F}} (\pm v^T L^O + \kappa \mathcal{E}(v)) = \min_{v \in Y} \max_{\lambda \in Y} \mathcal{L}^{\pm}(v, \lambda).$$

It is clear that the first minimization is a trivial statement, since there is only one member of Y, namely u, which satisfies the constraint. Nevertheless, we can view the this problem as a minimization problem with equality constraints, similar to that encountered in section 3.2, and define a corresponding augmented Lagrangian

$$\mathcal{L}^{\pm}(v, \lambda) = \pm v^T L^O + \kappa \mathcal{E}(v) + \lambda^T (Av - F).$$

Lower (Upper) Bounds Using the duality result (4), we have

$$\pm s = \min_{v \in Y} \max_{\lambda \in Y} \mathcal{L}^{\pm}(v, \lambda) \geq \max_{\lambda \in Y} \min_{v \in Y} \mathcal{L}^{\pm}(v, \lambda).$$

Clearly, for any $\bar{\lambda} \in Y$,

$$s \geq \min_{v \in Y} \mathcal{L}^{+}(v, \bar{\lambda}) \equiv s^{-},$$
$$-s \geq \min_{v \in Y} \mathcal{L}^{-}(v, \bar{\lambda}) \equiv -s^{+},$$

which implies that $s^{-} \leq s \leq s^{+}$

Similarly, if we are given an approximation $\tilde{u} \approx u$, $\tilde{u} \in \tilde{Y} \subset Y$,

$$\pm s = \pm \tilde{s} + \min_{v \in Y} \max_{\lambda \in Y} \mathcal{L}^{\pm err}(v, \lambda).$$

The corresponding Lagrangian will be

$$\mathcal{L}^{\pm err}(v, \lambda) = \pm v^T L^O + \kappa \mathcal{E}^{err}(v) + \lambda^T (A\, v - R),$$

and, for any $\bar{\lambda} \in Y$,

$$s \geq \tilde{s} + \min_{v \in Y} \mathcal{L}^{+err}(v, \bar{\lambda}) \equiv s^{-},$$
$$-s \geq \tilde{s} + \min_{v \in Y} \mathcal{L}^{-err}(v, \bar{\lambda}) \equiv -s^{+}.$$

We note that in order to compute bounds, in principle, we need to solve the following minimization problems,

$$\min_{v \in Y} \mathcal{L}^{\pm}(v, \bar{\lambda}), \tag{12}$$

or,

$$\min_{v \in Y} \mathcal{L}^{\pm err}(v, \bar{\lambda}). \tag{13}$$

However, it is clear that if instead we compute lower bounds for the above minimums, we will also obtain lower bounds to $\pm s$. The computation of lower bounds for (12) and (13) is an unconstrained minimization problem to which we can apply either of the three approaches discussed in section 4.1.

5 Symmetric Coercive Partial Differential Equations : Bounds for the Energy Norm

The abstract bound results presented in the previous section will now be used to yield bounds for the energy norm of the solution of a symmetric coercive partial differential equation. As we shall see, this will be the basis for computing bounds for more general equations and outputs. We will use Laplace's equation to illustrate the presentation,

$$-\nabla^2 u = f^I, \quad \text{in } \Omega, \quad \begin{aligned} u|_{\Gamma_D} &= 0, \\ u|_{\Gamma_N} &= h^I. \end{aligned} \tag{14}$$

The output of interest is the energy of the solution

$$s = \int_\Omega |\nabla u|^2 \, d\Omega.$$

It is well known that the solution, u, to the above problem (14), can be characterized as the function that minimizes the following problem

$$J(u) = \min_{v \in X} J(v), \tag{15}$$

where,

$$J(v) \equiv \frac{1}{2} \int_\Omega |\nabla v|^2 \, d\Omega \; - \int_\Omega f^I v \, d\Omega \; - \int_{\Gamma_N} h^I v \, ds.$$

The field variable u is defined over a domain $\Omega \subset \mathbb{R}^2$ with boundary $\Gamma = \Gamma_D \bigcup \Gamma_N$; u must reside in $X = \{v \in H^1(\Omega) \mid v = 0 \text{ on } \Gamma_D\}$. Here $H^1(\Omega)$ is the usual Sobolev space; Γ_D is the portion of the domain boundary on which the essential (Dirichlet) boundary data, is imposed; and Γ_N is the portion of Γ on which natural (normal derivative) boundary conditions are imposed.

The minimizer in (15), satisfies the stationarity conditions which can be expressed in the following variational statement: find $u \in X$ such that

$$a(u, v) = l(v), \qquad \forall v \in X.$$

Here, X is the space of variations which consists of the functions in $H^1(\Omega)$ satisfying $v|_{\Gamma_D} = 0$, and

$$a(w, v) \equiv \int_\Omega \nabla w \cdot \nabla v \, d\Omega, \qquad l(v) \equiv \int_\Omega f^I v \, d\Omega \; + \int_{\Gamma_N} h^I v \, ds.$$

Note that, the form $a(\cdot, \cdot)$ is symmetric and positive definite. That is $a(v, v) > 0, \quad \forall v \in X$ (v non zero). Finally,

$$J(u) = -\frac{1}{2} \int_\Omega |\nabla u|^2 \, d\Omega$$

and,

$$s = -2\, J(u).$$

5.1 Meshes and Spaces

Two triangulations of the computational domain Ω are considered: the working or design H-mesh, $\mathcal{T}_H$, consisting of K_H elements T_H; and the "truth" h-mesh, $\mathcal{T}_h$, consisting of K_h elements T_h. We require that $\mathcal{T}_h$ is a refinement of $\mathcal{T}_H$; the geometric requirements on $\mathcal{T}_H$ are discussed in the context of our refinement strategy.

To each of these meshes we associate regular piecewise–linear continuous finite element subspaces,

$$X_H = \{v \in X \mid v|_{T_H} \in \mathbf{P}_1(T_H),\ \forall T_H \in \mathcal{T}_H\},$$
$$X_h = \{v \in X \mid v|_{T_h} \in \mathbf{P}_1(T_h),\ \forall T_h \in \mathcal{T}_h\},$$

where $\mathbf{P}_1(T)$ denotes the space of linear polynomials over T. Note that, by construction, $X_H \subset X_h \subset X$.

We shall assume that, for all practical purposes, the triangulation $\mathcal{T}_h$ is conservatively fine so that any difference between s and the output computed from the h-mesh solution, s_h, is sufficiently small so that $s \approx s_h$. Then, our objective will be to compute bounds for s_h using only H-mesh global solves.

Here, we will only consider "truth" meshes that are obtained by h-refining a given coarse triangulation $\mathcal{T}_H$. However, the algorithms described in these notes can be extended to "truth" meshes obtained by p-refinement of the coarse grid triangulation with minimal changes.

The algorithms to be presented require that our spaces and forms be expressed as sums of contributions over the H-elements T_H. Towards this end, we introduce the subdomain local spaces $Z_H(T_H)$ and $Z_h(T_H)$,

$$Z_H(T_H) = \mathbf{P}_1(T_H), \quad \forall T_H \in \mathcal{T}_H,$$
$$Z_h(T_H) = \{v|_{T_h} \in \mathbf{P}_1(T_h),\ \forall T_h \in \mathcal{R}_{T_H}\} \cap H^1(T_H), \quad \forall T_H \in \mathcal{T}_H,$$

where $\mathcal{R}_{T_H}$, denotes the set of h-mesh elements contained in T_H. The global representation of $Z_H(T_H)$ and $Z_h(T_H)$ are the "broken" spaces $\hat{X}_H$ and $\hat{X}_h$,

$$\hat{X}_H = \{v \in L_2(\Omega) \mid v|_{T_H} \in Z_H(T_H),\ \forall T_H \in \mathcal{T}_H\},$$
$$\hat{X}_h = \{v \in L_2(\Omega) \mid v|_{T_H} \in Z_h(T_H),\ \forall T_H \in \mathcal{T}_H\},$$

where $L_2(\Omega)$ is the space of square-integrable functions over Ω. The different meshes and associated finite element spaces are illustrated in figure 7.

Finally, the bilinear — $a(\cdot,\cdot)$ — and linear — $l(\cdot)$, $l^O(\cdot)$ — forms are extended to accept discontinuous functions in the "broken" spaces by redefining these forms as a sum of H-element contributions. For instance, $a(\cdot,\cdot)$ is now written as

$$a(v,w) = \sum_{T_H \in \mathcal{T}_H} a_{T_H}(v|_{T_H}, w|_{T_H}), \quad \forall v, w \in \hat{X}_h;$$

similar expressions are obtained for l, and l^O.

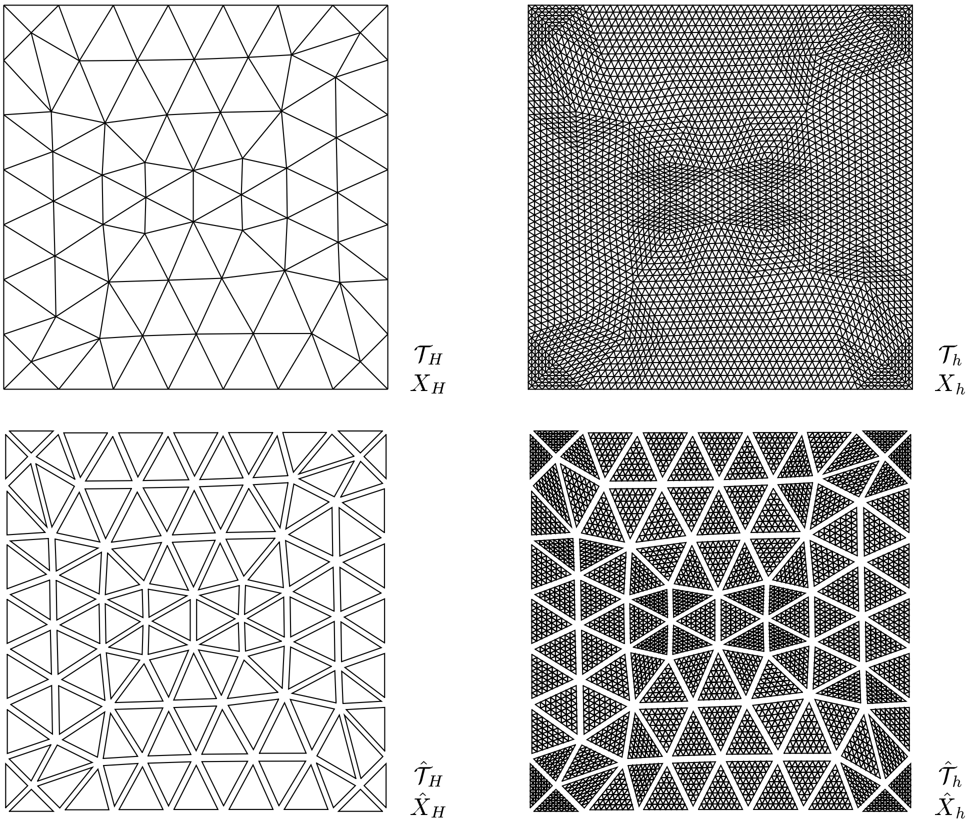


Fig. 7. Illustration of the different meshes and finite element spaces used in the bounds formulation.

5.2 Continuity constraint

Let $\mathcal{E}(\mathcal{T}_H)$ (respectively, $\mathcal{E}(\mathcal{T}_h)$) denote the set of open edges in the triangulation $\mathcal{T}_H$ (respectively, $\mathcal{T}_h$). We can then introduce the spaces of functions over the element edges $\gamma_H \in \mathcal{E}(\mathcal{T}_H)$.

$$\begin{aligned} Q_H &= \{q|_{\gamma_H} \in \mathbf{P}_1(\gamma_H),\ \forall \gamma_H \in \mathcal{E}(\mathcal{T}_H) \mid q|_{\Gamma_N} = 0\} \\ Q_h &= \{q|_{\gamma_h} \in \mathbf{P}_1(\gamma_h),\ \forall \gamma_h \in \mathcal{E}(\mathcal{T}_h) \cap \mathcal{E}(\mathcal{T}_H) \mid q|_{\Gamma_N} = 0\} \end{aligned}$$

It follows that $Q_H \subset Q_h \subset H^{-1/2}(\mathcal{E}(\mathcal{T}_H))$; the functions in these spaces can, of course, be discontinuous [7]. Now, we introduce the "jump" bilinear form

$b : \hat{X}_h \times Q_h \mapsto \mathbb{R}$

$$b(v,q) = \sum_{\gamma_H \in \mathcal{E}(\mathcal{T}_H)} \int_{\gamma_H} [v]_{\gamma_H}\ q|_{\gamma_H}\ ds,$$

where $[v]_{\gamma_H}$ is the *jump* in v across γ_H when γ_H is an interior edge, and the *trace* of v on γ_H when γ_H is on the boundary Γ. The form $b(\cdot,\cdot)$ can be used to enforce continuity on functions in $\hat{X}_H$ and $\hat{X}_h$; in particular,

$$X_H \equiv \{\hat{v} \in \hat{X}_H \mid b(\hat{v},q) = 0,\ \forall q \in Q_H\},$$

$$X_h \equiv \{\hat{v} \in \hat{X}_h \mid b(\hat{v},q) = 0,\ \forall q \in Q_h\}.$$

Note that $b(\cdot,\cdot)$ places no restriction on v on natural boundaries.

5.3 Lower Bound for s_h

Lower bounds for s are trivially computed. Since

$$J(u_h) = \min_{v \in X_h} J(v),$$

and $X_H \subset X_h$, we have

$$J(u_H) = \min_{v \in X_H} J(v) \geq J(u_h).$$

Therefore,

$$s^- \equiv s_H = -2J(u_H) \leq\ -2J(u_h) = s_h$$

and thus, a lower bound is trivially obtained.

5.4 Upper Bound for s_h

It is clear that in order to obtain an upper bound for s_h, the minimization needs to be carried out over a space larger than X_h so that a lower bound for $J(u_h)$ is obtained. To this end, we consider the "broken" space $\hat{X}_h$. We start by formulating the following trivial minimization statement

$$J(u_h) = \min_{v \in X_h} J(v) = \min_{\left\{\begin{array}{l} \hat{v} \in \hat{X}_h \text{ subject to} \\ b(\hat{v},q) = 0, \forall q \in Q_h \end{array}\right.} J(\hat{v}).$$

The above constrained minimization problem suggests the formation of the linear-quadratic Lagrangian $\mathcal{L}_h : \hat{X}_h \times Q_h \to \mathbb{R}$

$$\mathcal{L}_h(\hat{v},q) = J(\hat{v}) + b(\hat{v},q),$$

which allows us to express $J(u_h)$ as

$$J(u_h) = \min_{\hat{v} \in \hat{X}_h} \max_{q \in Q_h} \mathcal{L}_h(\hat{v},q).$$

Relaxation From duality results presented in (4), it follows that

$$\begin{aligned} J(u_h) &= \min_{\hat{v}\in\hat{X}_h} \max_{q\in Q_h} \mathcal{L}_h(\hat{v},q) \\ &\geq \max_{q\in Q_h} \min_{\hat{v}\in\hat{X}_h} \mathcal{L}_h(\hat{v},q) \\ &\geq \min\nolimits_{\hat{v}\in\hat{X}_h} \mathcal{L}_h(\hat{v},\bar{q}) = J^-(\bar{q}), \quad \forall \bar{q}\in Q_h. \end{aligned}$$

Thus, the upper bound s^+ is given by

$$s^+ \equiv -2J^-(\bar{q}) \geq -2J(u_h) = s_h$$

for any $\bar{q} \in Q_h$. It is important to note that, once a suitable $\bar{q}$ has been determined, the minimization required to compute $J^-(\bar{q})$ is carried out over functions which are discontinuous, and therefore can be effected independently for each macro-element $\mathcal{T}_H$.

Hybrid Flux Calculation We know that, at the saddle point (u_h, p_h),

$$J(u_h) = \mathcal{L}(u_h,p_h) = \min_{\hat{v}\in\hat{X}_H} \max_{q\in Q_H} \mathcal{L}(\hat{v},q) = \mathcal{L}(u_H,p_H).$$

We refer to q as the hybrid flux, and for this particular problem it can be identified with an approximation to the normal derivative of u on the edges.

We consider here the problem of determining H-mesh approximations to the hybrid flux p_h, $\bar{q}$, that will yield sharp bounds. For this purpose, we look for the saddle-point of the Lagrangian in the subspaces $\hat{X}_H \subset \hat{X}_h$ and $Q_H \subset Q_h$. That is we solve,

$$J(u_H) = \min_{v\in X_H} J(v) = \min_{\hat{v}\in\hat{X}_H} \max_{q\in Q_H} \mathcal{L}(\hat{v},q) = \mathcal{L}(u_H,p_H), \tag{16}$$

and set,

$$\bar{q} \equiv p_H.$$

The stationarity conditions corresponding to (16), are

$$\begin{aligned} a(u_H,\hat{v}) - l(\hat{v}) + b(\hat{v},p_H) &= 0, \qquad \forall \hat{v}\in\hat{X}_H \\ b(u_H,q) &= 0, \qquad \forall q \in Q_H. \end{aligned}$$

The second equation forces $u_H \in X_H$. The first equation must be satisfied for all $v \in \hat{X}_H$, and thus, for all $v \in X_H \subset \hat{X}_H$. Therefore, u_H can be obtained by solving the original problem

$$a(u_H,v) = l(v), \quad \forall v \in X_H,$$

which yields u_H. Now we can solve for p_H,

$$b(\hat{v}, p_H) = l(\hat{v}) - a(u_H, \hat{v}), \quad \forall \hat{v} \in \hat{X}. \tag{17}$$

The above equations represent a solvable but indeterminate system. The computation of an acceptable compatible solution is known as the *equilibration problem*. We follow here the approach proposed in [9,4,1], which involves the solution of an indeterminate system at each vertex of the $\mathcal{T}_H$ grid, the size of which is given by the number of edges that meet at the vertex. Details of this procedure can be found in [2].

There exist other approaches to obtaining implicit error bounds which circumvent the equilibration problem, typically at the expense of either some additional computational cost, or some limitations in applicability. We refer the reader to the partition-of-unity flux-free approach presented in [11], and also the method described in [4].

Local Problems Once suitable hybrid fluxes have been computed, we proceed to the solution of the local minimization problems. The minimizers, $\hat{u}$, will satisfy the following stationarity condition: find $\hat{u}_h \in \hat{X}_h$ such that

$$a(\hat{u}_h, \hat{v}) = l(\hat{v}) + b(\hat{v}, p_H), \qquad \forall \hat{v} \in \hat{X}_h.$$

This involves the solution of K_H independent *local* Neumman sub-problems, which are *semi-definite*, but solvable provided $\hat{X}_H$ and $\hat{X}_h$ share the same singular modes. In this case, this is guaranteed by the equilibration procedure.

Finally,

$$s^+ \equiv -2J^-(p_H) = a(\hat{u}_h, \hat{u}_h) \geq a(u, u) \equiv s_h \tag{18}$$

since

$$\min_{\hat{v} \in \hat{X}_h} \mathcal{L}_h(\hat{v}, p_H) = \mathcal{L}_h(\hat{u}_h, p_H) = \frac{1}{2} a(\hat{u}_h, \hat{u}_h) - l(\hat{u}_h) + b(\hat{u}_h, p_H) = -\frac{1}{2} a(\hat{u}_h, \hat{u}_h).$$

5.5 Energy Norm of the Error

For a given Galerkin solution $u_H \in X_H$, the above approach allows us to compute an upper bound of the error, s^{err}, in the energy norm by simply taking $s^{err+} = s^+ - s_H$, where $s_H = -2J(u_H)$. An alternative is to compute directly an upper bound for the energy norm of the error. This procedure is outlined below using an error formulation. As it turns out, the two approaches yield the same result for this particular problem.

Let $u_H \in X_H$ be such that

$$a(u_H, v) = l(v), \qquad \forall v \in X_H.$$

The error, e_h, is then, defined as $e_h = u_h - u_H \in X_h$, and the output of interest is $s_h^{err} = a(e_h, e_h)$. The error equation is given by

$$a(e_h, v) = R(v), \qquad \forall v \in X_h, \tag{19}$$

where the residual $R : X_h \mapsto \mathbb{R}$, is $R(v) = l(v) - a(u_H, v)$.

An alternative minimization formulation of (19), is obtained by introducing $J^{err}(v) = \frac{1}{2}a(v,v) - R(v)$,

$$J^{err}(e_h) = \min_{v \in X_h} J^{err}(v).$$

At the minimum,

$$J^{err}(e_h) = -\frac{1}{2}a(e_h, e_h)$$

and

$$s_h^{err} = -2J^{err}(e_h).$$

In order to compute an upper bound for s_h^{err}, we need a lower bound for $J^{err}(v)$ and this can be done following an identical path to that described previously for $J(u)$. Below we simply highlight the main steps.

Upper Bound for s^{err} Consider,

$$J^{err}(e_h) = \min_{v \in X_h} J^{err}(v) = \min_{\hat{v} \in \hat{X}_h} \max_{q \in Q_h} \mathcal{L}^{err}(\hat{v}, q),$$

where,

$$\mathcal{L}^{err}(\hat{v}, q) = J^{err}(\hat{v}) + b(\hat{v}, q).$$

Using duality (4) plus a coarse grid approximation (17) to the hybrid fluxes, we write

$$J^{err}(e_h) \geq \min_{\hat{v} \in \hat{X}_h} \mathcal{L}^{err}(\hat{v}, p_H).$$

The local problems: find $\hat{e}_h \in X_h$ such that

$$a(\hat{e}_h, \hat{v}) = R(\hat{v}) + b(\hat{v}, p_H), \quad \forall \hat{v} \in X_h,$$

are then solved to yield

$$J^{err}(e_h) \geq J^{err}(\hat{e}_h) = -\frac{1}{2}a(\hat{e}_h, \hat{e}_h),$$

and finally,

$$s^{err+} \equiv a(\hat{e}_h, \hat{e}_h) \geq a(e, e) \equiv s_h^{err}.$$

We point out that the upper bound s^{err+}, can be expressed as the sum of H-mesh elemental contributions

$$a(\hat{e}_h, \hat{e}_h) = \sum_{T_h \in \mathcal{T}_H} a_{T_H}(\hat{e}_h|_{T_H}, \hat{e}_h|_{T_H}).$$

In addition, since each of the elemental contributions is positive, we can identify the contribution of each element to the final bound thus providing a procedure for adaptivity. This concept will be further exploited in the next section.

6 Symmetric Coercive Partial Differential Equations : Bounds for Functional Outputs

The procedure developed in the previous section can now be extended to account for linear functional outputs. We will use an error formulation for the purposes of describing the approach but the formulation in terms of the total unknown follows analogous steps.

Let us consider a linear second order coercive (not necessarily symmetric) problem written in variational form as : find $u_H \in X^D(\Omega)$ such that

$$a(u, v) = l(v), \quad \forall v \in X.$$

The output of interest $s \in \mathbb{R}$, will be an affine functional of the solution which we write as,

$$s = l^O(u).$$

Here, $X = \{v \in H^1(\Omega) \mid v = 0 \quad \text{on} \quad \Gamma_D\}$, $X^D = \{v + u_D \mid v \in X\}$, $u_D|_{\Gamma_D} = g_D$, $\Omega \subset \mathbb{R}^2$ and $\Gamma = \Gamma_D \bigcup \Gamma_N$; Γ_D is the portion of the domain boundary on which the essential (Dirichlet) boundary data, g_D, is imposed; u_D is any function in $H^1(\Omega)$ satisfying the essential boundary conditions, that is, $u_D|_{\Gamma_D} = g_D$; and Γ_N is the portion of Γ on which natural (normal derivative) boundary conditions are imposed. Also, $a : H^1(\Omega) \times H^1(\Omega) \mapsto \mathbb{R}$ is a continuous bi-linear form, and $l : H^1(\Omega) \mapsto \mathbb{R}$ (respectively, $l^O : H^1(\Omega) \mapsto \mathbb{R}$) is a continuous linear (respectively, affine) functional.

The symmetric (coercive) part of $a(\cdot, \cdot)$, $a^S : X \times X \mapsto \mathbb{R}$, is given by

$$a^S(v, w) = \frac{1}{2}(a(v, w) + a(w, v)), \quad \forall v, w \in X,$$

and norm induced by it is represented as $|||\cdot|||$,

$$|||v|||^2 = a^S(v, v), \quad \forall v \in X.$$

6.1 $\mathcal{T}_H$ and $\mathcal{T}_h$ Problems

We consider the Galerkin finite element solution on $\mathcal{T}_H$, $u_H \in X_H^D$,

$$a(u_H, v) = l(v), \quad \forall v \in X_H \quad \text{and,} \quad s_H = l^O(u_H).$$

Analogously, on $\mathcal{T}_h$, $u_h \in X_h^D$, satisfies

$$a(u_h, v) = l(v), \quad \forall v \in X_h \quad \text{and,} \quad s_h = l^O(u_h).$$

In the smooth case, $|s - s_H| \sim \mathcal{O}(H^2)$, and $|s - s_h| \sim \mathcal{O}(h^2)$, where H and h are the diameters of $\mathcal{T}_H$ and $\mathcal{T}_h$, respectively. This is proved by standard Aubin-Nitsche duality methods.

It will be assumed that the calculation of u_H, and hence s_H, can be performed at moderate cost. However, the h-mesh will typically be much finer than the H-mesh, in order to ensure that $|s - s_h|$ is suitably small; the computation of u_h and s_h will thus, most likely, be impractical. Our objective is, therefore, to devise a procedure that will yield sharp upper and lower bounds for s_h in an affordable manner.

6.2 Lagrangian Formulation

We start by first deriving the so called "energy" equality which will be used to convexify out problem. Noting that $u_h - u_H \in X_h$, we have that

$$a(u_h, u_h - u_H) - l(u_h - u_H) = 0.$$

Using the bi-linearity of $a(\cdot,\cdot)$, we can write

$$a^S(u_h - u_H, u_h - u_H) - l(u_h - u_H) + a(u_H, u_h - u_H) = 0,$$

or,

$$a^S(e_h, e_h) - R(e_h) = 0.$$

Here $e_h = u_h - u_H \in X_h$ and $R(v) \equiv l(v) - a(v, u_H), \quad \forall v \in \hat{X}_h$. Note that $a^S(e_h, e_h)$ is simply $|||e_h|||^2$.

Consider now the following functional space $\mathcal{S} \subset \hat{X}_h$,

$$\mathcal{S} = \{v \in \hat{X}_h \mid a(v, w) = R(w),\ \forall w \in X_h;\ b(v, q) = 0,\ \forall q \in Q_h\}.$$

We observe that the second constraint enforces continuity and the homogeneous essential conditions, that is, $v \in X_h$, and the first constraint then forces $v = e_h$. Therefore, the space $\mathcal{S}$ consists of a single function, e_h. The following trivial minimization statement can then be written,

$$\pm s_h = \min_{v \in \mathcal{S}} \ \kappa(a^S(v, v) - R(v)) \ \pm\ l^O(u_H + v), \quad \forall \kappa \in \mathbb{R}^+,$$

where κ is a non-negative parameter that will be used later to optimize the computed bounds.

The above constrained minimization problem suggests the formation of the linear-quadratic Lagrangian $\mathcal{L} : \hat{X}_h \times X_h \times Q_h \mapsto \mathbb{R}$,

$$\begin{aligned}\mathcal{L}^{\pm}(v, \theta, q) = {}& \kappa\ (a^S(v, v) - R(v)) \pm l^O(u_H + v) \\ & + a(u_H + v, \theta) - l(\theta) + b(v, q)\end{aligned}$$

which allows us to express $\pm s_h$ as

$$\pm s_h = \min_{v \in \hat{X}_h} \ \max_{\theta \in X_h,\ q \in Q_h} \mathcal{L}^{\pm}(v, \theta, q).$$

We know from section 3.2 that, at the saddle–point $(e_h, \psi_h^{\pm}, p_h^{\pm})$,

$$\mathcal{L}^{\pm}(e_h, \psi_h^{\pm}, p_h^{\pm}) = \min_{v \in \hat{X}_h} \ [\max_{\theta \in X_h,\ q \in Q_h} \mathcal{L}^{\pm}(v, \theta, q)].$$

We shall refer to θ as the adjoint (dual) variable, and q as the hybrid flux.

Duality From linear-quadratic duality theory, [22] (see also section 4), it follows that, for any $\bar{\theta}_h^{\pm} \in X_h$ and any $\bar{q}_h^{\pm} \in Q_h$,

$$\begin{aligned} \pm s_h &= \max_{\theta \in X_h,\ q \in Q_h} \ \min_{v \in \hat{X}_h} \mathcal{L}^{\pm}(v, \theta, q) \\ &\geq \min_{v \in \hat{X}_h} \ \mathcal{L}^{\pm}(v, \bar{\theta}_h^{\pm}, \bar{q}_h^{\pm}), \end{aligned}$$

with equality for $(\bar{\theta}_h^{\pm}, \bar{p}_h^{\pm}) = (\psi_h^{\pm}, p_h^{\pm})$. Upper and lower bounds for the output s_h, s^+ and s^-, respectively, are now readily constructed as

$$s^- \equiv \min_{v \in \hat{X}_h} \mathcal{L}^+(v, \bar{\theta}_h^+, \bar{q}_h^+) \leq s_h \leq - \min_{v \in \hat{X}_h} \mathcal{L}^-(v, \bar{\theta}_h^-, \bar{q}_h^-) \equiv s^+. \tag{20}$$

We note that this expression holds for any functions $\bar{\theta}_h^{\pm} \in X_h$ and $\bar{q}_h^{\pm} \in Q_h$. In particular, we will choose these functions to be members of $X_H \subset X_h$ and $Q_H \subset Q_h$.

Adjoint and Hybrid Fluxes We consider here the problem of determining H-mesh approximations to the adjoint $\psi_h^{\pm}$, $\bar{\theta}_h^{\pm}$, and the hybrid flux $p_h^{\pm}$, $\bar{q}_h^{\pm}$, that will yield sharp bounds. Towards this end, we look for the saddle point of the Lagrangian in the subspaces $\hat{X}_H \subset \hat{X}_h$, $X_H \subset X_h$, and $Q_H \subset Q_h$. In particular we find $(e_H^{\pm}, \psi_H^{\pm}, p_H^{\pm}) \in \hat{X}_H \times X_H \times Q_H$ such that,

$$\mathcal{L}^{\pm}(e_H, \psi_H^{\pm}, p_H^{\pm}) = \min_{v \in \hat{X}_H} [\max_{\theta \in X_H,\ q \in Q_H} \mathcal{L}^{\pm}(v, \theta, q)],$$

and set $\bar{\theta}_h^{\pm} = \psi_H^{\pm}$ and $\bar{q}_h^{\pm} = p_H^{\pm}$. The stationarity conditions are obtained by requiring that the variations of $\mathcal{L}^{\pm}$ with respect to v, θ, and q vanish. Thus, we find $e_H^{\pm} \in \hat{X}_H$, $\psi_H^{\pm} \in X_H$, and $p_H^{\pm} \in Q_H$ such that

$$\begin{aligned} b(w, p_H^{\pm}) &= -\{\kappa\ (2a^S(e_H^{\pm}, w) - R(w)) \pm l^O(w) + a(w, \psi_H^{\pm})\}, \quad \forall w \in \hat{X}_H, \\ a(u_H + e_H^{\pm}, w) &= l(w), \qquad \forall w \in X_H, \\ b(e_H^{\pm}, q) &= 0, \qquad \forall q \in Q_H. \end{aligned}$$

First, the third equation forces $e_H^{\pm} \in X_H$, which combined with the second and the primal H-mesh problem implies that $e_H^{\pm} = 0$. Next, the first equation must be satisfied for all $w \in \hat{X}_H$, and thus, for all $w \in X_H \subset \hat{X}_H$. The adjoint, $\psi_H^{\pm} \in X_H$, can therefore be determined (precisely as in Aubin–Nitsche procedures [5]) as

$$a(w, \psi_H) = -l^O(w), \quad \forall w \in X_H,$$

with $\psi_H^{\pm} = \pm\psi_H$. Note that, in deriving this equation, we exploit, first, $R(w) = 0$, $\forall w \in X_H$, and second $b(w, p_H^{\pm}) = 0$, $\forall w \in X_H$. Finally, we can determine the hybrid fluxes $p_H^{\pm} \in Q_H$, such that

$$b(w, p_H^{\pm}) = \kappa R(w) \mp l^O(w) - a(w, \psi_H^{\pm}), \quad \forall w \in \hat{X}_H.$$

We set $p_H^{\pm} = \pm p_{0H} + \kappa p_{1H}$ and find $p_{0H} \in Q_H$ and $p_{1H} \in Q_H$ such that

$$\begin{aligned} b(w, p_{0H}) &= -l^O(w) - a(w, \psi_H), \quad \forall w \in \hat{X}_H, \\ b(w, p_{1H}) &= R(w) \ (= l(w) - a(u_H, w)), \quad \forall w \in \hat{X}_H. \end{aligned}$$

Note that, p_{0H} *equilibrates* the H-mesh adjoint problem, whereas p_{1H} *equilibrates* the original H-mesh primal solution. The above equations represent a solvable but indeterminate system which is solved using analogous procedures to those mentioned in section 5.4.

Local Problems We now turn to the solution of the minimization problems: find $\hat{e}_h^{\pm} \in \hat{X}_h$ such that

$$\mathcal{L}^{\pm}(\hat{e}_h^{\pm}, \psi_H^{\pm}, p_H^{\pm}) = \min_{\hat{v} \in \hat{X}_h} \mathcal{L}^{\pm}(v, \psi_H^{\pm}, p_H^{\pm}).$$

The corresponding stationarity condition is ,

$$2\kappa a^S(\hat{e}_h^{\pm}, w) = \kappa R(w) \mp l^O(w) - a(w, \psi_H^{\pm}) - b(w, p_H^{\pm}), \quad \forall w \in \hat{X}_h,$$

which we recognize as $2K_H$ independent T_H–local symmetric Neumann subdomain problems. Note that, although a^S is coercive over $X_h \times X_h$, it is typically only semi-definite over $\hat{X}_h \times \hat{X}_h$; however, thanks to the equilibration process, if the singular modes of $\hat{X}_h$ are shared by $\hat{X}_H$, the above system will be solvable, and our bounds finite. The equations for $\hat{e}_h^{+}$ and $\hat{e}_h^{-}$ can be rearranged into the following problems for $\hat{e}_{0h}$ and $\hat{e}_{1h}$,

$$\begin{aligned} 2a^S(\hat{e}_{0h}, w) &= -l^O(w) - a(w, \psi_H) - b(w, p_{0H}), \quad \forall w \in \hat{X}_h, \\ 2a^S(\hat{e}_{1h}, w) &= l(w) - a(u_H, w) - b(w, p_{1H}), \quad \forall w \in \hat{X}_h, \end{aligned}$$

where $p_H^{\pm} = \pm p_{0H} + \kappa p_{1H}$ and $\hat{e}_h^{\pm} = \pm\frac{1}{\kappa}\hat{e}_{0h} + \hat{e}_{1h}$.

Note that $\hat{e}_{0h}$ and $\hat{e}_{1h}$ are related to the error in the adjoint ψ_H, and the error in the solution u_H, respectively.

6.3 Bounds Calculation

To evaluate the bounds, it remains only to insert the computed values for $\hat{e}_h^{\pm}$, $\psi_H^{\pm}$, and $p_H^{\pm}$ into our Lagrangian.

$$\begin{aligned} s^{+}(\kappa) &\equiv -\mathcal{L}^{-}(\hat{e}_h^{-}, \psi_H^{-}, p_H^{-}) = l^O(u_H) + \kappa a^S(\hat{e}_h^{-}, \hat{e}_h^{-}) \\ s^{-}(\kappa) &\equiv \ \ \mathcal{L}^{+}(\hat{e}_h^{+}, \psi_H^{+}, p_H^{+}) = l^O(u_H) - \kappa a^S(\hat{e}_h^{+}, \hat{e}_h^{+}) \end{aligned}$$

where the arguments of s^+ and s^- indicate that the quality of the bounds depends, of course, on the working mesh $\mathcal{T}_H$, but also on the parameter κ. Note that our bounds for s_h are also bounds for s_H. Note also that, for the "compliance" case — $a(\cdot,\cdot)$ symmetric, $l^O = l$, homogeneous essential conditions, and $\kappa = 1$ — $\psi_H^+ = -u_H$ and $p_H^+ = 0$; it follows that $\hat{e}_h^+ = 0$, and thus $s^- = l^O(u_H)$.

We can also express our bounds in terms of $\hat{e}_{0h}$, $\hat{e}_{1h}$,

$$s^+(\kappa) = l^O(u_H) - 2a^S(\hat{e}_{0h}, \hat{e}_{1h}) + \frac{1}{\kappa}a^S(\hat{e}_{0h}, \hat{e}_{0h}) + \kappa a^S(\hat{e}_{1h}, \hat{e}_{1h})$$
$$s^-(\kappa) = l^O(u_H) - 2a^S(\hat{e}_{0h}, \hat{e}_{1h}) - \frac{1}{\kappa}a^S(\hat{e}_{0h}, \hat{e}_{0h}) - \kappa a^S(\hat{e}_{1h}, \hat{e}_{1h}),$$

which we recall is valid for any positive value of the parameter κ. The choice of κ will affect the (half) bound gap,

$$\Delta(\kappa) \equiv \frac{1}{2}(s^+ - s^-) = \frac{1}{\kappa}a^S(\hat{e}_{0h}, \hat{e}_{0h}) + \kappa a^S(\hat{e}_{1h}, \hat{e}_{1h}),$$

but will not affect the output predictor,

$$s_{pre} \equiv \frac{1}{2}(s^+ + s^-) = l^O(u_H) - 2a^S(\hat{e}_{0h}, \hat{e}_{1h}).$$

Note that $|s_h - s_{pre}| \leq \Delta$.

Optimum κ Since $\hat{e}_{0h}$ and $\hat{e}_{1h}$ do not depend on the choice of κ, we can readily find that κ which minimizes the bound gap, and hence renders the lower and upper bounds as sharp as possible. This procedure gives

$$\kappa^* = \sqrt{\frac{a^S(\hat{e}_{0h}, \hat{e}_{0h})}{a^S(\hat{e}_{1h}, \hat{e}_{1h})}}.$$

The optimized bounds are

$$s^+(\kappa^*) = l^O(u_H) - 2a^S(\hat{e}_{0h}, \hat{e}_{1h}) + 2\ |||\hat{e}_{0h}|||\ |||\hat{e}_{1h}|||, \tag{21}$$
$$s^-(\kappa^*) = l^O(u_H) - 2a^S(\hat{e}_{0h}, \hat{e}_{1h}) - 2\ |||\hat{e}_{0h}|||\ |||\hat{e}_{1h}|||, \tag{22}$$

and the optimized (half) bound gap is

$$\Delta(\kappa^*) \equiv \frac{1}{2}(s^+(\kappa^*) - s^-(\kappa^*)) = 2\ |||\hat{e}_{0h}|||\ |||\hat{e}_{1h}|||.$$

Note that, in the energy output case of section 5, $\kappa^* = 1$. We thus see that our procedure is a *bound formulation* of the classical Aubin–Nitsche result; an implicit counterpart to existing explicit schemes for linear functional outputs [3,5]. In the smooth case we expect that $|||\hat{e}_{0h}|||$ and $|||\hat{e}_{1h}|||$ will each be

$O(H)$, and thus $\Delta(\mathcal{T}_H) \sim O(H^2)$ will vanish at the same rate as $|s - s_H|$ as $H \to 0$.

For the special situation in which $\kappa = 1$ and $l^O(u_H + v) \equiv R(v)$, we see that our upper bound corresponds to an upper bound for the error in the energy norm, and thus in this case our method is a (rather trivial) extension of the method of [1] to nonsymmetric operators. (The lower bound is zero, consistent with our compliance result.) Note that, for both the symmetric and nonsymmetric problems, $\psi_H^- = 0$, however only for the symmetric problem is $\psi_h^- = 0$. It follows that only for the symmetric problem is e_h the minimizer of $\mathcal{L}^{\pm}(v, \psi_H^- = 0, \cdot\,)$ ($= 2J(v)$ of [1] *in this particular case*) over all $v \in X_h$. This highlights the distinction between our *energy functional*, $a^S(v,v) - R(v) \pm l^O(u_H + v)$, and a variational principle: the former is only required to *evaluate* to $l^O(u_h)$ for $v = e_h$, while the latter must attain its *minimum* $l^O(u_h)$ at $v = e_h$. We are able to consider more general objective functions since the task of enforcing $v = e_h$ has been moved to a constraint.

Finally, we briefly summarize computational complexity. On $\mathcal{T}_H$, we require two global calculations, one for u_H and one for ψ_H, and two (local) hybrid flux equilibrations. On $\mathcal{T}_h$ we need only invert $2K_H$ independent T_H-local *symmetric* Neumann operators. It follows from the superlinearity of most general solution algorithms that the effort to compute our bounds will be much less than the effort to directly compute u_h and s_h. Note that for each additional output of interest, we require only *one* additional global calculation on $\mathcal{T}_H$, and *one* additional equilibration, and K_H additional T_H-local Neumann problems.

7 Adaptivity

The procedure presented in the previous section to calculate bounds can be implemented on any pair of meshes, $\mathcal{T}_H$ and $\mathcal{T}_h$, which satisfy the requirement that $\mathcal{T}_h$ is a refinement of $\mathcal{T}_H$. Since the cost of computing the bounds is essentially a function of the number of elements K_H in $\mathcal{T}_H$, it is desirable to construct *optimized* triangulations that maximize the bound accuracy (minimize the bound gap) for a given number of degrees-of-freedom. In this section we present an algorithm for generating such optimized grids in an adaptive iterative manner.

We first note that the bound gap can be expressed as a sum of elemental contributions,

$$\Delta = \sum_{T_H \in \mathcal{T}_H} \Delta_{T_H},$$

$$\Delta_{T_H} = \frac{1}{\kappa^*} a^S_{T_H}(\hat{e}_{0h}, \hat{e}_{0h}) + \kappa^* a^S_{T_H}(\hat{e}_{1h}, \hat{e}_{1h}) \geq 0.$$

Note that $\Delta_{T_H}(\mathcal{T}_H)$ is non-negative and can thus be directly interpreted as the *contribution* to the bound gap from element T_H.

We can now describe our iterative adaptive strategy. Starting from an initial grid $\mathcal{T}_H^0$ with bound gap Δ^0, we generate a sequence of triangulations $\{\mathcal{T}_H^k,\ k = 1, 2, \ldots\}$, with corresponding bound gaps $\{\Delta^k,\ k = 1, 2, \ldots\}$, such that each triangulation $\mathcal{T}_H^k$ is a refinement of the preceding triangulation $\mathcal{T}_H^{k-1}$. The approach does not guarantee that $\Delta^k \leq \Delta^{k-1}$ for any particular k, but it does ensure that, for a sufficiently large k, $\Delta^k \leq \Delta^{targ}$, where $\Delta^{targ} > 0$ is a specified positive gap target.

In order to identify the elements in $\mathcal{T}_H^{k-1}$ that need to be refined, we first calculate the largest elemental contribution Δ_{max}^{k-1} to the bound gap Δ^{k-1},

$$\Delta_{max}^{k-1} = \max_{T_H \in \mathcal{T}_H^{k-1}} \Delta_{T_H}^{k-1},$$

and refine all those elements $T_H \in \mathcal{T}_H^{k-1}$ for which

$$\Delta_{T_H}^{k-1} \geq \alpha \Delta_{max}^{k-1}, \qquad 0 < \alpha < 1.$$

Here, α is a parameter which controls the fraction of elements to be refined at each iteration. At present, this parameter is specified *a priori*, and is independent of k. More sophisticated strategies for choosing α can, of course, be devised; for instance, we can determine α^k such that, based on *a priori* scalings, a given target gap $\Delta^{targ,k}$ should be satisfied by the refined grid $\mathcal{T}_H^k$.

The refinement process is terminated when a selected accuracy goal is achieved. In the applications that follow, the accuracy measure is the relative (half) bound gap r^k,

$$r^k = \frac{\Delta^k}{s_{pre}^k}.$$

The iteration is thus halted when $r^k \leq r^{obj}$ where r^{obj} is the prescribed accuracy.

7.1 Grid Refinement Strategy

The triangulation $\mathcal{T}_H^k$ is generated from $\mathcal{T}_H^{k-1}$ by refinement of the selected triangles T_H such that $\Delta_{T_H}^{k-1} > \alpha \Delta_{max}^{k-1}$. Each triangle that satisfies this condition is subdivided into *four* similar triangles. This refinement strategy leads to the generation of *hanging nodes* whenever only one of the two triangles sharing an edge is refined. Since an element T_H with a hanging node — a node on ∂T_H which is *not* a vertex of T_H — requires special computational treatment, we permit only one hanging node per element. We thus, first, refine all those elements satisfying equation the above threshold equation, and then, in addition, recursively refine those elements that have more than one hanging node. At the conclusion of the process, only two types of elements (having either zero or one hanging node) are present. This approach has the benefit that fairly smooth grids are generated, and that only one type of *special* element needs to be developed.

The unknowns corresponding to the hanging nodes appearing as a result of the grid refinement process are constrained to be the equal to the average of the unknown at the two end nodes of the associated edge. Details about this procedure can be found in [18].

7.2 Mesh Adaptivity Results

Linear Elasticity Results are presented for the two-dimensional generalized planar stress linear elasticity equations. Although here the solution field u is the displacement vector, the theory presented for the scalar problem can be trivially extended to handle this case. The particular problem considered is that of one-fourth of a symmetrically loaded plate with two rectangular holes.

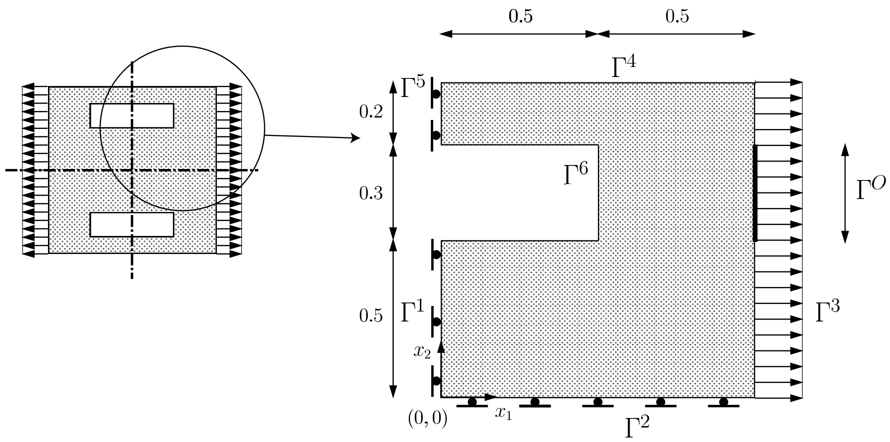


Fig. 8. Geometry for linear elasticity example: one–fourth of a symmetrically loaded plate with two square holes. A normal traction is applied on Γ^3; the outputs of interest are the normal force on Γ^5, and the average normal displacement over the segment $\Gamma^O \subset \Gamma^3$.

The domain boundaries are denoted by Γ^j, $j = 1, \ldots, 6$. A horizontal unit distributed force is applied along Γ^3, and homogeneous essential boundary conditions are enforced on the normal displacement on Γ^1, Γ^2, and Γ^5; stress-free condition apply on all other boundaries and components. The Young's modulus and Poisson ratio are taken to be 1 and 0.3, respectively.

Two outputs are considered: the normal force on the boundary segment Γ^5, denoted by s_1, and the average horizontal displacement over the segment Γ^O, denoted by s_2. These outputs can be expressed as $s_1 = l_1^O(u)$, $s_1 = l_1^O(u)$, where

$$l_1^O(v) = \int_\Omega \frac{\partial \chi_i}{\partial x_j} \sigma_{ij}(u) d\Omega, \quad l_2^O(v) = \frac{1}{|\Gamma^O|} \int_{\Gamma^O} v_1 ds.$$

We express the force output as a bounded linear functional [16], where summation over repeated indices is assumed $(i, j = 1, 2)$, $\sigma_{ij}(u)$ is the stress tensor, and χ_i is a vector function defined as

$$\chi_1 = \begin{cases} 1 - 2x_1 & : \quad (x_1, x_2) \in \Omega' \\ 0 & : \quad (x_1, x_2) \in \Omega/\Omega' \end{cases}, \qquad \chi_2 = 0.$$

Here $\Omega' = \{(x_1, x_2)|\ 0 \le x_1 \le 0.5, 0.8 \le x_2 \le 1.0\}$. Since there are no volume forces, or tractions applied at either Γ^4 or Γ^6, $l_1^O(v)$ is equivalent, for sufficiently smooth functions, to the integral along Γ^5 of the horizontal component of the traction vector.

The same initial triangulation $\mathcal{T}_H^0$ and "truth" mesh $\mathcal{T}_h$ have been considered for both outputs. The initial triangulation consists of 108 triangles. The "truth" mesh has been generated from this initial triangulation by four refinement interactions in which *all* triangles are subdivided; thus, the number of elements in the "truth" mesh is $27,648 = 108 \times 4^4$. In the adaptive process, elements were selected for refinement based on a threshold of $\alpha = 0.3$ for all the iterations. The accuracy goal, r^{obj}, for the relative (half) bound gap is 0.003 (or 0.3%).

For the force output s_1, 16 refinement iterations are required starting from a relative bound gap of 15.64%; only 12 refinement iterations are required for the displacement output s_2, for which the initial relative bound gap is 13.82%. Figures 9 and 10 show the relative bound gap as a function of the number of elements, at each adaptive iteration, for the force and displacement outputs, respectively. We also show the bound gap computed for a sequence of grids obtained by uniform refinement, the finest of which still has four times fewer elements than the "truth" mesh $\mathcal{T}_h$. For both outputs we observe that, the uniformly refined meshes require approximately four times as many elements as the adaptive meshes for comparable levels of accuracy.

We note that, say for the force output, $|(s_{1,pre}^{16} - s_{1,h})/s_{1,pre}^{16}| = 0.0001$, whereas $r_1^{16} = 0.003$, and thus the effectivity of the estimator in the classical sense is not as good as we may hope, even though the convergence *rate* of $\Delta(\mathcal{T}_H)$ is optimal. We emphasize, however, that our bounds provide *certainty* that $|(s_{1,pre}^{16} - s_{1,h})/s_{1,pre}^{16}| \le 0.003$ — and more importantly, that s_h lies between $s_1^{-,16} = -0.3230$ and $s_1^{+,16} = -0.3211$ — without calculating s_h on the uniformly highly refined (and hence expensive) "truth" mesh $\mathcal{T}_h$. Future work will address techniques by which to reduce the ratio $\Delta/|s_h - s_{pre}|$ as $H \to 0$; this, in turn, will permit to achieve r^{obj} on coarser meshes $\mathcal{T}_H$.

Finally, figures 11 and 12 show the initial, an intermediate, and the final adaptive triangulations obtained for the force and displacement outputs, respectively. Refinement around the singular re-entrant corners is apparent in both cases but, in addition, in the displacement output case, further refinement is observed near the points where l_2^O is "discontinuous".

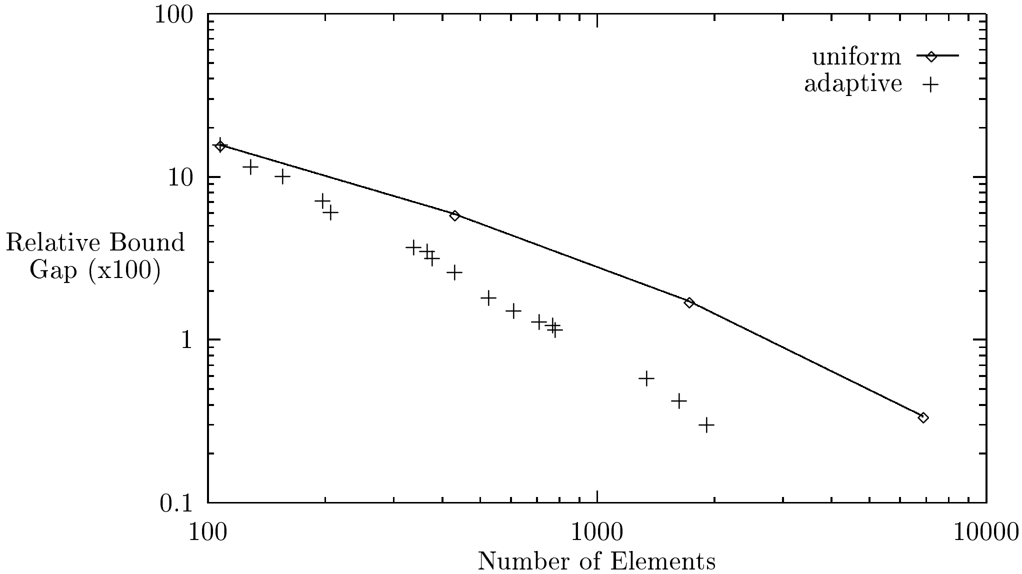


Fig. 9. Bound gap $\Delta(\mathcal{T}_H)$ as a function of the number of elements K_H, for the force output $s11$, for meshes generated using uniform and adaptive refinement.

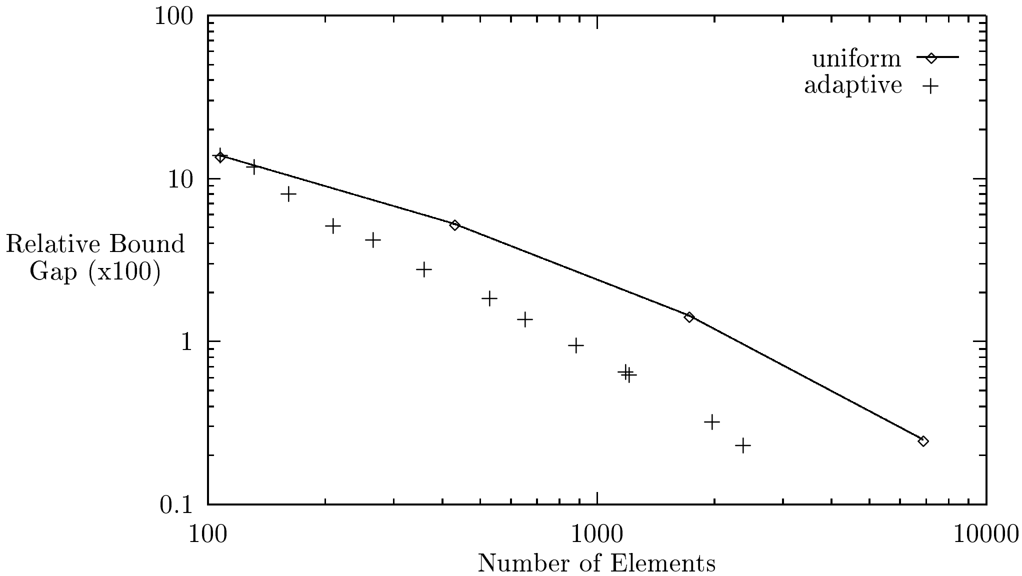


Fig. 10. Bound gap $\Delta(\mathcal{T}_H)$ as a function of the number of elements K_H, for the displacement output s_2, for meshes generated using uniform and adaptive refinement.

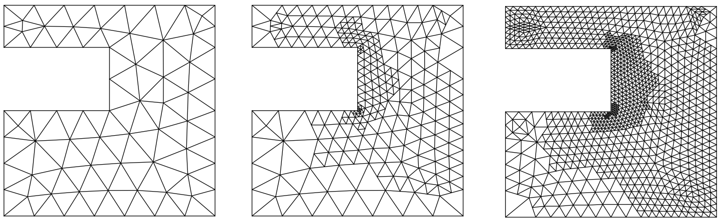


Fig. 11. Adaptively generated meshes for force output s_1. Figure shows $\mathcal{T}_H^0$, $\mathcal{T}_H^8$, and $\mathcal{T}_H^{16}$, with $K_H^0 = 108$, $K_H^8 = 432$ and $K_H^{16} = 1917$ elements; predicted outputs are $s_{1,pre}^0 = -0.3198$, $s_{1,pre}^8 = -0.3215$, and $s_{1,pre}^{16} = -0.3220$; and relative bound gaps are $r_1^8 = 15.64\%$, $r_1^8 = 2.59\%$ and $r_1^{16} = 0.30\%$.

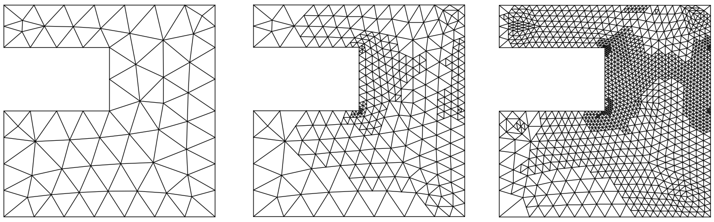


Fig. 12. Adaptively generated meshes for displacement output s^2. Figure shows $\mathcal{T}_H^0$, $\mathcal{T}_H^6$, and $\mathcal{T}_H^{12}$, with $K_H^0 = 108$, $K_H^6 = 531$ and $K_H^{12} = 2382$ elements; predicted outputs are $s_{2,pre}^0 = -0.4056$, $s_{2,pre}^6 = -0.4204$, and $s_{2,pre}^{12} = -0.4226$; and relative bound gaps are $r_2^0 = 13.82\%$, $r_2^6 = 1.83\%$ and $r_2^{12} = 0.23\%$.

8 Stabilized Methods

We now extend out bound procedure to stabilized formulations for convection dominated problems. The most important difference in this case is that due to the presence of the stabilization terms, the corresponding linear and bi-linear forms in the variational statement are a function of the mesh parameter $\delta = (H \text{ or } h)$. As a result standard orthogonality is lost and the bound procedure needs to be modified accordingly.

We consider for illustration purposes the stationary linear convection problem

$$U \cdot \nabla u + \sigma u = f^I \quad \text{in } \Omega, \quad \nabla \cdot U = 0$$

$$u = g^I \quad \text{on } \Gamma_i,$$

where Ω is a bounded domain in $\mathbb{R}^2$ with boundary Γ, σ is a real non–negative parameter, and $U = (U_1, U_2)$ is an incompressible prescribed velocity field. The boundary Γ is decomposed into an inflow boundary $\Gamma_i := \{\mathbf{x} \in \partial\Omega \mid \mathbf{n} \cdot \mathbf{U} < 0\}$ and an outflow boundary $\Gamma_o = \Gamma \backslash \Gamma_i$, where $\mathbf{n}$ is the unit outward normal on Γ. We assume that f^I and g^I are sufficiently regular so that the problem is well posed.

In order to introduce the stabilized weak formulation of this problem we define

$$(w,v) = \int_\Omega wv\,d\mathbf{x}, \qquad \langle w,v\rangle = \int_\Gamma \mathbf{U}\cdot\mathbf{n}\,wv\,ds,$$
$$\langle w,v\rangle_i = \int_{\Gamma_i} \mathbf{U}\cdot\mathbf{n}\,wv\,ds, \quad \langle w,v\rangle_o = \langle w,v\rangle - \langle w,v\rangle_i.$$

The output of interest, $s = l^O(u)$, will be of the form

$$l^O(v) = (f^O, v) + \langle g^O, v\rangle_o.$$

For $\delta = H, h$, we consider the following coarse and fine mesh problems : find $u_\delta \in X_\delta$ such that

$$a_\delta(u_\delta, v) = l_\delta(v), \quad \forall v \in X_\delta.$$

Here,

$$\begin{aligned} a_\delta(w,v) &\equiv (U\cdot\nabla w + \sigma w\ ,\ v + \tau_\delta(U\cdot\nabla w - \sigma w)) - \langle w,v\rangle_i \\ l_\delta(v) &\equiv (f^I\ ,\ v + \tau_\delta(U\cdot\nabla w - \sigma w)) - \langle g^I, v\rangle_i. \end{aligned}$$

Note that, since the boundary conditions are imposed naturally, the functions in X_H and X_h are, in this case, unrestricted on the boundary.

The stabilization parameter τ_δ is taken to be constant over each triangle $T_\delta \in \mathcal{T}_\delta$ and equal to

$$\tau_\delta = [|(\mathbf{l}_{12}\cdot\bar{\mathbf{U}})/(\mathbf{l}_{12}\cdot\mathbf{l}_{12})| + |(\mathbf{l}_{23}\cdot\bar{\mathbf{U}})/(\mathbf{l}_{23}\cdot\mathbf{l}_{23})| + |(\mathbf{l}_{31}\cdot\bar{\mathbf{U}})/(\mathbf{l}_{31}\cdot\mathbf{l}_{31})|]^{-1}.$$

Here $\bar{\mathbf{U}}$ is the average value of $\mathbf{U}$ in the element, and $\mathbf{l}_{ij} = \mathbf{x}_j - \mathbf{x}_i$, where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the coordinates of the element nodes i and j respectively. *A priori* error analysis of the resulting scheme (see [8] for the analysis of a similar algorithm) predicts an optimal convergence rate for the L_2–norm of the error in the $\mathbf{U}$–directional derivative, $\|\mathbf{U}\cdot\nabla(u-u_\delta)\|_0 \le C\delta\|u\|_2$, and a near–optimal convergence rate for the L_2–norm of the error in the solution, $\|u-u_\delta\|_0 \le C\delta^{3/2}\|u\|_2$. Here $\|\cdot\|_k$ denotes the usual Sobolev norms.

8.1 Lagrangian Bound Formulation

Repeating the steps of section 6.2, we construct the "energy" equality that will be used to convexify out problem. On mesh h we have

$$a_h^S(e_h, e_h) = R_h(e_h) \equiv l_h(e_h) - a_h(u_H, e_h),$$

where

$$\begin{aligned} a_h^S(w,v) \equiv\ & (U\cdot\nabla w\ ,\ \tau_h\, U\cdot\nabla v) + (\sigma w, v) \\ & - (\sigma w, \tau_h \sigma w) + \frac{1}{2}[\langle w,v\rangle_o - \langle w,v\rangle_i] \end{aligned}$$

which is coercive provided τ_h is sufficiently small — a condition which is not difficult to meet a-priori. Our Lagrangian $\mathcal{L}_h : \hat{X}_h \times X_h \times Q_h \mapsto \mathbb{R}$ is then,

$$\begin{aligned}\mathcal{L}_h^{\pm}(\hat{v},\theta,q) \equiv \kappa(a_h^S(\hat{v},\hat{v}) - R_h(\hat{v})) \pm l^O(v) \\ + a_h(\hat{v},\theta) - R_h(\theta) + b(\hat{v},q).\end{aligned}$$

Adjoint computation Our approach to computing a suitable adjoint is to derive the continuous adjoint equation and then approximate it on the coarse mesh. The continuous adjoint equation in this case reads,

$$-U\cdot\nabla\psi + \sigma\psi = -f^O \quad \text{in } \Omega, \quad \nabla\cdot U = 0$$

$$\psi = -g^O \quad \text{on } \Gamma_o.$$

The H-mesh approximation of this problem is : find $\psi_H \in X_H$,

$$\begin{aligned}(v - \tau_H\, U\cdot\nabla v\,, -U\cdot\nabla\psi_H + \sigma\psi_H) + \langle v,\psi_H\rangle_o \\ = (v - \tau_H\, U\cdot\nabla v\,, -f^O) + \langle v, -g^O\rangle_o, \quad \forall\, v \in X_H\end{aligned}$$

Note that the perturbation of the right hand side is necessary to obtain a high order approximation to the continuous adjoint problem; inattention to this detail yields a much less accurate adjoint ψ_H [12].

Bounds The procedures for equilibration and solution of the local problems are identical to those described in section 6.2. After calculation, the optimized bounds become,

$$\begin{aligned}s^+ &= l^O(u_H) - 2a_h^S(\hat{e}_{0h},\hat{e}_{1h}) - R_h(\psi_H) + 2\ |||\hat{e}_{0h}|||_h\ |||\hat{e}_{1h}|||_h,\\ s^- &= l^O(u_H) - 2a^S(\hat{e}_{0h},\hat{e}_{1h}) - R_h(\psi_H) - 2\ |||\hat{e}_{0h}|||_h\ |||\hat{e}_{1h}|||_h,\end{aligned}$$

where we recall that $|||\cdot|||^2 \equiv a_h^S(\cdot,\cdot)$.

The only difference between these expressions and (21), (22) is the term $R_h(\psi_H)$ which in this case is nonzero due to the lack of orthogonality alluded to earlier.

8.2 Application Example

We consider the numerical solution of the two–dimensional convection equation on a unit square $\Omega =]0,1[^2$. The convection velocity $\mathbf{U}$, is given by $\mathbf{U} = (-x_2, x_1)$; we take $\sigma = 0$ and $f^I = 0$. At the inflow boundary $\partial\Omega_i$, g^I is given by

$$g^I = \begin{cases} 5.0(x_1 - 0.2), & \text{for } 0.2 < x_1 \le 0.4, \quad x_2 = 0 \\ 1, & \text{for } 0.4 < x_1 \le 0.6, \quad x_2 = 0 \\ 1 - 5(x_1 - 0.6), & \text{for } 0.6 < x_1 \le 0.8, \quad x_2 = 0 \\ 0, & \text{otherwise} \end{cases} \qquad (23)$$

The output considered is the weighted integral of the solution on the outflow boundary $\partial\Omega_i$, that is, $f^O = 0$ and

$$g^O = \begin{cases} 10(x_1 - 0.6), & \text{for} \quad 0.6 \quad < x_2 \le 0.7, \quad x_1 = 0 \\ 1 - 10(x_1 - 0.7), & \text{for} \quad 0.7 \quad < x_2 \le 0.8, \quad x_1 = 0 \\ 0, & \text{otherwise} \quad . \end{cases} \tag{24}$$

The computational domain, together with the functions g and g^O, is sketched in figure 13.

A nearly uniform initial mesh consisting of 240 triangles is uniformly refined four times to generate a sequence of meshes with $240 \cdot 4^r$ elements for $r = 1, 2, 3$, and 4. The finest mesh so obtained, consisting of $61,440$ triangles, is taken to be the fine mesh $\mathcal{T}_h$. For each of the coarser meshes we calculate the upper and lower bounds; the results are shown in Fig. 14. The observed convergence rate is slightly greater than $O(H^2)$.

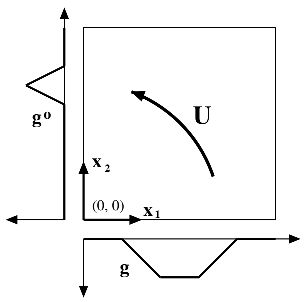


Fig. 13. Two–dimensional linear convection: computational domain with the definition of the functions g and g^O.

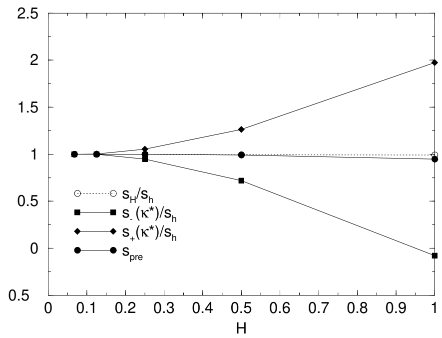


Fig. 14. Two–dimensional linear convection: computed upper and lower bounds as a function of H.

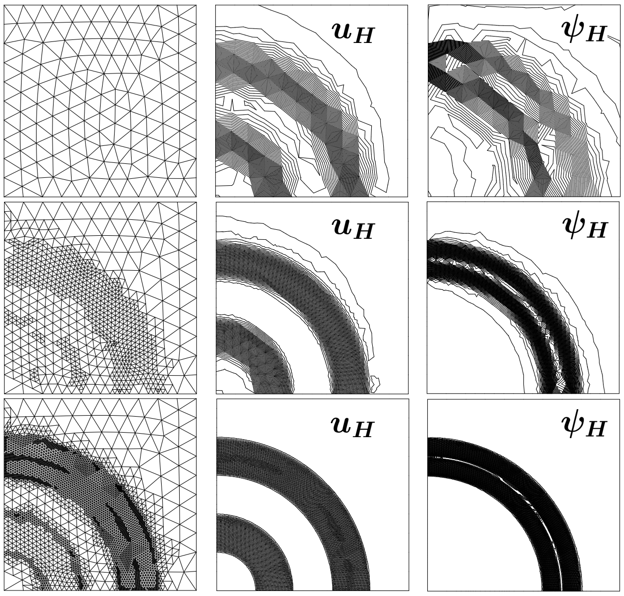


Fig. 15. Two–dimensional linear convection: adaptively generated mesh (left), u_H (middle), and ψ_H (right), for a) the initial mesh of 240 elements and relative bound gap 108%, b) an intermediate mesh, after 5 refinements, of 1, 557 elements and relative bound gap 7.10%, and c) the final mesh, after 10 refinements, of 7, 791 elements and relative bound gap 0.1%.

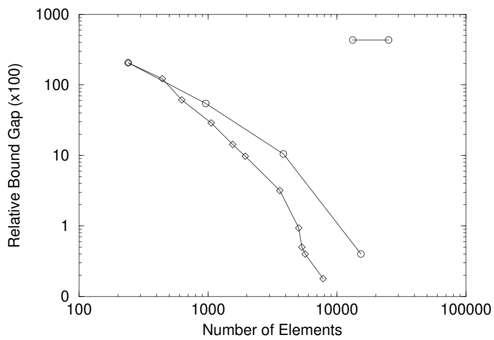


Fig. 16. Two–dimensional linear convection: bound gap $2\Delta/s_{pre}$ as a function of the number of elements in $\mathcal{T}_H$ for meshes generated using uniform and adaptive refinement.

We next implement the adaptive procedure described in section 7. The starting triangulation is the coarsest mesh considered above, which has 240 triangles. On this mesh the error, given by Δ/s_{pre}, is 1.08. In the subsequent adaptive process all those triangles whose contribution to the bound gap is larger than 0.2 that of the largest elemental contribution are refined; the strategy described in [18] is used in order to treat the "hanging nodes" generated during the refinement process. The accuracy goal for Δ/s_{pre} is set to 0.001; a total of 10 meshes are generated before this goal is achieved. The final adapted mesh has $7,791$ elements, which compares favorably with the uniform mesh that would be required to obtain a similar level of accuracy. In particular, a three–level uniformly refined mesh has $15,360$ triangles, but only produces an accuracy $\Delta/s_{pre} = 0.002$. Fig. 15 shows the mesh and the contours for the computed solution u_H and adjoint ψ_H, initially, after one intermediate refinement step, and at the end of the adaptive process; note the effect of the adjoint (and hence output) structure on the mesh. Finally, Fig. 16 presents $2\Delta/s_{pre}$ as a function of the number of elements in $\mathcal{T}_H$ for both adaptive and uniform refinement

9 Non-coercive problems

The bound procedure developed above is applicable whenever a Lagrangian can be found which has the required properties. When the differential operator considered is non-coercive, the convexification procedure based on the energy equality is not globally valid and therefore our standard approach fails. The basic method can be extended to deal with these more general cases but at the expense of obtaining only asymptotic bounds. That is, bounds are guaranteed once the coarse mesh is such that $H < H_0$. Unfortunately, H_0 is not known a priori and as a consequence some certainty is lost. In practice however, this approach has been found to work very well and only, in very rare circumstances and using very inaccurate H-mesh approximations, have we found that the computed bounds are not valid. This procedure is given in detail in [14], and here we only provide a very brief description.

We consider the problem of determining $u \in X$, such that

$$\mathcal{A}(u,v) = 0, \qquad \forall v \in X.$$

Here, $\mathcal{A} : X \times X \mapsto \mathbb{R}$, is linear in the second argument. We also consider an output $s = \mathcal{S}(u)$, defined by $\mathcal{S} : X \mapsto \mathbb{R}$. We introduce also the coarse grid solution $u_H \in X_H$, such that

$$\mathcal{A}(u_H,v) = 0, \quad \forall v \in X_H.$$

The forms $\mathcal{A}$ and $\mathcal{S}$ are expanded about u_H,

$$\mathcal{A}(u_H + w, v) - \mathcal{A}(u_H, v) = E(w,v) + F(w,w,v) + G(w,v)$$
$$\mathcal{S}(u_H + w) - \mathcal{A}(u_H) = L(w) + M(w,w) + N(w).$$

Here $L(w)$ is linear, $E(w,v)$ and $M(w,w)$ are bi-linear and $F(w,w,v)$ is tri-linear. The forms $G(w,v)$ and $N(w)$ correspond to the remainder in the expansions and contain the higher order terms.

We consider the symmetric part of $E(\cdot,\cdot)$ defined as

$$E^S(w,v) = \frac{1}{2}(E(w,v) + E(v,w)).$$

Furthermore, we introduce the following splitting of $E^S(w,v)$,

$$E^S(w,v) = E_c^S(w,v) + E_a^S(v,w), \quad E_c^S(w,w) > 0, \ \forall w \in X \ (w \neq 0),$$

that is, we identify the (or a) coercive part of $E^S(w,v)$, with $E_c^S(w,v)$, and $E_a^S(v,w)$ with the remainder.

9.1 Bound Algorithm

The bound algorithm in the non-coercive case comprises the following steps, given below in algorithmic form:

- Solve problem on the coarse mesh: find $u_H \in X_H : \mathcal{A}(u_H, v) = 0, \quad \forall v \in X_H$,
  and evaluate the primal residual $R^u(v) = -\mathcal{A}(u_H, v), \quad \forall v \in X_h$.
- Compute the coarse mesh adjoint : find $\psi_H \in X_H : E(v, \psi_H) = -L(v), \ \forall v \in X_H$, and determine the adjoint residual $R^\psi(v) = -L(v) - E(v, \psi_H), \quad \forall v \in X_h$
- Calculate the hybrid fluxes, $p_H^u \in Q$, $p_H^\psi \in Q$, using the equilibration procedure described in section 5.4.
- Compute the reconstructed error by solving the local problems: find $\hat{e}_h^u \in \hat{X}_h$, and $\hat{e}_h^\psi \in \hat{X}_h$ such that

$$\begin{aligned} 2E_c^S(\hat{e}_h^u, v) &= R^u(v) + b(v, p_H^u), \quad \forall v \in \hat{X}_h, \\ 2E_c^S(\hat{e}_h^\psi, v) &= R^\psi(v) + b(v, p_H^\psi), \quad \forall v \in \hat{X}_h. \end{aligned}$$

Finally, the bound average is computed as

$$\bar{s} = \mathcal{S}(u_H) - 2E_c^S(\hat{e}_h^u, \hat{e}_h^\psi) - R(\psi_H).$$

Note that, whenever the form $\mathcal{A}$ does not depend on h, the term $R(\psi_H)$ will vanish.

The bound gap is

$$\Delta = 2\sqrt{E_c^S(\hat{e}_h^u, \hat{e}_h^u) \ E_c^S(\hat{e}_h^\psi, \hat{e}_h^\psi)},$$

which, if required, can be written in terms of positive elemental contributions for adaptivity purposes [14]. The upper and lower bounds follow as $s^\pm = \bar{s} \pm \Delta$.

Bound Properties (s^-) Introducing $e = u_h - u_H$ and $\hat{e} = \hat{e}_h^u + \frac{1}{\kappa}\hat{e}_h^\psi$, it is possible to write [14],

$$s^- = \mathcal{S}(u_h) - \kappa E_c^S(e - \hat{e}, e - \hat{e}) + \mathcal{F}(e),$$

with $\kappa = \sqrt{E_c^S(\hat{e}_h^\psi, \hat{e}_h^\psi)/E_c^S(\hat{e}_h^u, \hat{e}_h^u)}$.

For coercive linear problems $\mathcal{F}(e) = 0$ and bounds will obtain for all H. For more general problems, if we can find a coercive splitting (9) such that $E_c^S(e - \hat{e}, e - \hat{e}) \gg \mathcal{F}(e)$ for H sufficiently small then asymmptotic bounds will be obtained. Typically, we will have expect $E_c^S(e-\hat{e}, e-\hat{e}) \sim O(H^p)$ and , $\mathcal{F}(e) \sim O(H^p)$, with $p > q$. This procedure has been applied to a number on non-coercive problems [19,14,21] including the Helmholtz equation, the Burgers equation and the symmetric eigenvalue problem. In all these cases one obtains, using linear elements, $p = 4$, $q = 2$. The approach has also been extended to the Navier-Stokes equations, but in this case further elaboration is necessary and the reader is referred to [13] for further details. Below we give some computational results for the symmetric eigenvalue problem and for the incompressible Navier-Stokes equations.

9.2 Eigenvalue Problem

We consider the following two dimensional eigenvalue problem

$$-\nabla^2 u = \lambda u, \quad \text{in } \Omega, \ \ u|_\Gamma = 0.$$

The computational domain is defined in figure 17 and consists of a square domain of unit side length with a rectangular cut out. Here, the outputs

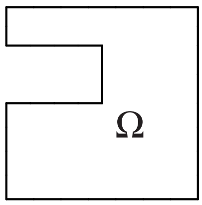


Fig. 17. Computational domain used for the eigenvalue problem.

of interest are the eigenvalues $s = \lambda$, but it is also possible to include the eigenvectors provided they are suitably normalized [14].

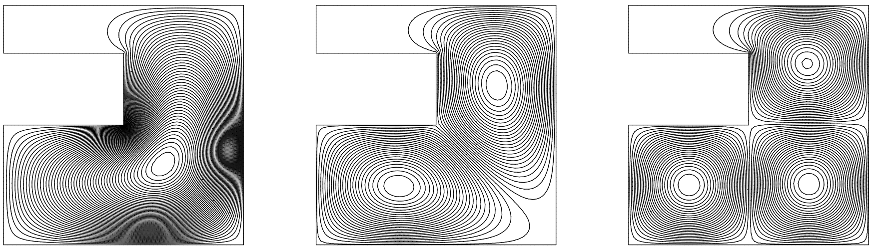
By defining

$$a(w, v) \equiv \int_\Omega \nabla w \cdot \nabla v \, d\Omega, \quad m(w, v) \equiv \int_\Omega w \nabla v \, d\Omega,$$

the variational problem becomes: find $(u, \lambda) \in X \times \mathbb{R}$ such that

$$a(u, v) - \lambda m(u, v) - \sigma(m(u, u) - 1) = 0, \quad \forall (v, \sigma) \in X \times \mathbb{R}.$$

In terms of the terminology introduced above we write : find $(u,\lambda) \in X \times \mathbb{R}$ such $\mathcal{A}((u,\lambda),(v,\sigma)) = 0$, $\forall (v,\sigma) \in X \times \mathbb{R}$. Here, $\mathcal{A}((w,\mu),(v,\sigma)) \equiv a(w,v) - \mu m(w,v) - \sigma(m(w,w) - 1)$. The corresponding bi-linear form becomes $E((w,\mu),(v,\sigma)) \equiv a(w,v) - \lambda_H m(w,v) - \mu m(u_H,v) + 2\sigma m(u_H,w)$ and we simply choose $E_c^S((w,\mu),(v,\sigma)) = a(w,v)$ as the coercive component of the symmetric part of $E(\cdot,\cdot)$. The results are summarized in figures 18 and 19. Figure 18 shows the contours of the first three eigenmodes and corresponding eigenvalues computed on a very fine mesh $\mathcal{T}_h$ consisting of 27,648 elements. This mesh is obtained by subdividing each element of the coarse mesh $\mathcal{T}_H$, shown in figure 19 into $4^4 = 256$ elements. Figure 19 shows the results computed for the first three eigenmodes on mesh $\mathcal{T}_H$ of 108 elements and mesh $\mathcal{T}_{H/4}$ of 1728 elements. In all cases bounds are obtained and the bound gap is found to converge at an approximate rate of $\mathcal{O}(H^2)$. For this problem, it can be shown that the eigenvalue computed on the coarse mesh, λ_H, is an upper bound for λ_h. Our bound procedure does in fact predict that $s^+ = \lambda_H$.



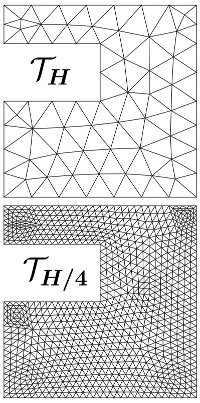
$\lambda_h^1 = 38.546$ $\lambda_h^2 = 60.473$ $\lambda_h^3 = 78.458$

Fig. 18. Contours of the first three eigenmodes and corresponding eigenvalues computed on the reference mesh $\mathcal{T}_h$.

9.3 Navier-Stokes Equations

We consider now the problem of natural convection in a two dimensional domain. This is the problem for which results were presented earlier in section 1.1. The incompressible Navier-Stokes equations are solved together with the energy equation. Boussinesq's approximation is used to model buoyancy effects. Thus, the set of partial equations to be solved are

$$-\frac{\partial}{\partial x_j}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) + \frac{\partial(u_j u_i)}{\partial x_j} + \frac{\partial p}{\partial x_i} = -Gr\,\theta \hat{g}_i$$
$$\frac{\partial u_i}{\partial u_i} = 0$$



$\lambda_h^1 = 38.546$	$\lambda_h^2 = 60.473$	$\lambda_h^3 = 78.458$
$s^+ = 42.023$ $s^- = 34.234$ $s^+ - s^- = 7.789$	$s^+ = 65.216$ $s^- = 56.448$ $s^+ - s^- = 14.504$	$s^+ = 87.346$ $s^- = 70.315$ $s^+ - s^- = 17.030$
$s^+ = 38.845$ $s^- = 38.164$ $s^+ - s^- = 0.681$	$s^+ = 60.779$ $s^- = 60.232$ $s^+ - s^- = 0.546$	$s^+ = 78.995$ $s^- = 78.033$ $s^+ - s^- = 0.961$

Fig. 19. Computed results for eigenvalue problem. Table shows the upper bound the lower bound and the bound gap for the first three eigenvalues and for each of the two meshes shown on the left consisting of 108 and 1728 linear triangular elements.

$$-\frac{1}{Pr}\frac{\partial^2\theta}{\partial x_j \partial x_j} + \frac{\partial(u_j\theta)}{\partial x_j} = 0,$$

where the summation convention over repeated indices applies, $\mathbf{u} = (u_1, u_2)$ is the velocity field, p is the pressure field, and θ is the temperature field, and $\hat{\mathbf{g}} = (\hat{g}_1, \hat{g}_2)$ is the unit vector in the direction of gravity. The flow is governed by two non-dimensional parameters: the Prandtl number, Pr, and the Grashof number, Gr.

The example considered is that used in the introduction of these notes. The computational boundary is shown in figure 20. The boundary conditions are $u = 0$ on all boundaries, and $\theta = 0$ on the right vertical boundary, a prescribed heat flux $\frac{1}{Pr}\frac{\partial\theta}{\partial n} = Q(= 1)$ on the boundaries shown, and adiabatic conditions on all other boundaries. For the simulations we have assumed that $Pr = 1$ and $Gr = 50,000$

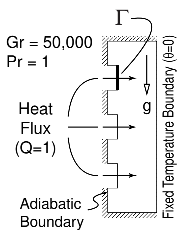


Fig. 20. Computational domain for the natural convection problem.

The output of interest is the average temperature over Γ, as shown in figure 20,

$$s(u, p, \theta) = \frac{1}{|\Gamma|} \int_{\Gamma} \theta \, ds.$$

The discretization of the Navier-Stokes equations is performed using the Crouzeix-Raviart element with linear discontinuos pressures and quadratic plus cubic bubble representation of the velocity components. The detail application of the bound procedure, including adaptivity, to this problem can be found in [13]. A prescribed accuracy of 1% is specified and the adaptive algorithm is initiated. Figure 3 shows the initial mesh $\mathcal{T}_H^0$ of 264 and the final mesh $\mathcal{T}_H^2$ obtained after two adaptations, of 680 elements, together with the computed isotherms and streamlines.

	$\mathcal{T}_H^0$	$\mathcal{T}_H^1$	$\mathcal{T}_H^2$	$\mathcal{T}_{H/2}$
# of elements	264	416	628	1056
s^-	0.253	0.270	0.273	0.272
s^+	0.296	0.282	0.279	0.279
Δ	0.0212	0.0057	0.0026	0.0036

Finally, the above table shows the number of elements and bounds computed on the initial mesh and two successively refined meshes. A reduction of the relative bound gap from $\approx 8\%$ for the initial mesh, to less than 1% for the final mesh, is achieved by optimal refinement with a final mesh which contains only slightly more than twice the number of elements of the initial mesh. These results contrast with the last column which shows the figures corresponding to a uniformly subdivided mesh. In the calculations reported here, the truth mesh is obtained by subdividing each coarse mesh element into 6^2 elements.

10 Conclusions

We have presented a rigorous procedure for computing bounds for functional outputs of solutions of partial differential equations. The procedure allows us to bound the find mesh solution at the cost of only solving coarse mesh global problems. In addition, the local problem solutions are symmetric and linear. For coercive problems the bounds are uniform whereas for general non-coercive and/or non-linear problems the bounds are only asymmptotic. In practice however, we have found that asymmptotic bounds are very reliable and only on very rare occasions, solving on very coarse meshes, have we not obtained bounds. The effectivities computed are of order $\mathcal{O}(1)$ (typically between 1 and 6-7) which we consider "good" if one takes into account the high level of certainty delivered. Finally, the procedure presented lends itself naturally to adaptive mesh refinement by allowing the bound gap to be decomposed into the sum of positive elemental contributions which can be used as refinement indicators.

Acknowledgments

These notes present a summary of research results undertaken by the authors which span a several years. We would like to acknowledge the collaboration with many colleagues, in particular, M. Paraschivoiu, Y. Maday, and L. Machiels. Over the years this work has been sponsored by several funding sources including NASA, DARPA, and the Singapore-MIT Alliance.

References

1. M. Ainsworth and T.J. Oden, *A unified approach to a posteriori error estimation based on element residual methods*, Numer. Math., 65:23-50, 1993.
2. M. Ainsworth and J.T. Oden, *A posteriori Error Estimation in Finite ELement Analysis*, Wiley-Interscience, 2000,
3. I. Babuška and W.C. Rheinboltd, *A posteriori error estimates for the finite element method*, Int. J. Num. Meth. in Egngr., 12:1597-1615, 1978.
4. R. Bank and A. Weiser, *Some a posteriori error estimators for elliptic partial differential equations*, Math. Comp., 44:283-301, 1985.
5. R. Becker and R. Rannacher, *Weighted a posteriori error control in finite element methods*, IWR Preprint 96-1 (SFB359), Heildelberg, 1996.
6. D. Bertsekas, *Nonlinear Programming*, Athena Scientific, 1995.
7. F. Brezzi and M. Fortin, *Mixed and hybrid finite element methods*, Springer-Verlag, New York, 1991.
8. C. Johnson, U. Nävert and J. Pitkäranta, *Finite Element Methods for Linear Hyperbolic Problems*, Comp. Meth. in Appl. Mech. and Engrg., 45 (1984), pp. 285–312.
9. P. Ladevèze and D. Leguillon, *Error estimate procedure in the finite element method and applications*, SIAM J. Numer. Anal., 20:485-509, 1983.
10. L.D. Landau and E.M. Lifshitz, *Fluid Mechanics*, Course of Theoretical Physiscs, vol. 6, Pergamon Press 1982.
11. L. Machiels, Y. Maday and A.T. Patera, *A "Flux-Free" nodal Neumann subproblem approach to output bounds for partial differential equations*, to appear in C.. Acad.Sci. Paris, 1999.
12. L. Machiels, A.T. Patera, J. Peraire and Y. Maday, *A general framework for finitie element a posteriori error control: application to linear and nonlinear convection-dominated problems*, ICFD Conference on Numerical Methods for Fluid Dynamics, Oxford, 1998.
13. L. Machiels, J. Peraire and A.T. Patera, *A posteriori finite element output bounds for the incompressible Navier-Stokes equations; Application to a natural convection problem*, J. Comp. Phys., 172, 401-425, 2001.
14. Y. Maday, A.T. Patera and J. Peraire, *A general formulation for a posteriori bounds for output functionals of partial differential equations; application to the eigenvalue problem*, C.. Acad.Sci. paris, t. 328, Serie I, p. 823-828, 1999.
15. M. Paraschivoiu and A.T. Patera, *A hierarchy duality approach to bounds for the outputs of partial differential equations*, Comput. Methods Appl. Mech. Engrg., 15 Ladevèze and J.T. Oden editors. Elsevier, 1998.

16. M. Paraschivoiu, J. Peraire and A.T. Patera, *A posteriori finite element bounds for linear-functional outputs of elliptic partial differential equations*, Comput. Methods Appl. Mech. Engrg., 150:289-312 (1997).
17. A.T. Patera and E. Ronquist, *A general output bound result: application to discretization and iteration error estimation and control*, Mathematical Models and Methods in Applied Science, to appear.
18. J. Peraire and A.T. Patera, *Bounds for linear-functional outputs of coercive partial differential equations: local indicators and adpative refinement*, in Advances in Adaptive Computational Methods in Mechanics, P. Ladevèze and J.T. Oden editors. Elsevier, 1998.
19. J. Peraire and A.T. Patera, *Asymmptotic a posteriori finite element bounds for the outputs of non-coercive problems: the Helmholtz and Burgers equations*, Comp. Meth. Appl. Mech. Engrg., 171, 77-86, 1999.
20. C. Prud'homme, D. Rovas, K. Veroy, L. Machiels, Y. Maday, A.T. Patera and G. Turinici, *Reliable real-time solution of parametrized partial differential equations: reduced-basis output bond methods*, Journal of Fluids Engineering, 2002 (to appear).
21. J. Sarrate, J. Peraire and A.T. Patera, *A posteriori bounds for non-linear outputs of solutions of the Helmholtz equation*, Int. J. Num. Meth. Fluids, 31, 17-36, 1999.
22. G. Strang, *Introduction to applied mathematics*, Wellesley-Cambridge Press, 1986.