Numer. Math. (2005) 101: 185–219
DOI 10.1007/s00211-005-0601-x

Numerische
Mathematik

I. Babuška · F. Nobile · R. Tempone

Worst case scenario analysis for elliptic problems with uncertainty

Received: 9 September 2004 / Published online: 10 June 2005

Abstract This work studies linear elliptic problems under uncertainty. The major emphasis is on the deterministic treatment of such uncertainty. In particular, this work uses the Worst Scenario approach for the characterization of uncertainty on functional outputs (quantities of physical interest). Assuming that the input data belong to a given functional set, eventually infinitely dimensional, this work proposes numerical methods to approximate the corresponding uncertainty intervals for the quantities of interest. Numerical experiments illustrate the performance of the proposed methodology.

Mathematics Subject Classification (2000) 35J20 · 35R60 · 65N30

Introduction

During the last few decades and influenced by the rapid development of digital computers, numerical simulations have become an essential tool in engineering, environmental sciences, biology, medicine, chemistry and many other fields. Furthermore, simulation tools more and more frequently are at the basis for decisions in engineering, public policy, etc.

In addition to classical deterministic computations, simulations taking into consideration various uncertainties and probabilities that may arise in the description of a physical problem are used widely today. Such simulations appear in civil engineering ([16,23,24,34]), nuclear engineering ([14,20,21,32]), ground flows [15] and in many other fields as the basis of risk analysis. Uncertainty Quantification

I. Babuška · R. Tempone
ICES, The University of Texas at Austin, 1 University Station CO200, TX 78712, USA
E-mail: {babuska, rtempone}@ices.utexas.edu

F. Nobile (✉)
MOX, Dipartimento di Matematica, Politecnico di Milano, via Bonardi 9, 20133 Milano, Italy
E-mail: fabio.nobile@mate.polimi.it

(UQ) is also a necessary step in assessing the reliability of computer simulations and, in this respect, it is also part of the broader area of *Validation and Verification* (we refer to the wide literature in the field: the guide [1], the survey articles [5,26, 27] and the book [33] where many other relevant references are given).

Computational analysis (simulation) relies, typically, on a *mathematical model* and its *input*, to obtain an *output* of a desired quantity of interest. By a mathematical model we mean a set of mathematical relations, usually based on physical principles, like conservation laws, Newton's gravitation law, etc. By the input we mean the data needed in the mathematical formulation, for example the physical domain, the coefficient functions, boundary conditions etc. These are natural inputs in boundary value problems.

Uncertainty may arise at different levels. It could appear in the mathematical model itself, for instance if we are not sure about the linear behavior of some physical system, or in the input data of the model. In this work we will discuss only the uncertainty in the input data, typically, coefficients and forcing terms in the mathematical model. An analysis of the effect of uncertainty in the domain can be found in [3,4]. Also, the works [30,29] on modeling error estimations can be seen as an attempt to quantify the uncertainty in the mathematical model.

Very likely, the most complete way to describe uncertainty in the input data is in a probabilistic setting. Suppose that for all the "uncertain" data of the model we know the associated probability distribution. In such a case, we can solve a *stochastic* problem to compute the probability distribution of the output quantity. This will allow us to predict information like the mean value and standard deviation of the quantity of interest, the probability that the output is larger than a critical value, etc.

However, in some applications, a full characterization of the probability distribution implies a huge amount of experiments and measurements, often unaffordable because of budget or time constraints. This constitutes a major limitation to the probabilistic approach for such applications.

Another limitation comes from the fact that, whenever the input data belong to an infinite dimensional space (they might be functions of position and/or time), their probabilistic characterization must include knowledge of the cross correlation of the values that the data can take at different points in space and time. In this case, the solution of a stochastic model becomes quickly too costly.

In this work, we consider an alternative and inexpensive way to characterize uncertainty in the output. This approach is particularly useful in those cases where we only know little information on the uncertainty in the input data, namely that the input data lie in a functional set (that might well be infinite dimensional). For instance, we may consider any load acting on a portion of a given structure, provided it does not exceed a maximum allowed value.

The methodology proposed relies on a perturbation technique around the *nominal* values of the data. We will present an algorithm to compute the first term in the expansion and we will provide rigorous bounds for the reminder of the expansion, valid for any size of the perturbations, provided some minimal requirements are preserved (like the coerciveness of the form for elliptic problems). In some cases, however, these bounds might be too pessimistic for large perturbations as it will be shown in one of the two numerical tests in Section 5. Also, we will present convergence results in case a finite element approximation is used to solve numerically the model.

To give some more details of the main idea, let us suppose that we are interested in computing a specific *quantity of interest* (q.o.i), Q, that is a function of the solution u of a partial differential equation which, on its turn, depends on the input parameters of the mathematical model, hereafter denoted by $\boldsymbol{\eta}$ (bold symbols will be used to indicate vector quantities). We will make the assumption that the quantity Q can be represented by a linear functional on the space V of admissible solutions.

The quantity of interest Q depends on the parameters $\boldsymbol{\eta}$ –which may be functions– through the solution $u(\boldsymbol{\eta})$ and possibly also explicitly. To highlight this dependence we introduce the notation

$$\psi(\boldsymbol{\eta}) = Q(\boldsymbol{\eta}; u(\boldsymbol{\eta})), \qquad \forall \boldsymbol{\eta} \in \mathcal{A}_\eta, \tag{1}$$

where $\mathcal{A}_\eta$ is an admissible set for the input $\boldsymbol{\eta}$. The situation we are trying to describe here is the one where the *true* parameter in the model is unknown and the only information that we dispose of it is that *it lies in the functional set* $\mathcal{A}_\eta$. The goal of this work is to develop a technique to bound the error introduced in the computation of the quantity of interest when we choose a parameter that might not be the true one. More precisely, suppose that we are able to compute the quantity of interest $\psi(\boldsymbol{\eta}_0) = Q(\boldsymbol{\eta}_0, u(\boldsymbol{\eta}_0))$ for some $\boldsymbol{\eta}_0 \in \mathcal{A}_\eta$. The goal is to estimate the maximum error

$$\Delta Q = \sup_{\boldsymbol{\eta} \in \mathcal{A}_\eta} |\psi(\boldsymbol{\eta}) - \psi(\boldsymbol{\eta}_0)| \tag{2}$$

corresponding to the *worst-case scenario* within all possible choices of the parameters $\boldsymbol{\eta}$ in $\mathcal{A}_\eta$. There is a large amount of literature – see e.g. [9,13,17] and the references therein – addressing the worst scenario approach for which many names are used. A recent and related approach is the so called “info-gap” theory where a family of sets $\mathcal{A}_\eta$ is used to describe uncertainty (see [8]). In addition, formulation (2) can be seen as an optimization (or *anti-optimization*) problem. Whenever the set $\mathcal{A}_\eta$ of admissible parameters is infinitely dimensional, as it is in all the situations we will address in this work, the computation of ΔQ can be challenging and extremely costly. The perturbation technique that we advocate can be applied with low computational cost for any size of the perturbations and it is endowed with a posteriori error estimates for discretization error control.

1 Examples

Let D be a bounded polygonal domain in $\mathbb{R}^d$, $d = 1, 2, 3$. We will make this assumption for the domains throughout the rest of the work. Further, assume that the boundary set ∂D is the disjoint union of subsets where different boundary conditions are imposed, i.e. Γ_N for Neumann and Robin type, and Γ_D for essential (e.g. Dirichlet) ones. We also assume that the set Γ_D has positive measure in ∂D to ensure the well posedness of the weak formulations.

*Example 1 (*Scalar isotropic diffusion equation with unknown coefficient*)* Consider the equation

$$\begin{cases} -\operatorname{div}(\beta \nabla u) = f, & \text{on } D \subset \mathbb{R}^3 \\ u = g, & \text{on } \Gamma_D \subset \partial D \\ \beta\, \partial_n u = h_1 - u\, h_2, & \text{on } \Gamma_N \subset \partial D \end{cases} \tag{3}$$

with smooth data, $f : D \to \mathbb{R}$, $g : \Gamma_D \to \mathbb{R}$ and $h_1, h_2 : \Gamma_N \to \mathbb{R}$, respectively. The uncertain parameter, in this example, will be the diffusivity coefficient, i.e. $\boldsymbol{\eta} = [\beta]$. In principle, the coefficient β is not constant over D. Provided that β is uniformly bounded away from zero, the solution u belongs to the affine subspace $H^1_{g,\Gamma_D}(D) = \{v \in H^1(D) : v = g \text{ on } \Gamma_D\}$. The quantity of interest might be, for instance, the average of the quantity u over a subset of the domain

$$Q(u) = \frac{1}{|S|} \int_S u\, d\mathbf{x}, \qquad S \subset D$$

or a similar quantity based on the gradient of the solution,

$$Q(\beta; u) = \frac{1}{|S|} \int_S \beta \nabla u \cdot \boldsymbol{\gamma}\, d\mathbf{x}, \qquad S \subset D.$$

where $\boldsymbol{\gamma}$ is a vector field in S.

Assuming that the material is homogeneous in D, the set of parameters could take the form $\mathcal{A}_\eta \subset \mathbb{R}$, $\mathcal{A}_\eta = [\beta_{min}, \beta_{max}]$, with $\beta_{min} > 0$. The two bounds β_{min} and β_{max} will typically come from experimental measurements.

In practical situations, though, we can not exclude a priori the presence of heterogeneities in the material. Yet, precise measurements of such heterogeneities are very hard to obtain if not impossible in general. As a pessimistic (*worst*) scenario, we might consider the set $\mathcal{A}_\eta \subset L^\infty(D)$, $\beta_{min} \leq \beta(\mathbf{x}) \leq \beta_{max}, \quad \forall \mathbf{x} \in D$.

*Example 2 (*Linear elasticity with unknown material properties*)* Consider the displacement filed $\mathbf{u}$ of a body occupying the region $D \subset \mathbb{R}^3$, described by the equations of linear elasticity

$$\begin{cases} -\operatorname{div}(\mathbf{C}(\boldsymbol{\beta})\nabla_s \mathbf{u}) = \mathbf{f}, & \text{in } D \subset \mathbb{R}^3 \\ \mathbf{u} = 0 & \text{on } \Gamma_D \subset \partial D \\ \mathbf{C}(\boldsymbol{\beta})\nabla_s \mathbf{u} = \mathbf{g} & \text{on } \Gamma_N \subset \partial D \end{cases} \tag{4}$$

where $\mathbf{C}(\boldsymbol{\beta})$ is the fourth order elasticity tensor and $\nabla_s \mathbf{u} = (\nabla \mathbf{u} + \nabla^T \mathbf{u})/2$ is the strain tensor. The solution $\mathbf{u}(\beta)$ belongs to $V = [H^1_{\Gamma_D}(D)]^3 \equiv \{\mathbf{v} \in [H^1(D)]^3, \quad \mathbf{v} = 0 \text{ on } \Gamma_D\}$. The coefficients in the model can be taken as the Young module, E, and Poisson ratio, ν, (i.e. $\boldsymbol{\beta} = [E, \nu]$) should the material be modeled as isotropic. If the material is orthotropic, then the vector of coefficients $\boldsymbol{\beta}$ consists of the 9 coefficients characterizing the elasticity tensor, whereas in the general anisotropic case, $\boldsymbol{\beta}$ consists of 21 coefficients functions. As in the previous example, we will take as uncertain parameters in the model the elasticity coefficients, i.e. $\boldsymbol{\eta} = \boldsymbol{\beta}$ and we will allow for pointwise perturbations within a given range, thus accounting for possible inhomogeneities of the material. Section 5.2

presents numerical results for a cantilever beam with rectangular cross section and, as quantities of interest Q_1 and Q_2, averaged displacement in the vertical and axial directions at the free end of the beam, respectively.

*Example 3 (*Linear elasticity with unknown forcing term*)* Let us consider again the previous example, but this time with uncertainty in the load terms. For instance, to describe the uncertainty in the volume force $\mathbf{f}$, we may consider the parameter set $\{\mathbf{f} \in [L^2(D)]^3 : \|\mathbf{f} - \mathbf{f}_0\|_{[L^2(D)]^3} < \epsilon_f\}$, for some forcing term $\mathbf{f}_0 \in [L^2(D)]^3$ and a real number $\epsilon_f > 0$. Similarly, to describe the uncertainty in the traction $\mathbf{g}$ applied on the portion Γ_N of the boundary, we can use the set $\{\mathbf{g} \in [L^\infty(\Gamma_N)]^3 : \|\mathbf{g} - \mathbf{g}_0\|_{[L^\infty(\Gamma_N)]^3} < \epsilon_g\}$, for some function $\mathbf{g}_0 \in [L^\infty(\Gamma_N)]^3$ and $\epsilon_g > 0$. Observe that the previous choices of $L^2(D)$ and $L^\infty(\Gamma_N)$ as underlying spaces for parameters is illustrative and not general –see Application 3–. This choice has to be motivated for each case by the physical information available.

2 Mathematical description of the problem

Let V be a Hilbert space and W a Banach space over the domain $D \subset \mathbb{R}^d$, $d = 1, 2, 3$, with norms denoted by $\|\cdot\|_V$ and $\|\cdot\|_W$, respectively. Moreover, we will denote by W' the dual space of W.

For a non negative integer s and $1 \le p \le +\infty$, let $W^{s,p}(D)$ be the Sobolev space of functions having generalized derivatives up to order s in the space $L^p(D)$, see [18]. The standard Sobolev norm of $v \in W^{s,p}(D)$ will be denoted by $\|v\|_{W^{s,p}(D)}$, $1 \le p \le +\infty$, and in the case $p = 2$ we shall write $H^s(D) \equiv W^{s,2}(D)$.

The space V is typically a Sobolev space of the form $(H^s(D))^n$. For instance, we have $n = 1$ for the diffusion equation (3) and $n = d$ for the linear elasticity case (4). In both cases, we have $s = 1$. Besides, we introduce the subspace V_0 of functions that satisfy essential (Dirichlet) homogeneous boundary conditions on $\Gamma_D \subset \partial D$. Similarly, we will denote by V_g the affine subspace of V of functions that satisfy the non-homogeneous Dirichlet datum g on Γ_D. For example, in the case $V = H^1(D)$ the subspace $V_0 \subset V$ is simply defined as $V_0 = H^1_{\Gamma_D} \equiv \{v \in V : v = 0, \forall x \in \Gamma_D\}$ and $V_g = \{v \in V : v = g \text{ on } \Gamma_D\}$.

Consider a bilinear form B on V that depends on some coefficients $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_m] \in \mathcal{A}_\beta \subseteq W$. We will call $\mathcal{A}_\beta$ the *set of coefficients* and W the *space of coefficients*. Hence B will take the form: $B : \mathcal{A}_\beta \times V \times V \to \mathbb{R}$, bilinear with respect to its last two arguments and supposed to satisfy the following hypotheses:

– *uniform coerciveness over* V_0. There exist $\alpha, \tilde{\alpha}(\boldsymbol{\beta}) > 0$, s.t.

$$B(\boldsymbol{\beta}; v, v) \ge \tilde{\alpha}(\boldsymbol{\beta})\|v\|_V^2 > \alpha\|v\|_V^2, \qquad \forall v \in V_0, \boldsymbol{\beta} \in \mathcal{A}_\beta. \tag{5}$$

– *uniform continuity.* There exist $M, \tilde{M}(\boldsymbol{\beta}) > 0$, s.t.

$$B(\boldsymbol{\beta}; v, w) \le \tilde{M}(\boldsymbol{\beta})\|v\|_V\|w\|_V < M\|v\|_V\|w\|_V, \qquad \forall v, w \in V, \boldsymbol{\beta} \in \mathcal{A}_\beta. \tag{6}$$

We also introduce a set of *admissible loads*, $\mathcal{A}_\mathcal{L} \subset {V_0}'$. Then, for any load $\mathcal{L} \in \mathcal{A}_\mathcal{L}$ and any coefficient $\boldsymbol{\beta} \in \mathcal{A}_\beta$ the Lax-Milgram Lemma (see e.g. [11],[10]) implies the existence and uniqueness of the solution $u(\boldsymbol{\beta}, \mathcal{L}) \in V_g$ to the variational problem:

$$B(\boldsymbol{\beta}; u(\boldsymbol{\beta}, \mathcal{L}), v) = \mathcal{L}(v), \ \forall v \in V_0. \tag{7}$$

We also introduce the standard *energy semi-norm* as

$$|v|^2_{E,\beta} \equiv B(\boldsymbol{\beta}; v, v), \forall v \in V.$$

Observe that this semi-norm depends on the choice of the coefficient $\boldsymbol{\beta}$. Moreover, for each $\boldsymbol{\beta} \in \mathcal{A}_\beta$, its restriction to the subspace V_0, is in fact equivalent to the norm $\| \cdot \|_V$ thanks to the uniform coerciveness and continuity assumptions (5)–(6) on the form B. In the light of problem (2), we consider in this work parameters $\boldsymbol{\eta}$ of the form $\boldsymbol{\eta} = [\boldsymbol{\beta}, \mathcal{L}]$ and corresponding sets $\mathcal{A}_\eta = \mathcal{A}_\beta \times \mathcal{A}_\mathcal{L}$, that is, we allow perturbations in the coefficients and in the loads of the equation to solve.

2.1 Perturbation technique for uncertainty quantification

In presence of small uncertainty in the parameters $\boldsymbol{\eta}$, we can employ a perturbation analysis around the nominal value $\boldsymbol{\eta}_0$ to estimate the interval ΔQ in (2). Under the assumption that the quantity of interest Q (and a fortiori the bilinear form B) is a regular function of the parameters $\boldsymbol{\eta}$, we can introduce the Fréchet derivative $D_\eta \psi(\boldsymbol{\eta})$ of the functional $\psi(\boldsymbol{\eta})$, as defined in (1), on the space of perturbations W, evaluated at the point $\boldsymbol{\eta}$:

$$< D_\eta \psi(\boldsymbol{\eta}), \delta\boldsymbol{\eta} >= \lim_{s \to 0^+} \frac{\psi(\boldsymbol{\eta} + s\delta\boldsymbol{\eta}) - \psi(\boldsymbol{\eta})}{s}, \quad \forall \delta\boldsymbol{\eta} \in W$$

and the second order Fréchet derivative $D^2_\eta \psi(\boldsymbol{\eta})$ (bilinear form on W) in $\boldsymbol{\eta}$:

$$D^2_\eta \psi(\boldsymbol{\eta})(\delta\boldsymbol{\eta}_1, \delta\boldsymbol{\eta}_2) = \lim_{s \to 0^+} < \frac{D_\eta \psi(\boldsymbol{\eta} + s\delta\boldsymbol{\eta}_2) - D_\eta \psi(\boldsymbol{\eta})}{s}, \delta\boldsymbol{\eta}_1 > .$$

Then, for all $\boldsymbol{\eta} = \boldsymbol{\eta}_0 + \delta\boldsymbol{\eta} \in \mathcal{A}_\eta$ the following expansion holds: there exists $\theta \in (0, 1)$ such that

$$\psi(\boldsymbol{\eta}) - \psi(\boldsymbol{\eta}_0) =< D_\eta \psi(\boldsymbol{\eta}_0), \delta\boldsymbol{\eta} > + \frac{1}{2} D^2_\eta \psi(\boldsymbol{\eta}_0 + \theta\delta\boldsymbol{\eta})(\delta\boldsymbol{\eta}, \delta\boldsymbol{\eta}). \tag{8}$$

Based on expansion (8) we can give the following bound of the uncertainty interval introduced in (2):

$$\Delta Q \leq \underbrace{\sup_{\delta\boldsymbol{\eta} \in \mathcal{A}_\eta - \boldsymbol{\eta}_0} | < D_\eta \psi(\boldsymbol{\eta}_0), \delta\boldsymbol{\eta} > |}_{\text{Linear term } (\Delta Q^{lin})} + \underbrace{\frac{1}{2} \sup_{\delta\boldsymbol{\eta} \in \mathcal{A}_\eta - \boldsymbol{\eta}_0} \sup_{\theta \in (0,1)} |D^2_\eta \psi(\boldsymbol{\eta}_0 + \theta\delta\boldsymbol{\eta})(\delta\boldsymbol{\eta}, \delta\boldsymbol{\eta})|}_{\text{Remainder } (\mathcal{R})} . \tag{9}$$

As we will see in the following sections, in most of the cases, the first term of the expansion (hereafter ΔQ^{lin}) can be computed quite accurately, whereas the computation of the second term is, in general, unaffordable and only bounds or rough approximations of it can be provided. Yet, in presence of small uncertainty, we should expect the remainder to be much smaller than the leading linear term.

3 Uncertainty in the coefficients

We concentrate in this Section on the uncertainty in the coefficients $\boldsymbol{\beta} \in \mathcal{A}_\beta$ of the bilinear form $B(\boldsymbol{\beta}; \cdot, \cdot)$ from (7). The analysis of the load uncertainty is simpler, and is postponed to Section 4. Besides, for the sake of readibility, in this section we will denote the set of coefficients by $\mathcal{A}$ instead of $\mathcal{A}_\beta$.

3.1 Computation of the leading order term

Since we are interested in the type of problems introduced in Section 1, we will restrict ourselves to perturb the coefficients of the partial differential equations in $L^\infty(D)$. Moreover, we will assume that the coefficients are uncorrelated. That is to say, at each point x of the domain, the set of coefficients is a hypercube of the form $\prod_{i=1}^m [\beta_i^{min}(x), \beta_i^{max}(x)]$, where m is the number of coefficients. The set of coefficients we will consider hereafter is

$$\mathcal{A} = \{\boldsymbol{\beta} \in [L^\infty(D)]^m, \;\; |\beta_i(x) - \beta_{0i}(x)| \leq \delta_i(x), \;\forall x \in D\} \tag{10}$$

with the assumption that the functions $\delta_i(x)$, $i = 1, \ldots, m$, are piecewise smooth.

For the purpose of characterizing the term ΔQ^{lin} we first introduce the *influence function* – sometimes also called adjoint or dual solution – $\varphi(\boldsymbol{\beta}) \in V_0$, associated to the quantity of interest Q, as the solution of the dual problem

$$\forall \boldsymbol{\beta} \in \mathcal{A}, \;\text{find } \varphi(\boldsymbol{\beta}) \in V_0 \;\text{ such that }\quad B(\boldsymbol{\beta}; v, \varphi) = Q(\boldsymbol{\beta}, v), \quad \forall v \in V_0. \tag{11}$$

The use of the dual problem to compute the gradient of the quantity of interest, Q, with respect to the coefficients $\boldsymbol{\beta}$ is a standard technique in the areas of sensitivity analysis [12], optimal control [19,31] and a posteriori error estimation [2,6,7]. Under the assumption that Q is a linear bounded functional on V, the solution $\varphi(\boldsymbol{\beta})$ exists and is unique in V_0.

Next, we define the following derivative:

– $\forall v, w \in V, \;\forall \boldsymbol{\beta} \in \mathcal{A}, \quad D_\beta B(\boldsymbol{\beta}; v, w) \in W'$ is such that

$$< D_\beta B(\boldsymbol{\beta}; v, w), \delta\boldsymbol{\beta} >= \lim_{s\to 0^+} \frac{1}{s} \left[B(\boldsymbol{\beta} + s\delta\boldsymbol{\beta}; v, w) - B(\boldsymbol{\beta}; v, w)\right], \quad \forall \delta\boldsymbol{\beta} \in W.$$

In typical boundary value problems, the bilinear form B has an integral representation

$$B(\boldsymbol{\beta}; v, w) = \int_D G(\boldsymbol{\beta}; v, w)\, d\mathbf{x} \tag{12}$$

for some function $G \in L^1(D)$ that depends on $\boldsymbol{\beta}, v, w$ and their derivatives. At least this is the case for the examples where only Neumann or Dirichlet-type boundary conditions are imposed. A boundary term should be added in the previous representation if Robin boundary conditions are considered. However, observe that in most of the practical cases this boundary term does not depend on the coefficient $\boldsymbol{\beta}$ and therefore, does not contribute to $D_\beta B$. The regularity assumptions under which the derivative of the bilinear form is well defined are:

Assumption 1 *For all $v, w \in V$ and all $\boldsymbol{\beta} \in \mathcal{A}$,*

$$\bullet\ \nabla_\beta G(\boldsymbol{\beta}; v, w) \in [L^1(D)]^m, \qquad \bullet\ \textit{for } i = 1, \dots, m, \quad \exists\, C_i(\boldsymbol{\beta}) > 0 \ \textit{s.t.} \tag{13a}$$

$$\int_D \left| \frac{\partial G}{\partial \beta_i}(\boldsymbol{\beta}; v, w) \right| d\mathbf{x} \le C_i(\boldsymbol{\beta}) |v|_{E,\beta} |w|_{E,\beta}, \tag{13b}$$

$$\bullet\ C_i^{sup} = \sup_{\boldsymbol{\beta} \in \mathcal{A}} C_i(\boldsymbol{\beta}) < \infty. \tag{13c}$$

Consequently,

$$< D_\beta B(\boldsymbol{\beta}; v, w), \delta\boldsymbol{\beta} >= \int_D \nabla_\beta G(\boldsymbol{\beta}; v, w) \cdot \delta\boldsymbol{\beta}\, d\mathbf{x},$$

and, for any $\boldsymbol{\beta} \in \mathcal{A}$ and $\delta\boldsymbol{\beta} \in W$, the quantity $< D_\beta B(\boldsymbol{\beta}; v, w), \delta\boldsymbol{\beta} >$ defines a bounded bilinear form on $V \times V$. In all the examples presented in Section 1, these regularity assumptions are satisfied. We will come back to this point later on in the text. In addition, we consider functionals Q of the form

$$Q(\boldsymbol{\beta}; v) = \int_D F(\boldsymbol{\beta}; v) d\mathbf{x} + \int_{\Gamma_Q \subseteq \Gamma_N} H(v) dA. \tag{14}$$

The function F depends on $\boldsymbol{\beta}$, v and its gradient and satisfies:

Assumption 2 *For all $v \in V$ and all $\boldsymbol{\beta} \in \mathcal{A}$*

$$\bullet\ \nabla_\beta F(\boldsymbol{\beta}; v) \in [L^1(D)]^m, \qquad \bullet\ \textit{for } i = 1, \dots, m, \quad \exists\, \bar{C}_i(\boldsymbol{\beta}) > 0 \ \textit{s.t.} \tag{15a}$$

$$\int_D \left| \frac{\partial F}{\partial \beta_i}(\boldsymbol{\beta}; v) \right| d\mathbf{x} \le \bar{C}_i(\boldsymbol{\beta}) |v|_{E,\beta}, \tag{15b}$$

$$\bullet\ \bar{C}_i^{sup} = \sup_{\boldsymbol{\beta} \in \mathcal{A}} \bar{C}_i(\boldsymbol{\beta}) < \infty. \tag{15c}$$

Similarly, we assume that for all $v \in V$ the function $H(v) \in L^1(\Gamma_Q)$. Observe that the quantities of interest introduced in the examples of Section 1 can be written in the form (14). The assumptions (15) yield

$$< D_\beta Q(\boldsymbol{\beta}; v), \delta\boldsymbol{\beta} >= \int_D \nabla_\beta F(\boldsymbol{\beta}; v) \cdot \delta\boldsymbol{\beta}\, d\mathbf{x}.$$

The purpose of the next example is to provide intuition on the constants introduced in assumptions (13) and (15).

Application 1 *Consider the scalar diffusion case introduced in Example 1. Here,* $B(\beta; v, w) = \int_D \beta \nabla v \cdot \nabla w$ *and therefore,* $\nabla_\beta G(\beta; v, w) = \nabla v \cdot \nabla w$*. Thus, bounding*

$$\int_D |\nabla_\beta G| \leq \|1/\beta\|_{L^\infty(D)} |v|_{E,\beta} |w|_{E,\beta},$$

we can take $C(\beta) = \|1/\beta\|_{L^\infty(D)}$*. On the other hand, with the quantity of interest* $Q(\beta; v) = \int_D \beta \nabla v \cdot \boldsymbol{\gamma}$*, with* $\boldsymbol{\gamma} \in (L^2(D))^3$*, we have* $\nabla_\beta F = \nabla v \cdot \boldsymbol{\gamma}$ *and*

$$\int_D |\nabla_\beta F| \leq \|1/\beta\|_{L^\infty(D)} \left(\int_D \beta |\boldsymbol{\gamma}|^2 d\mathbf{x} \right)^{1/2} |v|_{E,\beta},$$

so we can choose $\bar{C}(\beta) = \|1/\beta\|_{L^\infty(D)} \sqrt{\int_D \beta |\boldsymbol{\gamma}|^2 d\mathbf{x}}$*.*

We can introduce as well the derivative of the solution $u(\boldsymbol{\beta})$ of the primal problem (respectively $\varphi(\boldsymbol{\beta})$ of the dual problem), with respect to $\boldsymbol{\beta}$:

– $\forall \boldsymbol{\beta} \in \mathcal{A}, \quad D_\beta u(\boldsymbol{\beta}) : W \to V_0,$

$$D_\beta u(\boldsymbol{\beta})(\delta \boldsymbol{\beta}) = \lim_{s \to 0^+} \frac{1}{s} \left[u(\boldsymbol{\beta} + s \delta \boldsymbol{\beta}) - u(\boldsymbol{\beta}) \right], \qquad \forall \delta \boldsymbol{\beta} \in W.$$

It is easy to show that $D_\beta u(\boldsymbol{\beta})(\delta \boldsymbol{\beta}) \in V_0$ –note the homogeneous boundary conditions satisfied by $D_\beta u(\boldsymbol{\beta})(\delta \boldsymbol{\beta})$– satisfies the following variational problem:

$$B(\boldsymbol{\beta}; D_\beta u(\boldsymbol{\beta})(\delta \boldsymbol{\beta}), v) = - < D_\beta B(\boldsymbol{\beta}; u(\boldsymbol{\beta}), v), \delta \boldsymbol{\beta} >, \quad \forall v \in V_0. \tag{16}$$

The regularity assumptions (13) allow to state that such derivative is well defined. The following Lemma recalls the characterization of the differential $D_\beta \psi$ in terms of the dual solution φ and provides a formula for the leading order term ΔQ^{lin} as well as a characterization of the worst perturbation $\delta \boldsymbol{\beta}^*$, i.e. a perturbation that leads to the supremum of ΔQ^{lin}:

Lemma 1 *Under assumptions (13) and (15), there holds:*

a)
$$< D_\beta \psi(\boldsymbol{\beta}), \delta \boldsymbol{\beta} > = < D_\beta Q(\boldsymbol{\beta}; u(\boldsymbol{\beta})) - D_\beta B(\boldsymbol{\beta}; u(\boldsymbol{\beta}), \varphi(\boldsymbol{\beta})), \delta \boldsymbol{\beta} >, \quad \forall \boldsymbol{\beta} \in \mathcal{A}, \ \forall \delta \boldsymbol{\beta} \in \mathcal{A} - \boldsymbol{\beta}, \tag{17}$$

b)
$$\Delta Q^{lin} = \sum_{i=1}^{m} \int_D \left| \frac{\partial (F - G)}{\partial \beta_i} (\boldsymbol{\beta}_0; u(\boldsymbol{\beta}_0), \varphi(\boldsymbol{\beta}_0)) \right| \delta_i \, d\mathbf{x}, \tag{18}$$

c) the worst perturbation is $\delta \beta_i^* = \delta_i \operatorname{sign} \left(\dfrac{\partial (F - G)}{\partial \beta_i} \right)$.

Proof By hypothesis, the functional Q is linear with respect to $u \in V$. Hence, $\forall \boldsymbol{\beta} \in \mathcal{A}$, and $\forall \delta \boldsymbol{\beta} \in \mathcal{A} - \boldsymbol{\beta}$, we have

$$\begin{aligned} < D_\beta \psi(\boldsymbol{\beta}), \delta \boldsymbol{\beta} > &= Q(\boldsymbol{\beta}; D_\beta u(\boldsymbol{\beta})(\delta \boldsymbol{\beta})) + < D_\beta Q(\boldsymbol{\beta}; u(\boldsymbol{\beta})), \delta \boldsymbol{\beta} > \\ &= B(\boldsymbol{\beta}; D_\beta u(\boldsymbol{\beta})(\delta \boldsymbol{\beta}), \varphi(\boldsymbol{\beta})) + < D_\beta Q(\boldsymbol{\beta}; u(\boldsymbol{\beta})), \delta \boldsymbol{\beta} > \\ &= - < D_\beta B(\boldsymbol{\beta}; u(\boldsymbol{\beta}), \varphi(\boldsymbol{\beta})), \delta \boldsymbol{\beta} > + < D_\beta Q(\boldsymbol{\beta}; u(\boldsymbol{\beta})), \delta \boldsymbol{\beta} > \end{aligned}$$

and this proves part a). Observe that the second equality holds because the function $D_\beta u(\boldsymbol{\beta})(\delta\boldsymbol{\beta}) \in V_0$ and therefore, it can be used as a test function in the dual problem (11). Next, to any $\delta\boldsymbol{\beta}(x) = \boldsymbol{\beta}(x) - \boldsymbol{\beta}_0(x)$ we can associate the function $\widetilde{\delta\boldsymbol{\beta}}$ such that

$$\widetilde{\delta\beta}_i(x) = \begin{cases} \delta\beta_i(x)/\delta_i(x), & \text{if } \delta_i(x) \neq 0 \\ 0 & \text{if } \delta_i(x) = 0. \end{cases}$$

Hence, $\|\widetilde{\delta\beta}_i\|_{L^\infty(D)} \leq 1$, and (denoting by $u_0 = u(\boldsymbol{\beta}_0)$ and $\varphi_0 = \varphi(\boldsymbol{\beta}_0)$)

$$\begin{aligned} \Delta Q^{lin} &= \sup_{\delta\boldsymbol{\beta}\in\mathcal{A}-\boldsymbol{\beta}_0} < D_\beta Q(\boldsymbol{\beta}_0; u_0) - D_\beta B(\boldsymbol{\beta}_0; u_0, \varphi_0), \delta\boldsymbol{\beta} > \\ &= \sum_{i=1}^{m} \sup_{\substack{\widetilde{\delta\beta}_i \in L^\infty \\ \|\widetilde{\delta\beta}_i\|_{L^\infty} \leq 1}} \int_D \frac{\partial(F-G)}{\partial\beta_i}(\boldsymbol{\beta}_0; u_0, \varphi_0)\widetilde{\delta\beta}_i \delta_i \, d\mathbf{x} = \sum_{i=1}^{m} \left\| \frac{\partial(F-G)}{\partial\beta_i}\delta_i \right\|_{L^1(D)} \end{aligned}$$

the supremum being achieved for

$$\widetilde{\delta\beta}_i = \operatorname{sign}\left(\frac{\partial(F-G)}{\partial\beta_i}\right)$$

and this completes the proof. □

Let us point out the main result given in Lemma 1: whenever the perturbation of the coefficients is of the form (10), the computation of the bound on the leading order term implies the solution of the primal and dual problems as well as the evaluation of L^1-norms of the functions $\partial(F-G)/\partial\beta_i$. Lemma 1 also provides a useful characterization of the *worst distribution* $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0 \pm \delta\boldsymbol{\beta}^*$ of the coefficients, that maximizes the uncertainty interval in the quantity of interest.

The fact that the evaluation of ΔQ^{lin} can be recast to evaluating L^1-norms of the functions $\partial(F-G)/\partial\beta_i$ is a consequence of the choice of L^∞ as space of perturbations. Conversely, the fact that the uncertainty interval is simply the sum of intervals associated to each coefficient is a consequence of the assumption that the coefficients are uncorrelated, i.e. at each point of the domain the set of coefficients is a hypercube in $\mathbb{R}^m$.

*Remark 1 (*More general perturbations: correlated coefficients*)* In general, the uncertain coefficients may be correlated. In such a case it seems more realistic to use convex polyhedra to describe the set of allowed perturbations in each position $x \in D$. Hence, the set of coefficients will take the form $\mathcal{A} = \{\delta\boldsymbol{\beta} \in [L^\infty(D)]^m, \ (\boldsymbol{\beta}_0 + \delta\boldsymbol{\beta})(x) \in A(x) \subset \mathbb{R}^m, \ \forall x \in D\}$. The set function A gives, at each point x, convex polyhedra that vary piecewise smoothly with respect to x. Their diameter represents the level of uncertainty at each point, with a similar role as the functions δ_i have in the uncorrelated case. We can build a computable approximation based on uncoupled standard linear programs: given a suitable partition $\mathcal{T}$ of the domain (typically related to a finite element mesh), on each element of which the set function A is smooth, and considering only piecewise constant perturbations $\delta\boldsymbol{\beta}_K$, the quantity ΔQ^{lin} can be approximated as

$$\Delta Q^{lin} \approx \sum_{K\in\mathcal{T}} \max_{\delta\boldsymbol{\beta}_K \in A(x_K)-\boldsymbol{\beta}_0(x_K)} \sum_{i=1}^{m} \delta\beta_K^{(i)} \int_K \frac{\partial(F-G)}{\partial\beta_i}(\boldsymbol{\beta}_0; u(\boldsymbol{\beta}_0), \varphi(\boldsymbol{\beta}_0))dx$$

where x_K is an arbitrary point in K. The previous expression implies the solution of a simple linear programming problem on each element of the partition.

*Remark 2 (*More general perturbations: smoother perturbations*)* Whenever a priori regularity on the coefficients is available, e.g. if we have physical information that tells us that a given coefficient has a bounded first derivative w.r.t. x, a penalized version of the first term on the right hand side of (9) may be useful, i.e. given an m-dimensional parameter $\boldsymbol{\rho} > 0$ we may choose,

$$\delta\boldsymbol{\beta}^* = \underset{\delta\boldsymbol{\beta}\in\mathcal{A}-\boldsymbol{\beta}_0}{\arg\max} \left\{ < D_\beta Q(\boldsymbol{\beta}_0; u_0) - D_\beta B(\boldsymbol{\beta}_0; u_0, \varphi_0), \delta\boldsymbol{\beta} > -\frac{1}{2}\|\delta\boldsymbol{\beta}\|^2_{H^1_\rho(D)} \right\}$$

with $\|\delta\boldsymbol{\beta}\|^2_{H^1_\rho(D)} = \sum_{i=1}^m \rho_i \|\delta\beta_i\|^2_{H^1_0(D)}$. Thus, the resulting optimization problem is no longer separable in x and we have to use a multilevel algorithm to solve it.

Extra sensitivity information can be extracted from the representation (17): let us introduce the *sensitivity function*

$$\alpha^{(i)} = \beta_{0i} \frac{\partial(F-G)}{\partial\beta_i}(\boldsymbol{\beta}_0; u(\boldsymbol{\beta}_0), \varphi(\boldsymbol{\beta}_0)), \qquad i = 1, \dots, m \tag{19}$$

relative to each coefficient β_i. The function $\alpha^{(i)}$ represents a density per unit volume and unit relative perturbation associated to the $i-th$ coefficient and allows us to quantify the regions of the domain in which a unitary relative perturbation of the coefficient has a large influence on the uncertainty interval of the quantity of interest. In other words, such function can be used to identify the regions where the coefficients should be measured sharply.

3.1.1 Finite elements approximation

As stated in Lemma 1, the computation of the leading order term ΔQ^{lin} implies the solution of the primal and dual problems for the choice $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ of the coefficients. In practice, the exact solutions $u(\boldsymbol{\beta}_0)$ and $\varphi(\boldsymbol{\beta}_0)$ are not accessible and only approximations of them will be available. Let us indicate by $u_h(\boldsymbol{\beta}_0)$ and $\varphi_h(\boldsymbol{\beta}_0)$ some suitable finite element approximations to the exact solutions and by ΔQ_h^{lin} the approximation of ΔQ^{lin} based upon these finite elements solutions, that is

$$\Delta Q_h^{lin} = \sum_{i=1}^m \left\| \frac{\partial(F-G)}{\partial\beta_i}(\boldsymbol{\beta}_0; u_h(\boldsymbol{\beta}_0), \varphi_h(\boldsymbol{\beta}_0))\delta_i \right\|_{L^1(D)}. \tag{20}$$

The following Lemma states that the quantity ΔQ_h^{lin} converges to the true value at the same rate as the finite element approximations $u_h(\boldsymbol{\beta}_0)$ and $\varphi_h(\boldsymbol{\beta}_0)$ in the energy norm.

Lemma 2 *Under the regularity assumptions (13) and (15), there exists a constant $C > 0$, independent of the discretization parameter h, such that*

$$|\Delta Q^{lin} - \Delta Q_h^{lin}| \le C \left(|u(\boldsymbol{\beta}_0) - u_h(\boldsymbol{\beta}_0)|_{E,\beta_0} + |\varphi(\boldsymbol{\beta}_0) - \varphi_h(\boldsymbol{\beta}_0)|_{E,\beta_0} \right). \tag{21}$$

Proof Let $\tilde{G} = F - G$. In this proof we will use the shorthand notation $u = u(\boldsymbol{\beta}_0)$ and $u_h = u_h(\boldsymbol{\beta}_0)$ (similarly for φ and φ_h).

We first recall the inequality $\forall a, b \in \mathbb{R} \quad \big| |a| - |b| \big| \leq |a - b|$. Then,

$$
\begin{aligned}
|\Delta Q^{lin} - \Delta Q_h^{lin}| &= \left| \sum_{i=1}^{m} \int_D \left(\left| \frac{\partial \tilde{G}}{\partial \beta_i}(\boldsymbol{\beta}_0; u, \varphi)\delta_i \right| - \left| \frac{\partial \tilde{G}}{\partial \beta_i}(\boldsymbol{\beta}_0; u_h, \varphi_h)\delta_i \right| \right) d\mathbf{x} \right| \\
&\leq \sum_{i=1}^{m} \int_D \left| \left|\frac{\partial \tilde{G}}{\partial \beta_i}(\boldsymbol{\beta}_0; u, \varphi)\right| - \left|\frac{\partial \tilde{G}}{\partial \beta_i}(\boldsymbol{\beta}_0; u_h, \varphi_h)\right| \right| \delta_i d\mathbf{x} \\
&\leq \sum_{i=1}^{m} \int_D \left| \frac{\partial \tilde{G}}{\partial \beta_i}(\boldsymbol{\beta}_0; u, \varphi) - \frac{\partial \tilde{G}}{\partial \beta_i}(\boldsymbol{\beta}_0; u_h, \varphi_h) \right| \delta_i d\mathbf{x} \\
&\leq \sum_{i=1}^{m} \int_D \left(\left| \frac{\partial \tilde{G}}{\partial \beta_i}(\boldsymbol{\beta}_0; u - u_h, \varphi) \right| + \left| \frac{\partial \tilde{G}}{\partial \beta_i}(\boldsymbol{\beta}_0; u_h, \varphi - \varphi_h) \right| \right) \delta_i d\mathbf{x} \\
&\leq \sum_{i=1}^{m} [C_i(\boldsymbol{\beta}_0) + \bar{C}_i(\boldsymbol{\beta}_0)] \big(|u - u_h|_{E,\beta_0} |\varphi|_{E,\beta_0} \\
&\quad + |u_h|_{E,\beta_0} |\varphi - \varphi_h|_{E,\beta_0} \big) \|\delta_i\|_{L^\infty(D)},
\end{aligned}
$$

where in the last inequality we have used (13) and (15). The proof now follows observing that both $|\varphi|_{E,\beta_0}$ and $|u_h|_{E,\beta_0}$ are bounded quantities. □

Once the finite element approximations to the primal and dual solutions are available, the practical computation of ΔQ_h^{lin} still involves some technicalities due to the non smoothness of the functions to integrate. Let $\mathcal{T}_h$ be the finite element mesh employed in the computation of $u_h(\boldsymbol{\beta}_0)$ and $\varphi_h(\boldsymbol{\beta}_0)$. We make the assumption that $\mathcal{T}_h$ is aligned with eventual discontinuities of the function δ_i, if any. Then, on each element $K \in \mathcal{T}_h$ we can reasonably assume that the functions $\frac{\partial(F-G)}{\partial \beta_i}(\boldsymbol{\beta}_0, u_h(\boldsymbol{\beta}_0), \varphi_h(\boldsymbol{\beta}_0))\delta_i(x)$ are regular enough, such that they can be integrated with high accuracy by a quadrature formula. Yet, the absolute value of those functions will not be smooth, in general, and may present surfaces of non differentiability. This feature demands for an adaptive quadrature algorithm.

3.1.2 Implementation of the Adaptive Quadrature Algorithm

This subsection presents a detailed implementation of an adaptive quadrature algorithm, applying the theory and implementation devised in the work [25].

The goal of the adaptive algorithm described below is to construct, starting from the mesh $\mathcal{T}_h$, a refined mesh $\mathcal{T}_{h_q}$, with $h_q \leq h$, such that the quadrature error in the computation of (20) is smaller than a given error tolerance, TOL > 0. Besides, for efficiency reasons, the contributions to the total error from each of the elements in the refined mesh are approximately equidistributed. To this end, start the adaptive quadrature with the initial mesh of size $h_q[1] = h$, where h is the mesh size used to compute u_h and φ_h, and then specify iteratively a new mesh $h_q[k+1]$, from $h_q[k]$, using the following dividing strategy:

for all elements $n = 1, 2, \ldots, N[k]$
compute the error indicator $\bar{r}_n[k]$
if $\bar{r}_n[k] > \dfrac{\mathrm{TOL}}{N[k]}$ **then**
mark element n for division
endif
endfor
divide all marked elements into 2^d uniform sub elements. (22)

Here the error indicator, $\bar{r}_n[k]$, corresponding to the element n on the k mesh is obtained using the difference between the quadrature formula with the current mesh-size and another with half the mesh-size. With this dividing strategy, we use the stopping criterion:

$$\textbf{if} \quad \left(\max_{1\le n\le N[k]} \bar{r}_n[k] \le S_1 \frac{\mathrm{TOL}}{N[k]} \right) \quad \textbf{then} \quad \text{stop}. \tag{23}$$

The dividing strategy (22) is applied iteratively until the approximate solution is sufficiently resolved, i.e. the elements satisfy the stopping criterion (23). The constant S_1 is defined following [25].

Remark 3 Once the adaptive algorithm stops, the final mesh is well suited to represent the sensitivity functions $\alpha^{(i)}$ and the corresponding worst distribution of the coefficients β_i^*, $i = 1, \ldots, m$ by means of piecewise constant functions. This representation is of practical important value to visualize the functions $\alpha^{(i)}$ and β_i^* and thus identify the regions of the domain D that most influence the uncertainty in the quantity of interest Q.

3.2 Computational bounds for the remainder

The goal of this section is to present estimates for the remainder of the bound in (9), i.e.

$$\mathcal{R} \equiv \frac{1}{2} \sup_{\delta\boldsymbol{\beta}\in\mathcal{A}-\boldsymbol{\beta}_0} \sup_{\theta\in(0,1)} |D_\beta^2 \psi(\boldsymbol{\beta}_0 + \theta\delta\boldsymbol{\beta})(\delta\boldsymbol{\beta}, \delta\boldsymbol{\beta})|. \tag{24}$$

The second derivative, $D_\beta^2\psi$ can be characterized by the identity

$$\begin{aligned} D_\beta^2\psi(\boldsymbol{\beta})(\delta\boldsymbol{\beta}, \delta\boldsymbol{\beta}) = & -D_\beta^2 B(\boldsymbol{\beta}; u(\boldsymbol{\beta}), \varphi(\boldsymbol{\beta}))(\delta\boldsymbol{\beta}, \delta\boldsymbol{\beta}) \\ & -2 < D_\beta B(\boldsymbol{\beta}; D_\beta u(\boldsymbol{\beta})(\delta\boldsymbol{\beta}), \varphi(\boldsymbol{\beta})), \delta\boldsymbol{\beta} > \\ & +2 < D_\beta Q(\boldsymbol{\beta}; D_\beta u(\boldsymbol{\beta})(\delta\boldsymbol{\beta})), \delta\boldsymbol{\beta} > \\ & +D_\beta^2 Q(\boldsymbol{\beta}; u(\boldsymbol{\beta}))(\delta\boldsymbol{\beta}, \delta\boldsymbol{\beta}). \end{aligned} \tag{25}$$

which follows directly by taking variations on equation a) from Lemma 1. Observe that the computation of the second derivative $D_\beta^2\psi(\delta\boldsymbol{\beta}, \delta\boldsymbol{\beta})$ in the direction $\delta\boldsymbol{\beta}$ depends on the knowledge of the first derivative $D_\beta u(\delta\boldsymbol{\beta})$. Since we would like to maximize $D^2\psi$ over all possible directions $\delta\boldsymbol{\beta}$, and we consider infinite dimensional sets of coefficients, we would have to know the value of $D_\beta u(\delta\boldsymbol{\beta})$ for all $\delta\boldsymbol{\beta}$ which is practically unfeasible. Hence, we derive computable bounds for the remainder based on a priori energy estimates for $D_\beta u$. Of course, the price to pay is that those bounds might be pessimistic in certain situations.

Besides, we now restrict ourselves to regular symmetric problems, assuming that

Assumption 3 *For each $\boldsymbol{\beta} \in \mathcal{A}$ and $1 \leq i, j \leq m$ there exist non-negative constants $C_{ij}(\boldsymbol{\beta}), C^{\#}_{ij}(\boldsymbol{\beta}) > 0$ such that*

$$\frac{|D^2_\beta B(\boldsymbol{\beta}; u, v)(\delta\boldsymbol{\beta}, \delta\boldsymbol{\beta})|}{|u|_{E,\beta}|v|_{E,\beta}} \leq \left(\sum_{i,j=1}^{m} C_{ij}(\boldsymbol{\beta}) \, \|\delta\beta_i\|_{W_i} \, \|\delta\beta_j\|_{W_j} \right), \quad \forall u, v \in V, \ \forall \delta\boldsymbol{\beta} \in \mathcal{A} - \boldsymbol{\beta}, \tag{26}$$

and

$$\frac{|D^2_\beta Q(\boldsymbol{\beta}; u)(\delta\boldsymbol{\beta}, \delta\boldsymbol{\beta})|}{|u|_{E,\beta}} \leq \left(\sum_{i,j=1}^{m} C^{\#}_{ij}(\boldsymbol{\beta}) \, \|\delta\beta_i\|_{W_i} \, \|\delta\beta_j\|_{W_j} \right), \quad \forall u \in V, \ \forall \delta\boldsymbol{\beta} \in \mathcal{A} - \boldsymbol{\beta}. \tag{27}$$

These assumptions are closely related to (13,15). Based on (25) and the previous assumption we derive a bound for $D^2_\beta \psi(\delta\boldsymbol{\beta}, \delta\boldsymbol{\beta})$, at least for the case of symmetric problems, based on the energy norm of the primal and dual solutions, namely

Lemma 3 *Under assumptions (13–15) and (26–27) there holds*

$$\frac{|D^2_\beta \psi(\boldsymbol{\beta})(\delta\boldsymbol{\beta}, \delta\boldsymbol{\beta})|}{|u(\boldsymbol{\beta})|_{E,\beta}} \leq |\varphi(\boldsymbol{\beta})|_{E,\beta} \sum_{i,j=1}^{m} (C_{ij}(\boldsymbol{\beta}) + 2C_i(\boldsymbol{\beta})C_j(\boldsymbol{\beta})) \, \|\delta\beta_i\|_{W_i} \, \|\delta\beta_j\|_{W_j}$$
$$+ \sum_{i,j=1}^{m} (C^{\#}_{ij}(\boldsymbol{\beta}) + 2\,\bar{C}_i(\boldsymbol{\beta})C_j(\boldsymbol{\beta}))\|\delta\beta_i\|_{W_i} \, \|\delta\beta_j\|_{W_j}$$

Proof Apply the triangle inequality to (25), then bound the first and the last terms by direct application of (26) and (27). Let us now bound the second term using (13):

$$|< D_\beta B(\boldsymbol{\beta}; D_\beta u(\boldsymbol{\beta})(\delta\boldsymbol{\beta}), \varphi(\boldsymbol{\beta})), \delta\boldsymbol{\beta} >|$$
$$\leq |D_\beta u(\boldsymbol{\beta})(\delta\boldsymbol{\beta})|_{E,\beta}|\varphi(\boldsymbol{\beta})|_{E,\beta} \left(\sum_{i=1}^{m} C_i(\boldsymbol{\beta}) \, \|\delta\beta_i\|_{W_i} \right).$$

Recall that $D_\beta u(\boldsymbol{\beta})(\delta\boldsymbol{\beta})$ satisfies (16). Take the particular test function $v = D_\beta u(\boldsymbol{\beta})(\delta\boldsymbol{\beta})$ in (16) yielding

$$|D_\beta u(\boldsymbol{\beta})(\delta\boldsymbol{\beta})|^2_{E,\beta} = - < D_\beta B(\boldsymbol{\beta}; u(\boldsymbol{\beta}), D_\beta u(\boldsymbol{\beta})(\delta\boldsymbol{\beta})), \delta\boldsymbol{\beta} >$$
$$\leq \left(\sum_{i=1}^{m} C_i(\boldsymbol{\beta}) \, \|\delta\beta_i\|_{W_i} \right) |D_\beta u(\boldsymbol{\beta})(\delta\boldsymbol{\beta})|_{E,\beta}|u(\boldsymbol{\beta})|_{E,\beta}.$$

Combine the last two inequalities to obtain the desired bound for the second term, i.e.

$$
\begin{aligned}
&| < D_\beta B(\boldsymbol{\beta}; D_\beta u(\boldsymbol{\beta})(\delta\boldsymbol{\beta}), \varphi(\boldsymbol{\beta})), \delta\boldsymbol{\beta} > | \\
&\leq |u(\boldsymbol{\beta})|_{E,\beta} |\varphi(\boldsymbol{\beta})|_{E,\beta} \left(\sum_{i=1}^{m} C_i(\boldsymbol{\beta}) \, \|\delta\beta_i\|_{W_i} \right)^2 .
\end{aligned}
$$

The bound for the third term is obtained analogously. This finishes the proof. □

The previous Lemma yields a bound for the remainder that is almost suitable for computations: if the dependence of the various constants on $\boldsymbol{\beta}$ is known, we just need to relate the energy norm of $u(\boldsymbol{\beta})$ and $\varphi(\boldsymbol{\beta})$ with the energy norms of $u(\boldsymbol{\beta}_0)$ and $\varphi(\boldsymbol{\beta}_0)$, which are in turn approximated by the computable energy norms of $u_h(\boldsymbol{\beta}_0)$ and $\varphi_h(\boldsymbol{\beta}_0)$. The next two Lemmas answer this point and Theorem 1 states the final result. Let us introduce a coercivity related assumption, i.e.

Assumption 4 *There exists $k(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) > 0$ such that for all $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathcal{A}$ the norm bound*

$$
|\cdot|_{E,\beta_1} \leq k(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) |\cdot|_{E,\beta_2}
$$

holds. Besides, introduce the notation $k^{sup}(\boldsymbol{\beta}_2) = \sup_{\boldsymbol{\beta}_1 \in \mathcal{A}} k(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$.

Observe that for the scalar diffusion case we have $k(\beta_0, \beta) = \sqrt{\| \frac{\beta_0}{\beta} \|_{L^\infty}}$ and $k^{sup}(\beta) = \sqrt{\| \frac{\beta_{max}}{\beta} \|_{L^\infty}}$, with $\beta_{max}(x) = \sup_{\beta \in \mathcal{A}} \beta(x)$.

Lemma 4 (Energy norm estimates) *Assume that (13,15,26,27) and Assumption 4 hold. Let*

$$
\mathcal{I}_k(\boldsymbol{\beta}_0, \boldsymbol{\beta}) \equiv \max \left\{ k(\boldsymbol{\beta}_0, \boldsymbol{\beta}) \, , \; k^{sup}(\boldsymbol{\beta}_0) \left(1 + k^{sup}(\boldsymbol{\beta}) \sum_{i=1}^{m} C_i^{sup} \, \|\delta\beta_i\|_{W_i} \right) \right\} .
$$

With the notation $u_0 = u(\boldsymbol{\beta}_0)$ and $\varphi_0 = \varphi(\boldsymbol{\beta}_0)$, there holds

$$
|u(\boldsymbol{\beta})|_{E,\beta} \leq \mathcal{I}_k(\boldsymbol{\beta}_0, \boldsymbol{\beta}) |u_0|_{E,\beta_0} \tag{28a}
$$

$$
|\varphi(\boldsymbol{\beta})|_{E,\beta} \leq k(\boldsymbol{\beta}_0, \boldsymbol{\beta}) \, |\varphi_0|_{E,\beta_0} + k^{sup}(\boldsymbol{\beta}) \sum_{i=1}^{m} \bar{C}_i^{sup} \|\delta\beta_i\|_{W_i} . \tag{28b}
$$

Proof To prove the first inequality in (28) introduce the auxiliary solutions with non homogeneous boundary conditions, $\tilde{g}(\boldsymbol{\beta}) \in V_g$, such that

$$
B(\boldsymbol{\beta}; \tilde{g}(\boldsymbol{\beta}), v) = 0, \;\; \forall v \in V_0.
$$

Thus, this orthogonality condition implies

$$
|u(\boldsymbol{\beta})|^2_{E,\beta} = |\tilde{g}(\boldsymbol{\beta})|^2_{E,\beta} + |u(\boldsymbol{\beta}) - \tilde{g}(\boldsymbol{\beta})|^2_{E,\beta} . \tag{29}
$$

To bound the first term, observe that using the mean value Theorem

$$
\begin{aligned}
|B(\boldsymbol{\beta}; \tilde{g}(\boldsymbol{\beta}) - \tilde{g}(\boldsymbol{\beta}_0), v)| &= |B(\boldsymbol{\beta}_0; \tilde{g}(\boldsymbol{\beta}_0), v) - B(\boldsymbol{\beta}; \tilde{g}(\boldsymbol{\beta}_0), v)| \\
&= \left| \int_0^1 < D_\beta B(\boldsymbol{\beta}_0 + \theta\, \delta\boldsymbol{\beta}, \tilde{g}(\boldsymbol{\beta}_0), v), \delta\boldsymbol{\beta}) > d\theta \right| \\
&\leq \sum_{i=1}^m \|\delta\beta_i\|_{W_i} \int_0^1 |\tilde{g}(\boldsymbol{\beta}_0)|_{E,\beta_0+\theta\delta\beta}\, |v|_{E,\beta_0+\theta\delta\beta}\, C_i(\boldsymbol{\beta}_0 + \theta\delta\boldsymbol{\beta})\, d\theta \\
&\leq k^{sup}(\boldsymbol{\beta}_0)\, |\tilde{g}(\boldsymbol{\beta}_0)|_{E,\beta_0}\, k^{sup}(\boldsymbol{\beta})\, |v|_{E,\beta} \sum_{i=1}^m \|\delta\beta_i\|_{W_i} \int_0^1 C_i(\boldsymbol{\beta}_0 + \theta\delta\boldsymbol{\beta})\, d\theta \\
&\leq k^{sup}(\boldsymbol{\beta}_0)\, k^{sup}(\boldsymbol{\beta})\, |\tilde{g}(\boldsymbol{\beta}_0)|_{E,\beta_0}\, |v|_{E,\beta} \sum_{i=1}^m C_i^{sup}\, \|\delta\beta_i\|_{W_i}.
\end{aligned}
$$

Let $v = \tilde{g}(\boldsymbol{\beta}) - \tilde{g}(\boldsymbol{\beta}_0) \in V_0$ in the previous equation. This implies

$$
\begin{aligned}
|\tilde{g}(\boldsymbol{\beta})|_{E,\beta} &\leq |\tilde{g}(\boldsymbol{\beta}_0)|_{E,\beta} + |\tilde{g}(\boldsymbol{\beta}) - \tilde{g}(\boldsymbol{\beta}_0)|_{E,\beta} \\
&\leq |\tilde{g}(\boldsymbol{\beta}_0)|_{E,\beta_0}\, k^{sup}(\boldsymbol{\beta}_0) \left(1 + k^{sup}(\boldsymbol{\beta}) \sum_{i=1}^m C_i^{sup}\, \|\delta\beta_i\|_{W_i} \right). \qquad (30)
\end{aligned}
$$

Next, derive a bound for $|u(\boldsymbol{\beta}) - \tilde{g}(\boldsymbol{\beta})|_{E,\beta}$ observing that by construction of the auxiliary solutions $\tilde{g}(\boldsymbol{\beta})$, we have

$$B(\boldsymbol{\beta}; u(\boldsymbol{\beta}) - \tilde{g}(\boldsymbol{\beta}), v) = B(\boldsymbol{\beta}_0; u(\boldsymbol{\beta}_0) - \tilde{g}(\boldsymbol{\beta}_0), v), \forall v \in V.$$

Take $v = u(\boldsymbol{\beta}) - \tilde{g}(\boldsymbol{\beta})$ in the previous identity, apply Cauchy Schwartz to get

$$
\begin{aligned}
|u(\boldsymbol{\beta}) - \tilde{g}(\boldsymbol{\beta})|^2_{E,\beta} &\leq |u(\boldsymbol{\beta}_0) - \tilde{g}(\boldsymbol{\beta}_0)|_{E,\beta_0} |u(\boldsymbol{\beta}) - \tilde{g}(\boldsymbol{\beta})|_{E,\beta_0} \\
&\leq |u(\boldsymbol{\beta}_0) - \tilde{g}(\boldsymbol{\beta}_0)|_{E,\beta_0} k(\boldsymbol{\beta}_0, \boldsymbol{\beta}) |u(\boldsymbol{\beta}) - \tilde{g}(\boldsymbol{\beta})|_{E,\beta}.
\end{aligned}
$$

Combine the previous estimate with (29) and (30) to obtain

$$
\begin{aligned}
|u(\boldsymbol{\beta})|^2_{E,\beta} &= |\tilde{g}(\boldsymbol{\beta})|^2_{E,\beta} + |u(\boldsymbol{\beta}) - \tilde{g}(\boldsymbol{\beta})|^2_{E,\beta} \\
&\leq \max \left\{ k^2(\boldsymbol{\beta}_0, \boldsymbol{\beta}),\, \left[k^{sup}(\boldsymbol{\beta}_0) \left(1 + k^{sup}(\boldsymbol{\beta}) \sum_{i=1}^m C_i^{sup}\, \|\delta\beta_i\|_{W_i} \right) \right]^2 \right\} |u_0|^2_{E,\beta_0}
\end{aligned}
$$

from which the first inequality in (28) follows. Let us prove now the second inequality in (28). By the definition of the dual problem (11) we have

$$B(\boldsymbol{\beta}; v, \varphi(\boldsymbol{\beta})) = B(\boldsymbol{\beta}_0; v, \varphi_0) + Q(\boldsymbol{\beta}; v) - Q(\boldsymbol{\beta}_0; v),\ \forall v \in V.$$

Take $v = \varphi(\boldsymbol{\beta})$ in the above. Then, apply the triangle and Cauchy-Schwartz inequalities arriving at

$$|\varphi(\boldsymbol{\beta})|^2_{E,\beta} \leq |\varphi(\boldsymbol{\beta})|_{E,\beta_0} |\varphi_0|_{E,\beta_0} + |Q(\boldsymbol{\beta}; \varphi(\boldsymbol{\beta})) - Q(\boldsymbol{\beta}_0; \varphi(\boldsymbol{\beta}))|.$$

Now use the mean value theorem and (15) to bound the difference of functionals, yielding the second inequality in (28). □

Theorem 1 *Under the same assumptions and notation of Lemma 4, let* $\bar{\delta}_i \equiv \sup_{\delta\boldsymbol{\beta}\in\mathcal{A}-\boldsymbol{\beta}_0} \|\delta\beta_i\|_{W_i}$, $i = 1, \dots, m$. *There holds*

$$\begin{aligned}\mathcal{R} \leq \frac{|u_0|_{E,\beta_0}}{2} \sup_{\boldsymbol{\beta}\in\mathcal{A}} \Bigg\{ \mathcal{I}_k(\boldsymbol{\beta}_0, \boldsymbol{\beta}) \Bigg[\sum_{i,j=1}^{m} (C^{\#}_{ij}(\boldsymbol{\beta}) + 2\bar{C}_i(\boldsymbol{\beta}) C_j(\boldsymbol{\beta}))\, \bar{\delta}_i\, \bar{\delta}_j \\ + \left(k(\boldsymbol{\beta}_0, \boldsymbol{\beta})\, |\varphi_0|_{E,\beta_0} + k^{sup}(\boldsymbol{\beta}) \sum_{i=1}^{m} \bar{C}_i^{sup} \bar{\delta}_i \right) \\ \times \sum_{i,j=1}^{m} (C_{ij(\boldsymbol{\beta})} + 2C_i(\boldsymbol{\beta}) C_j(\boldsymbol{\beta}))\, \bar{\delta}_i\, \bar{\delta}_j \Bigg] \Bigg\}.\end{aligned} \tag{31}$$

Proof Recall (24). Since

$$\sup_{\delta\boldsymbol{\beta}\in\mathcal{A}-\boldsymbol{\beta}_0} \sup_{\theta\in(0,1)} |D^2_\beta \psi(\boldsymbol{\beta}_0 + \theta\delta\boldsymbol{\beta})(\delta\boldsymbol{\beta}, \delta\boldsymbol{\beta})| \leq \sup_{\boldsymbol{\beta}\in\mathcal{A}} \sup_{\delta\boldsymbol{\beta}\in\mathcal{A}-\boldsymbol{\beta}_0} |D^2_\beta \psi(\boldsymbol{\beta})(\delta\boldsymbol{\beta}, \delta\boldsymbol{\beta})|$$

it is enough to apply Lemma 3 and Lemma 4 to bound

$$\begin{aligned}|D^2_\beta \psi(\boldsymbol{\beta})(\delta\boldsymbol{\beta}, \delta\boldsymbol{\beta})| &\leq |u(\boldsymbol{\beta})|_{E,\beta} \sum_{i,j=1}^{m} (C^{\#}_{ij} + 2\bar{C}_i C_j)(\boldsymbol{\beta}) \|\delta\beta_i\|_{W_i} \|\delta\beta_j\|_{W_j} \\ &+ |u(\boldsymbol{\beta})|_{E,\beta} |\varphi(\boldsymbol{\beta})|_{E,\beta} \sum_{i,j=1}^{m} (C_{ij} + 2C_i C_j)(\boldsymbol{\beta})\, \|\delta\beta_i\|_{W_i} \|\delta\beta_j\|_{W_j} \\ &\leq \mathcal{I}_k(\boldsymbol{\beta}_0, \boldsymbol{\beta}) |u_0|_{E,\beta_0} \Bigg\{ \sum_{i,j=1}^{m} (C^{\#}_{ij} + 2\bar{C}_i C_j)(\boldsymbol{\beta}) \|\delta\beta_i\|_{W_i} \|\delta\beta_j\|_{W_j} \\ &+ \left(k(\boldsymbol{\beta}_0, \boldsymbol{\beta})\, |\varphi_0|_{E,\beta_0} + k^{sup}(\boldsymbol{\beta}) \sum_{i=1}^{m} \bar{C}_i^{sup} \|\delta\beta_i\|_{W_i} \right) \\ &\times \sum_{i,j=1}^{m} (C_{ij} + 2C_i C_j)(\boldsymbol{\beta})\, \|\delta\beta_i\|_{W_i} \|\delta\beta_j\|_{W_j} \Bigg\}.\end{aligned}$$

Now use the bounds for $\|\delta\beta_i\|_{W_i}$ to finish the proof. □

Remark 4 The previous estimate shows as expected that the remainder behaves like $\mathcal{O}(\delta^2)$, where δ represents the size of absolute perturbations in the coefficient $\boldsymbol{\beta}$. In cases where either $\boldsymbol{\beta}_0$ or δ are not constant over the domain D, sharper estimates may be obtained by means of relative perturbations, see Application 2 at the end of this Section.

*Remark 5 (*Particular cases*)* If the solution $u(\boldsymbol{\beta})$ has homogeneous Dirichlet boundary conditions the result in the last Proposition reduces to

$$\frac{|\mathcal{R}|}{|u(\boldsymbol{\beta}_0)|_{E,\beta_0}} \le \frac{1}{2} \sup_{\boldsymbol{\beta}\in\mathcal{A}} \left\{ k(\boldsymbol{\beta}_0,\boldsymbol{\beta}) \Big\{ \sum_{i,j=1}^{m} (C^{\#}_{ij} + 2\bar{C}_i C_j)(\boldsymbol{\beta})\bar{\delta}_i\,\bar{\delta}_j \right.$$
$$\left. + \left(k(\boldsymbol{\beta}_0,\boldsymbol{\beta})\;|\varphi_0|_{E,\beta_0} + \sum_{i=1}^{m} \bar{C}_i^{sup}\bar{\delta}_i \right) \sum_{i,j=1}^{m} (C_{ij} + 2C_i C_j)(\boldsymbol{\beta})\,\bar{\delta}_i\,\bar{\delta}_j \Big\} \right\}. \quad (32)$$

If in addition the functional Q does not depend on $\boldsymbol{\beta}$ explicitly, then

$$\frac{|\mathcal{R}|}{|u_0|_{E,\beta_0}|\varphi_0|_{E,\beta_0}} \le \frac{1}{2} \sup_{\boldsymbol{\beta}\in\mathcal{A}} \left\{ (k(\boldsymbol{\beta}_0,\boldsymbol{\beta}))^2 \sum_{i,j=1}^{m} (C_{ij} + 2C_i C_j)(\boldsymbol{\beta})\,\bar{\delta}_i\,\bar{\delta}_j \right\}. \quad (33)$$

If in the general case we disregard the third order terms in δ_i, $i = 1, \dots, m$, we get instead

$$\frac{|\mathcal{R}|}{|u_0|_{E,\beta_0}} \le \frac{1}{2} \sup_{\boldsymbol{\beta}\in\mathcal{A}} \left\{ \mathcal{I}_k(\boldsymbol{\beta}_0,\boldsymbol{\beta}) \Big\{ \sum_{i,j=1}^{m} (2C^{\#}_{ij} + \bar{C}_i C_j)(\boldsymbol{\beta})\bar{\delta}_i\,\bar{\delta}_j \right.$$
$$\left. + k(\boldsymbol{\beta}_0,\boldsymbol{\beta})\;|\varphi_0|_{E,\beta_0} \sum_{i,j=1}^{m} (C_{ij} + 2C_i C_j)(\boldsymbol{\beta})\,\bar{\delta}_i\,\bar{\delta}_j \Big\} \right\} + \text{h.o.t.}$$

*Remark 6 (*Approximation of the remainder*)* In some practical cases like the elasticity equations for nearly incompressible materials the bounds presented in this section which rely mostly upon Cauchy-Schwartz and triangle inequalities might be too pessimistic. In those cases and for general unsymmetric problems, we suggest to approximate the remainder by computing the second order term along the perturbation in the coefficient $\boldsymbol{\beta}$, $\delta\boldsymbol{\beta}^*$, that gives the largest first order perturbation, $\delta\boldsymbol{\beta}^*$. In other words, use identity (25) to compute the approximation

$$\mathcal{R} \approx \frac{1}{2} D^2_{\beta}\psi(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}^*, \delta\boldsymbol{\beta}^*) \quad (34)$$

with

$$\delta\boldsymbol{\beta}^* = \underset{\delta\boldsymbol{\beta}\in\mathcal{A}-\boldsymbol{\beta}_0}{\arg\max} \; < D_{\beta}\psi(\boldsymbol{\beta}_0), \delta\boldsymbol{\beta} >$$

Observe that this procedure entails the computation of an extra solution, $D_{\beta}u(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}^*)$, that satisfies (16) and we no longer guarantee to have an upper bound on the reminder. The quantity in (34) should only be taken as an indication of the order of magnitude of $\mathcal{R}$.

Application 2 (Scalar diffusion) *Similarly as in the previous results, we may also bound the remainder using the relative size of the perturbation* $\|\delta\beta/\beta\|_{L^\infty(D)}$, *instead of the absolute one, and changing the constants accordingly. In the case of the scalar diffusion equation, let* $\beta_{min}(x) = \inf_{\beta\in\mathcal{A}} \beta(x)$ *and* $\beta_{max}(x) = \sup_{\beta\in\mathcal{A}} \beta(x)$, *for* $x \in D$. *Then, for the quantity of interest* $Q(\beta; v) = \int_D \beta\nabla v \cdot \boldsymbol{\gamma}$ *and homogeneous Dirichlet boundary conditions, the constants become:*

$$C_i(\beta) = C_i^{sup} = 1 \qquad C_{ij}(\beta) = 0$$

$$\bar{C}_i(\beta) = \sqrt{\int_D \beta|\boldsymbol{\gamma}|^2\, d\mathbf{x}}, \qquad \bar{C}_i^{sup} = \sqrt{\int_D \beta_{max}|\boldsymbol{\gamma}|^2\, d\mathbf{x}}, \qquad C_{ij}^{\#} = 0$$

$$k(\beta_1,\beta_2) = \left\|\frac{\beta_1}{\beta_2}\right\|_{L^\infty(D)}^{\frac{1}{2}} \qquad k^{sup}(\beta) = \left\|\frac{\beta_{max}}{\beta}\right\|_{L^\infty(D)}^{\frac{1}{2}}.$$

and the reminder, upon (32), *can be bound as*

$$|\mathcal{R}| \le |u_0|_{E,\beta_0} \sup_{\beta\in\mathcal{A}} \sup_{\delta\beta\in\mathcal{A}-\beta_0} \left\{ \left\|\frac{\beta_0}{\beta}\right\|_{L^\infty(D)}^{\frac{1}{2}} \left(\sqrt{\int_D \beta|\boldsymbol{\gamma}|^2} + |\varphi_0|_{E,\beta_0}\right) + \left\|\frac{\delta\beta}{\beta}\right\|_{L^\infty(D)} \sqrt{\int_D \beta_{max}|\boldsymbol{\gamma}|^2} \right\} \left\|\frac{\delta\beta}{\beta}\right\|_{L^\infty(D)}^2.$$

We now introduce the maximum relative perturbation

$$\epsilon = \sup_{\delta\beta\in\mathcal{A}-\beta_0} \|\delta\beta/\beta_0\|_{L^\infty(D)}.$$

Therefore,

$$\left\|\frac{\delta\beta}{\beta}\right\|_{L^\infty(D)} \le \epsilon \left\|\frac{\beta_0}{\beta}\right\|_{L^\infty(D)}$$

and a computable bound for the reminder is

$$|\mathcal{R}| \le |u_0|_{E,\beta_0} \left\{ \left(1 + \left\|\frac{\beta_0}{\beta_{min}}\right\|_{L^\infty}^{\frac{1}{2}} \epsilon\right) \sqrt{\int_D \beta_{max}|\boldsymbol{\gamma}|^2} + |\varphi_0|_{E,\beta_0} \right\} \left\|\frac{\beta_0}{\beta_{min}}\right\|_{L^\infty}^{\frac{5}{2}} \epsilon^2. \tag{35}$$

4 Uncertainty in the forcing terms

In this section we analyze the effect on the quantity of interest Q, of the uncertainty in the forcing term of problem (7). We assume that the right hand side in (7) is the sum of several forcing terms defined on different regions of the domain or the boundary, i.e.

$$\mathcal{L}(v) = \sum_{i=1}^{m_1} \mathcal{L}_i(v), \qquad \forall v \in V_0.$$

The nominal value of the forcing terms will be denoted by $\mathcal{L}_{0i}$, $i = 1, \dots, m_1$. Let us introduce, now, a sequence of Banach spaces $W_i \supseteq V_0$, $i = 1, \dots, m_1$, equipped with the norms $\|\cdot\|_{W_i}$. We assume that the perturbations $\delta\mathcal{L}_i$ belong to the spaces $W_i' \subset V_0'$ and that V_0 is continuously embedded in W_i, i.e.

Assumption 5 *For $i = 1, \dots, m_1$ and for any $\boldsymbol{\beta} \in \mathcal{A}_\beta$, there exist constants $\check{C}(\boldsymbol{\beta})_i > 0$, depending also on D, V_0 and W_i, such that*

$$\|\cdot\|_{W_i} \leq \check{C}_i(\boldsymbol{\beta})|\cdot|_{E,\beta}. \tag{36}$$

Then, we consider the *set of admissible loads*

$$\mathcal{A}_\mathcal{L} = \bigcup_{i=1}^{m_1} \mathcal{A}_\mathcal{L}^i, \qquad \mathcal{A}_\mathcal{L}^i \equiv \{\mathcal{L}_i \in V_0',\ \ \delta\mathcal{L}_i = \mathcal{L}_i - \mathcal{L}_{0i} \in W_i',\ \ \|\delta\mathcal{L}_i\|_{W_i'} \leq \epsilon_i\} \tag{37}$$

Typical choices for the spaces W_i are $W_i = L^{q_i}$, $1 \leq q_i < \infty$ (hence $W_i' = L^{p_i}$, with $1 = 1/p_i + 1/q_i$).

To highlight the dependence of the quantity of interest Q on the forcing term, we introduce the notation $u(\mathcal{L})$ to indicate the solution to problem (7) and $\psi(\mathcal{L}) = Q(u(\mathcal{L}))$ to indicate the quantity of interest. Once more, our aim is to quantify the uncertainty interval

$$\Delta Q = \sup_{\mathcal{L}\in\mathcal{A}_\mathcal{L}} |\psi(\mathcal{L}) - \psi(\mathcal{L}_0)|.$$

The quantification of ΔQ, in this case, is much simpler than in the case of perturbations in the coefficients of the bilinear form in (7). Indeed, the solution $u(\mathcal{L})$ of (7) is an affine function of the forcing term and the quantity of interest Q does not depend explicitly on $\mathcal{L}$. The following Theorem gives the characterization of ΔQ in terms of the solution φ of the dual problem (11) (which does not depend on $\mathcal{L}$, either).

Theorem 2 *With the load perturbations* (37)*, we have*

a) $$\psi(\mathcal{L}) - \psi(\mathcal{L}_0) = \mathcal{L}(\varphi) - \mathcal{L}_0(\varphi), \qquad \forall \mathcal{L} \in \mathcal{A}_\mathcal{L}. \tag{38}$$

b) $$\Delta Q = \sum_{i=1}^{m_1} \epsilon_i \|\varphi\|_{W_i}. \tag{39}$$

c) There exists a worst perturbation $\delta\mathcal{L}^* = \sum_{i=1}^{m_1} \delta\mathcal{L}_i^*$, with $\delta\mathcal{L}_i^* \in W_i'$, that maximizes the uncertainty interval in the quantity of interest.

Proof Point a) comes immediately from the linearity of the form B as well as the quantity of interest Q. Indeed, since $u(\mathcal{L}) - u(\mathcal{L}_0) \in V_0$, we have

$$\psi(\mathcal{L}) - \psi(\mathcal{L}_0) = Q(u(\mathcal{L}) - u(\mathcal{L}_0)) = B(\boldsymbol{\beta}, u(\mathcal{L}) - u(\mathcal{L}_0), \varphi) = \mathcal{L}(\varphi) - \mathcal{L}_0(\varphi).$$

Point b) is also immediate observing that

$$\Delta Q = \sup_{\mathcal{L}\in\mathcal{A}_\mathcal{L}} |\psi(\mathcal{L}) - \psi(\mathcal{L}_0)| = \sum_{i=1}^{m_1} \sup_{\delta\mathcal{L}_i \in \mathcal{A}_\mathcal{L}^i - \mathcal{L}_{0i}} |\delta\mathcal{L}_i(\varphi)| = \sum_{i=1}^{m_1} \epsilon_i \|\varphi\|_{W_i}.$$

The fact that the supremum is attained for a particular choice of perturbation $\delta\mathcal{L}_i \in \mathcal{A}^i_{\mathcal{L}}$ is a consequence of the Hahn-Banach theorem (see e.g. [11, Chapter 1]). □

Observe that, being the quantity of interest Q linear with respect to $\mathcal{L}$, the characterization of ΔQ given in the theorem is exact and, whenever perturbing $\mathcal{L}$ alone, we do not have to bound any reminder of the expansion around $\mathcal{L}_0$. Section 4.1 gives indications on how to treat simultaneous uncertainty in loads and coefficients.

Application 3 (Linear elasticity problem) *Referring to Example 3, we could consider perturbations of the forcing term* $\mathbf{f}$ *in any space* $W' = [L^p(D)]^3$*, with* $6/5 \le p \le \infty$*. To allow for a different size of perturbation in each component of the vector* $\mathbf{f}$*, we endow the space* W' *with the weighted norm*

$$\|\mathbf{v}\|_{W'} = \left\{ \int_D \left(\sum_{i=1}^{3} (v_i/\epsilon_i)^2 \right)^{p/2} d\mathbf{x} \right\}^{1/p} \tag{40}$$

and define the set of admissible loads as $\mathcal{A}_{\mathcal{L}} = \{\mathbf{f} : \ \delta\mathbf{f} = \mathbf{f} - \mathbf{f}_0 \in W', \ \|\delta\mathbf{f}\|_{W'} \le 1\}$*. Here,* $\epsilon_i \ge 0$*,* $i = 1, 2, 3$*, characterizes the size of perturbation in each component. Observe that for strictly positive perturbations* ϵ_i*, the norm* $\|\cdot\|_{W'}$ *is equivalent to the standard* $[L^p(D)]^3$*-norm.*

In this case, the uncertainty interval is simply computed as

$$\Delta Q = \|\boldsymbol{\varphi}\|_W$$

where the dual norm $\|\cdot\|_W$ *is given by*

$$\|\mathbf{v}\|_W = \left\{ \int_D \left(\sum_{i=1}^{3} (\epsilon_i v_i)^2 \right)^{q/2} d\mathbf{x} \right\}^{1/q}, \qquad \frac{1}{q} = 1 - \frac{1}{p}.$$

The Sobolev embedding theorems guarantee that the quantity ΔQ*, previously introduced, is well defined. Moreover, we can characterize the worst distribution of the forcing term as*

$$\delta f_i^*(\mathbf{x}) = \frac{\epsilon_i^2}{\|\boldsymbol{\varphi}\|_W^{q/p}} \left(\sum_{j=1}^{3} (\epsilon_j v_j(\mathbf{x}))^2 \right)^{q/2-1} v_i(\mathbf{x}), \quad \forall \mathbf{x} \in D, \quad i = 1, 2, 3.$$

In a very similar way, we can consider perturbations of the traction $\mathbf{g}$ *on the boundary* Γ_N *in any space* $[L^{p_1}(\Gamma_N)]^3$*, with* $4/3 \le p_1 \le \infty$*.*

In some situations, we might want to consider only perturbations of the forcing term that preserve the total force*, i.e. such that* $\int_D \delta\mathbf{f} = \mathbf{0}$*. If we consider perturbations in* $[L^2(D)]^3$*, this constraint can be very easily taken into account. Let us define the unconstrained set of perturbations*

$$\delta\mathcal{A}_{\mathcal{L}} \equiv \{\delta\mathbf{f} = \mathbf{f} - \mathbf{f}_0 \in [L^2(D)]^3, \ \ \|\delta\mathbf{f}\|_{W'} \le 1\}$$

and the constrained set of perturbations

$$\delta\mathcal{A}_{\mathcal{L},0} \equiv \{\delta\mathbf{f} \in \delta\mathcal{A}_{\mathcal{L}}, \ \ \int_D \delta\mathbf{f}\, d\mathbf{x} = \mathbf{0}\}.$$

If we denote by $\bar{v}$ *the average over the domain of a function* $v \in L^1(D)$*, i.e.* $\bar{v} = \left(\int_D v\, d\mathbf{x}\right)/|D|$*, it is easy to show that*

$$\|\mathbf{v} - \bar{\mathbf{v}}\|_{W'} \leq \|\mathbf{v}\|_{W'}, \qquad \forall \mathbf{v} \in [L^2(D)]^3.$$

Hence, given any $\delta\mathbf{f} \in \delta\mathcal{A}_{\mathcal{L}}$*, we have* $(\delta\mathbf{f} - \bar{\delta\mathbf{f}}) \in \delta\mathcal{A}_{\mathcal{L},0}$ *and*

$$\begin{aligned}\Delta Q &= \sup_{\delta\mathbf{f}\in\delta\mathcal{A}_{\mathcal{L},0}} \left|\int_D \delta\mathbf{f}\,\boldsymbol{\varphi}\, d\mathbf{x}\right| = \sup_{\delta\mathbf{f}\in\delta\mathcal{A}_{\mathcal{L},0}} \left|\int_D \delta\mathbf{f}(\boldsymbol{\varphi} - \bar{\boldsymbol{\varphi}})\, d\mathbf{x}\right| \\ &= \sup_{\delta\mathbf{f}\in\delta\mathcal{A}_{\mathcal{L}}} \left|\int_D (\delta\mathbf{f} - \bar{\delta\mathbf{f}})(\boldsymbol{\varphi} - \bar{\boldsymbol{\varphi}})\, d\mathbf{x}\right| = \sup_{\delta\mathbf{f}\in\delta\mathcal{A}_{\mathcal{L}}} \left|\int_D \delta\mathbf{f}(\boldsymbol{\varphi} - \bar{\boldsymbol{\varphi}})\, d\mathbf{x}\right|\end{aligned}$$

Therefore, from the last equality we infer that the uncertainty interval in the constrained case can be computed as

$$\Delta Q = \|\boldsymbol{\varphi} - \bar{\boldsymbol{\varphi}}\|_W. \tag{41}$$

A similar argument holds for perturbations in the Neumann boundary conditions.

4.1 Simultaneous perturbation of coefficients and loads

We have already pointed out that the dependence of the quantity of interest, $\psi(\boldsymbol{\eta}) = Q(\boldsymbol{\beta}, u(\boldsymbol{\beta}, \mathcal{L}))$, with respect to the loads is linear. Hence, in particular, the second derivative $D^2_{\mathcal{L}}\psi$ vanishes. Whenever we perturb simultaneously the coefficients and the load, though, the remainder of the expansion (see (9)) will contain a term involving the cross derivative $D^2_{\beta,\mathcal{L}}\psi$ (besides the term $D^2_{\beta}\psi$, which has already been analyzed in Section 3.2).

To extend the result in Theorem 1 and Remark 6 we have then to bound this extra contribution:

$$\mathcal{R}_M \equiv \sup_{\delta\boldsymbol{\beta}\in\mathcal{A}-\boldsymbol{\beta}_0}\ \sup_{\delta\mathcal{L}\in\mathcal{A}_{\mathcal{L}}-\mathcal{L}_0}\ \sup_{\theta\in(0,1)} |D^2_{\beta,\mathcal{L}}\psi(\boldsymbol{\beta}_0 + \theta\delta\boldsymbol{\beta}, \mathcal{L}_0 + \theta\delta\mathcal{L})(\delta\boldsymbol{\beta}, \delta\mathcal{L})|, \tag{42}$$

which has to be added to $\mathcal{R}$ from (24) to get an upper bound for the total remainder. The mixed terms are characterized by the identity

$$\begin{aligned}D^2_{\beta,\mathcal{L}}\psi(\boldsymbol{\beta}, \mathcal{L})(\delta\boldsymbol{\beta}, \delta\mathcal{L}) = &- < D_\beta B(\boldsymbol{\beta}; D_{\mathcal{L}}u(\boldsymbol{\beta}, \mathcal{L})(\delta\mathcal{L}), \varphi(\boldsymbol{\beta})), \delta\boldsymbol{\beta} > \\ &+ < D_\beta Q(\boldsymbol{\beta}; D_{\mathcal{L}}u(\boldsymbol{\beta}, \mathcal{L})(\delta\mathcal{L})), \delta\boldsymbol{\beta} > .\end{aligned}$$

As in Remark 6, a practical approximation consists in evaluating the previous identity for the worst perturbations $\delta\boldsymbol{\beta}^*$ and $\delta\mathcal{L}^*$, respectively. This approach entails the computation of an auxiliary solution, $D_{\mathcal{L}}u(\boldsymbol{\beta}_0, \mathcal{L}_0)(\delta\mathcal{L}^*)$.

The other way, extending the results from Theorem 1, is to utilize (36), yielding to the a priori estimate for the energy norm of $D_{\mathcal{L}}u(\boldsymbol{\beta}, \mathcal{L})(\delta\mathcal{L})$,

$$|D_{\mathcal{L}}u(\boldsymbol{\beta}, \mathcal{L})(\delta\mathcal{L})|_{E,\beta} \leq \sum_{i=1}^{m_1} \check{C}_i(\boldsymbol{\beta})\|\delta\mathcal{L}_i\|_{W'_i}$$

and use the available assumptions on $D_\beta Q$ and $D_\beta B$, stated in (13)–(15). This leads to the estimate

Theorem 3 *With* (36), (37) *and the same assumptions and notation of Lemma 4, there holds*

$$\mathcal{R}_M \leq \sup_{\boldsymbol{\beta}\in\mathcal{A}} \left\{ \left(\sum_{i=1}^{m_1} \check{C}_i(\boldsymbol{\beta})\epsilon_i \right) \left[\sum_{i=1}^{m} \bar{C}_i(\boldsymbol{\beta})\bar{\delta}_i \right.\right.$$
$$\left.\left. + \left(k(\boldsymbol{\beta}_0, \boldsymbol{\beta})\, |\varphi_0|_{E,\beta_0} + k^{sup}(\boldsymbol{\beta}) \sum_{i=1}^{m} \bar{C}_i^{sup}\bar{\delta}_i \right) \left(\sum_{i=1}^{m} C_i(\boldsymbol{\beta})\bar{\delta}_i \right) \right] \right\}.$$

Moreover, when perturbations in both coefficients and loads are considered simultaneously, also the bound given in Theorem 1, relative to the term $D^2_\beta\psi$, changes slightly. In particular, the result (28.a) does not hold anymore. Indeed, in this case, we have to relate the norm $|u(\boldsymbol{\beta}, \mathcal{L})|_{E,\beta}$ with $|u(\boldsymbol{\beta}_0, \mathcal{L}_0)|_{E,\beta_0}$. A simple calculation shows that

$$|u(\boldsymbol{\beta}, \mathcal{L})|_{E,\beta} \leq \mathcal{I}_k(\boldsymbol{\beta}_0, \boldsymbol{\beta}) \left\{ |u(\boldsymbol{\beta}_0, \mathcal{L}_0)|_{E,\beta_0} + \sum_{i=1}^{m_1} \check{C}_i(\boldsymbol{\beta}) \|\delta\mathcal{L}\|_{W_i'} \right\}$$

which should replace (28.a). Hence, the result of Theorem 1 should be changed accordingly. Observe that this new estimate introduces only a third order correction in the result of Theorem 1.

5 Examples and numerical results

5.1 Scalar diffusion

In this example we consider the heat transfer over a unitary cube of low conductive material containing an inclusion with much higher conductivity (see Figure 1). The conductivity of the material is taken equal to 1 while the conductivity of the inclusion is 50 times larger. A unitary heat flux is imposed on the rear face Γ_3, while a constant temperature $u = 0$ is kept on the two stripes Γ_1 and Γ_2 on the upper and lower faces, respectively. The remaining portion of the boundary is insulated (i.e. zero heat flux is imposed).

This problem fits then within the class of problems described in Example 1 with $\Gamma_D = \Gamma_1 \cup \Gamma_2$, $\Gamma_N = \partial D \setminus \Gamma_D$, $f = g = h_2 = 0$ and $h_1 = 1$ on Γ_3 and zero elsewhere. The solution u represents the temperature of the body and the coefficient β the material conductivity.

The quantity we want to compute is the outward heat flux through Γ_1, namely

$$Q(\beta; u(\beta)) = -\int_{\Gamma_1} \beta\partial_n u(\beta)\, dS. \tag{43}$$

Observe that the functional $Q(\beta; v) = -\int_{\Gamma_1} \beta\partial_n v\, dS$, $\forall v \in H^1_{\Gamma_D}(D)$ is not bounded since we can not define the trace of the normal derivative of an H^1 function, in general. The quantity of interest $Q(u)$, though, is well defined because u has extra regularity, being the solution of the equation

$$\operatorname{div}(\beta\nabla u) = 0, \quad \text{in } D. \tag{44}$$

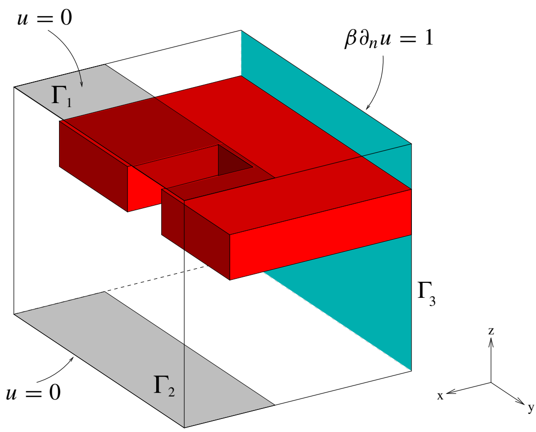

Fig. 1 Domain and boundary conditions used in the heat transfer test case. In red is shown the inclusion with higher conductivity

Indeed, the quantity $\beta\nabla u \in H_{div}(D) \equiv \{v \in (L^2(D))^3,\ \operatorname{div} v \in L^2(D)\}$ and its trace on the boundary is well defined –at least as an $H^{-1/2}$ functional–.

In order to obtain a bounded H^{-1} functional as the right hand side for the dual problem, we need to rewrite the quantity (43) in a more suitable way: let us take a particular function $\hat{v} \in H^1(D)$ such that

$$\hat{v} = \begin{cases} 1 & \text{on } \Gamma_1, \\ 0 & \text{on } \Gamma_2 \cup \Gamma_3. \end{cases}$$

If we multiply equation (44) by $\hat{v}$ and integrate by parts, we easily obtain

$$Q(\beta; u) = -\int_{\Gamma_1} \beta \partial_n u \, dS = -\int_D \beta \nabla u \cdot \nabla \hat{v} \, d\mathbf{x},$$

and the right hand side is now well defined for all functions $u \in H^1_{\Gamma_D}(D)$. Hence, the dual problem reads: *find* $\varphi(\beta) \in H^1_{\Gamma_D}(D)$ *s.t.*

$$\int_D \beta \nabla \varphi(\beta) \cdot \nabla v \, d\mathbf{x} = -\int_D \beta \nabla \hat{v} \cdot \nabla v \, d\mathbf{x} \qquad \forall v \in H^1_{\Gamma_D}(D). \tag{45}$$

Clearly, the choice of the auxiliary function $\hat{v}$ is not unique. In this example we have chosen $\hat{v} = \frac{5}{3}xz$ for $0 \le x \le \frac{3}{5}$ and $\hat{v} = z$ for $\frac{3}{5} < x \le 1$. Observe that the dual solution $\varphi(\boldsymbol{\beta})$ depends on the choice of $\hat{v}$, yet this does not affect the uncertainty analysis, see Remark 7.

The computed quantity for the nominal value of the coefficients is

$$Q(\beta_0; u(\beta_0)) \simeq 0.597,$$

obtained with a goal-oriented refined mesh with 7535 degrees of freedom, using $\mathbb{Q}^2$ finite elements. The discretization error in the computation of the quantity of interest is less than 1% and has been controlled using the goal-oriented estimator

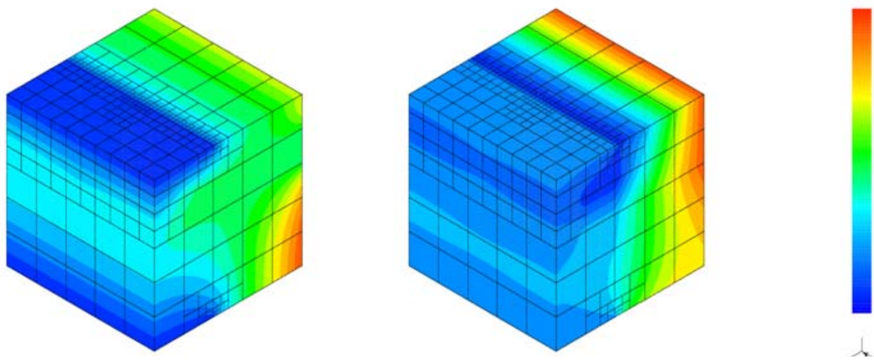

Fig. 2 Primal (left) and dual (right) solution for the heat transfer test case

η^L_{eeu} introduced in [28], which relates the error estimation in the quantity of interest to errors in the energy norm of the primal and dual solutions via the parallelogram law. Energy norm error estimates have been obtained with the subdomain residual method proposed in [22], which requires the solution of local problems on patches of elements. To meet the tolerance of 1% we have used an adaptive algorithm consisting in refining uniformly (h-refinement) all the elements that have a local error indicator larger that 50% of the maximum local error indicator in the current mesh.

Figure 2 shows the primal and dual solutions as well as the refined mesh used in the computation.

Observe that the mesh refinement procedure captures the singularity (of type $\sqrt{r}$) that appears in both the primal and dual solutions at the interface between the Dirichlet and Neumann boundaries.

Uncertainty analysis. We have considered an uncertainty in the conductivity coefficient of 10% both in the material and the inclusion and we allow for pointwise perturbations (L^∞ perturbations). The computation of the leading order term in the uncertainty interval for the quantity $Q(u)$ gives

$$\Delta Q_h^{lin} = 0.0446 = 7.47\%\ Q(\beta_0; u(\beta_0))$$

with an absolute error in the adaptive quadrature formula less than 2×10^{-5}. To achieve this tolerance the mesh has been further refined according to the algorithm described in Section 3.1.2. Figure 3 shows the worst distribution β^* of the coefficients leading to the maximum uncertainty interval while Figure 4 shows the sensitivity function α for the heat transfer test case.

Both functions have been represented with piecewise constants on the adapted mesh obtained at the final step of the adaptive algorithm described in Section 3.1.2.

Remark 7 Observe that the uncertainty interval ΔQ^{lin} as well as the worst distribution $\delta\beta^*$ of the coefficients and the sensitivity function α do not depend on the auxiliary function $\hat{v}$. To show this, let us first observe that these quantities depend only on the difference $(F - G)$ and, for this example

$$(F - G)(\beta_0, u(\beta_0), \varphi(\beta_0)) = -\beta \nabla u(\beta_0) \cdot \nabla(\varphi(\beta_0) + \hat{v}).$$

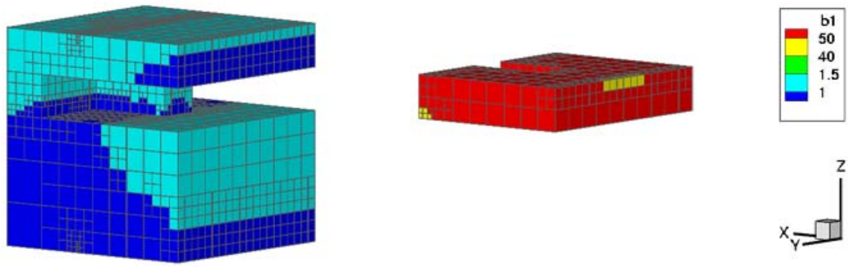

Fig. 3 Worst distribution β^* of the conductivity coefficient in the low conductivity material (left) and in the inclusion (right)

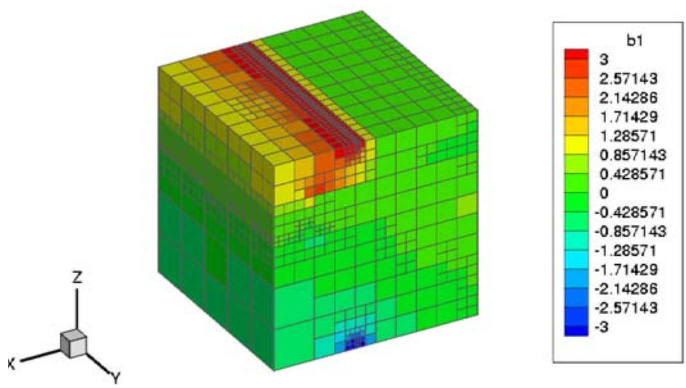

Fig. 4 Sensitivity function per unit volume and unitary relative perturbation of the conductivity coefficient

Hence, they depend only on the sum $\mathcal{E} = \varphi(\beta_0) + \hat{v}$. It is easy to show that $\mathcal{E}$ solves the problem

$$\begin{cases} \operatorname{div} \beta \nabla \mathcal{E} = 0, & \text{in } D \\ \mathcal{E} = 1 & \text{on } \Gamma_1 \\ \mathcal{E} = 0 & \text{on } \Gamma_2 \\ \beta \partial_n \mathcal{E} = 0 & \text{on } \Gamma_N. \end{cases} \tag{46}$$

Therefore, $\mathcal{E}$ is independent of $\hat{v}$ and it is the harmonic extension of the Dirichlet datum that takes 1 on Γ_1 and 0 on Γ_2.

We point out that the adaptivity procedure for error control in the quantity of interest is also independent of the choice of $\hat{v}$ as long as $\hat{v}$ itself belongs to the finite element space (as it is the case in our example). Indeed, the error estimator that has been empolyed is of residual type and the residual of the dual problem, given by

$$\mathcal{R}^{dual}(v) = -\int_D \beta \nabla(\varphi_h + \hat{v}) \cdot \nabla v \, d\mathbf{x}, \qquad \forall v \in H^1_{\Gamma_D}(D)$$

depends only on $\mathcal{E}_h = (\varphi_h + \hat{v})$, which is the finite element solution of (46) whenever $\hat{v}$ is exactly represented in the finite element space. Henceforth, the residual depends only on $\mathcal{E}_h$ and so does the adaptivity procedure.

Finally, the estimate (35) for the reminder of the expansion (with $\boldsymbol{\gamma} = \nabla \hat{v}$) gives

$$|\mathcal{R}| < 6.5\epsilon^2 + O(\epsilon^3)$$

where $\epsilon = \sup_{\delta\beta \in \mathcal{A}-\beta_0} \|\delta\beta/\beta_0\|_{L^\infty}$. In this case $\epsilon = 0.1$ and the reminder is of the same order of magnitude as the leading term ΔQ^{lin}. This estimate might be pessimistic, since it relies on a priori estimates and Cauchy-Schwartz inequalities, though it provides a guaranteed bound for the uncertainty interval, valid for any size of perturbation that preserves the coercivity of the bilinear form.

The previous bound can be improved by taking $\hat{v} = \mathcal{E}$. Indeed, with this choice, the dual solution $\varphi = \mathcal{E} - \hat{v} = 0$ and the function $\mathcal{E}$ minimizes the energy term $\sqrt{\beta}|\nabla \hat{v}|$ in (35) over all possible functions $\hat{v}$ that satisfy the constraint $\hat{v} = 1$ on Γ_1 and $\hat{v} = 0$ on Γ_2.

5.2 Elasticity

We consider, now, a "bulky" prismatic, linear elastic isotropic aluminum bar, of dimensions $20 \times 50 \times 10\, cm$ cantilevered at one end and loaded by a uniform shear force per unit area of $1\, MPa$ at its free end (see Figure 5). Hence, the mathematical problem is the one given in Example 2, where $\mathbf{u} = \mathbf{u}(x, y, z)$ represents the displacement field in the body, (x, y, z) being the Cartesian coordinate system shown in Figure 5 and $\boldsymbol{\beta} = [E, \nu]$ the coefficients - Young's modulus and Poisson's ratio, respectively - characterizing the material properties. As quantities of interest, we take the average displacements in the z and y directions at the free end of the beam (see Figure 6), namely

$$Q_1(\mathbf{u}) = \int_0^{20}\int_0^{10} u_z(x, 50, z)\, dxdz, \qquad Q_2(\mathbf{u}) = \int_0^{20}\int_5^{10} u_y(x, 50, z)\, dxdz.$$

The material is assumed to be isotropic. Based on laboratory tests on aluminum, the nominal values of the coefficients are $E = 0.68 \times 10^6\, MPa$ and $\nu = 0.33$ while the deviation with respect to the nominal values are 1% in Young's modulus and 8% in Poisson's ratio.

The elasticity tensor is then given by $C_{ijkl}(\boldsymbol{\beta}) = \lambda(\boldsymbol{\beta})\delta_{ij}\delta_{kl} + \mu(\boldsymbol{\beta})(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}),\ 1 \le i, j, k, l \le 3$, where $\lambda = \beta_1\beta_2/(1+\beta_2)(1-2\beta_2),\ \mu = \beta_1/2(1+\beta_2)$, so that $\beta_1 = E$ and $\beta_2 = \nu$.

The primal and dual solutions (for both the quantities Q_1 and Q_2) for the nominal values of the coefficients were obtained with hexahedral finite elements with forth-order polynomial shape functions using a mesh manually graded and anisotropically refined around the built-in section. Each of the solutions has 84867 degrees of freedom and is believed to be a highly accurate approximation of the exact one.

Table 1 Elasticity test case: Computed quantities Q, ΔQ^{lin}, $|\mathcal{R}|$, $D^2_{\boldsymbol{\beta}}\psi(\delta\beta^*, \delta\beta^*)$

Quantity of Interest	$\Delta Q^{lin}/Q$	$\lvert\mathcal{R}\rvert/Q$	$D^2_{\boldsymbol{\beta}}\psi/Q$
$Q_1 = -7.36135\,10^{-4}$	1.536%	> 15%	0.027%
$Q_2 = 5.3964\,10^{-5}$	1.594%	> 27.6%	0.016%

The computed values of the two quantities of interest are shown in Table 1. In both cases the numerical error is below 0.1% and it has been estimated by comparing the solutions obtained with different polynomial degrees ($p = 2, 3, 4$).

Uncertainty analysis. We have considered L^∞ perturbations of the coefficients E and ν of size 1% and 8% repectively. The computed value of the leading order term is given in Table 1. For both quantities of interest, the error in the adaptive quadrature formula, relative to the size of the leading term, is less than 10^{-4}.

Figure 7 shows the worst distribution of the Young's modulus E (left column) and Poisson's ratio ν (right column) that give the maximum uncertainty interval for the quantity Q_1. The red region corresponds to the higher coefficient and the blue region to the lower one. The first row in the Figure gives a global view while the second and third rows give the split view of the regions with positive, respectively negative, perturbation.

Figure 8 shows, instead, the sensitivity functions, as defined in (19), corresponding to E and ν. Finally, Figures 9 and 10 present the analogous results for the quantity of interest Q_2.

To bound the reminder we can use, in this case, estimate (33). The estimation of the constants can be pursued in the following way: assume that we can give a 6×6 matrix representation of the elasticity tensor $\mathbf{C}(\boldsymbol{\beta})$ (see e.g. [35]), which we will still denote by $\mathbf{C}$ with a little abuse of notation, for any given $\boldsymbol{\beta} \in \mathcal{A}_\beta$. Then compute its Cholesky factorization $\mathbf{C}(\boldsymbol{\beta}) = \mathbf{R}(\boldsymbol{\beta})\mathbf{R}^T(\boldsymbol{\beta})$. This yields, for instance, for the estimate of the constant $C_i(\boldsymbol{\beta})$ introduced in (13b)

$$\begin{aligned}
\int_D \left| \frac{\partial G}{\partial \beta_i}(\boldsymbol{\beta}; \mathbf{u}, \mathbf{v}) \right| d\mathbf{x} &= \int_D \left| \nabla_s \mathbf{u} \frac{\partial \mathbf{C}(\boldsymbol{\beta})}{\partial \beta_i} \nabla_s \mathbf{v} \right| d\mathbf{x} \\
&= \int_D \left| \nabla_s \mathbf{u}\, \mathbf{R}(\boldsymbol{\beta})\mathbf{R}^{-1}(\boldsymbol{\beta})\, \frac{\partial \mathbf{C}(\boldsymbol{\beta})}{\partial \beta_i}\, \mathbf{R}^{-T}(\boldsymbol{\beta})\mathbf{R}^T(\boldsymbol{\beta})\, \nabla_s \mathbf{v} \right| d\mathbf{x} \\
&\le \operatorname*{ess\,sup}_{\mathbf{x}\in D} \left\| \mathbf{R}^{-1}(\boldsymbol{\beta}) \frac{\partial \mathbf{C}(\boldsymbol{\beta})}{\partial \beta_i} \mathbf{R}^{-T}(\boldsymbol{\beta}) \right\|_2 |\mathbf{u}|_{E,\boldsymbol{\beta}} |\mathbf{v}|_{E,\boldsymbol{\beta}},
\end{aligned}$$

where we have denoted by $\|\mathbf{A}\|_2$ the euclidean norm of a matrix $\mathbf{A} \in \mathbb{R}^{6\times 6}$. Hence

$$C_i(\boldsymbol{\beta}) = \operatorname*{ess\,sup}_{\mathbf{x}\in D} \left\| \mathbf{R}^{-1}(\boldsymbol{\beta}) \frac{\partial \mathbf{C}(\boldsymbol{\beta})}{\partial \beta_i} \mathbf{R}^{-T}(\boldsymbol{\beta}) \right\|_2$$

and an analogous expression holds for $C_{ij}(\boldsymbol{\beta})$ (defined in (26)) with the second derivative of the matrix $\mathbf{C}$ replacing the first one. Moreover, observe that

$$\begin{aligned}|\mathbf{u}|^2_{E,\boldsymbol{\beta}} &= \int_D \nabla_s \mathbf{u}\, \mathbf{C}(\boldsymbol{\beta})\, \nabla_s \mathbf{u}\, d\mathbf{x} \\ &= \int_D \nabla_s \mathbf{u}\, \mathbf{R}(\boldsymbol{\beta}_0)\mathbf{R}^{-1}(\boldsymbol{\beta}_0)\, \mathbf{C}(\boldsymbol{\beta})\, \mathbf{R}^{-T}(\boldsymbol{\beta}_0)\mathbf{R}^{T}(\boldsymbol{\beta}_0)\, \nabla_s \mathbf{u}\, d\mathbf{x} \\ &\geq \operatorname*{ess\,inf}_{\mathbf{x}\in D} \lambda_{min}\left[\mathbf{R}^{-1}(\boldsymbol{\beta}_0)\mathbf{C}(\boldsymbol{\beta})\mathbf{R}^{-T}(\boldsymbol{\beta}_0)\right] |\mathbf{u}|^2_{E,\boldsymbol{\beta}_0},\end{aligned}$$

where we have denoted by $\lambda_{min}[\mathbf{A}]$ the minimum eigenvalue of the positive definite matrix $\mathbf{A} \in \mathbb{R}^{6\times 6}$. Therefore, the constant $k(\boldsymbol{\beta}_0, \boldsymbol{\beta})$ is given by

$$k(\boldsymbol{\beta}_0, \boldsymbol{\beta}) = \left\{\operatorname*{ess\,inf}_{\mathbf{x}\in D} \lambda_{min}\left[\mathbf{R}^{-1}(\boldsymbol{\beta}_0)\mathbf{C}(\boldsymbol{\beta})\mathbf{R}^{-T}(\boldsymbol{\beta}_0)\right]\right\}^{-\frac{1}{2}}.$$

The computation of those constants simplifies for the special case where $\boldsymbol{\beta}_0$ is constant all over the domain, as in our example. Indeed, if we denote by $\mathcal{I} \subset \mathbb{R}^m$ the set $\mathcal{I} = \prod_{i=1}^m [\beta_{0i} - \bar{\delta}_i, \beta_{0i} + \bar{\delta}_i]$, where $\bar{\delta}_i$ is the maximum possible perturbation of the i-th coefficient on the domain, then, given any $\boldsymbol{\beta} \in \mathcal{A}$, clearly $\boldsymbol{\beta}(\mathbf{x}) \in \mathcal{I}$, $\forall \mathbf{x} \in D$ and

$$k(\boldsymbol{\beta}_0, \boldsymbol{\beta}) \leq \tilde{k}(\boldsymbol{\beta}_0), \quad \text{where } \tilde{k}(\boldsymbol{\beta}_0) = \left\{\inf_{\boldsymbol{\eta}\in\mathcal{I}} \lambda_{min}\left[\mathbf{R}^{-1}(\boldsymbol{\beta}_0)\mathbf{C}(\boldsymbol{\eta})\mathbf{R}^{-T}(\boldsymbol{\beta}_0)\right]\right\}^{-\frac{1}{2}}$$

and

$$C_i(\boldsymbol{\beta}) \leq \tilde{C}_i, \quad \text{where } \tilde{C}_i = \sup_{\boldsymbol{\eta}\in\mathcal{I}} \left\| \mathbf{R}^{-1}(\boldsymbol{\eta}) \frac{\partial \mathbf{C}(\boldsymbol{\eta})}{\partial \beta_i} \mathbf{R}^{-T}(\boldsymbol{\eta}) \right\|_2.$$

Similar results hold for $C_{ij}(\boldsymbol{\beta})$. Since it is always possible to find a particular function $\beta(x) \in \mathcal{A}$ whose image covers the set $\mathcal{I}$, the following equality holds

$$\sup_{\beta\in\mathcal{A}} \left\{ (k(\beta_0, \beta))^2 \sum_{i,j=1}^m (C_{ij}+2C_iC_j)(\beta)\, \bar{\delta}_i\, \bar{\delta}_j \right\} = \tilde{k}(\beta_0)^2 \sum_{i,j=1}^m (\tilde{C}_{ij}+2\tilde{C}_i\tilde{C}_j)\, \bar{\delta}_i\, \bar{\delta}_j.$$

The computation of the constants $\tilde{k}$, $\tilde{C}_i$ and $\tilde{C}_{i,j}$ implies a simple optimization of a function on a hypercube of $\mathbb{R}^m$. Henceforth, these constants can be evaluated a priori and used in the bound (33) for the residual once the norms of $\mathbf{u}_0$ and $\boldsymbol{\varphi}_0$ are available.

Unfortunately, it turns out that this bound is extremely pessimistic when the Poisson's ratio is not close to zero. The principal reason is that there is a cancellation of the term involving "div $\mathbf{u}$" between the first two terms appearing on the right hand side of (25), which is not taken into account in the bound.

As an alternative to this a-priori bound, we have computed the second derivative of the quantity of interest along the worst perturbation $\delta\boldsymbol{\beta}^*$ (see Remark 6). The third and fourth columns in Table 1 show the computed values of the bound on the reminder and the coresponding value obtained from evaluating the second order term in the "worst direction". As it can be observed, this second order term is of order $\bar{\delta}^2$ while the guaranteed bound for the remainder is about three orders of magnitude larger.

We finally conclude this section considering perturbations in the shear force applied at the end of the bar (surface S in Figure 6).

The nominal value of the traction is $\mathbf{g}_0 = [0, 0, g_z]$, with $g_z = 1\,MPa$. We have considered perturbations in all three components of the traction, either in $[L^2(S)]^3$ or $[L^\infty(S)]^3$. In both cases we have used the weighted norm introduced in (40) with

$$\epsilon_x = \epsilon_y = 0.01\|g_z\|_*, \quad \epsilon_z = 0.05\|g_z\|_*$$

where the *-norm is $L^2(S)$, respectively $L^\infty(S)$. Moreover, we have also considered perturbations in $[L_0^2(S)]^3 \equiv \{\mathbf{g} \in [L^2(S)]^3, \ \int_S \mathbf{g} = \mathbf{0}\}$, which preserve the total force (see Application 3). Tables 2 and 3 summarize the uncertainty interval due to perturbations in the traction for the two quantities Q_1 and Q_2. The first (resp. second, third) column shows the resulting uncertainty interval when only the first (resp. second, third) component of the traction is perturbed; i.e. $\epsilon_y = \epsilon_z = 0$. The forth column, instead, shows the total uncertainty interval when all the components are perturbed simultaneously. As it can be observed, there is not a large difference between perturbations in $L^2(S)$ and $L^\infty(S)$. Yet, if we enforce the perturbation to preserve the total force, the uncertainty interval decreases dramatically for both quantities of interest.

Finally, Figure 11 shows the worst distribution of the traction, for both quantities of interest, in the case where the traction is perturbed in $[L^2(S)]^3$.

Conclusions

In this work we have proposed a methodology, based on perturbation analysis and duality techniques, to compute the worst-case scenario for elliptic problems whose coefficients and/or forcing terms lie in infinite dimensional spaces.

We have shown that the worst-case scenario can be computed inexpensively for certain classes of infinite dimensional perturbations in coefficients and loads, by postprocessing the solutions of the primal and dual problems computed for the nominal values of the parameters. This postprocessing entailed, in the case of uncertainty in the coefficients, the use of an adaptive quadrature algorithm.

We have also analyzed and quantified the error in the computation of the worst scenario bound due to truncation of the Taylor expansion around the nominal values of the parameters, giving computable bounds for symmetric problems and suggesting a possible approximation of it in the more general case.

Finally, we point out that uncertainty quantification should always come along with verification of the numerical solution, whenever the primal and dual solutions are computed approximately, e.g. by means of a finite element discretization. By verification, we mean control of the discretization error. With this respect, in our numerical examples, we have controlled the error in the computation of the quantity of interest with a posteriori error estimation techniques and we have provided an a priori bound for the error in the computation of the uncertainty interval. What is left to do is to derive a posteriori error estimators and adaptive techniques for the

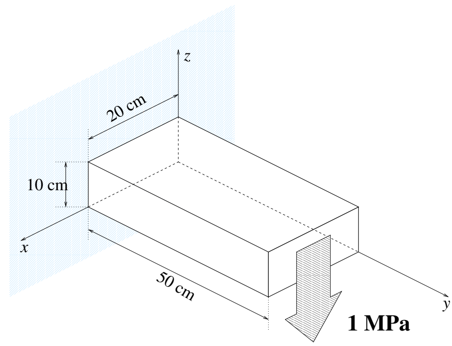

Fig. 5 Geometry of “bulky” prismatic body

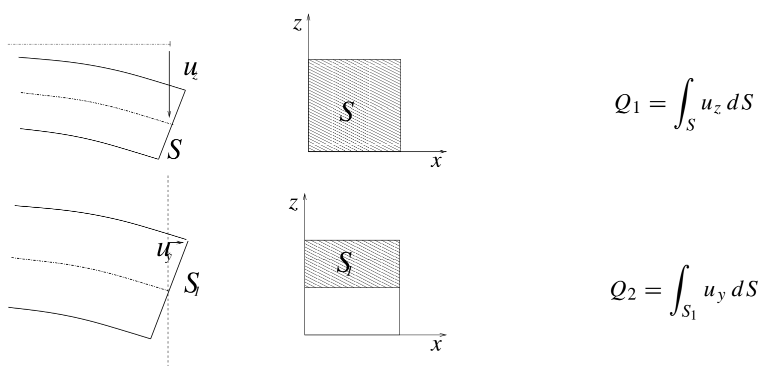

Fig. 6 Quantities of Interest - Average displacements at the free end of the beam

Table 2 Relative uncertainty interval in the quantity Q_1 ($\Delta Q_1^{lin}/Q_1$) for perturbations in each component of the traction separately (first three columns) and simultaneous perturbations in all the components (fourth column)

	δg_x	δg_y	δg_z	δg_{tot}
L^2	0.0005%	0.085%	5%	5.0007%
L_0^2	0.0005%	0.085%	0.004%	0.085%
L^∞	0.0004%	0.073%	5%	5.0007%

Table 3 Relative uncertainty interval in the quantity Q_2 ($\Delta Q_2^{lin}/Q_2$) for perturbations in each component of the traction separately (first three columns) and simultaneous perturbations in all the components. (fourth column)

	δg_x	δg_y	δg_z	δg_{tot}
L^2	0.007%	0.18%	5%	5.003%
L_0^2	0.007%	0.12%	0.024%	0.12%
L^∞	0.005%	0.15%	5%	5.003%

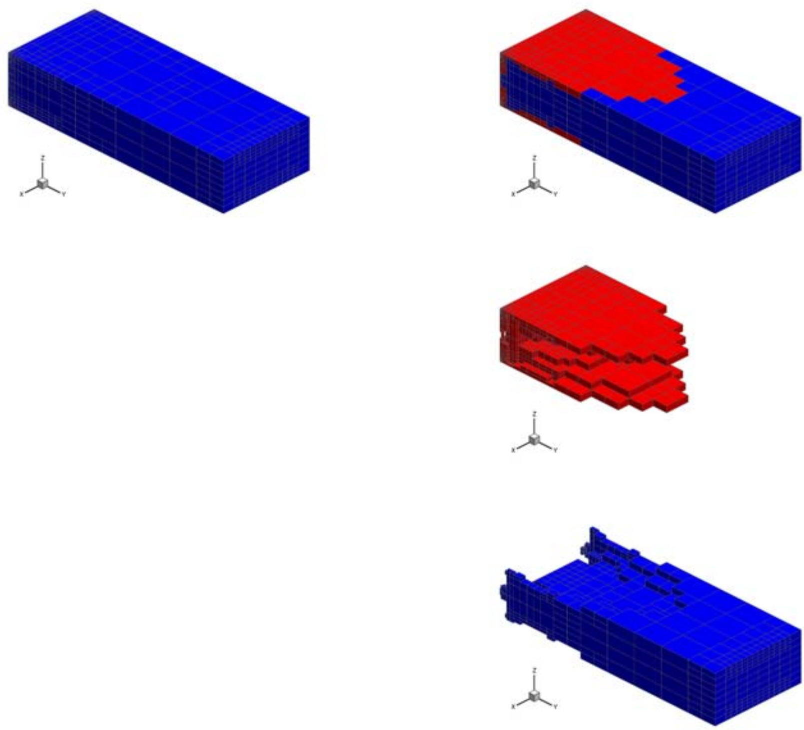

Fig. 7 Quantity of Interest Q_1: worst distribution of E (left) and ν (right). The red region corresponds to an increased coefficient, the blue region corresponds to a decreased one

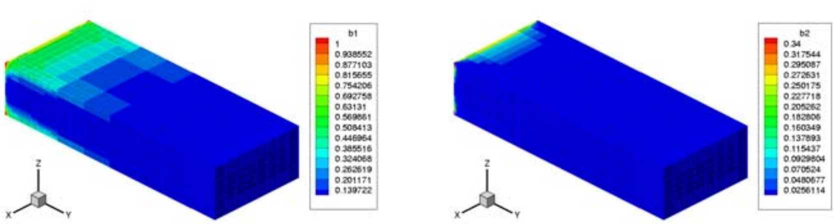

Fig. 8 Quantity of Interest Q_1: sensitivity functions α^E (left) and α^ν (right)

computation of the uncertainty interval as well. This issue is the subject of ongoing research.

Another source of uncertainty that has not been considered in this paper but is very important in many applications is the uncertainty in the non-homogeneous Dirichlet boundary datum. This issue will be analyzed in a forthcoming work, too.

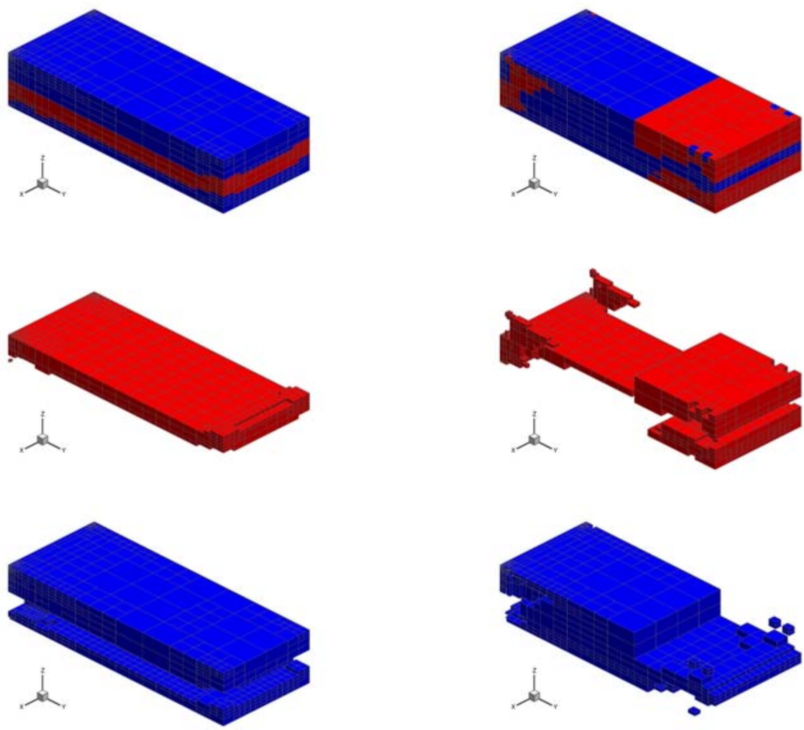

Fig. 9 Quantity of Interest Q_2: worst distribution of E (left) and ν (right). The red region corresponds to an increased coefficient, the blue region corresponds to a decreased one

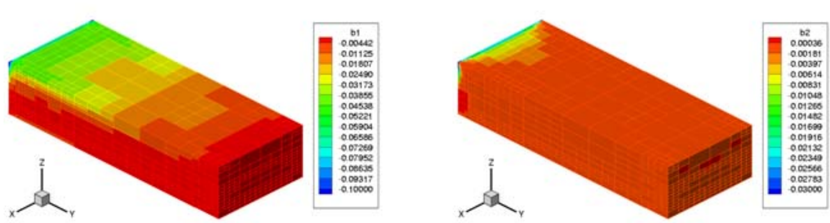

Fig. 10 Quantity of Interest Q_2: sensitivity functions α^E (left) and α^ν (right)

Acknowledgements The second and third authors were supported by the ICES Postdoctoral Fellowship Program. The three authors were also partially supported by the Sandia National Labs project # 268687, and NSF grant EIA-0205181.

The third author was also partially supported by the UdelaR in Uruguay, and by the European Network HYKE (HYperbolic and Kinetic Equations: Asymptotics, Numerics, Analysis), funded by the EC (Contract HPRN-CT-2002-00282).

The implementation of uncertainty quantification algorithms has been done within the *hp* 3D finite element code developed by Dr. Leszek Demkowicz, whose help and advice is gratefully acknowledged.

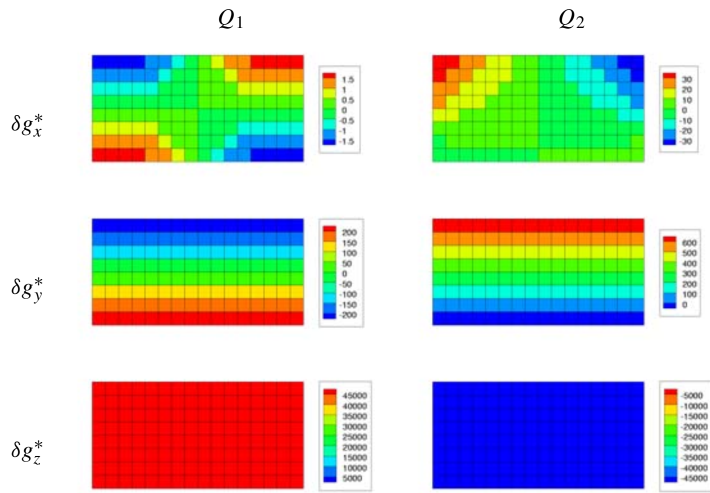

Fig. 11 Worst distribution of the traction applied at the free end of the bar for perturbations in $[L^2(S)]^3$. The left column corresponds to the quantity of interest Q_1, the right column to Q_2

References

1. Guide for Verification and Validation of Computational Fluid dynamics simulation. Technical Report AIAA G-077-1998, American Institute of Aeronautics and Astronautics, 1998
2. Ainsworth, M., Oden, J-T.: A posteriori Error Estimation in Finite Element Analysis. Wiley, 2000
3. Babuška, I., Chleboun, J.: Effects of uncertainties in the domain on the solution of Neumann boundary value problems in two spatial dimensions. Math. Comp. **71**(240), 1339–1370 (electronic) (2002)
4. Babuška, I., Chleboun, J.: Effects of uncertainties in the domain on the solution of Dirichlet boundary value problems. Numer. Math. **93**(4), 583–610 (2003)
5. Babuška, I., Oden, J.T.: Verification and validation in computational engineering and science. Part I: Basic concepts. Comp. Meth. Appl. Mech. Engrg. **193**, 4057-4066 (2004)
6. Babuška, I., Stroboulis, T.: The Finite Element Method and its Reliability. Oxford Sci. Publ., 2001
7. Bangerth, W., Rannacher, R.: Adaptive finite element methods for differential equations. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2003
8. Ben-Haim, Y.: Information-gap decision theory. Series on Decision and Risk. Academic Press Inc., San Diego CA, 2001. Decisions under severe uncertainty
9. Ben-Haim, Y., Elishakoff, I.: Convex models of uncertainty in applied mechanics. Elsevier, Amsterdam, 1990
10. Brenner, S.C., Scott, L.R.: The Mathematical Theory of Finite Element Methods. Springer–Verlag, 1994
11. Brezis, H.: Analyse fonctionnelle: Théorie et applications. Collection Mathématiques Appliquées pour la Maîtrise. Masson, Paris, 1983
12. Cacuci, D.: Sensitivity and Uncertainty analysis: theory. Volume **1**. Chapman & Hall 2003
13. Achenbach, J., Hlaváček, I., Chleboun, J., Babuška, I.: Uncertain input data problems and the worst scenario method. Elsevier, 2004
14. National Research Council. Science and Judgment in Risk Assessment. National Academy Press, 1994
15. Cullen, A.C., Frey, H.Ch.: Probabilistic Techniques Exposure Assessment. Plenum Press, 1999

16. Elishakoff, I.: Probabilistic Methods in the Theory of Structures. Dover, Second edition, 1999
17. Elishakoff, I. (ed.): Whys and hows in uncertainty modelling. Springer–Verlag, 1999
18. Evans, L.: Partial Differential Equations. Volume **19** of Graduate Studies in Mathematics. AMS, 1998
19. Garreau, S., Guillaume, P., Masmoudi, M.: The topological asymptotic for PDE systems: the elasticity case. SIAM J. Control Optim. 39 (6):1756–1778 (electronic), 2001.
20. Helton, J.C.: Analysis in the presence of stochastic and subjective uncertainties. J. Stochastical Computations and Simulations **57**, 3–76 (1997)
21. Helton, J.C.: Uncertainty and Sensitivity Analysis in performance assessment for waste isolation Plant. Comput. Phys. Commun. **117**, 156–180 (1999)
22. Machiels, L., Maday, Y., Patera, A.T.: A "flux-free" nodal Neumann subproblem approach to output bounds for partial differential equations. C. R. Acad. Sci. Paris Sér. I Math. **330**(3), 249–254 (2000)
23. Melchers, R.E.: Structural Reliability Analysis and Prediction. Wiley, 1999
24. Modarres, M., What Every Engineer Should know about Reliability and Risk Analysis. Dekker, 1993
25. Moon, K.-S., Von Schwerin, E., Szepessy, A., Tempone, R.: Convergence Rates for an adaptive dual weighted residual finite element algorithm. ICES Report 04-19, ICES University of Texas at Austin, April 2004
26. Oberkampf, W.L., Trucano, T.G.: Verification and validation in computational fluid dynamics. Progress in Aerospace Sciences **38**, 209–272 (2002)
27. Oberkampf, W.L., Trucano, T.G., Hirsch, Ch.: Verification, validation and predictive capabilities in computational engineering and physics. SANDIA Report SAND2003-3769, Sandia National Laboratories, 2003
28. Oden, J.T., Prudhomme, S.: Goal-oriented error estimation and adaptivity for the finite element method. Comput. Math. Appl. **41**(5–6), 735–756 (2001)
29. Oden, J.T., Prudhomme, S.: Estimation of modeling error in computational mechanics. J. Comput. Phys. **182**, 496–515 (2002)
30. Oden, J.T., Prudhomme, S., Hammerand, D.C., Kuczma, M.S.: Modeling error and adaptivity in nonlinear continuum mechanics. Comput. Meth. Appl. Mech. Engrg. **190**, 6663–6684 (2001)
31. Pironneau, O.: Optimal shape design for elliptic systems. Springer Series in Computational Physics. Springer-Verlag, New York, 1984
32. Ramsey, Ch.B., Modarres, M.: Commercial Nuclear Power. Assuring Safety for the Future. Wiley, 1997
33. Roache, P.J.: Verification and Validation in Computational Science and Engineering. Hermosa Publishers, 1998
34. Spanos, P.D., Deodatis, D. (eds.): Computational Stochastic Mechanics. Proceeding of the Fourth International Conference on Computational Mechanics, Corfu Greece June 9–12, 2002. Millpress, 2003
35. Szabó, B., Babuška, I.: Finite element analysis. Wiley, New York, 1991