ESAIM: Control, Optimisation and Calculus of Variations June 2002, Vol. 8, 1007–1028
URL: http://www.emath.fr/cocv/ DOI: 10.1051/cocv:2002041

A POSTERIORI ERROR ESTIMATION FOR REDUCED-BASIS APPROXIMATION OF PARAMETRIZED ELLIPTIC COERCIVE PARTIAL DIFFERENTIAL EQUATIONS: "CONVEX INVERSE" BOUND CONDITIONERS

Karen Veroy[1], Dimitrios V. Rovas[2] and Anthony T. Patera[3]

Abstract. We present a technique for the rapid and reliable prediction of linear-functional outputs of elliptic coercive partial differential equations with affine parameter dependence. The essential components are (*i*) (provably) rapidly convergent global reduced-basis approximations – Galerkin projection onto a space W_N spanned by solutions of the governing partial differential equation at N selected points in parameter space; (*ii*) *a posteriori* error estimation – relaxations of the error-residual equation that provide inexpensive bounds for the error in the outputs of interest; and (*iii*) off-line/on-line computational procedures – methods which decouple the generation and projection stages of the approximation process. The operation count for the on-line stage – in which, given a new parameter value, we calculate the output of interest and associated error bound – depends only on N (typically very small) and the parametric complexity of the problem; the method is thus ideally suited for the repeated and rapid evaluations required in the context of parameter estimation, design, optimization, and real-time control. In our earlier work we develop a rigorous *a posteriori* error bound framework for reduced-basis approximations of elliptic coercive equations. The resulting error estimates are, in some cases, quite sharp: the ratio of the estimated error in the output to the true error in the output, or *effectivity*, is close to (but always greater than) unity. However, in other cases, the necessary "bound conditioners" – in essence, operator preconditioners that (*i*) satisfy an additional spectral "bound" requirement, and (*ii*) admit the reduced-basis off-line/on-line computational stratagem – either can not be found, or yield unacceptably large effectivities. In this paper we introduce a new class of improved bound conditioners: the critical innovation is the direct approximation of the parametric dependence of the *inverse* of the operator (rather than the operator itself); we thereby accommodate higher-order (*e.g.*, piecewise linear) effectivity constructions while simultaneously preserving on-line efficiency. Simple convex analysis and elementary approximation theory suffice to prove the necessary bounding and convergence properties.

Mathematics Subject Classification. 35J50, 65N15.

Received February 13, 2002.

Keywords and phrases: Elliptic partial differential equations, reduced-basis methods, output bounds, Galerkin approximation, *a posteriori* error estimation, convex analysis.

[1] Massachusetts Institute of Technology, Department of Civil and Environmental Engineering, Room 3-264, Cambridge, MA 02139-4307, U.S.A.

[2] Massachusetts Institute of Technology, Department of Mechanical Engineering, Room 3-264, Cambridge, MA 02139-4307, U.S.A.

[3] Massachusetts Institute of Technology, Department of Mechanical Engineering, Room 3-266, Cambridge, MA 02139-4307, U.S.A.; e-mail: patera@MIT.EDU

Résumé. Nous présentons une technique pour la prédiction rapide et sûre de sorties – fonctionnelles linéaires – d'équations coercives aux dérivées partielles avec une dépendance affine en fonction des paramètres. Les composantes essentielles sont (*i*) approximations globales par bases-réduites rapidement convergentes – projection de Galerkin sur un espace W_N engendré par les solutions de l'équation aux dérivées partielles à N points sélectionnés dans l'espace des paramètres ; (*ii*) estimation d'erreur *a posteriori* – relaxations de l'équation de l'erreur qui fournissent des bornes peu coûteuses pour l'erreur effectuée sur la sortie d'intérêt ; et (*iii*) procédures de calcul en différé/en ligne – méthodes qui découplent l'étape de génération de l'étape de projection de l'approximation. Le décompte des opérations pour l'étape en ligne – dans laquelle, étant donnée une nouvelle valeur du paramètre, nous calculons la sortie d'intérêt et les bornes de l'erreur associées – dépend uniquement de N (typiquement très petit) et de la complexité paramétrique du problème ; la méthode est ainsi idéalement applicable pour des évaluations répétées et rapides dans un contexte d'estimation de paramètre, de design, d'optimisation, et de contrôle temps réel. Dans nos travaux précédents, nous avons développé un cadre rigoureux *a posteriori* pour les bornes de l'erreur due à l'approximation par bases-réduites d'équations elliptiques coercives. Les estimations d'erreur résultantes sont, dans certains cas, très précis : le rapport entre l'erreur estimée et la véritable erreur effectuée sur la sortie, encore appelée *efficacité*, est proche de (mais toujours plus grande que) l'unité. Cependant, dans d'autres contextes, les "conditioneurs pour les bornes" – essentiellement des opérateurs/préconditioneurs qui (*i*) satisfont une condition spectrale "borne" supplémentaire, et (*ii*) admettent le stratagème de calcul bases-réduites en différé/en ligne – peuvent soit ne pas être trouvés soit impliquent des efficacités larges inacceptables. Dans ce papier, nous introduisons une nouvelle classe de conditioneurs pour les bornes améliorés : l'innovation essentielle est l'approximation directe de la dépendance paramétrique de l'*inverse* de l'opérateur (plutôt que celle de l'opérateur elle-même) ; de ce fait nous facilitons la construction d'ordre élevée (*e.g.* linéaires par morceaux) de l'efficacité tout en préservant les performances de l'étape en ligne. Une analyse de convexité simple et un usage élémentaire de théorie de l'approximation sont suffisantes à prouver les propriétés nécessaires de convergence et de bornes.

Introduction

The optimization, control, and characterization of an engineering component or system requires the prediction of certain "quantities of interest", or performance metrics, which we shall denote *outputs* – for example deflections, maximum stresses, maximum temperatures, heat transfer rates, flowrates, or lifts and drags. These outputs are typically expressed as functionals of field variables associated with a parametrized partial differential equation which describes the physical behavior of the component or system. The parameters, which we shall denote *inputs*, serve to identify a particular "configuration" of the component: these inputs may represent design or decision variables, such as geometry – for example, in optimization studies; actuator variables, such as throttle power – for example in real-time control applications; or characterization variables, such as physical properties – for example in inverse problems. We thus arrive at an implicit *input–output* relationship, evaluation of which demands solution of the underlying partial differential equation.

Our goal is the development of computational methods that permit *rapid* and *reliable* evaluation of this partial-differential-equation-induced input-output relationship *in the limit of many queries* – that is, in the design and optimization, control, and characterization contexts. The "many query" limit has certainly received considerable attention: from "fast loads" or multiple right-hand side notions (*e.g.* [7,8]) to matrix perturbation theories (*e.g.* [1,19]) to continuation methods (*e.g.* [2,17]). Our particular approach is based on the reduced-basis method, first introduced in the late 1970s for nonlinear structural analysis [3, 13], and subsequently developed more broadly in the 1980s and 1990s [5, 6, 9, 14, 15, 18]. The reduced-basis method recognizes that the field variable is not, in fact, some arbitrary member of the infinite-dimensional solution space associated with the partial differential equation; rather, the field variable resides, or "evolves", on a much lower-dimensional manifold induced by the parametric dependence.

The reduced-basis approach as earlier articulated is local in parameter space in both practice and theory. To wit, Lagrangian or Taylor approximation spaces for the low-dimensional manifold are typically defined relative

to a particular parameter point; and the associated *a priori* convergence theory relies on asymptotic arguments in sufficiently small neighborhoods [9]. As a result, the computational improvements – relative to conventional (say) finite element approximation – are often quite modest [15]. Our work [10,12,16] differs from these earlier efforts in several important ways: first, we develop (in some cases, provably) *global* approximation spaces; second, we introduce rigorous *a posteriori error estimators*; and third, we exploit *off-line/on-line* computational decompositions (see [5] for an earlier application of this strategy within the reduced–basis context). These three ingredients allow us – for the restricted but important class of "parameter-affine" problems – to reliably decouple the generation and projection stages of reduced-basis approximation, thereby effecting computational economies of several orders of magnitude.

In our earlier work we develop a rigorous *a posteriori* error bound framework for reduced-basis approximations of elliptic coercive equations. The resulting error estimates are, in some cases, quite sharp: the ratio of the estimated error in the output to the true error in the output, or *effectivity*, is close to (but always greater than) unity. However, in other cases, the necessary "bound conditioners" – in essence, operator preconditioners that (*i*) satisfy an additional spectral "bound" requirement, and (*ii*) admit the reduced-basis off-line/on-line computational stratagem – either can not be found, or yield unacceptably large effectivities. In this paper we introduce a new class of improved bound conditioners: the critical innovation is the direct approximation of the parametric dependence of the *inverse* of the operator (rather than the operator itself); we thereby accommodate higher-order (*e.g.*, piecewise-linear) effectivity constructions while simultaneously preserving on-line efficiency. Simple convex analysis and elementary approximation theory suffice to prove the necessary bounding and convergence properties.

In Section 1 we present the problem statement, and demonstrate the monotonicity and convexity results on which our new bound conditioner formulation is constructed. In Section 2 we describe the new *a posteriori* error estimation framework, and prove the requisite *a posteriori* bound results. In Section 3 we develop the *a priori* convergence theory for our output bounds for the special case of a single parameter. Finally, in Section 4, we present numerical results for several illustrative "model-problem" examples.

1. Problem formulation

1.1. Exact statement

We first introduce a Hilbert space Y, and associated inner product and norm $(\cdot,\cdot)$ and $\|\cdot\| \equiv (\cdot,\cdot)^{1/2}$, respectively. We next introduce the dual space of Y, Y', and the associated duality pairing between Y and Y', ${}_{Y'}\langle\cdot,\cdot\rangle_Y \equiv \langle\cdot,\cdot\rangle$.

We then define, for any $\mu \in \mathcal{D}^\mu \subset \mathbb{R}^P$, the parametrized (distributional) operator $\mathcal{A}(\mu)\colon Y \to Y'$. We assume that $\mathcal{A}(\mu) = A(\Theta(\mu))$, where, for any $\theta \in \mathbb{R}^Q_+$, $A(\theta)\colon Y \to Y'$ is given by

$$A(\theta) = A_0 + \sum_{q=1}^{Q} \theta_q\, A_q,$$

and the $\Theta_q\colon \mathcal{D}^\mu \to \mathbb{R}_+$, $q = 0,\ldots,Q$, are non-negative functions (for future reference, we also define $\Theta_0 \equiv 1$). Here $\mathbb{R}_+$ refers to the non-negative real numbers. The range of Θ is denoted $\mathcal{D}^\theta$; and we define $\theta^{\min}$ (≥ 0), $\theta^{\max}$ (assumed finite), and $\mathcal{D}^\theta_{\rm box} \subset \mathbb{R}^Q_+$ as

$$\theta_q^{\min} \equiv \sup t_{\{t\in\mathbb{R}_+ \,|\, \Theta_q(\mu)\geq t,\ \forall \mu\in\mathcal{D}^\mu\}}, \qquad q = 1,\ldots,Q,$$

$$\theta_q^{\max} \equiv \inf t_{\{t\in\mathbb{R}_+ \,|\, \Theta_q(\mu)\leq t,\ \forall \mu\in\mathcal{D}^\mu\}}, \qquad q = 1,\ldots,Q,$$

and $\mathcal{D}^\theta_{\rm box} \equiv \Pi_{q=1}^Q [\theta_q^{\min}, \theta_q^{\max}]$, respectively.

Finally, we require that A_0 is continuous, symmetric, and coercive, and that the A_q, $q = 1, \ldots, Q$, are continuous, symmetric, and positive-semidefinite ($\langle A_q v, v\rangle \geq 0$, $\forall\, v \in Y$); it follows that $A(\theta)$ (respectively, $\mathcal{A}(\mu)$) is continuous, symmetric, and coercive for all θ in $\mathcal{D}^\theta_{\rm box}$ (respectively, for all μ in $\mathcal{D}^\mu$).

Our problem can then be stated as: given a $\mu \in \mathcal{D}^\mu$, and linear functional $F \in Y'$, evaluate the output

$$s(\mu) = \langle F, u(\mu)\rangle,$$

where $u(\mu) \in Y$ is the unique solution of $A(\Theta(\mu))\, u(\mu) = F$; we shall interpret the latter as

$$\langle A(\Theta(\mu))\, u(\mu), v\rangle = \langle F, v\rangle, \quad \forall\, v \in Y. \tag{1.1}$$

Note that $s(\mu)$ may also be interpreted as the energy of the solution; $s(\mu) = \langle F, u(\mu)\rangle = \langle A(\Theta(\mu))\, u(\mu), u(\mu)\rangle$ – and is hence strictly positive. (In this paper, the output $s(\mu)$ is "compliant", and the operator $A(\theta)$ is symmetric; however, our formulation is readily extended [16] to treat both noncompliant outputs, $s(\mu) = \langle L, u(\mu)\rangle$ for given $L \in Y'$, and non-symmetric (but still coercive) operators.)

We may also express our output as

$$s(\mu) = \langle F, A^{-1}(\Theta(\mu))F\rangle \cdot \tag{1.2}$$

Here, for any $\theta \in \mathcal{D}^\theta_{\rm box}$, $A^{-1}(\theta)\colon Y' \to Y$ is the (continuous, symmetric, coercive) inverse of $A(\theta)$; in particular, $\forall\, G \in Y'$, $\langle A(\theta)\, A^{-1}(\theta)\, G, v\rangle = \langle G, v\rangle$, $\forall\, v \in Y$.

1.2. "Truth" approximation

The $u(\mu)$ of (1.1) are, in general, not calculable. In order to construct our reduced-basis space we will therefore require a *finite*-dimensional "truth" approximation to Y, which we shall denote $\tilde{Y}$; $\tilde{Y}$ is an $\mathcal{N}$-dimensional subspace of Y. For example, for $\Omega \subset \mathbb{R}^{d=1,\,2,\,\text{or}\,3}$, and $Y \subset H^1(\Omega) \equiv \{v \in L^2(\Omega),\ \nabla v \in (L^2(\Omega))^d\}$ (here $L^2(\Omega)$ is the space of square-integrable functions over Ω), $\tilde{Y}$ will typically be a finite element approximation space associated with a very fine triangulation of Ω. In general, we expect that $\mathcal{N}$ will be very large.

Our (Galerkin) truth approximation can be stated as: given a $\mu \in \mathcal{D}^\mu$, evaluate the output

$$\tilde{s}(\mu) = \langle F, \tilde{u}(\mu)\rangle, \tag{1.3}$$

where $\tilde{u}(\mu) \in \tilde{Y}$ is the unique solution of

$$\langle A(\Theta(\mu))\, \tilde{u}(\mu), v\rangle = \langle F, v\rangle, \quad \forall\, v \in \tilde{Y}. \tag{1.4}$$

As before, the output can be expressed as a (strictly positive) energy: $\tilde{s}(\mu) = \langle F, \tilde{u}(\mu)\rangle = \langle A(\Theta(\mu))\, \tilde{u}(\mu), \tilde{u}(\mu)\rangle$.

It shall prove convenient to express (1.3, 1.4) in terms of a (in fact, any) basis for $\tilde{Y}$, $\{\phi_i,\ i = 1, \ldots, \mathcal{N}\}$. To wit, we first introduce the matrices $\underline{\tilde{A}}_q \in \mathbb{R}^{\mathcal{N}\times\mathcal{N}}$, $q = 0, \ldots, Q$, as $\tilde{A}_{q\,i\,j} = \langle A\phi_j, \phi_i\rangle$, $1 \leq i, j \leq \mathcal{N}$; it is readily shown that $\underline{\tilde{A}}_0$ (respectively, $\underline{\tilde{A}}_q$, $q = 1, \ldots, Q$) is symmetric positive-definite (respectively, symmetric positive-semidefinite). For any $\theta \in \mathcal{D}^\theta_{\rm box}$, we then define $\underline{\tilde{A}}(\theta) \in \mathbb{R}^{\mathcal{N}\times\mathcal{N}}$ as

$$\underline{\tilde{A}}(\theta) = \underline{\tilde{A}}_0 + \sum_{q=1}^{Q} \theta_q\, \underline{\tilde{A}}_q;$$

$\underline{\tilde{A}}(\theta)$ is symmetric positive-definite for all $\theta \in \mathcal{D}^\theta_{\rm box}$. In a similar fashion we introduce $\underline{\tilde{F}} \in \mathbb{R}^{\mathcal{N}}$ as $\tilde{F}_i = \langle F, \phi_i\rangle$, $1 \leq i \leq \mathcal{N}$.

Our truth approximation can then be restated as: given a $\mu \in \mathcal{D}^\mu$, evaluate the output

$$\tilde{s}(\mu) = \underline{\tilde{F}}^T \underline{\tilde{u}}(\mu),$$

where $\underline{\tilde{u}}(\mu) \in \mathbb{R}^{\mathcal{N}}$ is the unique solution of

$$\underline{\tilde{A}}(\Theta(\mu))\, \underline{\tilde{u}}(\mu) = \underline{\tilde{F}}\ ; \tag{1.5}$$

here ^{T} refers to the algebraic transpose. Note that $\tilde{u}(\mu)$ and $\underline{\tilde{u}}(\mu) = (\tilde{u}_1, \ldots, \tilde{u}_{\mathcal{N}})$ are related *via*

$$\tilde{u}(\mu) = \sum_{j=1}^{\mathcal{N}} \tilde{u}_j(\mu)\, \phi_j.$$

As always, our compliance output can be expressed as an energy:

$$\tilde{s}(\mu) = \underline{\tilde{u}}^T(\mu)\, \underline{\tilde{A}}(\Theta(\mu))\, \underline{\tilde{u}}(\mu),$$

or, equivalently,

$$\tilde{s}(\mu) = \underline{\tilde{F}}^T\, \underline{\tilde{A}}^{-1}(\Theta(\mu))\, \underline{\tilde{F}}, \tag{1.6}$$

where $\underline{\tilde{A}}^{-1}(\theta)$ is the (symmetric, positive-definite) inverse of $\underline{\tilde{A}}(\theta)$. Note that since $\mathcal{N}$ is large, solution of (1.5), and hence evaluation of $\tilde{s}(\mu)$, will be computationally expensive.

1.3. **Monotonicity and convexity of the inverse**

In this section we prove that the quadratic forms associated with $A^{-1}(\theta)$ and $\underline{\tilde{A}}^{-1}(\theta)$ are monotonic and convex in the parameter θ. To begin, we define $\mathcal{J} : \mathcal{D}^\theta_{\text{box}} \times Y' \to \mathbb{R}$ and $\tilde{\mathcal{J}} : \mathcal{D}^\theta_{\text{box}} \times \mathbb{R}^{\mathcal{N}} \to \mathbb{R}$ as

$$\mathcal{J}(\theta, G) = \langle G, A^{-1}(\theta)\, G\rangle,$$

$$\tilde{\mathcal{J}}(\theta, \underline{G}) = \underline{G}^T\, \underline{\tilde{A}}^{-1}(\theta)\, \underline{G}\ . \tag{1.7}$$

Also, given $\theta^1 \in \mathcal{D}^\theta_{\text{box}}$, $\theta^2 \in \mathcal{D}^\theta_{\text{box}}$, and $\tau \in [0,1]$, we define $\mathcal{J}_{\text{seg}}(\tau; \theta^1, \theta^2; G) = \mathcal{J}(\theta^1 + \tau(\theta^2 - \theta^1), G)$, and $\tilde{\mathcal{J}}_{\text{seg}}(\tau; \theta^1, \theta^2; \underline{G}) = \tilde{\mathcal{J}}(\theta^1 + \tau(\theta^2 - \theta^1), \underline{G})$.

We also define $A_{\text{seg}}(\tau; \theta^1, \theta^2) = A(\theta^1 + \tau(\theta^2 - \theta^1))$ and $\underline{\tilde{A}}_{\text{seg}}(\tau; \theta^1, \theta^2) = \underline{\tilde{A}}(\theta^1 + \tau(\theta^2 - \theta^1))$. We can then write

$$A_{\text{seg}}(\tau; \theta^1, \theta^2) = A_0 + \sum_{q=1}^{Q} \theta^1_q\, A_q + \tau \sum_{q=1}^{Q} (\theta^2_q - \theta^1_q)\, A_q,$$

$$\underline{\tilde{A}}_{\text{seg}}(\tau; \theta^1, \theta^2) = \underline{\tilde{A}}_0 + \sum_{q=1}^{Q} \theta^1_q\, \underline{\tilde{A}}_q + \tau \sum_{q=1}^{Q} (\theta^2_q - \theta^1_q)\, \underline{\tilde{A}}_q. \tag{1.8}$$

Note that $\mathcal{J}_{\text{seg}}(\tau; \theta^1, \theta^2; G) = \langle G, A^{-1}_{\text{seg}}(\tau; \theta^1, \theta^2)\, G\rangle$ and $\tilde{\mathcal{J}}_{\text{seg}}(\tau; \theta^1, \theta^2; \underline{G}) = \underline{G}^T\, \underline{\tilde{A}}^{-1}_{\text{seg}}(\tau; \theta^1, \theta^2)\, \underline{G}$.

We first consider monotonicity in

Proposition 1.1. *$\mathcal{J}(\theta, G)$ and $\tilde{\mathcal{J}}(\theta, \underline{G})$ are non-increasing functions: for any $\theta^1 \in \mathcal{D}^\theta_{\text{box}}$, $\theta^2 \in \mathcal{D}^\theta_{\text{box}}$, such that $\theta^2 \geq \theta^1$ (i.e., $\theta^2_q \geq \theta^1_q$, $q = 1, \ldots, Q$), $\mathcal{J}(\theta^2, G) \leq \mathcal{J}(\theta^1, G)$ for any $G \in Y'$, and $\tilde{\mathcal{J}}(\theta^2, \underline{G}) \leq \tilde{\mathcal{J}}(\theta^1, \underline{G})$ for any $\underline{G} \in \mathbb{R}^{\mathcal{N}}$.*

Proof. We give the proof for $\tilde{\mathcal{J}}(\theta, \underline{G})$; similar arguments apply to $\mathcal{J}(\theta, G)$.

We need only demonstrate that, for any (fixed) θ^1, θ^2 such that $\theta^2 \geq \theta^1$, and any (fixed) $\underline{G} \in \mathbb{R}^{\mathcal{N}}$,

$$\frac{\mathrm{d}\tilde{\mathcal{J}}_{\text{seg}}(\tau; \theta^1, \theta^2; \underline{G})}{\mathrm{d}\tau} \leq 0, \quad \forall\, \tau \in [0, 1].$$

To evaluate $\mathrm{d}\tilde{\mathcal{J}}_{\text{seg}}/\mathrm{d}\tau$, we note that

$$\frac{\mathrm{d}\tilde{\mathcal{J}}_{\text{seg}}(\tau; \theta^1, \theta^2; \underline{G})}{\mathrm{d}\tau} = \underline{G}^T \, \frac{\mathrm{d}}{\mathrm{d}\tau} \left(\underline{\tilde{A}}^{-1}_{\text{seg}}(\tau; \theta^1, \theta^2) \right) \underline{G};$$

it thus remains only to show that $\mathrm{d}(\underline{\tilde{A}}^{-1}_{\text{seg}}(\tau; \theta^1, \theta^2))/\mathrm{d}\tau$ is symmetric negative-semidefinite.

To this end, we note that $\underline{\tilde{A}}^{-1}_{\text{seg}}(\tau; \theta^1, \theta^2)\, \underline{\tilde{A}}_{\text{seg}}(\tau; \theta^1, \theta^2) = \underline{\text{Id}}$ (the identity), and thus

$$\frac{\mathrm{d}}{\mathrm{d}\tau} \left(\underline{\tilde{A}}^{-1}_{\text{seg}}(\tau; \theta^1, \theta^2) \right) \underline{\tilde{A}}_{\text{seg}}(\tau; \theta^1, \theta^2) + \underline{\tilde{A}}^{-1}_{\text{seg}}(\tau; \theta^1, \theta^2) \, \frac{\mathrm{d}}{\mathrm{d}\tau} \, \left(\underline{\tilde{A}}_{\text{seg}}(\tau; \theta^1, \theta^2) \right) = \underline{0}.$$

Application of (1.8) then yields

$$\frac{\mathrm{d}}{\mathrm{d}\tau} \left(\underline{\tilde{A}}^{-1}_{\text{seg}}(\tau; \theta^1, \theta^2) \right) = -\underline{\tilde{A}}^{-1}_{\text{seg}}(\tau; \theta^1, \theta^2) \left(\sum_{q=1}^{Q} (\theta^2_q - \theta^1_q)\, \underline{\tilde{A}}_q \right) \underline{\tilde{A}}^{-1}_{\text{seg}}(\tau; \theta^1, \theta^2);$$

the desired result then directly follows, since $\theta^2_q \geq \theta^1_q$, and the $\underline{\tilde{A}}_q$ are symmetric positive-semidefinite. □

We next consider convexity in

Proposition 1.2. *$\mathcal{J}(\theta, G)$ and $\tilde{\mathcal{J}}(\theta, \underline{G})$ are convex functions of θ: for any $\theta^1 \in \mathcal{D}^\theta_{\text{box}}$, $\theta^2 \in \mathcal{D}^\theta_{\text{box}}$, and for all $\tau \in [0, 1]$, $\mathcal{J}(\theta^1 + \tau(\theta^2 - \theta^1), G) \leq (1 - \tau)\, \mathcal{J}(\theta^1, G) + \tau \mathcal{J}(\theta^2, G)$ for any $G \in Y'$, and $\tilde{\mathcal{J}}(\theta^1 + \tau(\theta^2 - \theta^1), \underline{G}) \leq (1 - \tau)\, \tilde{\mathcal{J}}(\theta^1, \underline{G}) + \tau \tilde{\mathcal{J}}(\theta^2, \underline{G})$ for any $\underline{G} \in \mathbb{R}^{\mathcal{N}}$.*

Proof. We give the proof for $\tilde{\mathcal{J}}(\theta, \underline{G})$; similar arguments apply to $\mathcal{J}(\theta, G)$.

We need to demonstrate that, for any $\theta^1 \in \mathcal{D}^\theta_{\text{box}}$, $\theta^2 \in \mathcal{D}^\theta_{\text{box}}$, and $\tau \in [0, 1]$, $\tilde{\mathcal{J}}_{\text{seg}}(\tau; \theta^1, \theta^2, \underline{G}) \leq (1 - \tau)\tilde{\mathcal{J}}_{\text{seg}}(0; \theta^1, \theta^2; \underline{G}) + \tau \tilde{\mathcal{J}}_{\text{seg}}(1; \theta^1, \theta^2; \underline{G})$ for any $\underline{G} \in \mathbb{R}^{\mathcal{N}}$. From standard results in convex analysis [4] it suffices to show that, for any (fixed) $\underline{G} \in \mathbb{R}^{\mathcal{N}}$,

$$\frac{\mathrm{d}^2 \tilde{\mathcal{J}}_{\text{seg}}(\tau; \theta^1, \theta^2; \underline{G})}{\mathrm{d}\tau^2} \geq 0, \quad \forall\, \tau \in [0, 1].$$

From the definition of $\tilde{\mathcal{J}}_{\text{seg}}(\tau; \theta^1, \theta^2; \underline{G})$, it thus remains only to show that $\mathrm{d}^2(\underline{\tilde{A}}^{-1}_{\text{seg}}(\tau; \theta^1, \theta^2))/\mathrm{d}\tau^2$ is symmetric positive-semidefinite.

To this end, we continue the differentiation of Proposition 1.1 to obtain

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}\left(\underline{\tilde{A}}_{\mathrm{seg}}^{-1}(\tau;\theta^1,\theta^2)\right) &= -\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\underline{\tilde{A}}_{\mathrm{seg}}^{-1}(\tau;\theta^1,\theta^2)\right)\left(\sum_{q=1}^{Q}(\theta_q^2-\theta_q^1)\underline{\tilde{A}}_q\right)\underline{\tilde{A}}_{\mathrm{seg}}^{-1}(\tau;\theta^1,\theta^2)\\ &\quad -\underline{\tilde{A}}_{\mathrm{seg}}^{-1}(\tau;\theta^1,\theta^2)\left(\sum_{q=1}^{Q}(\theta_q^2-\theta_q^1)\,\underline{\tilde{A}}_q\right)\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\underline{\tilde{A}}_{\mathrm{seg}}^{-1}(\tau;\theta^1,\theta^2)\right)\\ &= 2\underline{\tilde{A}}_{\mathrm{seg}}^{-1}(\tau;\theta^1,\theta^2)\Big(\textstyle\sum_{q=1}^{Q}(\theta_q^2-\theta_q^1)\,\underline{\tilde{A}}_q\Big)\,\underline{\tilde{A}}_{\mathrm{seg}}^{-1}(\tau;\theta^1,\theta^2)\Big(\textstyle\sum_{q=1}^{Q}(\theta_q^2-\theta_q^1)\,\underline{\tilde{A}}_q\Big)\,\underline{\tilde{A}}_{\mathrm{seg}}^{-1}(\tau;\theta^1,\theta^2).\end{aligned}$$

The desired result then directly follows since $\underline{\tilde{A}}_{\mathrm{seg}}^{-1}(\tau;\theta^1,\theta^2)$ is symmetric positive-definite. □

Note that we can deduce from Propositions 1.1 and 1.2, and the relations (1.2) and (1.6), various properties of the parametric dependence of the output: for example, in the simple case in which $P = Q$, $\mathcal{D}^\mu$ is a convex set in $\mathbb{R}^P_+$, and $\Theta_q(\mu) = \mu_q$, $q = 1,\ldots,Q$, we directly obtain the result that $s(\mu)$ and $\tilde{s}(\mu)$ are non-increasing, convex functions of μ. The true value of Propositions 1.1 and 1.2, however, will be in constructing bound conditioners.

2. Reduced-basis output bounds

2.1. **Preliminaries**

We first introduce a "μ" sample $S_N^\mu = \{\mu^1,\ldots,\mu^N\}$, where $\mu^n \in \mathcal{D}^\mu$, $n = 1,\ldots,N$. We then define our reduced-basis space $W_N = \text{span}\,\{\tilde{\zeta}_n,\ n = 1,\ldots,N\}$, where $\tilde{\zeta}_n = \tilde{u}(\mu^n)$, $n = 1,\ldots,N$. Recall that $\tilde{u}(\mu^n)$ is the solution of (1.4) for $\mu = \mu^n$. For future reference we denote $\underline{\tilde{\zeta}}_n = \underline{\tilde{u}}(\mu^n)$, $n = 1,\ldots,N$.

We next introduce a "θ" sample $S_M^\theta = \{\theta^1,\ldots,\theta^M\}$, where $\theta^m \in \mathcal{D}^\theta_{\mathrm{box}}$, $m = 1,\ldots,M$. To each μ in $\mathcal{D}^\mu$ we then associate (i) a set of $|\mathcal{E}(\mu)|$ indices $\mathcal{E}(\mu) \subset \{1,\ldots,M\}$, and (ii) a point in $\mathcal{D}^\theta_{\mathrm{box}}$, $\overline{\theta}(\mu) \le \Theta(\mu)$, such that

$$\overline{\theta}(\mu) = \sum_{j\in\mathcal{E}(\mu)} \alpha_j(\mu)\,\theta^j$$

for a given set of coefficients $\alpha_j(\mu)$ satisfying $0 \le \alpha_j(\mu) \le 1$, $\forall\, j \in \mathcal{E}(\mu)$, and $\sum_{j\in\mathcal{E}(\mu)}\ \alpha_j(\mu) = 1$. We implicitly assume that S_M^θ is chosen such that, for all $\mu \in \mathcal{D}^\mu$, such a construction is possible; a deficient sample S_M^θ can always be rendered compliant simply by replacing one point with $\theta^{\min}$.

We now introduce our bound conditioner $\underline{\tilde{B}}(\mu) \in \mathbb{R}^{\mathcal{N}\times\mathcal{N}}$ as

$$\underline{\tilde{B}}(\mu) = \left(\sum_{j\in\mathcal{E}(\mu)} \alpha_j(\mu)\underline{\tilde{A}}^{-1}(\theta^j)\right)^{-1}. \tag{2.1}$$

Clearly, $\underline{\tilde{B}}^{-1}(\mu)$ and hence $\underline{\tilde{B}}(\mu)$ are symmetric positive-definite. In words, $\underline{\tilde{B}}^{-1}(\mu)$ is an approximation to $\underline{\tilde{A}}^{-1}(\Theta(\mu))$ constructed as a convex combination of $\underline{\tilde{A}}^{-1}$ at "neighboring" θ. We shall consider three different bound conditioners in this paper.

The first is a single-point conditioner, and will be labeled SP. Here we set $M = 1$, $S_M^\theta = \{\theta^{\min}\}$, $|\mathcal{E}(\mu)| = 1$, $\mathcal{E}(\mu) = \{1\}$, and $\overline{\theta}(\mu) = \theta^{\min}$. This conditioner is a special case of our earlier bound conditioner formulation [10,16], in which we take $\underline{\tilde{B}}(\mu) = g(\mu)\,\underline{\hat{A}}$ ($\underline{\hat{A}}$ independent of μ); SP corresponds to $g(\mu) = 1$, $\underline{\hat{A}} = \underline{\tilde{A}}(\theta^{\min})$. Note in our earlier work we typically choose not SP, but rather a different single-point conditioner given by $g(\mu) = \min(1, \min(\Theta_q(\mu),\ q = 1,\ldots,Q))$, $\underline{\hat{A}} = \underline{\tilde{A}}(\theta = (1,\ldots,1))$. We do not consider the development of this "$\min(1,\theta)$" conditioner any further in this paper, since first, it does not readily fit into the current "convex

approximation" context, and second, except for $\Theta(\mu)$ close to $(1,\ldots,1)$, it yields *worse* results than SP – in particular for $\min(\Theta_q(\mu),\ q=1,\ldots,Q)$ small. However, in the numerical experiments of Section 4, we will include the "$\min(1,\theta)$" conditioner results (labeled as SP$'$).

The second bound conditioner we develop here is piecewise-constant, and will be labeled PC. Now we set $M \geq 1$, $S_M^\theta = \{\theta^1 = \theta^{\min}, \theta^2, \ldots, \theta^M\}$, and $|\mathcal{E}(\mu)| = 1$, and choose $\mathcal{E}(\mu) = \{j_1(\mu)\}$ such that $\overline{\theta}(\mu) \equiv \theta^{j_1(\mu)} \leq \Theta(\mu)$. There will often be many possible choices for $j_1(\mu)$; we can either establish a definition of closeness, or alternatively consider all possible candidates and select the best (in the sense of yielding the lowest upper bound as defined in Sect. 2.2).

The third bound conditioner we develop here is piecewise-linear, and will be labeled PL. Now we set $M \geq Q+1$, $S_M^\theta = \{\theta^1, \ldots, \theta^M\}$, and $|\mathcal{E}(\mu)| = Q+1$, and choose $\mathcal{E}(\mu)$ such that the θ^j, $j \in \mathcal{E}(\mu)$, form a $(Q+1)$-simplex containing $\overline{\theta}(\mu) \equiv \Theta(\mu)$. Again, there will often be several choices for the index set $\mathcal{E}(\mu)$ and associated simplex; we can either establish an *a priori* criterion for goodness (*e.g.*, related to simplex size), or instead evaluate all candidates and select the best in the sense of "lowest upper bound". Note that, for μ for which S_M^θ contains no $\Theta(\mu)$-containing $(Q+1)$-simplex, we must accept a lower-order simplex and $\overline{\theta}(\mu) < \Theta(\mu)$ (*e.g.*, in the worst case, we revert to PC).

2.2. **Two-step approximation**

The importance of this two-step procedure will become clearer in Section 3. In the first step we compute our *predictor*, $s_N(\mu)$; in the second step we compute our *bounds*, $s_N^-(\mu) \leq \tilde{s}(\mu) \leq s_N^+(\mu)$. Although the latter may be gainfully interpreted as *a posteriori* estimators, we prefer to view the bounds as "improved" predictors imbued with a sense of direction – and hence certainty.

2.2.1. *Predictor*

In the first step, given a $\mu \in \mathcal{D}^\mu$, we find $s_N(\mu) = \langle F, u_N(\mu)\rangle$, where $u_N(\mu) \in W_N$ satisfies

$$\langle A(\Theta(\mu))\, u_N(\mu), v\rangle = \langle F, v\rangle, \quad \forall\, v \in W_N.$$

We may also express the output as an energy, $s_N(\mu) = \langle A(\Theta(\mu))\, u_N(\mu), u_N(\mu)\rangle$.

In terms of our basis functions, we can define the symmetric positive-definite matrix $\underline{A}_N(\mu) \in \mathbb{R}^{N\times N}$ as $A_{N\,i\,j}(\mu) = \langle A(\Theta(\mu))\,\tilde{\zeta}_j, \tilde{\zeta}_i\rangle$, $1 \leq i,j \leq N$, and the vector $\underline{F}_N \in \mathbb{R}^N$ as $F_{N\,i} = \langle F, \tilde{\zeta}_i\rangle$, $1 \leq i \leq N$. It is a simple matter to observe that

$$\underline{A}_N(\theta) = \underline{A}_{N\,0} + \sum_{q=1}^{Q} \theta_q\, \underline{A}_{N\,q}, \tag{2.2}$$

where $(A_{Nq})_{i\,j} = \langle A_q\,\tilde{\zeta}_j, \tilde{\zeta}_i\rangle$, $1 \leq i,j \leq N$, $0 \leq q \leq Q$; note that the $\underline{A}_{Nq} \in \mathbb{R}^{N\times N}$, $0 \leq q \leq Q$, are *independent* of θ.

Our first step can then be restated as: given a $\mu \in \mathcal{D}^\mu$, find $s_N(\mu) = \underline{F}_N^T \underline{u}_N(\mu)$, where $\underline{u}_N(\mu) \in \mathbb{R}^N$ is the unique solution to

$$\underline{A}_N(\Theta(\mu))\, \underline{u}_N(\mu) = \underline{F}_N.$$

Note that $u_N(\mu) = \sum_{j=1}^N u_{N\,j}(\mu)\,\tilde{\zeta}_j$. The output may also be expressed as $s_N(\mu) = \underline{u}_N^T(\mu)\, \underline{A}_N(\Theta(\mu))\, \underline{u}_N(\mu)$ $= \underline{F}_N^T\, \underline{A}_N^{-1}(\Theta(\mu))\, \underline{F}_N$.

2.2.2. *Lower and upper bounds*

We first define our residual $R \in Y'$ as $\langle R(\mu), v\rangle \equiv \langle F - A(\Theta(\mu))\, u_N(\mu), v\rangle$, $\forall\, v \in Y$; and then $\underline{\tilde{R}}(\mu) \in \mathbb{R}^{\mathcal{N}}$ as $\tilde{R}_i(\mu) = \langle R(\mu), \phi_i\rangle$ $i = 1,\ldots,\mathcal{N}$. We note for future reference that

$$\underline{\tilde{R}}(\mu) = \underline{\tilde{F}} - \underline{\tilde{A}}(\Theta(\mu))\, \underline{\tilde{u}}_N(\mu), \tag{2.3}$$

where $\underline{\tilde{u}}_N(\mu) \in \mathbb{R}^{\mathcal{N}}$ is given by

$$\underline{\tilde{u}}_N(\mu) = \sum_{n=1}^{N} u_{Nn}(\mu)\, \underline{\tilde{\zeta}}_n; \tag{2.4}$$

by construction, $u_N(\mu) = \sum_{i=1}^{\mathcal{N}} \tilde{u}_{N\,i}(\mu)\, \phi_i$.

We now find $\underline{\hat{e}}(\mu) \in \mathbb{R}^{\mathcal{N}}$ such that

$$\underline{\tilde{B}}(\mu)\, \underline{\hat{e}}(\mu) = \underline{\tilde{R}}(\mu); \tag{2.5}$$

this equation will of course have a unique solution since $\underline{\tilde{B}}(\mu)$ is symmetric positive-definite.

We can now define our lower and upper bounds as

$$s_N^-(\mu) = s_N(\mu),$$

and

$$s_N^+(\mu) = s_N(\mu) + \Delta_N(\mu),$$

where $\Delta_N(\mu)$, the bound gap, is given by

$$\begin{aligned}
\Delta_N(\mu) &\equiv \underline{\hat{e}}^T(\mu)\, \underline{\tilde{B}}(\mu)\, \underline{\hat{e}}(\mu) \\
&= \underline{\tilde{R}}^T(\mu)\, \underline{\tilde{B}}^{-1}(\mu)\, \underline{\tilde{R}}(\mu) \\
&= \underline{\tilde{R}}^T(\mu)\, \underline{\hat{e}}(\mu).
\end{aligned}$$

The first two expressions (respectively, third expression) for the bound gap will prove useful in the theoretical (respectively, computational) context.

2.3. **Bounding properties**

It remains to demonstrate our claim that $s_N^-(\mu) \le \tilde{s}(\mu) \le s_N^+(\mu)$ for all $N \ge 1$. We first consider

Proposition 2.1. *For all $\mu \in \mathcal{D}^\mu$, and all $N \ge 1$, $s_N^-(\mu) \le \tilde{s}(\mu)$.*

Proof. We have that

$$\begin{aligned}
\tilde{s}(\mu) - s_N(\mu) &= \langle F, \tilde{u}(\mu) - u_N(\mu) \rangle \\
&= \langle A(\Theta(\mu))\, \tilde{u}(\mu), \tilde{u}(\mu) - u_N(\mu) \rangle \\
&= \langle A(\Theta(\mu))\, (\tilde{u}(\mu) - u_N(\mu)), \tilde{u}(\mu) \rangle \\
&= \langle A(\Theta(\mu))\, (\tilde{u}(\mu) - u_N(\mu)), \tilde{u}(\mu) - u_N(\mu) \rangle \qquad (2.6) \\
&\ge 0
\end{aligned}$$

from the definition of $\tilde{s}(\mu)$, equation (1.4), symmetry of A, Galerkin orthogonality, and coercivity, respectively. □

This lower bound proof is a standard result in variational approximation theory. We now turn to the less trivial upper bound in

Proposition 2.2. *For all $\mu \in \mathcal{D}^\mu$, and all $N \geq 1$, $s_N^+(\mu) \geq \tilde{s}(\mu)$.*

Proof. We first define $\underline{\tilde{e}} \in \mathbb{R}^{\mathcal{N}}$ as $\underline{\tilde{e}} = \underline{\tilde{u}} - \underline{\tilde{u}}_N$; we then note from (1.5) and (2.3) that

$$\underline{\tilde{A}}(\Theta(\mu))\, \underline{\tilde{e}}(\mu) = \underline{\tilde{R}}(\mu), \tag{2.7}$$

which is the usual error-residual relationship. It then follows from (2.6) of Proposition 2.1 that

$$\begin{aligned} \tilde{s}(\mu) - s_N(\mu) &= \underline{\tilde{e}}^T(\mu)\, \underline{\tilde{A}}(\Theta(\mu))\, \underline{\tilde{e}}(\mu) \\ &= \underline{\tilde{R}}^T(\mu)\, \underline{\tilde{A}}^{-1}(\Theta(\mu))\, \underline{\tilde{R}}(\mu). \end{aligned}$$

It thus only remains to prove that

$$\eta_N(\mu) \equiv \frac{s_N^+(\mu) - s_N(\mu)}{\tilde{s}(\mu) - s_N(\mu)} = \frac{\Delta_N(\mu)}{\tilde{s}(\mu) - s_N(\mu)} = \frac{\underline{\tilde{R}}^T(\mu)\, \underline{\tilde{B}}^{-1}(\mu)\, \underline{\tilde{R}}(\mu)}{\underline{\tilde{R}}^T(\mu)\, \underline{\tilde{A}}^{-1}(\Theta(\mu))\, \underline{\tilde{R}}(\mu)} \tag{2.8}$$

is greater than unity; note $\eta_N(\mu)$ is denoted the *effectivity*.

From the definitions (1.7) and (2.1) we immediately note that

$$\eta_N(\mu) = \frac{\displaystyle\sum_{j \in \mathcal{E}(\mu)} \alpha_j(\mu)\, \tilde{\mathcal{J}}(\theta^j, \underline{\tilde{R}}(\mu))}{\tilde{\mathcal{J}}(\Theta(\mu), \underline{\tilde{R}}(\mu))}.$$

But from the construction of the $\alpha_j(\mu)$, the choice of $\mathcal{E}(\mu)$, Proposition 1.2, classical results in convex analysis, and Proposition 1.1, it directly follows that, for any $\underline{G} \in \mathbb{R}^{\mathcal{N}}$ (and therefore for $\underline{G} = \underline{\tilde{R}}(\mu)$),

$$\sum_{j \in \mathcal{E}(\mu)} \alpha_j(\mu)\, \tilde{\mathcal{J}}(\theta^j, \underline{G}) \geq \tilde{\mathcal{J}}(\overline{\theta}(\mu), \underline{G}) \geq \tilde{\mathcal{J}}(\Theta(\mu), \underline{G}),$$

which concludes the proof. □

We must now address the computation of $s_N^-(\mu)$ and $s_N^+(\mu)$.

2.4. **Computational procedure: Off-line/on-line decomposition**

2.4.1. *The predictor $s_N(\mu)$*

We review here arguments given in great detail in [16]; early applications of this approach may be found in [5].

In an *off-line* stage, we find the $\underline{\tilde{\zeta}}_n$, $n = 1, \ldots, N$ (N $\underline{\tilde{A}}$-solves), and form the $\underline{A}_{N\,q}$, $0 \leq q \leq Q$ ($(Q+1)N^2$ $\underline{\tilde{A}}$-inner products), and $\underline{F}_N$ ($N\mathcal{N}$ operations). In the *on-line* stage – given any new μ – we need only form $\underline{A}_N(\mu)$ from the $\underline{A}_{Nq}$ ($(Q+1)N^2$ operations), find $\underline{u}_N(\mu)$ ($O(N^3)$ operations), and evaluate $s_N(\mu)$ (N operations). The essential point is that the on-line complexity (and storage – $O(QN^2)$) is *independent* of the very large dimension of the truth space $\tilde{Y}$, $\mathcal{N}$; in particular, since N is typically very small (see the *a priori* results of Sect. 3 and the numerical results of Sect. 4), "real-time" response is obtained.

2.4.2. *The upper bound $s_N^+(\mu)$*

The arguments here differ slightly from those presented in [16] for our simpler bound conditioners. We first note from (2.1–2.5) that

$$\underline{\hat{e}}(\mu) = \sum_{j\in\mathcal{E}(\mu)} \alpha_j(\mu)\, \underline{\tilde{A}}^{-1}(\theta^j) \left[\underline{\tilde{F}} - \sum_{q=0}^{Q}\sum_{n=1}^{N} \Theta_q(\mu)\, u_{Nn}(\mu)\, \underline{\tilde{A}}_q\, \underline{\tilde{\zeta}}_n\right];$$

recall that $\Theta_0 = 1$. It follows that we may express $\underline{\hat{e}}(\mu)$ as

$$\underline{\hat{e}}(\mu) = \sum_{j\in\mathcal{E}(\mu)} \alpha_j(\mu) \left[\underline{\tilde{z}}_{00}^j + \sum_{q=0}^{Q}\sum_{n=1}^{N} \Theta_q(\mu)\, u_{Nn}(\mu)\, \underline{\tilde{z}}_{qn}^j\right],$$

where for all $j \in \{1,\ldots,M\}$, $\underline{\tilde{A}}(\theta^j)\underline{\tilde{z}}_{00}^j = \underline{\tilde{F}}$, and $\underline{\tilde{A}}(\theta^j)\underline{\tilde{z}}_{qn}^j = -\underline{\tilde{A}}_q\, \underline{\tilde{\zeta}}_n$, $0 \le q \le Q$, $1 \le n \le N$. We may thus express our bound gap $\Delta_N(\mu)$ as

$$\Delta_N(\mu) = \underline{\tilde{R}}^T(\mu)\, \underline{\hat{e}}(\mu) = \sum_{j\in\mathcal{E}(\mu)} \alpha_j(\mu) \left[\underline{\tilde{F}} - \sum_{q=0}^{Q}\sum_{n=1}^{N} \Theta_q(\mu)\, u_{Nn}(\mu)\, \underline{\tilde{A}}_q\, \underline{\tilde{\zeta}}_n\right]^T \left[\underline{\tilde{z}}_{00}^j + \sum_{q'=0}^{Q}\sum_{n'=1}^{N} \Theta_{q'}(\mu)\, u_{Nn'}(\mu)\, \underline{\tilde{z}}_{q'n'}^j\right]$$

$$= \sum_{j\in\mathcal{E}(\mu)} \alpha_j(\mu) \left[c^j + \sum_{q=0}^{Q}\sum_{n=1}^{N} \Theta_q(\mu)\, u_{Nn}(\mu)\, \Lambda_{qn}^j \right.$$

$$\left. + \sum_{q=0}^{Q}\sum_{n=1}^{N}\sum_{q'=0}^{Q}\sum_{n'=1}^{N} \Theta_q(\mu)\, \Theta_{q'}(\mu)\, u_{Nn}(\mu)\, u_{Nn'}(\mu)\, \Gamma_{qq'nn'}^j\right], \tag{2.9}$$

where for all $j \in \{1,\ldots,M\}$, $c^j = \underline{\tilde{F}}^T\, \underline{\tilde{z}}_{00}^j$, $\Lambda_{qn}^j = \underline{\tilde{F}}^T\, \underline{\tilde{z}}_{qn}^j - \underline{\tilde{\zeta}}_n^T\, \underline{\tilde{A}}_q\, \underline{\tilde{z}}_{00}^j$ for $0 \le q \le Q$, $1 \le n \le N$, and $\Gamma_{qq'nn'}^j = -\underline{\tilde{\zeta}}_n^T\, \underline{\tilde{A}}_q\, \underline{\tilde{z}}_{q'n'}^j$ for $0 \le q, q' \le Q$, $1 \le n, n' \le N$.

The off-line/on-line decomposition is now clear. In the *off-line* stage we compute the $\underline{\tilde{z}}_{00}^j$ and $\underline{\tilde{z}}_{qn}^j$ ($M((Q+1)N+1)$ $\underline{\tilde{A}}$-solves) and the c^j, Λ_{qn}^j, and $\Gamma_{qq'nn'}^j$ (predominated by M $((Q+1)^2N^2+(Q+1)N)$ $\underline{\tilde{A}}$-inner products). In the on-line stage we need "only" perform the sum (2.9), which requires $|\mathcal{E}(\mu)|((Q+1)^2N^2 + (Q+1)N + 1)$ operations. The essential point is that the on-line complexity (and storage – $O(M(Q+1)^2N^2)$) is independent of $\mathcal{N}$. It is true, however, that the Q scaling is not too appealing, in particular for the piecewise-linear bound conditioner (PL) for which $|\mathcal{E}(\mu)| = Q+1$. However, in general, for Q not too large, real-time (on-line) response is not compromised; indeed, for $Q \ll N$, the on-line cost is dominated by the calculation of $\underline{u}_N$ ($O(N^3)$ inversion of $\underline{A}_N$), and there is thus little (on-line) reason *not* to choose the more accurate PL conditioner. (As regards on-line storage, we shall have more to say about M in Sect. 4.3.)

We note that the off-line/on-line decomposition depends critically on the "separability" of $\underline{\tilde{B}}^{-1}$ as a sum of products of parameter-*dependent* functions (the $\alpha_j(\mu)$) and parameter-*independent* operators (the $\underline{\tilde{A}}^{-1}(\theta^j)$). In turn, it is the direct approximation of $\underline{\tilde{A}}^{-1}(\Theta(\mu))$ (*i.e.*, by a convex combination of $\underline{\tilde{A}}^{-1}(\theta^j)$) rather than of $\underline{\tilde{A}}(\Theta(\mu))$ (*e.g.*, by a convex combination of $\underline{\tilde{A}}(\theta^j)$) that permits us to achieve this separability while simultaneously pursuing a "high-order" bound conditioner. In particular, a computationally efficient (on-line complexity independent of $\mathcal{N}$) formulation of a piecewise-linear bound conditioner is *not* possible if we insist – as is the case, *de facto*, in the "$g(\mu)$" formulation – on direct approximation of $\underline{\tilde{A}}(\Theta(\mu))$.

Of course, the purpose of higher order bound conditioners is to achieve some fixed (known, certain) accuracy – as measured by $\Delta_N(\mu)$ – at lower computational effort; we must therefore understand the convergence properties

of $\Delta_N(\mu)$ for our different bound conditioners. In Section 3 we present an *a priori* theory for $\Delta_N(\mu)$ for the particular case $P = Q = 1$. And in Section 4 we present numerical results that corroborate our $P = Q = 1$ theory, and that provide empirical evidence that the method continues to perform well even for $P > 1$, $Q > 1$. In Sections 3 and 4 we also re-address computational complexity.

3. *A PRIORI* THEORY: $P = Q = 1$

3.1. General framework

We first introduce an *a priori* framework for the general case; we then proceed to the case $P = Q = 1$ in which we can obtain all the necessary estimates.

Depending on the context and application, we will either invoke the lower bound ($s_N^-(\mu)$) or upper bound ($s_N^+(\mu)$) as our estimator for $\tilde{s}(\mu)$. For example, in an optimization exercise in which $\tilde{s}(\mu)$ enters as a constraint $\tilde{s}(\mu) \le s^{\max}$ (respectively, $\tilde{s}(\mu) \ge s^{\min}$), we will replace this condition with $s_N^+(\mu) \le s^{\max}$ (respectively, $s_N^-(\mu) \ge s^{\min}$) so as to ensure satisfaction/feasibility even in the presence of approximation errors. The rigorous bounding properties proven in Section 2.3 provide the requisite certainty.

But we of course also require accuracy: if, in the optimization context cited above, $s_N^+(\mu)$ or $s_N^-(\mu)$ is not close to $\tilde{s}(\mu)$, then our design may be *seriously suboptimal*. Since $|s_N^+(\mu) - \tilde{s}(\mu)| \le |s_N^+(\mu) - s_N^-(\mu)| = \Delta_N(\mu)$ and $|\tilde{s}(\mu) - s_N^-(\mu)| \le |s_N^+(\mu) - s_N^-(\mu)| = \Delta_N(\mu)$, it is the convergence of $\Delta_N(\mu)$ to zero as a function of N that we must understand. In particular, from (2.8) and (2.6) we may write

$$\begin{aligned}\Delta_N(\mu) = s_N^+(\mu) - s_N^-(\mu) &= (\tilde{s}(\mu) - s_N(\mu)) \left(\frac{s_N^+(\mu) - s_N^-(\mu)}{\tilde{s}(\mu) - s_N(\mu)} \right) \\ &= \langle A(\Theta(\mu))\, \tilde{e}(\mu), \tilde{e}(\mu) \rangle\, \eta_N(\mu),\end{aligned}$$

where $\tilde{e}(\mu) = \tilde{u}(\mu) - u_N(\mu)$. In some sense, the first factor, $\langle A(\Theta(\mu))\tilde{e}(\mu), \tilde{e}(\mu)\rangle$, measures the error in the solution $\tilde{u}(\mu) - u_N(\mu)$, while the second factor, the effectivity $\eta_N(\mu)$, measures the ratio of the actual and estimated errors; the former should be small, while the latter should be close to unity (of course approaching from above, as guaranteed by Prop. 2.2). As we shall see, this two-step factorization is important not only as a theoretical construct: it is this factorization which permits us to achieve high accuracy while simultaneously honoring our bound requirements.

The effectivity analysis is facilitated by the introduction of the following generalized eigenvalue problem: given a $\mu \in \mathcal{D}^\mu$, find $(\tilde{\underline{\xi}}_i(\mu) \in \mathbb{R}^{\mathcal{N}}, \rho_i(\mu) \in \mathbb{R})$, $i = 1, \ldots, \mathcal{N}$, such that

$$\underline{\tilde{A}}(\mu)\, \tilde{\underline{\xi}}_i(\mu) = \rho_i(\mu)\, \underline{\tilde{B}}(\mu)\, \tilde{\underline{\xi}}_i(\mu), \tag{3.1}$$

with normalization $\tilde{\underline{\xi}}_i^T(\mu) \underline{\tilde{B}}(\mu) \tilde{\underline{\xi}}_i(\mu) = c_i(\mu)$ (the constant is not important). The eigenvalues are real and positive; we denote the minimum and maximum eigenvalues as $\rho_{\min}(\mu)$ and $\rho_{\max}(\mu)$, respectively.

It is then standard to show that

$$\begin{aligned}\rho_{\min}(\mu) = \min_{\underline{v} \in \mathbb{R}^{\mathcal{N}}} \frac{\underline{v}^T\, \underline{\tilde{A}}(\Theta(\mu))\, \underline{v}}{\underline{v}^T\, \underline{\tilde{B}}(\mu)\, \underline{v}} &= \min_{\underline{w} \in \mathbb{R}^{\mathcal{N}}} \frac{\underline{w}^T\, \underline{\tilde{B}}^{-1/2}(\mu)\, \underline{\tilde{A}}(\Theta(\mu))\, \underline{\tilde{B}}^{-1/2}(\mu)\, \underline{w}}{\underline{w}^T\, \underline{w}} \\ &= \min_{\underline{w} \in \mathbb{R}^{\mathcal{N}}} \frac{\underline{w}^T\, \underline{w}}{\underline{w}^T\, \underline{\tilde{B}}^{1/2}(\mu)\, \underline{\tilde{A}}^{-1}(\Theta(\mu))\, \underline{\tilde{B}}^{1/2}(\mu)\, \underline{w}} \\ &= \min_{\underline{z} \in \mathbb{R}^{\mathcal{N}}} \frac{\underline{z}^T\, \underline{\tilde{B}}^{-1}(\mu)\, \underline{z}}{\underline{z}^T \underline{\tilde{A}}^{-1}(\Theta(\mu))\, \underline{z}} \ge 1,\end{aligned}$$

where the last inequality follows from Proposition 2.2; indeed, $\rho_{\min}(\mu)$ is a *lower bound* for the effectivity $\eta_N(\mu)$, and hence $\rho_{\min}(\mu) \geq 1$ is our (sufficient) condition for $s_N^+(\mu) \geq \tilde{s}(\mu)$ (this can also be motivated very simply from variational arguments). Note that in this paper we exploit monotonicity and convexity to implicitly demonstrate $\rho_{\min}(\mu) \geq 1$; but the former are certainly not necessary conditions for the latter – the bounding properties of the SP′ "$\min(1,\theta)$" $g(\mu)$ conditioner are most easily proven by direct appeal to the Rayleigh quotient expression for $\rho_{\min}(\mu)$.

As might be expected, $\rho_{\max}(\mu) = \max_{\underline{v}\in\mathbb{R}^{\mathcal{N}}}(\underline{v}^T \underline{\tilde{A}}(\Theta(\mu))\, \underline{v})/(\underline{v}^T \underline{\tilde{B}}(\mu)\, \underline{v}) = \max_{\underline{z}\in\mathbb{R}^{\mathcal{N}}}(\underline{z}^T \underline{\tilde{B}}^{-1}(\mu)\, \underline{z})/(\underline{z}^T \underline{\tilde{A}}^{-1}(\Theta(\mu))\, \underline{z})$ is an upper bound for the effectivity. We can also derive this from (2.5) and (2.7):

$$\begin{aligned}
\underline{\hat{e}}^T(\mu)\, \underline{\tilde{B}}(\mu)\, \underline{\hat{e}}(\mu) &= \underline{\hat{e}}^T(\mu)\, \underline{\tilde{R}}(\mu) = \underline{\hat{e}}^T(\mu)\, \underline{\tilde{A}}(\Theta(\mu))\, \underline{\tilde{e}}(\mu) \\
&\leq (\underline{\tilde{e}}^T(\mu)\, \underline{\tilde{A}}(\Theta(\mu))\, \underline{\tilde{e}}(\mu))^{1/2} (\underline{\hat{e}}^T(\mu)\, \underline{\tilde{A}}(\Theta(\mu))\, \underline{\hat{e}}(\mu))^{1/2} \\
&\leq \rho_{\max}^{1/2}(\mu) (\underline{\tilde{e}}^T(\mu)\, \underline{\tilde{A}}(\Theta(\mu))\, \underline{\tilde{e}}(\mu))^{1/2}\, (\underline{\hat{e}}^T(\mu)\, \underline{\tilde{B}}(\mu)\, \underline{\hat{e}}(\mu))^{1/2};
\end{aligned}$$

it follows that $s_N^+(\mu) - s_N(\mu) = \underline{\hat{e}}^T(\mu) \underline{\tilde{B}}(\mu)\, \underline{\hat{e}}(\mu) \leq \rho_{\max}(\mu)\, (\underline{\tilde{e}}^T(\mu)\, \underline{\tilde{A}}(\Theta(\mu))\, \underline{\tilde{e}}(\mu)) = \rho_{\max}(\mu)\, (\tilde{s}(\mu) - s_N(\mu))$, or equivalently, $\eta_N(\mu) \leq \rho_{\max}(\mu)$. Clearly, we wish $\rho_{\max}(\mu)$ to be as close to unity, and hence as close to $\rho_{\min}(\mu)$, as possible: we thus see that good bound conditioners are similar to good (iterative) preconditioners – both satisfy $\rho_{\max}(\mu)/\rho_{\min}(\mu) \cong 1$ – except that bound conditioners must satisfy the additional spectral requirement $\rho_{\min}(\mu) \geq 1$. (Of course, our bound conditioners would not be appropriate in the iterative solution context since our off-line/on-line computational stratagem would not be relevant.)

3.2. $P = Q = 1$ model problem

We would thus like to understand the convergence of $\Delta_N(\mu)$ to zero as a function of N. Unfortunately, we do not yet have a general theory; we can, at present, treat completely only the case $P = Q = 1$. In particular, we consider the case in which $A(\mu) = A_0 + \mu A_1$ (and hence $\Theta_1(\mu) = \mu$), and $\mu \in \mathcal{D}^\mu \equiv [0, \mu^{\max}]$. From our continuity and coercivity assumptions, there exists a positive real constant γ_1 such that

$$\langle A_1 v, v\rangle \leq \gamma_1 \langle A_0 v, v\rangle; \tag{3.2}$$

it thus follows that $\langle A(\mu)v, v\rangle \leq (1 + \mu^{\max}\gamma_1)\, \langle A_0 v, v\rangle$. Defining $\|\cdot\|^2 \equiv \langle A_0 \cdot, \cdot\rangle$, we may thus write

$$\Delta_N(\mu) \leq (1 + \mu^{\max}\gamma_1)\, \|\tilde{u}(\mu) - u_N(\mu)\|^2\, \eta_N(\mu). \tag{3.3}$$

It remains to bound $\|\tilde{u}(\mu) - u_N(\mu)\|$ and $\eta_N(\mu)$; and, in particular, to understand the convergence rate of $\|\tilde{u}(\mu) - u_N(\mu)\| \to 0$ and $\eta_N(\mu) \to 1$ (or at least a constant) as N increases.

The proofs for both $\|\tilde{u}(\mu) - u_N(\mu)\|$ [11] and $\eta_N(\mu)$ implicate a particular "optimal" logarithmic point distribution which we thus impose *a priori*. In particular, we introduce an upper bound for γ_1, γ, and a "log increment" $\delta_N = (\ln(\gamma\mu^{\max} + 1))/(N-1)$; we then define

$$\mu^n = \exp\{-\ln\gamma + (n-1)\delta_N\} - \gamma^{-1}, \quad 1 \leq n \leq N, \tag{3.4}$$

and take $S_N^\mu = \{\mu^1, \ldots, \mu^N\}$. Clearly, $\ln(\mu^n + \gamma^{-1})$ is uniformly distributed.

For our bound conditioners we shall consider SP, SP′, PC, and PL. For SP, $\theta^{\min} = 0$; and for SP′, $g(\mu) = \min(1, \mu)$ (SP′ is in fact not defined for $\mu = 0$, and $\eta_N^{\mathrm{SP}'}(\mu)$ will become increasingly poor as $\mu \to 0$). For PC and PL we choose $M = N$ and $S_M^\theta = \{(\theta^1 =)\, \mu^1, \ldots, (\theta^M =)\, \mu^N\} = S_N^\mu$ ("staggered" $S_N^\mu - S_M^\theta$ meshes are considered in Sect. 4). For PC, we take $\mathcal{E}(\mu) = \{j_1(\mu)\}$ such that $\overline{\theta}(\mu) \equiv \theta^{j_1(\mu)} \leq \mu \in [\theta^{j_1(\mu)}, \theta^{j_1(\mu)+1}]$ (*i.e.*, $\overline{\theta}(\mu)$ is the largest $\theta^j \in S_M^\theta$ such that $\theta^j \leq \mu$). Finally, for PL, $\mathcal{E}(\mu) = \{j_1(\mu), j_2(\mu) = j_1(\mu) + 1\}$ such that

$\overline{\theta}(\mu) \equiv \Theta(\mu) = \mu \in [\theta^{j_1(\mu)}, \theta^{j_2(\mu)}]$ – the vertices of our 2-simplex (*i.e.*, segment) are the two points nearest to μ in $(S_M^\theta =)\ S_N^\mu$.

Finally, in the proofs below, we shall require the following generalized eigenvalue problem: find ($\underline{\tilde{\chi}}_k \in \mathbb{R}^{\mathcal{N}}$, $\lambda_k \in \mathbb{R}$), $k = 1, \ldots, \mathcal{N}$, satisfying $\underline{\tilde{A}}_1\, \underline{\tilde{\chi}}_k = \lambda_k\, \underline{\tilde{A}}_0\, \underline{\tilde{\chi}}_k$, $\underline{\tilde{\chi}}_k^T\, \underline{\tilde{A}}_0\, \underline{\tilde{\chi}}_k = 1$. We shall order the (perforce real and non-negative) eigenvalues as $0 < \lambda_1 \leq \cdots \leq \lambda_{\mathcal{N}} \leq \gamma_1$, where the last inequality follows directly from the Rayleigh quotient and (3.2). The $\underline{\tilde{\chi}}_k$, $k = 1, \ldots, \mathcal{N}$, are of course a complete basis for $\mathbb{R}^{\mathcal{N}}$. (Note that for the corresponding eigenvalue problem defined over the *infinite-dimensional* space Y we must anticipate, for many $\underline{A}_1$, a continuous spectrum).

3.3. **Convergence proofs**

We begin by restating the main result of [11, 12] in

Lemma 3.1. *For $N \geq N_{\rm crit} \equiv 1 + \mathrm{e}\ \ln(\gamma\mu^{\max} + 1)$ and all $\mu \in \mathcal{D}^\mu$,*

$$\|\tilde{u}(\mu) - u_N(\mu)\| \leq (1 + \mu^{\max}\gamma_1)^{1/2}\, \|\tilde{u}(0)\|\, \mathrm{e}^{-\left(\frac{N}{N_{\rm crit}}\right)},$$

where we recall that $\|\cdot\|^2 = \langle A_0 \cdot, \cdot\rangle$.

Proof. See Theorem 3 of [11] (for $c^* = 1$). □

We see that we obtain *exponential* convergence, *uniformly* for all μ in $\mathcal{D}^\mu$. Furthermore, the convergence threshold parameter $N_{\rm crit} = 1 + \mathrm{e}\ \ln(\gamma\mu^{\max} + 1)$, and the exponential convergence rate $1/N_{\rm crit}$, depend only weakly – *logarithmically* – on γ_1 and $\mu^{\max}$ (which together comprise the continuity-coercivity ratio). In short, we expect extremely rapid convergence even for large parameter ranges. The sensitivity of these results to the point distribution is not too great; in [11, 12] we consider a "log distribution on the average" with little detriment to the final result.

To obtain a bound for $\eta_N(\mu)$ we need to obtain a bound for $\rho_{\max}(\mu)$. We do this in

Lemma 3.2. *For all $\mu \in \mathcal{D}^\mu$,*

$$\eta_N^{\rm SP}(\mu) \leq 1 + \gamma_1\mu^{\max} \equiv \overline{\eta}_N^{\rm SP}(\mu), \tag{3.5}$$

$$\eta_N^{\rm PC}(\mu) \leq \mathrm{e}^{\delta_N} \equiv \overline{\eta}_N^{\rm PC}(\mu), \tag{3.6}$$

$$\eta_N^{\rm PL}(\mu) \leq 1 + \frac{(\mathrm{e}^{\delta_N} - 1)^2}{4\mathrm{e}^{\delta_N}} \equiv \overline{\eta}_N^{\rm PL}(\mu). \tag{3.7}$$

Proof. We first rewrite the eigenvalue problem (3.1) as

$$\left(\sum_{j \in \mathcal{E}(\mu)} \alpha_j(\mu)\, \underline{\tilde{A}}^{-1}(\mu^j)\right) \underline{\tilde{A}}(\mu)\, \underline{\tilde{\xi}}_k(\mu) = \rho_k(\mu)\, \underline{\tilde{\xi}}_k(\mu), \quad k = 1, \ldots, N.$$

We then claim that $\underline{\tilde{\xi}}_k = \underline{\tilde{\chi}}_k$, and

$$\rho_k(\mu) = \sum_{j \in \mathcal{E}(\mu)} \frac{1 + \mu\lambda_k}{1 + \mu^j\lambda_k}\, \alpha_j(\mu).$$

To show this, we note that $\underline{\tilde{A}}(\mu)\underline{\tilde{\chi}}_k = (1+\mu\lambda_k)\underline{\tilde{A}}_0\underline{\tilde{\chi}}_k$, and thus $\underline{\tilde{A}}^{-1}(\mu)\underline{\tilde{A}}_0\underline{\tilde{\chi}}_k = (1+\mu\lambda_k)^{-1}\underline{\tilde{\chi}}_k$; applying first the former and then the latter (for $\mu = \mu^j$) yields

$$\left(\sum_{j\in\mathcal{E}(\mu)} \alpha_j(\mu)\,\underline{\tilde{A}}^{-1}(\mu^j)\right)\underline{\tilde{A}}(\mu)\,\underline{\tilde{\chi}}_k = \left(\sum_{j\in\mathcal{E}(\mu)} \frac{1+\mu\lambda_k}{1+\mu^j\lambda_k}\,\alpha_j(\mu)\right)\underline{\tilde{\chi}}_k, \quad k = 1,\ldots,\mathcal{N},$$

as desired.

For the SP case, $\rho_{\max}(\mu) \le \max_{\lambda\in[0,\gamma_1]}(1+\mu\lambda)$, and our result (3.5) then directly follows. For the PC case, we obtain

$$\rho^{\max}(\mu) \le \max_{n\in\{1,\ldots,N-1\}} \max_{\lambda\in[0,\gamma_1]} \frac{1+\mu^{n+1}\lambda}{1+\mu^n\lambda}.$$

But (3.6) then directly follows, since from (3.2), $\gamma \ge \gamma_1 \ge \lambda_{\mathcal{N}}$, and (3.4),

$$\begin{aligned}\frac{1+\mu^{n+1}\lambda}{1+\mu^n\lambda} = 1 + \frac{\mu^{n+1}-\mu^n}{\lambda^{-1}+\mu^n} &\le 1 + \frac{\mu^{n+1}-\mu^n}{\gamma^{-1}+\mu^n} \\ &= 1 + \frac{\exp\{-\ln\gamma + n\delta_N\} - \exp\{-\ln\gamma + (n-1)\delta_N\}}{\exp\{-\ln\gamma + (n-1)\delta_N\}} \\ &= \mathrm{e}^{\delta_N}.\end{aligned}$$

Turning now to our piecewise-linear bound conditioner, we can write

$$\eta_N(\mu) \le \rho_{\max}(\mu) \le \max_{n\in\{1,\ldots,N-1\}} \max_{\lambda\in[0,\gamma_1]} \max_{\tau\in[0,1]} \mathcal{F}^n(\tau,\lambda),$$

where, for $n = 1,\ldots,N-1$,

$$\mathcal{F}^n(\tau,\lambda) = (1 + (\mu^n + \tau(\mu^{n+1}-\mu^n))\lambda)\left[\frac{1-\tau}{1+\mu^n\lambda} + \frac{\tau}{1+\mu^{n+1}\lambda}\right].$$

It is a simple matter to show that $\mathcal{F}^n(\tau,\lambda)$ is maximized at $\tau = 1/2$ (independent of n and λ), and that

$$\mathcal{F}^n\left(\frac{1}{2},\lambda\right) = 1 + \frac{1}{4}\,\frac{(\mu^{n+1}-\mu^n)^2\lambda^2}{(1+\mu^{n+1}\lambda)(1+\mu^n\lambda)}.$$

The desired result (3.7) then directly follows, since from (3.2), $\gamma \ge \gamma_1 \ge \lambda_{\mathcal{N}}$, and (3.4),

$$\begin{aligned}\frac{(\mu^{n+1}-\mu^n)^2\lambda^2}{(1+\mu^{n+1}\lambda)(1+\mu^n\lambda)} &\le \frac{(\mu^{n+1}-\mu^n)^2}{(\gamma^{-1}+\mu^{n+1})(\gamma^{-1}+\mu^n)} \\ &= \frac{(\exp\{-\ln\gamma + n\delta_N\} - \exp\{-\ln\gamma + (n-1)\delta_N\})^2}{(\exp\{-\ln\gamma + n\delta_N\})(\exp\{-\ln\gamma + (n-1)\delta_N\})} \\ &= \frac{(\mathrm{e}^{\delta_N}-1)^2}{\mathrm{e}^{\delta_N}}.\end{aligned}$$

This concludes the proof.

We note that, for both the PC and PL conditioner, $\eta_N(\mu^n) = \Delta_N(\mu^n)/(\tilde{s}(\mu^n) - s_N(\mu^n)) = 0/0$, $n = 1,\ldots,N-1$, since $\underline{\tilde{e}}(\mu) = \underline{\hat{e}}(\mu) = 0$ for $\mu = \mu^n$, $n = 1,\ldots,N-1$ (and, in fact, for $n = N$ for PL). But if we expand $\underline{\tilde{R}}(\mu^n+\varepsilon) = \varepsilon\,\underline{\tilde{R}}_\mu(\mu^n) + \cdots$, then it is a simple matter to show that $\rho^{\min}(\mu^n+\varepsilon) + O(\varepsilon) \le \eta_N(\mu^n+\varepsilon) \le \rho^{\max}(\mu^n+\varepsilon) + O(\varepsilon)$, and hence that $\eta_N(\mu^n) = 1$ since $\rho^{\min}(\mu^n) = \rho^{\max}(\mu^n) = 1$, $n = 1,\ldots,N-1$. □

We see, first, how both the PC and PL proofs directly implicate the log point distribution – we wish to keep $(\mu^{n+1} - \mu^n)/\mu^{n+1}$ roughly constant. (Note the appearance of the log in the proof of Lem. 3.1 is not quite as transparent.) Second, we see that, as expected, the PC and PL bound conditioners yield linear $(\exp(\delta_N)-1 \sim \delta_N \sim O(1/(N-1))$ as $N \to \infty$) and quadratic $((\exp(\delta_N)-1)^2/\exp(\delta_N) \sim (\delta_N)^2 \sim O(1/(N-1)^2)$ as $N \to \infty$) convergence to unity, respectively. Third, we see that, even for modest N, our PC and certainly PL bound conditioners should yield $\eta_N(\mu)$ close to unity – even for large γ and $\mu^{\max}$. Fourth, and finally, we can further improve (at no additional on-line cost) the piecewise-linear conditioner by better choice of S_M^θ: in particular, a staggered $S_N^\mu - S_M^\theta$ promises better effectivities; this is demonstrated empirically in Section 4.

We can now prove (say, for the PL conditioner)

Proposition 3.3. *For $N \geq N_{\rm crit} \equiv 1 + \mathrm{e}\,\ln(\gamma\mu^{\max}+1)$, our piecewise-linear (PL) bound conditioner yields*

$$\Delta_N(\mu) \leq (1+\mu^{\max}\gamma_1)^2\,\|\tilde{u}(0)\|^2\,\mathrm{e}^{-(\frac{2N}{N_{\rm crit}})}\left\{1+\frac{1}{4}\left(\mathrm{e}^{\frac{\ln(\gamma\mu_{\max}+1)}{N-1}}-1\right)^2\right\} \tag{3.8}$$

for all $\mu \in \mathcal{D}^\mu$.

Proof. The result directly follows from (3.3), Lemmas 3.1 and 3.2, and $\mathrm{e}^{\delta_N} > 1$. □

Similar results apply to the SP and PC cases. We note that the only $\mathcal{N}$ dependence in (3.8), through $\|\tilde{u}(0)\|$, is readily removed, thus demonstrating stability with respect to the fineness of the (finite element) truth approximation (*i.e.*, the limit $\mathcal{N} \to \infty$).

We can now understand the importance of the two-step approximation of Section 2.2. In particular, for (say) PL, we can easily construct an upper bound for $\tilde{s}(\mu)$ directly as

$$s_N^{+,\rm direct}(\mu) = \underline{\tilde{F}}^T \left(\sum_{j\in\mathcal{E}(\mu)} \alpha_j(\mu)\,\underline{\tilde{A}}^{-1}(\theta^j)\right)\underline{\tilde{F}}, \tag{3.9}$$

which certainly satisfies $s_N^{+,\,\rm direct}(\mu) \geq \tilde{s}(\mu)$ by virtue of (1.6) and Proposition 1.2. However, the convergence rate of this combined "predictor-*and*-bound" will be only N^{-2}, *versus* the $\mathrm{e}^{-(\frac{2N}{N_{\rm crit}})}N^{-2}$ convergence rate of our "predictor-*then*-bound". The latter performs much better than the former because our perforce lower-order bound construction is not for the *output itself*, but rather for the estimate of the *error in the output*; whereas a (say) 20% error in the output is *not* acceptable, a 20% error in the (exponentially small) error in the output *is* acceptable. In essence, it is best to separate the accuracy and bounding requirements and approximations.

We note also that (3.9) is, in fact, a trivial application of convexity:

$$s_N^{+,\,\rm direct}(\mu) = \underline{\tilde{F}}^T \left(\sum_{j\in\mathcal{E}(\mu)} \alpha_j(\mu)\,\underline{\tilde{A}}^{-1}(\theta^j)\right)\underline{\tilde{F}} = \sum_{j\in\mathcal{E}(\mu)} \alpha_j(\mu)\underline{\tilde{F}}^T\,\underline{\tilde{A}}^{-1}(\theta^j)\,\underline{\tilde{F}} = \sum_{j\in\mathcal{E}(\mu)} \alpha_j(\mu)\,\tilde{s}(\mu^j);$$

we are really just linearly interpolating the (convex) output. In contrast,

$$\Delta_N(\mu) = \underline{\tilde{R}}^T(\mu)\left(\sum_{j\in\mathcal{E}(\mu)} \alpha_j(\mu)\,\underline{\tilde{A}}^{-1}(\theta^j)\right)\underline{\tilde{R}}(\mu)$$

is *not* just interpolating the error (which is not convex, and in fact is *zero* at the $\theta^j = \mu^j$); rather, we are truly interpolating the inverse with subsequent application to μ-dependent data (the residual).

4. Numerical results

4.1. Example I: $P = Q = 1$

We consider $-u_{xx} + \mu u = 0$ on a domain $\Omega \equiv]0,1[$ with a Neumann boundary condition $u_x = -1$ at $x = 0$ and a Dirichlet boundary condition $u = 0$ at $x = 1$; our output of interest is $s(\mu) = u(\mu)|_{x=0}$ for $\mu \in \mathcal{D}^\mu \equiv [0.01, 10^4]$. Our problem can then be formulated as: given a $\mu \in \mathcal{D}^\mu$, find $s(\mu) = \langle F, u(\mu)\rangle$, where $u(\mu) \in Y = \{v \in H^1(\Omega) \mid v|_{x=1} = 0\}$ is the solution to (1.1); for our example,

$$\langle A(\Theta(\mu))w, v\rangle = \underbrace{\int_0^1 v_x\, w_x}_{\langle A_0 w, v\rangle} + \underbrace{\mu}_{\Theta_1(\mu)} \underbrace{\int_0^1 v\, w}_{\langle A_1 w, v\rangle}, \quad \forall\, w, v \in Y,$$

$$\langle F, v\rangle = v|_{x=0}\,, \quad \forall\, v \in Y,$$

$P = Q = 1$, and $\Theta_1(\mu) = \Theta(\mu) = \mu$. We choose for our truth approximation space $\tilde{Y}$ a linear finite element space of dimension $\mathcal{N} = 1000$.

We choose for our "μ" sample, S_N^μ, the logarithmic point distribution of Section 3.2. We present in Table 1 the error in our predictor (and lower bound), $s_N(\mu)$ $(= s_N^-(\mu))$, as a function of N, for $\mu = 7,500$. We observe the exponential convergence implied by (2.6) and Lemma 3.1.

Table 1. Error and effectivities (for SP′, SP, PC, and PL) as a function of N for a representative point $\mu = 7,500$.

N	$(\tilde{s}(\mu) - s_N(\mu))/\tilde{s}(\mu)$	$\eta_N^{\mathrm{SP}'}(\mu) - 1$	$\eta_N^{\mathrm{SP}}(\mu) - 1$	$\eta_N^{\mathrm{PC}}(\mu) - 1$	$\eta_N^{\mathrm{PL}}(\mu) - 1$
2	9.55×10^{-3}	30.44	32.81	32.81	8.10
3	5.78×10^{-3}	25.17	26.57	6.89	1.64
4	2.51×10^{-3}	18.68	19.27	2.81	0.64
5	9.19×10^{-4}	14.19	14.44	1.63	0.36
6	2.98×10^{-4}	11.09	11.21	1.10	0.24
7	8.77×10^{-5}	8.91	8.97	0.81	0.17
8	2.36×10^{-5}	7.33	7.37	0.63	0.13
9	5.84×10^{-6}	6.15	6.18	0.51	0.10
10	1.33×10^{-6}	5.25	5.27	0.41	0.08

We now examine the effectivity for the SP′, SP, PC, and PL bound conditioners described in Section 3.2. For PC and PL we consider two different θ-samples: a *non-staggered* grid for which $M = N$ and $\theta^n = \mu^n, n = 1, \ldots, N$ (and hence $S_M^\theta = S_N^\mu$); and a *staggered* grid with $M = N+1$, $\theta^1 = \mu^1$, $\theta^{N+1} = \mu^N$, and $\ln(\theta^m + \gamma^{-1}) = \frac{1}{2}\left[\ln(\mu^{m-1} + \gamma^{-1}) + \ln(\mu^m + \gamma^{-1})\right], m = 2, \ldots, N$. All of our results are for a particular "representative" point $\mu = 7,500$.

We begin with the results for the non-staggered grid. We present in Table 1 $\eta_N^{\mathrm{SP}'}(\mu) - 1$, $\eta_N^{\mathrm{SP}}(\mu) - 1$, $\eta_N^{\mathrm{PC}}(\mu) - 1$, and $\eta_N^{\mathrm{PL}}(\mu) - 1$ as a function of N. The conditioners SP and SP′ behave in roughly the same fashion (since $\mu \gg 1$); for neither SP′ nor SP does the effectivity converge to unity as $N \to \infty$. The PC conditioner performs considerably better than SP′ or SP; and clearly $\eta_N^{\mathrm{PC}}(\mu) \to 1$ as $N \to \infty$, roughly as $1/(N-1)$ (see below). The PL conditioner is even better than PC; and $\eta_N^{\mathrm{PL}}(\mu) \to 1$ as $N \to \infty$ now roughly as $1/(N-1)^2$ (see below).

TABLE 2. Ratio of the effectivities and our *a priori* upper bound for the effectivities as a function of N for PC and PL.

N	$\frac{\left(\overline{\eta}_N^{\rm PC}(\mu)-1\right)}{\left(\eta_N^{\rm PC}(\mu)-1\right)}$	$\frac{\left(\overline{\eta}_N^{\rm PL}(\mu)-1\right)}{\left(\eta_N^{\rm PL}(\mu)-1\right)}$
2	247.09	250.22
3	12.92	13.42
4	6.79	7.08
5	5.20	5.28
6	4.57	4.44
7	4.29	3.95
8	4.17	3.64
9	4.16	3.43
10	4.20	3.29

To better ascertain the convergence rates, we present in Table 2 $\left(\overline{\eta}_N^{\rm PC}(\mu)-1\right)/\left(\eta_N^{\rm PC}(\mu)-1\right)$ and $\left(\overline{\eta}_N^{\rm PL}(\mu)-1\right)/\left(\eta_N^{\rm PL}(\mu)-1\right)$ as a function of N. We observe that our *a priori* bounds for the PC and PL conditioners are relatively precise as regards rate, though clearly somewhat pessimistic as regards amplitude.

We now turn to the results for the staggered grid. In particular, we present in Table 3 $(\eta_N^{\rm PC}(\mu)-1)^{\rm stag}/(\eta_N^{\rm PC}(\mu)-1)^{\rm non\text{-}stag}$ and $(\eta_N^{\rm PL}(\mu)-1)^{\rm stag}/(\eta_N^{\rm PL}(\mu)-1)^{\rm non\text{-}stag}$ as a function of N. As expected, the extra "zeroes" associated with the staggered arrangement yield both better effectivities for fixed N, and, it would appear, more rapid convergence of the effectivity to unity as N increases.

TABLE 3. Ratio of the effectivities for a non-staggered and staggered grid as a function of N for PC and PL.

N	$\frac{\left(\eta_N^{\rm PC}(\mu)-1\right)^{\rm stag}}{\left(\eta_N^{\rm PC}(\mu)-1\right)^{\rm non\text{-}stag}}$	$\frac{\left(\eta_N^{\rm PL}(\mu)-1\right)^{\rm stag}}{\left(\eta_N^{\rm PL}(\mu)-1\right)^{\rm non\text{-}stag}}$
2	0.30206	0.29661
3	0.25996	0.24534
4	0.29905	0.27985
5	0.31683	0.29775
6	0.32087	0.30371
7	0.31567	0.30117
8	0.30399	0.29154
9	0.28700	0.27664
10	0.26546	0.25697

If we were to perform a "minimum on-line complexity at fixed error ($\Delta_N(\mu)$)" analysis, no doubt (staggered) PC would be preferred. (In practice, of course, we must also consider the off-line complexity and on-line storage.) In particular, given the rapid convergence of $\Delta_N(\mu)$ to zero as N increases, $N_\varepsilon^{\rm PC}(\mu) \equiv \{N \mid \Delta_N^{\rm PC}(\mu) = \varepsilon\}$ will be only very slightly larger than $N_\varepsilon^{\rm PL}(\mu) \equiv \{N \mid \Delta_N^{\rm PL}(\mu) = \varepsilon\}$: the additional "$N$" work for PC will thus be less

than the additional "Q" work for PL. (Recall that for SP′, SP, and PC (respectively, PL) the on-line complexity to compute $s_N^+(\mu)$ is roughly $4N^2$ (respectively, $8N^2$).)

4.2. **Example II: $P = Q = 2$**

We now consider $-u_{xx} + \mu_1 u = 0$ in a domain $\Omega \equiv]0,1[$ with a Neumann boundary condition $u_x = -1$ at $x = 0$ and a Robin boundary condition $u_x + \mu_2 u = 0$ at $x = 1$; our output of interest is $s(\mu) = u(\mu)|_{x=0}$ for $\mu = (\mu_1, \mu_2) \in \mathcal{D}^\mu \equiv [1, 1000] \times [0.001, 0.1]$. Our problem can then be formulated as: given a $\mu \in \mathcal{D}^\mu$, find $s(\mu) = \langle F, u(\mu)\rangle$, where $u(\mu) \in Y = H_1(\Omega)$ is the solution to (1.1); for our example,

$$\langle A(\Theta(\mu))w, v\rangle = \underbrace{\int_0^1 v_x w_x}_{\langle A_0 w, v\rangle} + \underbrace{\mu_1}_{\Theta_1(\mu)} \underbrace{\int_0^1 v\, w}_{\langle A_1 w, v\rangle} + \underbrace{\mu_2}_{\Theta_2(\mu)} \underbrace{(v\, w)|_{x=1}}_{\langle A_2 w, v\rangle}, \quad \forall\, w, v \in Y,$$

$$\langle F, v\rangle = v|_{x=0}\,, \quad \forall\, v \in Y,$$

$P = Q = 2$, and $\Theta_q(\mu) = \mu_q, q = 1, 2$. We choose for our truth approximation space $\tilde{Y}$ a linear finite element space of dimension $\mathcal{N} = 1000$.

We choose for our "μ" sample, S_N^μ, a random bi-logarithmic point distribution [16]. We present in Table 4 the error in our predictor (and lower bound), $s_N(\mu)$ $(= s_N^-(\mu))$, as a function of N for $\mu = (200, 0.06)$. We observe that very rapid convergence is still obtained even for $P > 1$.

TABLE 4. Error and effectivities (for SP′, SP, PC, and PL) as a function of N for $\mu = (200, 0.06)$.

N	$(\tilde{s}(\mu) - s_N(\mu))/\tilde{s}(\mu)$	$\eta_N^{\mathrm{SP}'}(\mu) - 1$	$\eta_N^{\mathrm{SP}}(\mu) - 1$	$\eta_N^{\mathrm{PC}}(\mu) - 1$	$\eta_N^{\mathrm{PL}}(\mu) - 1$
3	3.73×10^{-3}	191.26	19.68	4.52	1.57
4	5.30×10^{-4}	92.17	8.30	3.21	1.15
5	2.77×10^{-5}	70.80	5.62	2.64	0.96
6	3.60×10^{-8}	49.42	2.96	1.85	0.69
7	2.53×10^{-9}	31.27	1.01	0.60	0.10
8	5.75×10^{-10}	24.68	0.57	0.40	0.05

We now examine the effectivity for SP′, SP, PC, and PL. We present results for a non-staggered mesh, $M = N$ and $S_M^\theta = S_N^\mu$; note, however, that a staggered mesh does, indeed, again improve the results, in particular for PC. For our numerical tests, we consider $\mu = (200, 0.06)$. There are often several points in S_M^θ such that $\theta^j \le \mu$ – we choose (for PC) the point which yields the lowest upper bound; and there are often several simplices (triangles) in S_M^θ that contain μ – we choose (for PL) the simplex that yields the lowest upper bound.

We present in Table 4 $\eta_N^{\mathrm{SP}'}(\mu) - 1$, $\eta_N^{\mathrm{SP}}(\mu) - 1$, $\eta_N^{\mathrm{PC}}(\mu) - 1$, and $\eta_N^{\mathrm{PL}}(\mu) - 1$ as a function of N. The conditioner SP′ now performs very poorly due to the small value of μ_2; however, SP performs quite well, at least for larger N. As in our $P = Q = 1$ example, PC is better than SP, in particular for smaller N; and PL is considerably better than PC, in particular for larger N. We note that, due to the usual curse of dimensionality, $\left|\bar{\theta}(\mu) - \Theta(\mu)\right|$ for PC, and the size of our simplex for PL, will grow (for fixed M) as $P = Q$ increases; the good effectivities of Table 4 are thus somewhat surprising. Future tests must consider "worst case" effectivities for all $\mu \in \mathcal{D}^\mu$.

As for $P = Q = 1$, the "minimum on-line complexity at fixed error $(\Delta_N(\mu))$" analysis no doubt again prefers PC; PL (which scales as $27N^2$) is now even more expensive relative to PC (which scales as $9N^2$).

4.3. Example III: $P = 1, Q = 2$; "θ patterns"

There are many cases, in particular involving geometric variations, in which Q is larger than P. If this implies M (much) larger than N, the off-line complexity and on-line storage for PC and PL could become prohibitive. However, since $\mathcal{D}^\theta$ is the image of $\mathcal{D}^\mu$ under $\Theta(\mu)$, $\mathcal{D}^\theta$ will be a low-dimensional manifold in $\mathbb{R}^Q_+$ – and we can thus hope that $M = \text{Const}(\text{independent of } N)N$ will suffice. We present here an example which supports this claim.

We consider $-\nabla^2 u = \frac{1}{\mu}$ in a domain $\Omega_0 \equiv]0,1[\times]0,\mu[$ with homogeneous Dirichlet conditions on the boundary, Γ_0; our output of interest is $s(\mu) = \frac{1}{\mu}\int_{\Omega_0} u(\mu)$ for $\mu \in \mathcal{D}^\mu \equiv [\mu_{\min}, 1] = [0.1, 1]$. Our problem can then be formulated (after affine mapping $\Omega_0 \to \Omega =]0,1[\times]0,1[$) as: given a $\mu \in \mathcal{D}^\mu$, find $s(\mu) = \langle F, u(\mu)\rangle$, where $u(\mu) \in Y = H^1_0(\Omega) = \{v \in H^1(\Omega) \mid v|_\Gamma = 0\}$ is the solution to (1.1); for our example,

$$\langle A(\Theta(\mu))w, v\rangle = \underbrace{\mu_{\min}\int_\Omega v_x w_x + v_y w_y}_{\langle A_0 w, v\rangle} + \underbrace{(\mu - \mu_{\min})}_{\Theta_1(\mu)}\underbrace{\int_\Omega v_x w_x}_{\langle A_1 w, v\rangle} + \underbrace{\left(\frac{1}{\mu} - \mu_{\min}\right)}_{\Theta_2(\mu)}\underbrace{\int_\Omega v_y w_y}_{\langle A_2 w, v\rangle}, \quad \forall w, v \in Y,$$

$$\langle F, v\rangle = \int_\Omega v, \quad \forall v \in Y,$$

$P = 1$, $Q = 2$, $\Theta_1(\mu) = \mu - \mu_{\min}$, and $\Theta_2(\mu) = \frac{1}{\mu} - \mu_{\min}$.

We take for S^μ_N our usual logarithmic distribution over the interval $\mathcal{D}^\mu$. We present in Table 5 the error in our predictor (and lower bound), $s_N(\mu)$ $(= s^-_N(\mu))$, as a function of N for $\mu = 0.11$. We again observe very rapid convergence.

TABLE 5. Error and effectivities (for SP′, SP, PC, and PL) as a function of N for $\mu = 0.11$.

N	$(\tilde{s}(\mu) - s_N(\mu))/\tilde{s}(\mu)$	$\eta^{\text{SP}'}_N(\mu) - 1$	$\eta^{\text{SP}}_N(\mu) - 1$	$\eta^{\text{PC}}_N(\mu) - 1$	$\eta^{\text{PL}}_N(\mu) - 1$
2	1.06×10^{-4}	98.12	2.05	2.05	0.18
3	5.43×10^{-5}	141.27	2.53	1.37	0.12
4	1.12×10^{-5}	29.47	1.37	0.73	0.05
5	1.48×10^{-6}	146.34	2.60	0.75	0.06
6	6.06×10^{-7}	71.22	1.48	0.49	0.04
7	1.61×10^{-7}	136.59	3.32	0.61	0.05
8	1.91×10^{-8}	41.35	1.25	0.35	0.02

We now construct our "θ" sample, S^θ_M. We first note that the $\mu^n \in S^\mu_N$ map to the "$\circ$" points on the curve $\mathcal{D}^\theta$ of Figure 1. For PC we consider $M = N - 1$, and take S^θ_M to be the "$\bullet$" points of Figure 1; for any $\mu \in \mathcal{D}^\mu (\Rightarrow \Theta(\mu) \in \mathcal{D}^\theta)$ there is a unique $\theta^j \in S^\theta_M$ such that $\theta^j \le \Theta(\mu)$. For PL we consider $M = 2N - 1$, and take S^θ_M to be the "$\circ$" points and "$\bullet$" points of Figure 1; for any $\mu \in \mathcal{D}^\mu (\Rightarrow \Theta(\mu) \in \mathcal{D}^\theta)$, $\mathcal{E}(\mu)$ is then chosen such that the resulting simplex is the (unique) right triangle containing $\Theta(\mu)$. We conduct our tests for $\mu = 0.11$, which we mark as "$*$" in Figure 1; the associated PL simplex is indicated in Figure 1 as dashed lines.

We present in Table 5 $\eta^{\text{SP}'}_N(\mu) - 1$, $\eta^{\text{SP}}_N(\mu) - 1$, $\eta^{\text{PC}}_N(\mu) - 1$, and $\eta^{\text{PL}}_N(\mu) - 1$ as a function of N. As for the previous example, SP′ performs poorly; SP, PC, and PL all perform reasonably well, with PC better than SP, and PL better than PC. To the extent that this problem is representative, we conclude that $M = \text{Const}\, N$ is indeed sufficient even for $Q > P$.

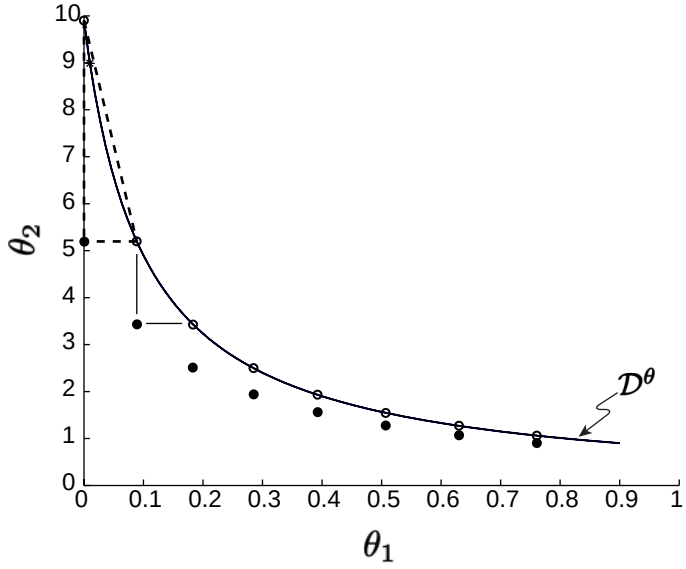

FIGURE 1. Points in $\mathcal{D}^\theta_{\rm box}$ which serve to construct S^θ_M for PC and PL.

It is admittedly disappointing that, even for this last example, PC is probably preferred over PL. However, there are problems for which $\tilde{s}(\mu) - s_N(\mu)$ will converge more slowly with N, in which case PL will perhaps be redeemed: N will be larger, and hence the "N work" (to find $u_N(\mu)$) may dominate the "Q work" (to find $\Delta_N(\mu)$); and $N^{\rm PC}_\varepsilon(\mu)/N^{\rm PL}_\varepsilon(\mu)$ will be larger, and hence the "N work" for PC may dominate the "N work" for PL. In any event, we have provided here a general and unified framework for the construction and evaluation of a wide variety of reduced-basis bound conditioners; future work must apply this framework to a more realistic suite of problems.

Many of the ideas in this paper originate in our joint work with Professor Yvon Maday of University of Paris VI, and Dr. Gabriel Turicini of ASCI-CNRS Orsay and INRIA Rocquencourt. We also thank Dr. Christophe Prud'homme of MIT, Dr. Luc Machiels of McKinsey Corporation, and Professor Jaime Peraire of MIT for helpful discussions on reduced-basis methods and *a posteriori* error estimation. This work was supported by the Singapore-MIT Alliance, by DARPA and AFOSR under Grant F49620-01-1-0458, by DARPA and ONR under Grant N00014-01-1-0523 (Subcontract 340-6218-3), and by NASA under Grant NAG-1-1978. This work was performed while D.R. was in residence at University of Paris VI, partially supported by a Chateaubriand Fellowship.

REFERENCES

[1] M.A. Akgun, J.H. Garcelon and R.T. Haftka, Fast exact linear and non-linear structural reanalysis and the Sherman–Morrison–Woodbury formulas. *Int. J. Numer. Meth. Engrg.* **50** (2001) 1587-1606.
[2] E. Allgower and K. Georg, Simplicial and continuation methods for approximating fixed-points and solutions to systems of equations. *SIAM Rev.* **22** (1980) 28-85.
[3] B.O. Almroth, P. Stern and F.A. Brogan, Automatic choice of global shape functions in structural analysis. *AIAA J.* **16** (1978) 525-528.
[4] M. Avriel, *Nonlinear Programming: Analysis and Methods*. Prentice-Hall, Inc., Englewood Cliffs, NJ (1976).
[5] E. Balmes, Parametric families of reduced finite element models. Theory and applications. *Mech. Systems and Signal Process.* **10** (1996) 381-394.
[6] A. Barrett and G. Reddien, On the Reduced Basis Method. *Z. Angew. Math. Mech.* **75** (1995) 543-549.
[7] T.F. Chan and W.L. Wan, Analysis of projection methods for solving linear systems with multiple right-hand sides. *SIAM J. Sci. Comput.* **18** (1997) 1698.

[8] C. Farhat, L. Crivelli and F.X. Roux, Extending substructure based iterative solvers to multiple load and repeated analyses. *Comput. Meth. Appl. Mech. Engrg.* **117** (1994) 195-209.
[9] J.P. Fink and W.C. Rheinboldt, On the error behavior of the reduced basis technique for nonlinear finite element approximations. *Z. Angew. Math. Mech.* **63** (1983) 21-28.
[10] L. Machiels, Y. Maday, I.B. Oliveira, A.T. Patera and D.V. Rovas, Output bounds for reduced-basis approximations of symmetric positive definite eigenvalue problems. *C. R. Acad. Sci. Paris Sér. I Math.* **331** (2000) 153-158.
[11] Y. Maday, A.T. Patera and G. Turinici, Global *a priori* convergence theory for reduced-basis approximations of single-parameter symmetric coercive elliptic partial differential equations. *C. R. Acad. Sci. Paris Sér. I Math.* (submitted).
[12] Y. Maday, A.T. Patera and G. Turinici, *A priori* convergence theory for reduced-basis approximations of single-parameter elliptic partial differential equations. *J. Sci. Comput.* (accepted).
[13] A.K. Noor and J.M. Peters, Reduced basis technique for nonlinear analysis of structures. *AIAA J.* **18** (1980) 455-462.
[14] J.S. Peterson, The reduced basis method for incompressible viscous flow calculations. *SIAM J. Sci. Stat. Comput.* **10** (1989) 777-786.
[15] T.A. Porsching, Estimation of the error in the reduced basis method solution of nonlinear equations. *Math. Comput.* **45** (1985) 487-496.
[16] C. Prud'homme, D. Rovas, K. Veroy, Y. Maday, A.T. Patera and G. Turinici, Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. *J. Fluids Engrg.* **124** (2002) 70-80.
[17] W.C. Rheinboldt, Numerical analysis of continuation methods for nonlinear structural problems. *Comput. & Structures* **13** (1981) 103-113.
[18] W.C. Rheinboldt, On the theory and error estimation of the reduced basis method for multi-parameter problems. *Nonlinear Anal. Theor. Meth. Appl.* **21** (1993) 849-858.
[19] E.L. Yip, A note on the stability of solving a rank-p modification of a linear system by the Sherman–Morrison–Woodbury formula. *SIAM J. Sci. Stat. Comput.* **7** (1986) 507-513.