Reliable Real-Time Solution of Parametrized Partial Differential Equations: Reduced-Basis Output Bound Methods

Christophe Prud'Homme, Dimitrios V. Rovas, Karen Veroy, Luc Machiels, Yvon Maday, Anthony T. Patera, Gabriel Turinici

Reliable Real-Time Solution of Parametrized Partial Differential Equations: Reduced-Basis Output Bound Methods

C. Prud'homme D. V. Rovas

K. Veroy

L. Machiels

Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Y. Maday

Laboratoire d'Analyse Numérique, Université Pierre et Marie Curie, Boîte courrier 187, 75252 Paris, Cedex 05, France

A. T. Patera

Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

G. Turinici

ASCI-CNRS Orsay, and INRA Rocquencourt M3N, B.P. 105, 78153 LeChesnay Cedex France

We present a technique for the rapid and reliable prediction of linear-functional outputs of elliptic (and parabolic) partial differential equations with affine parameter dependence. The essential components are (i) (provably) rapidly convergent global reduced-basis approximations—Galerkin projection onto a space WN spanned by solutions of the gov-erning partial differential equation at N selected points in parameter space; (ii) a poste-riori error estimation—relaxations of the error-residual equation that provide inexpensive yet sharp and rigorous 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.

1 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 lift 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; control variables, such as actuator power—for example in real-time 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, optimization, control, and characterization contexts. The ''many queries'' limit has certainly received considerable attention: from ''fast loads'' or multiple right-hand side notions (e.g., Chan and Wan [1], Farhat et al. [2]) to matrix perturbation theories (e.g., Akgun et al. [3], Yip [4]) to continuation methods (e.g., Allgower and Georg [5], Rheinboldt [6]). Our particular approach is based on the reduced-basis method, first introduced in the late 1970s for nonlinear structural analysis (Almroth et al. [7], Noor and Peters [8]), and subsequently developed more broadly in the 1980s and 1990s (Balmes [9], Barrett and Reddien [10], Fink and Rheinboldt [11], Peterson [12], Porsching [13], Rheinboldt [14]). 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, it 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 (Fink and Rheinboldt [11]). As a result, the computational improvements—relative to conventional (say) finite element approximation—are often quite modest (Porsching [13]). Our work 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 Balmes [9] 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 this expository review paper we focus on these new ingredients. In Section 2 we introduce an abstract problem formulation and several illustrative instantiations. In Section 3 we describe, for coercive symmetric problems and ''compliant'' outputs, the reduced-basis approximation; and in Section 4 we present the associated *a posteriori* error estimation procedures. In Section 5 we

consider the extension of our approach to noncompliant outputs and nonsymmetric operators; eigenvalue problems; and, more briefly, noncoercive operators, parabolic equations, and nonaffine problems. A description of the system architecture in which these numerical objects reside may be found in Veroy et al. [15].

2 Problem Statement

2.1 Abstract Formulation. We consider a suitably regular domain $\Omega \subset \mathbb{R}^d$, $d=1$, 2, or 3, and associated function space $X \subset H^1(\Omega)$, where $H^1(\Omega)=\{v \in L^2(\Omega),\ \nabla v \in (L^2(\Omega))^d\}$, and $L^2(\Omega)$ is the space of square-X integrable functions over Ω. The inner product and norm associated with X are given by $(\cdot,\cdot)_X$ and $\|\cdot\|_X=(\cdot,\cdot)_X^{1/2}$, respectively. We also define a parameter set $\mathcal{D} \in \mathbb{R}^P$, a particular point in which will be denoted μ. Note that Ω does *not* depend on the parameter.

We then introduce a bilinear form a: $X \times X \times \mathcal{D} \to \mathbb{R}$, and linear forms f: $X \to \mathbb{R}$, ℓ: $X \to \mathbb{R}$. We shall assume that a is continuous, $a(w,v;\mu) \leq \gamma(\mu)\|w\|_X\|v\|_X \leq \gamma_0\|w\|_X\|v\|_X$, $\forall \mu \in \mathcal{D}$; furthermore, in Sections 3 and 4, we assume that a is coercive,

$$0 < \alpha_0 \leq \alpha(\mu) = \inf_{w \in X} \frac{a(w,w;\mu)}{\|w\|_X^2}, \quad \forall \mu \in \mathcal{D}, \tag{1}$$

and symmetric, $a(w,v;\mu)=a(v,w;\mu)$; $\forall w,v \in X$, $\forall \mu \in \mathcal{D}$. We also require that our linear forms f and ℓ be bounded; in Sections 3 and 4 we additionally assume a "compliant" output, $f(v)=\ell(v)$, $\forall v \in X$.

We shall also make certain assumptions on the parametric dependence of a, f, and ℓ. In particular, we shall suppose that, for some finite (preferably small) integer Q, a may be expressed as

$$a(w,v;\mu) = \sum_{q=1}^{Q} \sigma^q(\mu) a^q(w,v), \quad \forall w,v \in X, \forall \mu \in \mathcal{D}, \tag{2}$$

for some σ^q: $\mathcal{D} \to \mathbb{R}$ and a^q: $X \times X \to \mathbb{R}$, $q=1,\ldots,Q$. This "separability," or "affine," assumption on the parameter dependence is crucial to computational efficiency; however, certain relaxations are possible—see Section 5.3.3. For simplicity of exposition, we assume that f and ℓ do not depend on μ; in actual practice, affine dependence is readily admitted.

Our abstract problem statement is then: for any $\mu \in \mathcal{D}$, find $s(\mu) \in \mathbb{R}$ given by

$$s(\mu) = \ell(u(\mu)), \tag{3}$$

where $u(\mu) \in X$ is the solution of

$$a(u(\mu),v;\mu) = f(v), \quad \forall v \in X. \tag{4}$$

In the language of the Introduction, a is our partial differential equation (in weak form), μ is our parameter, $u(\mu)$ is our field variable, and $s(\mu)$ is our output. For simplicity of exposition, we may on occasion suppress the explicit dependence on μ.

2.2 Particular Instantiations. We indicate here a few instantiations of the abstract formulation; these will serve to illustrate the methods (for coercive, symmetric problems) of Sections 3 and 4.

2.2.1 A Thermal Fin. In this example, we consider the two- and three-dimensional thermal fins shown in Fig. 1; these examples may be (interactively) accessed on our web site.[1] The fins consist of a vertical central "post" of conductivity $\tilde{k}_0$ and four horizontal "subfins" of conductivity $\tilde{k}^i$, $i=1,\ldots,4$. The fins conduct heat from a prescribed uniform flux source $\tilde{q}''$ at the root $\tilde{\Gamma}_{\text{root}}$ through the post and large-surface-area subfins to the surrounding flowing air; the latter is characterized by a sink temperature $\tilde{u}_0$ and prescribed heat transfer coefficient $\tilde{h}$. The physical model is simple conduction: the temperature field in the fin, $\tilde{u}$, satisfies

$$\sum_{i=0}^{4} \int_{\tilde{\Omega}_i} \tilde{k}^i \tilde{\nabla}\tilde{u} \cdot \tilde{\nabla}\tilde{v} + \int_{\partial\tilde{\Omega} \setminus \tilde{\Gamma}_{\text{root}}} \tilde{h}(\tilde{u}-\tilde{u}_0)\tilde{v} = \int_{\tilde{\Gamma}_{\text{root}}} \tilde{q}''\tilde{v},$$

$$\forall \tilde{v} \in \tilde{X} \equiv H^1(\tilde{\Omega}), \tag{5}$$

where $\tilde{\Omega}_i$ is that part of the domain with conductivity $\tilde{k}^i$, and $\partial\tilde{\Omega}$ denotes the boundary of $\tilde{\Omega}$.

We now (*i*) nondimensionalize the weak equations (5), and (*ii*) apply a continuous piecewise-affine transformation from $\tilde{\Omega}$ to a fixed (μ-independent) reference domain Ω (Maday et al. [16]). The abstract problem statement (4) is then recovered for $\mu=\{k^1,k^2,k^3,k^4,\text{Bi},L,t\}$, $\mathcal{D}=[0.1,10.0]^4 \times [0.01,1.0] \times [2.0,3.0] \times [0.1,0.5]$, and $P=7$; here $k^1,\ldots,k^4$ are the thermal conductivities of the "subfins" (see Fig. 1) relative to the thermal conductivity of the fin base; Bi is a nondimensional form of the heat transfer coefficient; and, L, t are the length and thickness of each of the "subfins" relative to the length of the fin root $\tilde{\Gamma}_{\text{root}}$. It is readily verified that a is continuous, coercive, and symmetric; and that the "affine" assumption (2) obtains for $Q=16$ (two-

[1] FIN2D: http://augustine.mit.edu/fin2d/fin2d.pdf and FIN3D: http://augustine.mit.edu/fin3d_1/fin3d_1.pdf

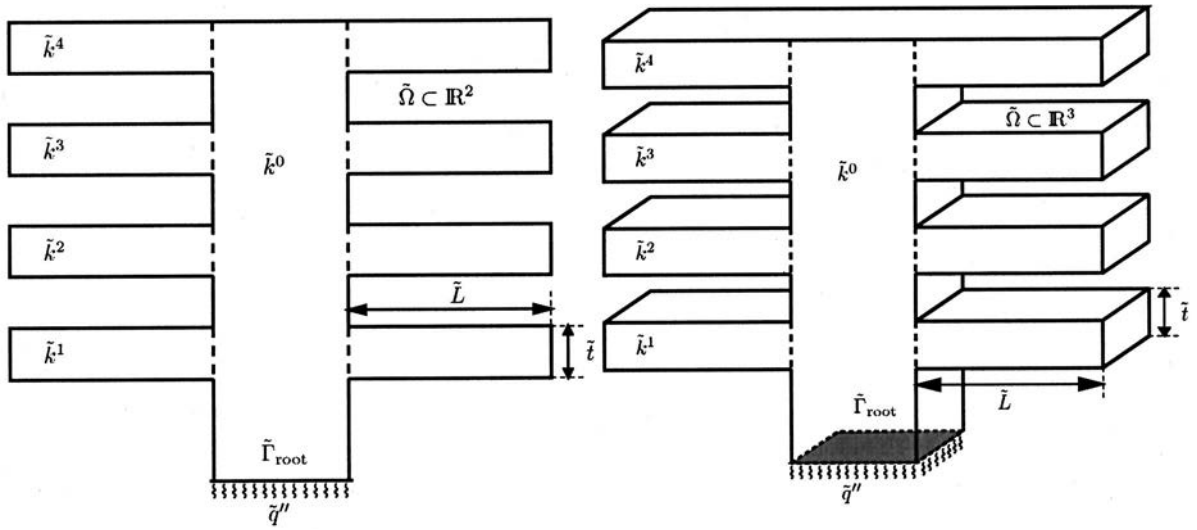

Fig. 1 Two- and three-dimensional thermal fins

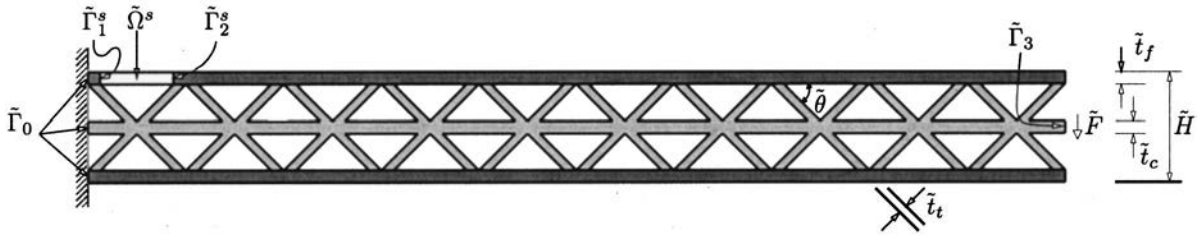

Fig. 2 A truss structure

dimensional case) and $Q=25$ (three-dimensional case). Note that the geometric variations are reflected, via the mapping, in the $\sigma^q(\mu)$.

For our output of interest, $s(\mu)$, we consider the average temperature of the root of the fin nondimensionalized relative to $\tilde{q}''$, $\tilde{k}^0$, and the length of the fin root. This output may be expressed as $s(\mu)=\ell(u(\mu))$, where $\ell(v)=\int_{\Gamma_{\text{root}}} v$. It is readily shown that this output functional is bounded and also "compliant": $\ell(v)=f(v)$, $\forall v\in X$.

2.2.2 A Truss Structure. We consider a prismatic microtruss structure (Evans et al. [17], Wicks and Hutchinson [18]) shown in Fig. 2; this example may be (interactively) accessed on our web site.[2] The truss consists of a frame (upper and lower faces, in dark gray) and a core (trusses and middle sheet, in light gray). The structure transmits a force per unit depth $\tilde{F}$ uniformly distributed over the tip of the middle sheet $\tilde{\Gamma}_3$ through the truss system to the fixed left wall $\tilde{\Gamma}_0$. The physical model is simple plane-strain (two-dimensional) linear elasticity: the displacement field u_i, $i=1,2$, satisfies

$$\int_{\tilde{\Omega}} \frac{\partial \tilde{v}_i}{\partial \tilde{x}_j} \tilde{E}_{ijkl} \frac{\partial \tilde{u}_k}{\partial \tilde{x}_l} = -\left(\frac{\tilde{F}}{\tilde{t}_c}\right) \int_{\tilde{\Gamma}_3} \tilde{v}_2, \quad \forall v \in \tilde{X}, \tag{6}$$

where $\tilde{\Omega}$ is the truss domain, $\tilde{E}_{ijkl}$ is the elasticity tensor, and $\tilde{X}$ refers to the set of functions in $H^1(\tilde{\Omega})$ which vanish on $\tilde{\Gamma}_0$. We assume summation over repeated indices.

We now (*i*) nondimensionalize the weak equations (6), and (*ii*) apply a continuous piecewise-affine transformation from $\tilde{\Omega}$ to a fixed (μ-independent) reference domain Ω. The abstract problem statement (4) is then recovered for $\mu=\{t_f,t_t,H,\theta\}$, $\mathcal{D}=[0.08,1.0]\times[0.2,2.0]\times[4.0,10.0]\times[30.0°,60.0°]$, and $P=4$. Here t_f and t_t are the thicknesses of the frame and trusses (normalized relative to $\tilde{t}_c$), respectively; H is the total height of the microtruss (normalized relative to $\tilde{t}_c$); and θ is the angle between the trusses and the faces. The Poisson's ratio, $\nu=0.3$, and the frame and core Young's moduli, $E_f=75$ GPa and $E_c=200$ GPa, respectively, are held fixed. It is readily verified that a is continuous, coercive, and symmetric; and that the "affine" assumption (2) obtains for $Q=44$.

Our outputs of interest are (*i*) the average downward deflection (compliance) at the core tip, Γ_3, nondimensionalized by $\tilde{F}/\tilde{E}_f$; and (*ii*) the average normal stress across the critical (yield) section denoted Γ_1^s in Fig. 2. These compliance and noncompliance outputs can be expressed as $s^1(\mu)=\ell^1(u(\mu))$ and $s^2(\mu)=\ell^2(u(\mu))$, respectively, where $\ell^1(v)=-\int_{\Gamma_3} v_2$, and

$$\ell^2(v)=\frac{1}{t_f}\int_{\Omega^s} \frac{\partial \chi_i}{\partial x_j} E_{ijkl} \frac{\partial u_k}{\partial x_l}$$

are bounded linear functionals; here χ_i is any suitably smooth function in $H^1(\Omega^s)$ such that $\chi_i \hat{n}_i=1$ on Γ_1^s and $\chi_i \hat{n}_i=0$ on Γ_2^s, where $\hat{n}$ is the unit normal. Note that $s^1(\mu)$ is a compliant output, whereas $s^2(\mu)$ is "noncompliant."

3 Reduced-Basis Approach

We recall that in this section, as well as in Section 4, we assume that a is continuous, coercive, symmetric, and affine in μ—see (2); and that $\ell(v)=f(v)$, which we denote "compliance."

3.1 Reduced-Basis Approximation. We first introduce a sample in parameter space, $S_N=\{\mu_1,\ldots,\mu_N\}$, where $\mu_i\in\mathcal{D}$, $i=1,\ldots,N$; see Section 3.2.2 for a brief discussion of point distribution. We then define our Lagrangian (Porsching [13]) reduced-basis approximation space as $W_N=\text{span}\{\zeta_n\equiv u(\mu_n), n=1,\ldots,N\}$, where $u(\mu_n)\in X$ is the solution to (4) for $\mu=\mu_n$. In actual practice, $u(\mu_n)$ is replaced by an appropriate finite element approximation on a suitably fine truth mesh; we shall discuss the associated computational implications in Section 3.3. Our reduced-basis approximation is then: for any $\mu\in\mathcal{D}$, find $s_N(\mu)=\ell(u_N(\mu))$, where $u_N(\mu)\in W_N$ is the solution of

$$a(u_N(\mu),v;\mu)=\ell(v), \quad \forall v\in W_N. \tag{7}$$

Non-Galerkin projections are briefly described in Section 5.3.1.

3.2 *A Priori* Convergence Theory.

3.2.1 Optimality. We consider here the convergence rate of $u_N(\mu)\to u(\mu)$ and $s_N(\mu)\to s(\mu)$ as $N\to\infty$. To begin, it is standard to demonstrate optimality of $u_N(\mu)$ in the sense that

$$\|u(\mu)-u_N(\mu)\|_X \le \sqrt{\frac{\gamma(\mu)}{\alpha(\mu)}} \inf_{w_N\in W_N} \|u(\mu)-w_N\|_X. \tag{8}$$

(We note that, in the coercive case, stability of our ("conforming") discrete approximation is not an issue; the noncoercive case is decidedly more delicate (see Section 5.3.1).) Furthermore, for our compliance output,

$$\begin{aligned} s(\mu)&=s_N(\mu)+\ell(u-u_N)=s_N(\mu)+a(u,u-u_N;\mu)\\ &=s_N(\mu)+a(u-u_N,u-u_N;\mu) \end{aligned} \tag{9}$$

from symmetry and Galerkin orthogonality. It follows that $s(\mu)-s_N(\mu)$ converges as the square of the error in the best approximation and, from coercivity, that $s_N(\mu)$ is a lower bound for $s(\mu)$.

3.2.2 Best Approximation. It now remains to bound the dependence of the error in the best approximation as a function of N. At present, the theory is restricted to the case in which $P=1$, $\mathcal{D}=[0,\mu_{\max}]$, and

$$a(w,v;\mu)=a_0(w,v)+\mu a_1(w,v), \tag{10}$$

where a_0 is continuous, coercive, and symmetric, and a_1 is continuous, positive semi-definite ($a_1(w,w)\ge 0$, $\forall w\in X$), and symmetric. This model problem (10) is rather broadly relevant, for

[2] TRUSS: http://augustine.mit.edu/simple_truss/simple_truss.pdf

Table 1 Error, error bound (Method I), and effectivity as a function of N, at a particular representative point $\mu \in \mathcal{D}$, for the two-dimensional thermal fin problem (compliant output)

N	$\lvert s(\mu)-s_N(\mu)\rvert/s(\mu)$	$\Delta_N(\mu)/s(\mu)$	$\eta_N(\mu)$
10	1.29×10^{-2}	8.60×10^{-2}	2.85
20	1.29×10^{-3}	9.36×10^{-3}	2.76
30	5.37×10^{-4}	4.25×10^{-3}	2.68
40	8.00×10^{-5}	5.30×10^{-4}	2.86
50	3.97×10^{-5}	2.97×10^{-4}	2.72
60	1.34×10^{-5}	1.27×10^{-4}	2.54
70	8.10×10^{-6}	7.72×10^{-5}	2.53
80	2.56×10^{-6}	2.24×10^{-5}	2.59

Table 2 Error, error bound (Method II), and effectivity as a function of N, at a particular representative point $\mu \in \mathcal{D}$, for the truss problem (compliant output)

N	$\lvert s(\mu)-s_N(\mu)\rvert/s(\mu)$	$\Delta_N(\mu)/s(\mu)$	$\eta_N(\mu)$
10	3.26×10^{-2}	6.47×10^{-2}	1.98
20	2.56×10^{-4}	4.74×10^{-4}	1.85
30	7.31×10^{-5}	1.38×10^{-4}	1.89
40	1.91×10^{-5}	3.59×10^{-5}	1.88
50	1.09×10^{-5}	2.08×10^{-5}	1.90
60	4.10×10^{-6}	8.19×10^{-6}	2.00
70	2.61×10^{-6}	5.22×10^{-6}	2.00
80	1.19×10^{-6}	2.39×10^{-6}	2.00

example to variable orthotropic conductivity, variable rectilinear geometry, variable piecewise-constant conductivity, and variable Robin boundary conditions.

We now suppose that the μ_n, $n=1,\ldots,N$, are logarithmically distributed in the sense that

$$\ln(\bar{\lambda}\mu_n+1)=\frac{n-1}{N-1}\ln(\bar{\lambda}\mu_{\max}+1),\quad n=1,\ldots,N, \tag{11}$$

where $\bar{\lambda}$ is an upper bound for the maximum eigenvalue of a_1 relative to a_0. (Note $\bar{\lambda}$ is perforce bounded thanks to our assumption of continuity and coercivity; the possibility of a continuous spectrum does not, in practice, pose any problems.) We can then prove (Maday et al. [19]) that, for $N>N_{\text{crit}}\equiv e\,\ln(\bar{\lambda}\mu_{\max}+1)$,

$$\inf_{w_N\in W_N}\|u(\mu)-w_N\|_X\leq(1+\mu_{\max}\bar{\lambda})\|u(0)\|_X \times\exp\left\{\frac{-(N-1)}{(N_{\text{crit}}-1)}\right\},\quad \forall\,\mu\in\mathcal{D}. \tag{12}$$

We observe exponential convergence, uniformly (globally) for all μ in $\mathcal{D}$, with only very weak (logarithmic) dependence on the range of the parameter ($\mu_{\max}$). (Note the constants in (12) are for the particular case in which $(\cdot,\cdot)_X=a_0(\cdot,\cdot)$.)

The proof exploits a parameter-space (nonpolynomial) interpolant as a surrogate for the Galerkin approximation. As a result, the bound is not always ''sharp:'' we observe many cases in which the Galerkin projection is considerably better than the associated interpolant; optimality (8) chooses to ''illuminate'' only certain points μ_n, automatically selecting a best ''subapproximation'' among all (combinatorially many) possibilities. We thus see why reduced-basis *state-space* approximation of $s(\mu)$ via $u(\mu)$ is preferred to simple *parameter-space* interpolation of $s(\mu)$ (''connecting the dots'') via $(\mu_n, s(\mu_n))$ pairs. We note, however, that the logarithmic point distribution (11) implicated by our interpolant-based arguments is *not* simply an artifact of the proof: in numerous numerical tests, the logarithmic distribution performs considerably (and in many cases, provably) better than other more obvious candidates, in particular for large ranges of the parameter. Fortunately, the convergence rate is not *too* sensitive to point selection: the theory only requires a log ''on the average'' distribution (Maday et al. [19]); and, in practice, $\bar{\lambda}$ need not be a sharp upper bound.

The result (12) is certainly tied to the particular form (10) and associated regularity of $u(\mu)$. However, we do observe similar exponential behavior for more general operators; and, most importantly, the exponential convergence rate degrades only very slowly with increasing parameter dimension, P. We present in Table 1 the error $\lvert s(\mu)-s_N(\mu)\rvert/s(\mu)$ as a function of N, at a particular representative point μ in $\mathcal{D}$, for the two-dimensional thermal fin problem of Section 2.2.1; we present similar data in Table 2 for the truss problem of Section 2.2.2. In both cases, since tensor-product grids are prohibitively profligate as P increases, the μ_n are chosen ''log-randomly'' over $\mathcal{D}$: we sample from a multivariate uniform probability density on $\log(\mu)$. We observe that, for both the thermal fin ($P=7$) and truss ($P=4$) problems, the error is remarkably small even for very small N; and that, in both cases, very rapid convergence obtains as $N\to\infty$. We do not yet have any theory for $P>1$. But certainly the Galerkin optimality plays a central role, automatically selecting ''appropriate'' scattered-data subsets of S_N and associated ''good'' weights so as to mitigate the curse of dimensionality as P increases; and the log-random point distribution is also important—for example, for the truss problem of Table 2, a *non-logarithmic* uniform random point distribution for S_N yields errors which are larger by factors of 20 and 10 for $N=30$ and 80, respectively.

3.3 Computational Procedure. The theoretical and empirical results of Sections 3.1 and 3.2 suggest that N may, indeed, be chosen very small. We now develop off-line/on-line computational procedures that exploit this dimension reduction.

We first express $u_N(\mu)$ as

$$u_N(\mu)=\sum_{j=1}^{N} u_{Nj}(\mu)\zeta_j=(\underline{u}_N(\mu))^T\underline{\zeta}, \tag{13}$$

where $\underline{u}_N(\mu)\in\mathbb{R}^N$; we then choose for test functions $v=\zeta_i$, $i=1,\ldots,N$. Inserting these representations into (7) yields the desired algebraic equations for $\underline{u}_N(\mu)\in\mathbb{R}^N$,

$$\underline{A}_N(\mu)\underline{u}_N(\mu)=\underline{F}_N, \tag{14}$$

in terms of which the output can then be evaluated as $s_N(\mu)=\underline{F}_N^T\underline{u}_N(\mu)$. Here $\underline{A}_N(\mu)\in\mathbb{R}^{N\times N}$ is the SPD matrix with entries $A_{N\ i,j}(\mu)\equiv a(\zeta_j,\zeta_i;\mu)$, $1\leq i,j\leq N$, and $\underline{F}_N\in\mathbb{R}^N$ is the ''load'' (and ''output'') vector with entries $F_{N\ i}\equiv f(\zeta_i)$, $i=1,\ldots,N$.

We now invoke (2) to write

$$A_{N\ i,j}(\mu)=a(\zeta_j,\zeta_i;\mu)=\sum_{q=1}^{Q}\sigma^q(\mu)a^q(\zeta_j,\zeta_i), \tag{15}$$

or

$$\underline{A}_N(\mu)=\sum_{q=1}^{Q}\sigma^q(\mu)\underline{A}_N^q,$$

where the $\underline{A}_N^q\in\mathbb{R}^{N\times N}$ are given by $A_{N\ i,j}^q=a^q(\zeta_j,\zeta_i)$, $i\leq i,j\leq N$, $1\leq q\leq Q$. The off-line/on-line decomposition is now clear. In the *off-line* stage, we compute the $u(\mu_n)$ and form the $\underline{A}_N^q$ and $\underline{F}_N$: this requires N (expensive) ''a'' finite element solutions and $O(QN^2)$ finite-element-vector inner products. In the *on-line* stage, for any given new μ, we first form $\underline{A}_N$ from (15), then solve (14) for $\underline{u}_N(\mu)$, and finally evaluate $s_N(\mu)=\underline{F}_N^T\underline{u}_N(\mu)$: this requires $O(QN^2)+O(2/3\ N^3)$ operations and $O(QN^2)$ storage.

Thus, as required, the incremental, or marginal, cost to evaluate $s_N(\mu)$ for any given new μ—as proposed in a design, optimization, or inverse-problem context—is very small: first, because N is very small, typically $O(10)$—thanks to the good convergence properties of W_N; and second, because (14) can be very rapidly

assembled and inverted—thanks to the off-line/on-line decomposition (see Balmes [9] for an earlier application of this strategy within the reduced-basis context). For the problems discussed in this paper, the resulting computational savings relative to standard (well-designed) finite-element approaches are significant, at least $O(10)$, typically $O(100)$, and often $O(1000)$ or more.

4 A Posteriori Error Estimation: Output Bounds

From Section 3 we know that, in theory, we can obtain $s_N(\mu)$ very inexpensively: the on-line computational effort scales as $O(2/3\ N^3)+O(QN^2)$; and N can, *in theory*, be chosen quite small. However, *in practice*, we do not know *how* small N can be chosen: this will depend on the desired accuracy, the selected output(s) of interest, and the particular problem in question; in some cases $N=5$ may suffice, while in other cases, $N=100$ may still be insufficient. In the face of this uncertainty, either too many or too few basis functions will be retained: the former results in computational inefficiency; the latter in unacceptable uncertainty---particularly egregious in the decision contexts in which reduced-basis methods typically serve. We thus need *a posteriori* error estimators for s_N. Surprisingly, *a posteriori* error estimation has received relatively little attention within the reduced-basis frame-work (Noor and Peters [8]), even though reduced-basis methods are particularly in need of accuracy assessment: the spaces are *ad hoc* and pre-asymptotic, thus admitting relatively little intuition, "rules of thumb," or standard approximation notions.

Recall that, in this section, we continue to assume that a is coercive and symmetric, and that ℓ is "compliant."

4.1 Method I. The approach described in this section is a particular instance of a general "variational" framework for *a posteriori* error estimation of outputs of interest. However, the reduced-basis instantiation described here differs significantly from earlier applications to finite element discretization error (Maday et al. [20], Machiels et al. [21]) and iterative solution error (Patera and Rønquist [22]) both in the choice of (energy) relaxation and in the associated computational artifice.

4.1.1 Formulation. We assume that we are given a positive function $g(\mu):\mathcal{D}\rightarrow\mathbb{R}_+$, and a continuous, coercive, symmetric (μ-independent) bilinear from $\hat{a}:X\times X\rightarrow\mathbb{R}$, such that

$$\underline{\alpha}_0\|v\|_X^2\leqslant g(\mu)\hat{a}(v,v)\leqslant a(v,v;\mu),\quad \forall v\in X,\forall\mu\in\mathcal{D} \tag{16}$$

for some positive real constant $\underline{\alpha}_0$. We then find $\hat{e}(\mu)\in X$ such that

$$g(\mu)\hat{a}(\hat{e}(\mu),v)=R(v;u_N(\mu);\mu),\quad \forall v\in X, \tag{17}$$

where for a given $w\in X$, $R(v;w;\mu)=\ell(v)-a(w,v;\mu)$ is the weak form of the residual. Our lower and upper output estimators are then evaluated as

$$s_N^-(\mu)\equiv s_N(\mu),\ \text{ and }\ s_N^+(\mu)\equiv s_N(\mu)+\Delta_N(\mu), \tag{18}$$

respectively, where

$$\Delta_N(\mu)\equiv g(\mu)\hat{a}(\hat{e}(\mu),\hat{e}(\mu)) \tag{19}$$

is the estimator gap.

4.1.2 Properties. We shall prove in this section that $s_N^-(\mu)\leqslant s(\mu)\leqslant s_N^+(\mu)$, and hence that $|s(\mu)-s_N(\mu)|=s(\mu)-s_N(\mu)\leqslant\Delta_N(\mu)$. Our lower and upper output estimators are thus lower and upper output *bounds*; and our output estimator gap is thus an output *bound* gap—a rigorous bound for the error in the output of interest. It is also critical that $\Delta_N(\mu)$ be a relatively *sharp* bound for the true error: a poor (overly large) bound will encourage us to refine an approximation which is, in fact, already adequate—with a corresponding (unnecessary) increase in off-line and on-line computational effort. We shall prove in this section that $\Delta_N(\mu)\leqslant(\gamma_0/\underline{\alpha}_0)(s(\mu)-s_N(\mu))$, where γ_0 and $\underline{\alpha}_0$ are the *N-independent* a-continuity and $g(\mu)\hat{a}$-coercivity constants defined earlier. Our two results of this section can thus be summarized as

$$1\leqslant\eta_N(\mu)\leqslant\text{Const},\quad\forall N, \tag{20}$$

where

$$\eta_N(\mu)=\frac{\Delta_N(\mu)}{s(\mu)-s_N(\mu)} \tag{21}$$

is the *effectivity*, and Const is a constant independent of N. We shall denote the left (bounding property) and right (sharpness property) inequalities of (20) as the lower effectivity and upper effectivity inequalities, respectively.

We first prove the lower effectivity inequality (bounding property): $s_N^-(\mu)\leqslant s(\mu)\leqslant s_N^+(\mu)$, $\forall\mu\in\mathcal{D}$, for $s_N^-(\mu)$ and $s_N^+(\mu)$ defined in (18). The lower bound property follows directly from Section 3.2.1. To prove the upper bound property, we first observe that $R(v;u_N;\mu)=a(u(\mu)-u_N(\mu),v;\mu)=a(e(\mu),v;\mu)$, where $e(\mu)\equiv u(\mu)-u_N(\mu)$; we may thus rewrite (17) as $g(\mu)\hat{a}(\hat{e}(\mu),v)=a(e(\mu),v;\mu),\forall v\in X$. We thus obtain

$$\begin{aligned}g(\mu)\hat{a}(\hat{e},\hat{e})&=g(\mu)\hat{a}(\hat{e}-e,\hat{e}-e)+2g(\mu)\hat{a}(\hat{e},e)-g(\mu)\hat{a}(e,e)\\&=g(\mu)\hat{a}(\hat{e}-e,\hat{e}-e)+(a(e,e;\mu)-g(\mu)\hat{a}(e,e))\\&\quad+a(e,e;\mu)\\&\geqslant a(e,e;\mu)\end{aligned} \tag{22}$$

since $g(\mu)\hat{a}(\hat{e}(\mu)-e(\mu),\hat{e}(\mu)-e(\mu))\geqslant 0$ and $a(e(\mu),e(\mu);\mu)-g(\mu)\hat{a}(e\mu),e(\mu))\geqslant 0$ from (16). Invoking (9) and (22), we then obtain $s(\mu)-s_N(\mu)=a(e(\mu),e(\mu);\mu)\leqslant g(\mu)\hat{a}(\hat{e}(\mu),\hat{e}(\mu))$; and thus $s(\mu)\leqslant s_N(\mu)+g(\mu)\hat{a}(\hat{e}(\mu),\hat{e}(\mu))\equiv s_N^+(\mu)$, as desired.

We next prove the upper effectivity inequality (sharpness property):

$$\eta_N(\mu)=\frac{\Delta_N(\mu)}{s(\mu)-s_N(\mu)}\leqslant\frac{\gamma_0}{\underline{\alpha}_0},\quad\forall N.$$

To begin, we appeal to a-continuity and $g(\mu)\hat{a}$-coercivity to obtain

$$a(\hat{e}(\mu),\hat{e}(\mu);\mu)\leqslant\frac{\gamma_0 g(\mu)}{\underline{\alpha}_0}\hat{a}(\hat{e}(\mu),\hat{e}(\mu)). \tag{23}$$

But from the modified error equation (17) we know that $g(\mu)\hat{a}(\hat{e}(\mu),\hat{e}(\mu))=a(e(\mu),\hat{e}(\mu);\mu)$. Invoking the Cauchy-Schwartz inequality, we obtain

$$\begin{aligned}g(\mu)\hat{a}(\hat{e},\hat{e})&=a(e,\hat{e};\mu)\\&\leqslant(a(\hat{e},\hat{e};\mu))^{1/2}(a(e,e;\mu))^{1/2}\\&\leqslant\left(\frac{\gamma_0}{\underline{\alpha}_0}\right)^{1/2}(g(\mu)\hat{a}(\hat{e},\hat{e}))^{1/2}(a(e,e;\mu))^{1/2};\end{aligned}$$

the desired result then directly follows from (19) and (9).

We now provide empirical evidence for (20). In particular, we present in Table 1 the bound gap and effectivities for the thermal fin example. Clearly, $\eta_N(\mu)$ is always greater than unity for any N, and bounded---indeed, quite close to unity---as $N\rightarrow\infty$.

4.1.3. Computational Procedure. Finally, we turn to the computational artifice by which we can efficiently compute $\Delta_N(\mu)$ in the on-line stage of our procedure. We again exploit the affine parameter dependence, but now in a less transparent fashion. To begin, we rewrite the "modified" error equation, (17), as

$$\hat{a}(\hat{e}(\mu),v)=\frac{1}{g(\mu)}\left(\ell(v)-\sum_{q=1}^{Q}\sum_{j=1}^{N}\sigma^q(\mu)u_{N\ j}(\mu)a^q(\zeta_j,v)\right),\quad\forall v\in X,$$

where we have appealed to our reduced-basis approximation (13) and the affine decomposition (2). It is immediately clear from linear superposition that we can express $\hat{e}(\mu)$ as

$$\hat{e}(\mu) = \frac{1}{g(\mu)} \left(\hat{z}_0 + \sum_{q=1}^{Q} \sum_{j=1}^{N} \sigma^q(\mu) u_{N\ j}(\mu) \hat{z}_j^q \right), \qquad (24)$$

where $\hat{z}_0 \in X$ satisfies $\hat{a}(\hat{z}_0, v) = \ell(v)$, $\forall v \in X$, and $\hat{z}_j^q \in X$, $j = 1, \ldots, N$, $q = 1, \ldots, Q$, satisfies $\hat{a}(\hat{z}_j^q, v) = -a^q(\zeta_j, v)$, $\forall v \in X$. Inserting (24) into our expression for the upper bound, $s_N^+(\mu) = s_N(\mu) + g(\mu)\hat{a}(\hat{e}(\mu), \hat{e}(\mu))$, we obtain

$$s_N^+(\mu) = s_N(\mu) + \frac{1}{g(\mu)} \left(c_0 + 2 \sum_{q=1}^{Q} \sum_{j=1}^{N} \sigma^q(\mu) u_{N\ j}(\mu) \Lambda_j^q + \sum_{q=1}^{Q} \sum_{q'=1}^{Q} \sum_{j=1}^{N} \sum_{j'=1}^{N} \sigma^q(\mu) \sigma^{q'}(\mu) u_{N\ j}(\mu) u_{N\ j'}(\mu) \Gamma_{jj'}^{qq'} \right) \qquad (25)$$

where $c_0 = \hat{a}(\hat{z}_0, \hat{z}_0)$, $\Lambda_j^q = \hat{a}(\hat{z}_0, \hat{z}_j^q)$, and $\Gamma_{jj'}^{qq'} = \hat{a}(\hat{z}_j^q, \hat{z}_{j'}^{q'})$.

The off-line/on-line decomposition should now be clear. In the *off-line* stage we compute $\hat{z}_0$ and $\hat{z}_j^q$, $j = 1, \ldots, N$, $q = 1, \ldots, Q$, and then form c_0, Λ_j^q, and $\Gamma_{jj'}^{qq'}$: this requires $QN + 1$ (expensive) "$\hat{a}$" finite element solutions, and $O(Q^2N^2)$ finite-element-vector inner products. In the *on-line* stage, for any given new μ, we evaluate s_N^+ as expressed in (25): this requires $O(Q^2N^2)$ operations and $O(Q^2N^2)$ storage (for c_0, Λ_j^q, and $\Gamma_{jj'}^{qq'}$). As for the computation of $s_N(\mu)$, the marginal cost for the computation of $s_N^{\pm}(\mu)$ for any given new μ is quite small—in particular, it is *independent* of the dimension of the truth finite element approximation space X.

There are a variety of ways in which the off-line/on-line decomposition and output error bounds can be exploited. A particularly attractive mode incorporates the error bounds into an on-line adaptive process, in which we successively approximate $s_N(\mu)$ on a sequence of approximation spaces $W_{N_j'} \subset W_N$, $N_j' = N_0 2^j$---for example, $W_{N_j'}$ may contain the N_j' samples points of S_N closest to the new μ of interest---until $\Delta_{N_j'}$ is less than a specified error tolerance. This procedure both minimizes the on-line computational effort and reduces conditioning problems—while simultaneously ensuring accuracy and certainty.

The essential advantage of the approach described in this section is the guarantee of rigorous bounds. There are, however, certain disadvantages. The first set of disadvantages relates to the choice of $g(\mu)$ and $\hat{a}$. In many cases, simple inspection suffices: for example, in our thermal fin problem of Section 2.2.1, $g(\mu) = \min_{q=1,\ldots,Q} \sigma^q(\mu)$ and $\hat{a}(w, v) = \Sigma_{q=1}^{Q} a^q(w, v)$ yields the very good effectivities summarized in Table 1. In other cases, however, there is no self-evident (or readily computed, Maday et al. [23]) good choice: for example, for the truss problem of Section 2.2.2, the existence of almost-pure rotations renders $g(\mu)$ very small relative to $\gamma(\mu)$, with corresponding detriment to $\eta_N(\mu)$. The second set of disadvantages relates to the computational expense—the $O(Q)$ off-line and the $O(Q^2)$ on-line scaling induced by (24) and (25), respectively. Both of these disadvantages are eliminated in the "Method II" to be discussed in the next section; however "Method II" only provides *asymptotic* bounds as $N \to \infty$. The choice thus depends on the relative importance of absolute certainty and computational efficiency.

4.2. Method II. As already indicated, Method I has certain limitations; we discuss here a Method II which addresses these limitations, albeit at the loss of complete certainty.

4.2.1. Formulation. To begin, we set $M > N$, and introduce a parameter sample $S_M = \{\mu_1, \ldots, \mu_M\}$ and associated reduced-basis approximation space $W_M = \text{span}\{\zeta_m \equiv u(\mu_m), m = 1, \ldots, M\}$; for both theoretical and practical reasons we require $S_N \subset S_M$ and therefore $W_N \subset W_M$. The procedure is very simple: we first find $u_M(\mu) \in W_M$ such that $a(u_M(\mu), v; \mu) = f(v)$, $\forall v \in W_M$; we then evaluate $s_M(\mu) = \ell(u_M(\mu))$; and, finally, we compute our upper and lower output estimators as

$$s_{N,M}^-(\mu) = s_N(\mu), \quad s_{N,M}^+(\mu) = s_N(\mu) + \Delta_{N,M}(\mu), \qquad (26)$$

where $\Delta_{N,M}(\mu)$, the estimator bound gap, is given by

$$\Delta_{N,M}(\mu) = \frac{1}{\tau}(s_M(\mu) - s_N(\mu)) \qquad (27)$$

for some $\tau \in (0,1)$. The effectivity of the approximation is defined as

$$\eta_{N,M}(\mu) = \frac{\Delta_{N,M}(\mu)}{s(\mu) - s_N(\mu)}. \qquad (28)$$

For our purposes here, we shall consider $M = 2N$.

4.2.2. Properties. As for Method I, we would like to prove the effectivity inequality $1 \leq \eta_{N,2N}(\mu) \leq \text{Const}$, $\forall N$. However, we will only be able to demonstrate an asymptotic form of this inequality. Furthermore, the latter shall require, and we shall make, the hypothesis that

$$\varepsilon_{N,2N}(\mu) \equiv \frac{s(\mu) - s_{2N}(\mu)}{s(\mu) - s_N(\mu)} \to 0, \quad \text{as } N \to \infty. \qquad (29)$$

We note that the assumption (29) is certainly plausible: if our *a priori* bound of (12) in fact reflects asymptotic behavior, then $s(\mu) - s_N(\mu) \sim c_1 e^{-c_2 N}$, $s(\mu) - s_{2N}(\mu) \sim c_1 e^{-2c_2 N}$, and hence $\varepsilon_{N,2N}(\mu) \sim e^{-c_2 N}$, as desired.

We first prove the lower effectivity inequality (bounding property): $s_{N,2N}^-(\mu) \leq s(\mu) \leq s_{N,2N}^+(\mu)$, as $N \to \infty$. To demonstrate the lower bound we again appeal to (9) and the coercivity of a; indeed, this result (still) obtains for *all* N. To demonstrate the upper bound, we write

$$s_{N,2N}^+(\mu) = s(\mu) + \left(\frac{1}{\tau} - 1\right)(s(\mu) - s_N(\mu)) - \frac{1}{\tau}(s(\mu) - s_{2N}(\mu)) \qquad (30)$$

$$= s(\mu) + \left(\frac{1}{\tau}[1 - \varepsilon_{N,2N}(\mu)] - 1\right)(s(\mu) - s_N(\mu)). \qquad (31)$$

We now recall that $s(\mu) - s_N(\mu) \geq 0$, and that $0 < \tau < 1$—that is, $1/\tau > 1$; it then follows from (31) and our hypothesis (29) that there exists a finite N^* such that

$$s_{N,2N}^+(\mu) - s(\mu) \geq 0, \quad \forall N > N^*. \qquad (32)$$

This concludes the proof: we obtain *asymptotic* bounds.

We now prove the upper effectivity inequality (sharpness property). From the definitions of $\eta_{N,2N}(\mu)$, $\Delta_{N,2N}(\mu)$ and $\varepsilon_{N,2N}(\mu)$, we directly obtain

$$\eta_{N,2N}(\mu) = \frac{1}{\tau} \frac{s_{2N}(\mu) - s_N(\mu)}{s(\mu) - s_N(\mu)}$$

$$= \frac{1}{\tau} \frac{(s_{2N}(\mu) - s(\mu)) - (s_N(\mu) - s(\mu))}{(s(\mu) - s_N(\mu))} \qquad (33)$$

$$= \frac{1}{\tau}(1 - \varepsilon_{N,2N}(\mu)). \qquad (34)$$

It is readily shown that $\eta_{N,2N}(\mu)$ is bounded from above by $1/\tau$ for all N: we know from (9) that $\varepsilon_{N,2N}(\mu)$ is strictly non-negative. It can also readily be shown that $\eta_{N,2N}(\mu)$ is non-negative: since

$W_N \subset W_{2N}$, it follows from (8) (for $(\cdot,\cdot)_X = a(\cdot,\cdot;\mu)$) and (9) that $s(\mu) \geq s_{2N}(\mu) \geq s_N(\mu)$, and hence $\varepsilon_{N,2N}(\mu) \leq 1$. We thus conclude that $0 \leq \eta_{N,2N}(\mu) \leq 1/\tau$ for all N. Furthermore, from our hypothesis on $\varepsilon_{N,2N}(\mu)$, (29), we know that $\eta_{N,2N}(\mu)$ will *tend* to $1/\tau$ as N increases.

The essential approximation enabler is exponential convergence: we obtain bounds even for rather small N and relatively large τ. We thus achieve both "near" certainty *and* good effectivities. We demonstrate this claim in Table 2, in which we present the bound gap and effectivity for our truss example of Section 2.2.2; the results tabulated correspond to the choice $\tau = 1/2$. We clearly obtain bounds for all N; and we observe that $\eta_{N,2N}(\mu)$ does, indeed, rather quickly approach $1/\tau$.

4.2.3. Computational Procedure. Since the error bounds are based entirely on evaluation of the output,we can directly adapt the off-line/on-line procedure of Section 3.3. Note that the calculation of the output approximation $s_N(\mu)$ and the output bounds are now integrated: $\underline{A}_N(\mu)$ and $\underline{F}_N(\mu)$ (yielding $s_N(\mu)$) are a sub-matrix and sub-vector of $\underline{A}_{2N}(\mu)$ and $\underline{F}_{2N}(\mu)$ (yielding $s_{2N}(\mu)$, $\Delta_{N,2N}(\mu)$, and $s^{\pm}_{N,2N}(\mu)$), respectively. In the *off-line* stage, we compute the $u(\mu_n)$ and form the $\underline{A}^q_{2N}$ and $\underline{F}_{2N}$: this requires $2N$ (expensive) "a" finite element solutions, and $O(4QN^2)$ finite-element-vector inner products. In the *on-line* stage, for any given new μ, we first form $\underline{A}_N(\mu)$, $\underline{F}_N$ and $\underline{A}_{2N}(\mu)$, $\underline{F}_{2N}$, then solve for $\underline{u}_N(\mu)$ and $\underline{u}_{2N}(\mu)$, and finally evaluate $s^{\pm}_{N,2N}(\mu)$: this requires $O(4QN^2) + O(16/3\ N^3)$ operations and $O(4QN^2)$ storage. The on-line effort for this Method II predictor/error estimator procedure (based on $s_N(\mu)$ and $s_{2N}(\mu)$) will thus require eightfold more operations than the "predictor-only" procedure of Section 3.

Method II is in some sense very naïve: we simply replace the true output $s(\mu)$ with a finer-approximation surrogate $s_{2N}(\mu)$. (There are more obscure ways to describe the method—in terms of a reduced-basis approximation for the error—however, there is little to be gained from these alternative interpretations.) The essential computation enabler is again exponential convergence, which permits us to choose $M = 2N$—hence controlling the additional computational effort attributable to error estimation—while simultaneously ensuring that $\varepsilon_{N,2N}(\mu)$ tends rapidly to zero. Exponential convergence also ensures that the cost to compute both $s_N(\mu)$ and $s_{2N}(\mu)$ is "negligible." In actual practice, since $s_{2N}(\mu)$ is available, we can of course take $s_{2N}(\mu)$, rather than $s_N(\mu)$, as our output prediction; this greatly improves not only accuracy, but also certainty—$\Delta_{N,2N}(\mu)$ is almost surely a bound for $s(\mu) - s_{2N}(\mu)$, albeit an exponentially conservative bound as N tends to infinity.

5. Extensions

5.1. Noncompliant Outputs and Nonsymmetric Operators. In Sections 3 and 4 we formulate the reduced-basis method and associated error estimation procedure for the case of compliant outputs, $\ell(v) = f(v)$, $\forall v \in X$. We briefly summarize here the formulation and theory for more general linear bounded output functionals; moreover, the assumption of symmetry (but not yet coercivity) is relaxed, permitting treatment of a wider class of problems---a representative example is the convection-diffusion equation, in which the presence of the convective term renders the operator nonsymmetric. We first present the reduced-basis approximation, now involving a dual or adjoint problem; we then formulate the associated *a posteriori* error estimators; and we conclude with a few illustrative results.

As a preliminary, we first generalize the abstract formulation of Section 2.1. As before, we define the "primal" problem as in (4), however we of course no longer require symmetry. But we also introduce an associated adjoint or "dual" problem: for any $\mu \in X$, find $\psi(\mu) \in X$ such that

$$a(v, \psi(\mu); \mu) = -\ell(v), \quad \forall v \in X; \tag{35}$$

recall that $\ell(v)$ is our output functional.

5.1.1. Reduced-Basis Approximation. To develop the reduced-basis space, we first choose, randomly or log-randomly as described in Section 3.2, a sample set in parameter space, $S_{N/2} = \{\mu_1, \ldots, \mu_{N/2}\}$, where $\mu_i \in \mathcal{D}$, $i = 1, \ldots, N/2$ (N even). We next define an "integrated" Lagrangian reduced-basis approximation space, $W_N = \text{span}\{(u(\mu_n), \psi(\mu_n)), n = 1, \ldots, N/2\}$.

For any $\mu \in \mathcal{D}$, our reduced basis approximation is then obtained by standard Galerkin projection onto W_N (though for highly nonsymmetric operators minimum residual and Petrov-Galerkin projections are attractive—stabler—alternatives). To wit, for the primal problem, we find $u_N(\mu) \in W_N$ such that $a(u_N(\mu), v; \mu) = f(v)$, $\forall v \in W_N$; and for the adjoint problem, we define (though, unless otherwise indicated, do *not* compute) $\psi_N(\mu) \in W_N$ such that $a(v, \psi_N(\mu); \mu) = -\ell(v)$, $\forall v \in W_N$. The reduced-basis output approximation is then calculated from $s_N(\mu) = \ell(u_N(\mu))$.

Turning now to the *a priori* theory, it follows from standard arguments that $u_N(\mu)$ and $\psi_N(\mu)$ are "optimal" in the sense that

$$\|u(\mu) - u_N(\mu)\|_X \leq \left(1 + \frac{\gamma(\mu)}{\alpha(\mu)}\right) \inf_{w_N \in W_N} \|u(\mu) - w_N\|_X,$$

$$\|\psi(\mu) - \psi_N(\mu)\|_X \leq \left(1 + \frac{\gamma(\mu)}{\alpha(\mu)}\right) \inf_{w_N \in W_N} \|\psi(\mu) - w_N\|_X.$$

The best approximation analysis is then similar to that presented in Section 3.2. As regards our output, we now have

$$\begin{aligned} |s(\mu) - s_N(\mu)| &= |\ell(u(\mu)) - \ell(u_N(\mu))| \\ &= |a(u - u_N, \psi; \mu)| \\ &= |a(u - u_N, \psi - \psi_N; \mu)| \leq \gamma_0 \|u - u_N\|_X \|\psi - \psi_N\|_X \end{aligned} \tag{36}$$

from Galerkin orthogonality, the definition of the primal and the adjoint problems, and the Cauchy-Schwartz inequality. We now understand why we include the $\psi(\mu_n)$ in W_N: to ensure that $\|\psi(\mu) - \psi_N(\mu)\|_X$ is small. We thus recover the "square" effect in the convergence rate of the output, albeit (and unlike the symmetric case) at the expense of some additional computational effort—the inclusion of the $\psi(\mu_n)$ in W_N; typically, even for the very rapidly convergent reduced-basis approximation, the "fixed error-minimum cost" criterion favors the adjoint enrichment.

For simplicity of exposition (and to a certain extent, implementation),we present here the "integrated" primal-dual approximation space. However, there are significant computational and conditioning advantages associated with a "nonintegrated" approach, in which we introduce *separate* primal ($u(\mu_n)$) and dual ($\psi(\mu_n)$) approximation spaces for $u(\mu)$ and $\psi(\mu)$, respectively. Note in the "nonintegrated" case we are obliged to compute $\psi_N(\mu)$, since to preserve the output error "square effect" we must modify our predictor with a residual correction, $f(\psi_N(\mu)) - a(u_N(\mu), \psi_N(\mu); \mu)$ (Maday et al. [23]). Both the "integrated" and "nonintegrated" approaches admit an off-line/on-line decomposition similar to that described in Section 3.3 for the compliant, symmetric problem; as before, the on-line complexity and storage are independent of the dimension of the very fine ("truth") finite element approximation.

5.1.2. Method I A Posteriori Error Estimators. We extend here the method developed in Section 4.1.2 to the more general case of noncompliant and nonsymmetric problems. We begin with the formulation.

We first find $\hat{e}^{\text{pr}}(\mu) \in X$ such that

$$g(\mu)\hat{a}(\hat{e}^{\text{pr}}(\mu), v) = R^{\text{pr}}(v; u_N(\mu); \mu), \quad \forall v \in X,$$

where $R^{\text{pr}}(v; w; \mu) \equiv f(v) - a(w, v; \mu)$, $\forall v \in X$; and $\hat{e}^{\text{du}}(\mu) \in X$ such that

$$g(\mu)\hat{a}(\hat{e}^{\rm du}(\mu),v)=R^{\rm du}(v;\psi_N(\mu);\mu),\quad \forall v\in X,$$

where $R^{\rm du}(v;w;\mu)\equiv-\ell(v)-a(v,w;\mu)$, $\forall v\in X$. We then define

$$\overline{s_N}(\mu)=s_N(\mu)-\frac{g(\mu)}{2}\hat{a}(\hat{e}^{\rm pr}(\mu),\hat{e}^{\rm du}(\mu)),\quad\text{and}\qquad(37)$$

$$\Delta_N(\mu)=\frac{g(\mu)}{2}[\hat{a}(\hat{e}^{\rm pr}(\mu),\hat{e}^{\rm pr}(\mu))]^{1/2}[\hat{a}(\hat{e}^{\rm du}(\mu),\hat{e}^{\rm du}(\mu))]^{1/2}.\qquad(38)$$

Finally, we evaluate our lower and upper estimators as $s_N^{\pm}(\mu)=\overline{s_N}(\mu)\pm\Delta_N(\mu)$. Note that, as before, $g(\mu)$ and $\hat{a}$ still satisfy (16); and that, furthermore, (16) will only involve the *symmetric* part of a. We define the effectivity as

$$\eta_N(\mu)=\frac{\Delta_N(\mu)}{|s(\mu)-s_N(\mu)|};\qquad(39)$$

note that $s(\mu)-s_N(\mu)$ now has no definite sign.

We now prove that our error estimators are bounds (the lower effectivity inequality): $s_N^-(\mu)\leq s(\mu)\leq s_N^+(\mu)$, $\forall N$. To begin, we define $\hat{e}^{\pm}(\mu)=\hat{e}^{\rm pr}(\mu)\mp 1/\kappa\ \hat{e}^{\rm du}(\mu)$, and note that, from the coercivity of $\hat{a}$,

$$\begin{aligned}\kappa g(\mu)\hat{a}&\left(e^{\rm pr}-\frac{1}{2}\hat{e}^{\pm},e^{\rm pr}-\frac{1}{2}\hat{e}^{\pm}\right)\\&=\kappa g(\mu)\hat{a}(e^{\rm pr},e^{\rm pr})\\&\quad+\frac{\kappa g(\mu)}{4}\hat{a}(\hat{e}^{\pm},\hat{e}^{\pm})-\kappa g(\mu)\hat{a}(\hat{e}^{\pm},e^{\rm pr})\geq 0,\end{aligned}\qquad(40)$$

where $e^{\rm pr}(\mu)=u(\mu)-u_N(\mu)$, $e^{\rm du}(\mu)=\psi(\mu)-\psi_N(\mu)$, and κ is a positive real number. From the definition of $\hat{e}^{\pm}(\mu)$ and $\hat{e}^{\rm pr}(\mu)$, $\hat{e}^{\rm du}(\mu)$, we can express the "cross-term" as

$$\begin{aligned}g(\mu)\hat{a}(\hat{e}^{\pm},e^{\rm pr})&=R^{\rm pr}(e^{\rm pr};u_N;\mu)\mp\frac{1}{\kappa}R^{\rm du}(e^{\rm pr};\psi_N;\mu)\\&=a(e^{\rm pr},e^{\rm pr};\mu)\mp\frac{1}{\kappa}a(e^{\rm pr},e^{\rm du};\mu)\\&=a(e^{\rm pr},e^{\rm pr};\mu)\pm\frac{1}{\kappa}(s(\mu)-s_N(\mu)),\end{aligned}\qquad(41)$$

since $R^{\rm pr}(e^{\rm pr};u_N;\mu)=a(u,e^{\rm pr};\mu)-a(u_N,e^{\rm pr};\mu)=a(e^{\rm pr},e^{\rm pr};\mu)$, $R^{\rm du}(e^{\rm pr};\psi_N;\mu)=a(e^{\rm pr},\psi;\mu)-a(e^{\rm pr},\psi_N;\mu)=a(e^{\rm pr},e^{\rm du};\mu)$, and $\ell(u)-\ell(u_N)=-a(u-u_N,\psi;\mu)=-a(u-u_N,\psi-\psi_N;\mu)$ (by Galerkin orthogonality)$=-a(e^{\rm pr},e^{\rm du};\mu)$. We then substitute (41) into (40) to obtain

$$\begin{aligned}\pm(s(\mu)-s_N(\mu))&\leq-\kappa(a(e^{\rm pr},e^{\rm pr};\mu)-g(\mu)\hat{a}(e^{\rm pr},e^{\rm pr}))\\&\quad+\frac{\kappa g(\mu)}{4}\hat{a}(\hat{e}^{\pm},\hat{e}^{\pm})\leq\frac{\kappa g(\mu)}{4}\hat{a}(\hat{e}^{\pm},\hat{e}^{\pm}),\end{aligned}$$

since $\kappa>0$ and $a(e^{\rm pr}(\mu),e^{\rm pr}(\mu);\mu)-g(\mu)\hat{a}(e^{\rm pr}(\mu),e^{\rm pr}(\mu))\geq 0$ from (16).

Expanding $\hat{e}^{\pm}(\mu)=\hat{e}^{\rm pr}(\mu)\mp 1/\kappa\ \hat{e}^{\rm du}(\mu)$ then gives

$$\begin{aligned}&\pm(s(\mu)-s_N(\mu))\\&\leq\frac{g(\mu)}{4}\left[\kappa\hat{a}(\hat{e}^{\rm pr},\hat{e}^{\rm pr})+\frac{1}{\kappa}\hat{a}(\hat{e}^{\rm du},\hat{e}^{\rm du})\mp 2\hat{a}(\hat{e}^{\rm pr},\hat{e}^{\rm du})\right],\end{aligned}$$

or

$$\begin{aligned}&\pm\left(s(\mu)-\left(s_N(\mu)-\frac{g(\mu)}{2}\hat{a}(\hat{e}^{\rm pr},\hat{e}^{\rm du})\right)\right)\\&\leq\frac{\kappa g(\mu)}{4}\hat{a}(\hat{e}^{\rm pr},\hat{e}^{\rm pr})+\frac{g(\mu)}{4\kappa}\hat{a}(\hat{e}^{\rm du},\hat{e}^{\rm du}).\end{aligned}\qquad(42)$$

We now choose $\kappa(\mu)$ as

$$\kappa(\mu)=\left(\frac{\hat{a}(\hat{e}^{\rm du}(\mu),\hat{e}^{\rm du}(\mu))}{\hat{a}(\hat{e}^{\rm pr}(\mu),\hat{e}^{\rm pr}(\mu))}\right)^{1/2}$$

so as to minimize the right-hand side (42); we then obtain

$$|s(\mu)-\overline{s_N}(\mu)|\leq\Delta_N(\mu),\qquad(43)$$

and hence $s_N^-(\mu)\leq s(\mu)\leq s_N^+(\mu)$.

We now turn to the upper effectivity inequality (sharpness property). If the primal and dual errors are a-orthogonal, or become increasingly orthogonal as N increases, then the effectivity will not, in fact, be bounded as $N\to\infty$. However, if we make the (plausible) hypothesis that $|s(\mu)-s_N(\mu)|\geq\underline{C}\|e^{\rm pr}(\mu)\|_X\|e^{\rm du}(\mu)\|_X$, then it is simple to demonstrate that

$$\eta_N(\mu)\leq\frac{\gamma_0^2}{2\underline{C}\underline{\alpha}_0}.\qquad(44)$$

In particular, it is an easy matter to demonstrate that $g^{1/2}(\mu)\times(\hat{a}(\hat{e}^{\rm pr}(\mu),\hat{e}^{\rm pr}(\mu)))^{1/2}\leq\gamma_0/\underline{\alpha}_0^{1/2}\|e^{\rm pr}(\mu)\|_X$ (note we lose a factor of $\gamma_0^{1/2}$ relative to the symmetric case); similarly, $g^{1/2}(\mu)\times(\hat{a}(\hat{e}^{\rm du}(\mu),\hat{e}^{\rm du}(\mu)))^{1/2}\leq\gamma_0/\underline{\alpha}_0^{1/2}\|e^{\rm du}(\mu)\|_X$. The desired result then directly follows from the definition of $\Delta_N(\mu)$ and our hypothesis on $|s(\mu)-s_N(\mu)|$.

Finally, turning to computational issues, we note that the off-line/on-line decomposition described in Section 4.1 for compliant symmetric problems directly extends to the noncompliant, nonsymmetric case—except that we must compute the norm of both the primal and dual "modified errors," with a concomitant doubling of computational effort.

5.1.3 Method II A Posteriori Error Estimators. We discuss here the extension of Method II of Section 4.2 to noncompliant outputs and nonsymmetric operators.

To begin, we set $M>N$, M even, and introduce a parameter sample $S_{M/2}=\{\mu_1,\ldots,\mu_{M/2}\}$ and associated "integrated" reduced-basis approximation space $W_M=\text{span}\{u(\mu_m),\psi(\mu_m),m=1,\ldots,M/2\}$. We first find $u_M(\mu)\in W_M$ such that $a(u_M(\mu),v;\mu)=f(v)$, $\forall v\in W_M$; we then evaluate $s_M(\mu)=\ell(u_M(\mu))$; and finally, we compute our upper and lower output estimators as

$$s_{N,M}^{\pm}(\mu)=s_N(\mu)+\frac{1}{2\tau}(s_M(\mu)-s_N(\mu))\pm\frac{1}{2}\Delta_{N,M}(\mu),\qquad(45)$$

$$\Delta_{N,M}(\mu)=\frac{1}{\tau}|s_M(\mu)-s_N(\mu)|,\qquad(46)$$

for $\tau\in(0,1)$. The effectivity of the approximation is defined as

$$\eta_{N,M}(\mu)=\frac{\Delta_{N,M}(\mu)}{|s(\mu)-s_N(\mu)|}.\qquad(47)$$

We shall again only consider $M=2N$.

As in Section 4.2, we would like to prove that $1\leq\eta_{N,2N}(\mu)\leq$Const for sufficiently large N; and, as in Section 4.2, we must again make the hypothesis (29). We first consider the lower effectivity inequality (bounding property), and prove that

$$s_{N,2N}^-(\mu)\leq s(\mu)\leq s_{N,2N}^+(\mu),\quad\text{as}\quad N\to\infty.\qquad(48)$$

In particular, simple algebraic manipulations yield

$$\begin{aligned}s_{N,2N}^-(\mu)=s(\mu)-\frac{1}{1-\varepsilon_{N,2N}}|s_N(\mu)-s_{2N}(\mu)|\\\times\begin{cases}1 & s_{2N}(\mu)\geq s_N(\mu)\\ \dfrac{1}{\tau}(1-\varepsilon_{N,2N})-1 & s_{2N}(\mu)<s_N(\mu)\end{cases},\end{aligned}\qquad(49)$$

Table 3 Error, error bound (Method II), and effectivity as a function of N, at a particular representative point $\mu \in \mathcal{D}$, for the truss problem (noncompliant output)

N	$\lvert s(\mu)-s_N(\mu)\rvert/s(\mu)$	$\Delta_{N,2N}(\mu)/s(\mu)$	$\eta_{N,2N}(\mu)$
20	2.35×10^{-2}	4.67×10^{-2}	1.99
40	1.74×10^{-4}	3.19×10^{-4}	1.83
60	5.59×10^{-5}	1.06×10^{-4}	1.90
80	1.44×10^{-5}	2.73×10^{-5}	1.89
100	7.45×10^{-6}	1.40×10^{-5}	1.88

$$s^+_{N,2N}(\mu) = s(\mu) + \frac{1}{1-\varepsilon_{N,2N}}\lvert s_N(\mu)-s_{2N}(\mu)\rvert \times \begin{cases} \frac{1}{\tau}(1-\varepsilon_{N,2N})-1 & s_{2N}(\mu)\geq s_N(\mu) \\ 1 & s_{2N}(\mu)<s_N(\mu) \end{cases}. \quad (50)$$

The desired result then directly follows from our hypothesis on $\varepsilon_{N,2N}$, (29), and the range of τ.

The proof for the upper effectivity inequality (sharpness property) parallels the derivation of Section 4.2.2. In particular, we write

$$\eta_{N,2N}(\mu) = \frac{\frac{1}{\tau}\lvert s_{2N}-s_N\rvert}{\lvert s-s_N\rvert} = \frac{\frac{1}{\tau}\lvert s_{2N}-s+s-s_N\rvert}{\lvert s-s_N\rvert} \quad (51)$$

$$= \frac{1}{\tau}\lvert 1-\varepsilon_{N,2N}\rvert; \quad (52)$$

from our hypothesis (29) we may thus conclude that $\eta_{N,2N}(\mu) \rightarrow 1/\tau$ as $N\rightarrow\infty$. Note in the noncompliant, nonsymmetric case we can make no stronger statement.

We demonstrate our effectivity claims in Table 3, in which we present the error, bound gap, and effectivity for the noncompliant output ($s^2(\mu)$, average stress) of the truss example of Section 2.2.2; the results tabulated correspond to the choice $\tau=1/2$. We clearly obtain bounds for all N; and the effectivity rather quickly approaches $1/\tau$ (for $N\geq 120$, $\eta_{N,2N}(\mu)$ remains fixed at $1/\tau=2.0$).

5.2 Eigenvalue Problems. We next consider the extension of our approach to symmetric positive definite eigenvalue problems. The eigenvalues of appropriately defined partial-differential-equation eigenproblems convey critical information about a physical system: in linear elasticity, the critical buckling load; in dynamic analysis of structures, the resonant modes; in conduction heat transfer, the equilibrium timescales. Solution of large-scale eigenvalue problems is computationally intensive: the reduced-basis method is thus very attractive.

The abstract statement of our eigenvalue problem is: find $(u_i(\mu),\lambda_i(\mu)) \in X\times\mathbb{R}$, $i=1,\ldots,$ such that

$$a(u_i(\mu),v;\mu) = \lambda_i(\mu)m(u_i(\mu),v;\mu), \quad \forall v\in X,$$
$$\text{and} \quad m(u_i(\mu),u_i(\mu);\mu)=1. \quad (53)$$

Here a is the continuous, coercive, symmetric form introduced earlier, and m is (say) the L^2 inner product over Ω. The assumptions on a and m imply the eigenvalues $\lambda_i(\mu)$ will be real and positive. We order the eigenvalues (and corresponding eigenfunctions u_i) such that $0<\lambda_1(\mu)<\lambda_2(\mu)\leq\ldots$; we shall assume that $\lambda_1(\mu)$ and $\lambda_2(\mu)$ are distinct. We suppose that our output of interest is the minimum eigenvalue,

$$s(\mu) = \lambda_1(\mu); \quad (54)$$

other outputs may also be considered.

Following (Machiels et al. [24]), we present here a reduced-basis predictor and a Method I error estimator for symmetric positive-definite eigenvalue problems; we also briefly describe the simpler Method II approach.

5.2.1 Reduced-Basis Approximation. We sample, randomly or log-randomly, our design space $\mathcal{D}$ to create the parameter sample $S_{N/2}=\{\mu_1,\ldots,\mu_{N/2}\}$; we then introduce the reduced-basis space $W_N=\text{span}\{u_1(\mu_1),u_2(\mu_1),\ldots,u_1(\mu_{N/2}),u_2(\mu_{N/2})\}$, where we recall that $u_1(\mu)$ and $u_2(\mu)$ are the eigenfunctions associated with the first (smallest) and second eigenvalues $\lambda_1(\mu)$ and $\lambda_2(\mu)$, respectively. Note that W_N has good approximation properties both for the first and second lowest eigenfunctions, and hence eigenvalues; this is required by the Method I error estimator to be presented below. Our reduced-order approximation is then: find $(u_{Ni}(\mu),\lambda_{Ni}(\mu)) \in W_N\times\mathbb{R}$, $i=1,\ldots,N$, such that

$$a(u_{Ni}(\mu),v;\mu) = \lambda_{Ni}(\mu)m(u_{Ni}(\mu),v;\mu), \quad \forall v\in W_N,$$
$$\text{and} \quad m(u_{Ni}(\mu),u_{Ni}(\mu);\mu)=1; \quad (55)$$

the output approximation is then $s_N(\mu)=\lambda_{N1}(\mu)$.

The formulation admits an on-line/off-line decomposition (Machiels et al. [24]) very similar to the approach described for equilibrium problems in Section 3.

5.2.2 Method I A Posteriori Error Estimators. As before, we assume that we are given a positive function $g(\mu):\mathcal{D}\rightarrow\mathbb{R}_+$ and a continuous, coercive, symmetric bilinear form $\hat{a}(w,v):X\times X\rightarrow\mathbb{R}$, that satisfy the inequality (16). We then find $\hat{e}(\mu)\in X$ such that

$$g(\mu)\hat{a}(\hat{e}(\mu),v) = [\lambda_{N1}m(u_{N1}(\mu),v;\mu) - a(u_{N1}(\mu),v;\mu)], \quad \forall v\in X, \quad (56)$$

in which the right-hand side is the eigenproblem equivalent of the residual. We then evaluate our estimators as

$$s^+_N(\mu) = \lambda_{N1}(\mu), \quad s^-_N(\mu) = \lambda_{N1}(\mu) - \Delta_N(\mu),$$
$$\Delta_N(\mu) = \frac{g(\mu)}{\tau\delta(\mu)}\hat{a}(\hat{e}(\mu),\hat{e}(\mu)),$$

where $\delta(\mu)=1-\lambda_{N1}(\mu)/\lambda_{N2}(\mu)$ and $\tau\in(0,1)$. The effectivity is defined as $\eta_N(\mu)=\Delta_N(\mu)/(\lambda_{N1}(\mu)-\lambda_1(\mu))$.

We now consider the lower and upper effectivity inequalities. As regards the lower effectivity inequality (bounding property), we of course obtain $s^+_N(\mu)\geq\lambda_1(\mu)$, $\forall N$. The difficult result is the lower bound: it can be proven (Machiels et al. [24]) that there exists an $N^*(S_{N/2},\mu)$ such that $s^-_N(\mu)\leq\lambda_1(\mu)$, $\forall N>N^*$. In practice, $N^*=1$, due to the good (theoretically motivated) choice for $\delta(\mu)$; there is thus very little uncertainty in our (asymptotic) bounds. We also prove in Machiels et al. [24] a result related to the upper effectivity inequality (sharpness property); in, practice, very good effectivities are obtained. To demonstrate these claims we consider the eigenvalue problem associated with (the homogeneous version) of our two-dimensional thermal fin example of Section 2.2.1. We present in Table 4 the error, error

Table 4 Error, error bound (Method I), and effectivities as a function of N, at a particular representative point $\mu\in\mathcal{D}$, for the thermal fin eigenproblem

N	$\lvert\lambda_1(\mu)-\lambda_{N1}(\mu)\rvert/\lambda_1(\mu)$	$\Delta_N(\mu)/\lambda_1(\mu)$	$\eta_N(\mu)$
10	1.19×10^{-2}	6.66×10^{-2}	5.63
20	1.08×10^{-3}	7.19×10^{-3}	6.65
30	6.20×10^{-4}	3.19×10^{-3}	5.17
40	1.72×10^{-4}	1.55×10^{-3}	9.44
50	3.47×10^{-5}	4.06×10^{-4}	11.74

bound, and effectivity as a function of N at a particular point $\mu \in \mathcal{D}$. We observe rapid convergence, bounds for all N, and good effectivities.

Finally, we note that our output estimator admits an off-line/on-line decomposition similar to that for equilibrium problems; the additional terms in (56) are readily treated through our affine expansion/linear superposition procedure.

5.2.3 Method II A Posteriori Error Estimators. For Method II, we no longer require an estimate for the second eigenvalue. We may thus define $S_N = \{\mu_1, \ldots, \mu_N\}$, $W_N = \text{span}\{u_1(\mu_i), i = 1, \ldots, N\}$, and (for $M = 2N$) $S_{2N} = \{\mu_1, \ldots, \mu_{2N}\} \supset S_N$, $W_{2N} = \text{span}\{u_1(\mu_i), i = 1, \ldots, 2N\} \supset W_N$. The reduced basis approximation now takes the form (53), yielding $s_N(\mu) = \lambda_{N1}(\mu)$ and (for $N \rightarrow 2N$) $s_{2N}(\mu) = \lambda_{2N1}(\mu)$. Our estimators are then given by

$$s^+_{N,2N}(\mu) = \lambda_{N1}(\mu), \quad s^-_{N,2N} = \lambda_{N1}(\mu) - \Delta_{N,2N}(\mu),$$

$$\Delta_{N,2N}(\mu) = \frac{1}{\tau}(s_N(\mu) - s_{2N}(\mu)) \tag{57}$$

for $\tau \in (0,1)$. The effectivity $\eta_{N,2N}(\mu)$ is defined as for Method I.

For the lower effectivity inequality (bounding property), we of course retain $s^+_{N,2N}(\mu) \geq \lambda_1(\mu)$, $\forall N$. We also readily derive $s^-_{N,2N}(\mu) = \lambda_1 - (\lambda_{N1} - \lambda_1)(1/\tau(1 - \varepsilon_{N,2N}) - 1)$; under our hypothesis (29), we thus obtain asymptotic bounds as $N \rightarrow \infty$. For the upper effectivity inequality (sharpness property), we directly obtain $\eta_{N,2N} = 1/\tau(1 - \varepsilon_{N,2N})$. By variational arguments it is readily shown that $0 \leq \varepsilon_{N,2N} \leq 1$: we thus conclude that $0 \leq \eta_{N,2N} \leq 1/\tau$, $\forall N$. Additionally, under hypothesis (29), we deduce that $\eta_{N,2N} \rightarrow 1/\tau$ as $N \rightarrow \infty$.

5.3 Further Generalizations. In this section we briefly describe several additional extensions of the methodology. In each case we focus on the essential new ingredient; further details (in most cases) may be found in the referenced literature.

5.3.1 Noncoercive Linear Operators. The archetypical noncoercive linear equation is the Helmholtz, or reduced-wave, equation; many (e.g., inverse scattering) applications of this equation arise, for example, in acoustics and electromagnetics. The essential new mathematical ingredient is the loss of coercivity of a. In particular, well-posedness is now ensured only by the inf-sup condition: there exists positive β_0, $\beta(\mu)$, such that

$$0 < \beta_0 \leq \beta(\mu) = \inf_{w \in X} \sup_{v \in X} \frac{a(w,v;\mu)}{\|w\|_X \|v\|_X}, \quad \forall \mu \in \mathcal{D}. \tag{58}$$

Two numerical difficulties arise due to this ''weaker'' stability condition.

The first difficulty is preservation of the inf-sup stability condition for finite dimensional approximation spaces. To wit, although in the coercive case restriction to the space W_N actually increases stability, in the noncoercive case restriction to the space W_N can easily decrease stability: the relevant supremizers may not be adequately represented. Loss of stability can, in turn, lead to poor approximations—the inf-sup parameter enters in the denominator of the *a priori* convergence result. The second numerical difficulty is estimation of the inf-sup parameter, which for noncoercive problems plays the role of $g(\mu)$ in Method I *a posteriori* error estimation techniques. In particular, $\beta(\mu)$ can not typically be deduced analytically, and thus must be evaluated (via an eigenvalue formulation) as part of the reduced-basis approximation. Our resolution of both these difficulties involves two elements (Maday et al. [23]): first, we consider projections other than standard Galerkin; and second, we consider ''enriched'' approximation spaces.

In one approach (Maday et al. [23]), we pursue a minimum-residual projection: the (low-dimensional) infimizing space contains both the solution u and also the inf-sup infimizer at the μ_n sample points; and the (high-dimensional) supremizing space is taken to be X. Stability is ensured and rigorous (sharp) error bounds are obtained—though technically the bounds are only asymptotic due to the approximation of the inf-sup parameter; and, despite the presence of X, the on-line complexity remains independent of the dimension of X—as in Section 3.3, we exploit affine parameter dependence and linear superposition to precompute the necessary inversions. In a second suite of much simpler and more general approaches (see Maday et al. [23] for one example in the symmetric case), we exploit minimum-residual or Petrov-Galerkin projections with infimizer-supremizer enriched, but still very low-dimensional, infimizing and supremizing spaces. Plausible but not yet completely rigorous arguments, and empirical evidence, suggest that stability is ensured and rigorous asymptotic (and sharp) error bounds are obtained.

In Maday et al. [23] we focus entirely on Method I *a posteriori* error estimator procedures; but Method II techniques are also appropriate. In particular, Method II approaches do not require accurate estimation of the inf-sup parameter; we thus need be concerned only with stability in designing our reduced-basis spaces.

5.3.2 Parabolic Partial Differential Equations. The next extension considered is the treatment of parabolic partial differential equations of the form $m(u_t, v; \mu) = a(u, v; \mu)$; typical examples are time-dependent problems such as unsteady heat conduction—the ''heat'' or ''diffusion'' equation. The essential new ingredient is the presence of the time variable, t.

The reduced-basis approximation and error estimator procedures are similar to those for noncompliant nonsymmetric problems, except that we now include the time variable as an additional parameter. Thus, as in certain other time-domain model-order-reduction methods (Antoulas and Sorensen [25], Sirovich and Kirby [26]), the basis functions are ''snapshots'' of the solution at selected time instants; however, in our case, we construct an *ensemble* of such series corresponding to different points in the non-time parameter domain $\mathcal{D}$. For rapid convergence of the output approximation, the solutions to an adjoint problem, which evolves *backward* in time, must also be included in the reduced-basis space.

For the temporal discretization method, many possible choices are available. The most appropriate method, although not the only choice, is the discontinuous Galerkin method (Machiels et al. [27]). The variational origin of the discontinuous Galerkin approach leads naturally to rigorous output bounds for Method I *a posteriori error* estimators; the Method II approach is also directly applicable. Under our affine assumption, off-line/on-line decompositions can be readily crafted; the complexity of the on-line stage (calculation of the output predictor and associated bound gap) is, as before, independent of the dimension of X.

5.3.3 Locally Nonaffine Parameter Dependence. An important restriction of our methods is the assumption of affine parameter dependence. Although many property, boundary condition, load, and even geometry variations can indeed be expressed in the required form (2) for reasonably small Q, there are many problems, for example, general boundary shape variations, which do not admit such a representation. One simple approach to the treatment of this more difficult class of nonaffine problems is (*i*) in the off-line stage, store the $\zeta_n \equiv u(\mu_n)$, and (*ii*) in the on-line stage, directly evaluate the reduced-basis stiffness matrix as $a(\zeta_j, \zeta_i, \mu)$. Unfortunately, the operation count (respectively, storage) for the on-line stage will now scale as $O(N^2 \dim(X))$ (respectively, $O(N \dim(X))$, where $\dim(X)$ is the dimension of the truth (very fine) finite element approximation space: the resulting method may no longer be competitive with advanced iterative techniques; and, in any event, ''real-time'' response may be compromised.

We prefer an approach which is slightly less general but potentially much more efficient. In particular, we note that in many cases—for example, boundary geometry modification—the non-

affine parametric dependence can be restricted to a small subdomain of Ω, Ω_{II}. We can then express our bilinear form a as an affine/nonaffine sum,

$$a(w,v;\mu) = a_I(w,v;\mu) + a_{II}(w,v;\mu). \tag{59}$$

Here a_I, defined over Ω_I, the majority of the domain, is affinely dependent on μ; and a_{II}, defined over Ω_{II}, a small portion of the domain, is not affinely dependent on μ. It immediately follows that the reduced-basis stiffness matrix can be expressed as the sum of two stiffness matrices corresponding to contributions from a_I and a_{II}, respectively; that the stiffness matrix associated with a_I admits the usual on-line/off-line decomposition described in Section 3.3; and that the stiffness matrix associated with a_{II} requires storage (and inner product evaluation) *only of* $\zeta_i|_{\Omega_{II}}$ (ζ_i restricted to Ω_{II}). The nonaffine contribution to the on-line computational complexity thus scales only as $O(N^2 \dim(X|_{\Omega_{II}}))$, where $\dim(X|_{\Omega_{II}})$ refers (in practice) to the number of finite-element nodes located within Ω_{II}, often extremely small. We thus recover a method that is (almost) independent of $\dim(X)$, though clearly the on-line code will be more complicated than in the purely affine case.

In the above we focus on approximation. As regards *a posteriori* error estimation, the nonaffine dependence of a (even locally) precludes the precomputation and linear superposition strategy required by Method I (unless domain decomposition concepts are exploited (Machiels et al. [28]); however, Method II directly extends to the locally nonaffine case.

Acknowledgments

We would like to thank Mr. Thomas Leurent (formerly) of MIT for his many contributions to the work described in this paper; thanks also to Shidrati Ali of the Singapore-MIT Alliance and Yuri Solodukhov of MIT for very helpful discussions. We would also like to acknowledge our longstanding collaborations with Professor Jaime Peraire of MIT and Professor Einar Rønquist of the Norwegian University of Science and Technology. 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.

References

[1] Chan, T. F., and Wan, W. L., 1997, ''Analysis of projection methods for solving linear systems with multiple right-hand sides,'' SIAM J. Sci. Comput. (USA), **18**, No. 6, p. 1698.

[2] Farhat, C., Crivelli, L., and Roux, F., 1994, ''Extending substructure based iterative solvers to multiple load and repeated analyses,'' Comput. Methods Appl. Mech. Eng., **117**, No. 1–2, pp. 195–209.

[3] Akgun, M. A., Garcelon, J. H., and Haftka, R. T., 2001, ''Fast exact linear and non-linear structural reanalysis and the Sherman-Morrison-Woodbury formulas,'' Int. J. Numer. Methods Eng., **50**, No. 7, pp. 1587–1606.

[4] Yip, E. L., 1986, ''A note on the stability of solving a rank-*p* modification of a linear system by the Sherman-Morrison-Woodbury formula,'' SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput., **7**, No. 2, pp. 507–513.

[5] Allgower, E., and Georg, K., 1980, ''Simplicial and continuation methods for approximating fixed-points and solutions to systems of equations,'' SIAM Rev., **22**, No. 1, pp. 28–85.

[6] Rheinboldt, W., 1981, ''Numerical analysis of continuation methods for nonlinear structural problems,'' Comput. Struct., **13**, No. 1–3, pp. 103–113.

[7] Almroth, B. O., Stern, P., and Brogan, F. A., 1978, ''Automatic choice of global shape functions in structural analysis,'' AIAA J., **16**, pp. 525–528.

[8] Noor, A. K., and Peters, J. M., 1980, ''Reduced basis technique for nonlinear analysis of structures,'' AIAA J., **18**, No. 4, pp. 455–462.

[9] Balmes, E., 1996, ''Parametric families of reduced finite element models theory and applications,'' Mech. Syst. Signal Process., **10**, No. 4, pp. 381–394.

[10] Barrett, A., and Reddien, G., 1995, ''On the reduced basis method,'' Z. Angew. Math. Mech., **75**, No. 7, pp. 543–549.

[11] Fink, J. P., and Rheinboldt, W. C., 1983, ''On the error behaviour of the reduced basis technique for nonlinear finite element approximations,'' Z. Angew. Math. Mech., **63**, pp. 21–28.

[12] Peterson, J. S., 1989, ''The reduced basis method for incompressible viscous flow calculations,'' SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput., **10**, No. 4, pp. 777–786.

[13] Porsching, T. A., 1985, ''Estimation of the error in the reduced basis method solution of nonlinear equations,'' Math. Comput., **45**, No. 172, pp. 487–496.

[14] Rheinboldt, W., 1993, ''On the theory and error estimation of the reduced basis method for multi-parameter problems,'' Nonlinear Analysis, Theory, Methods and Applications, **21**, No. 11, pp. 849–858.

[15] Veroy, K., Leurent, T., Prud'homme, C., Rovas, D., and Patera, A., 2002, ''Reliable real-time solution of parametrized elliptic partial differential equations: Application to elasticity,'' *Proceedings SMA Symposium 2002*.

[16] Maday, Y., Machiels, L., Patera, A. T., and Rovas, D. V., 2000, ''Blackbox reduced-basis output bound methods for shape optimization,'' *Proceedings 12th International Domain Decomposition Conference*, eds. T. Chan, et al., ddm.org, pp. 429–436.

[17] Evans, A. G., Hutchinson, J. W., Fleck, N. A., Ashby, M. F., and Wadley, H. N. G., 2001, ''The topological design of multifunctional cellular metals,'' Prog. Mater. Sci., **46**, No. 3–4, pp. 309–327.

[18] Wicks, N., and Hutchinson, J. W., 2001, ''Optimal truss plates,'' Int. J. Solids Struct., **38**, No. 30–31, pp. 5165–5183.

[19] Maday, Y., Patera, A., and Turinici, G., ''Global a priori convergence theory for reduced-basis approximation of single-parameter symmetric coercive elliptic partial differential equations,'' C. R. Acad. Sci. Paris Série I. Submitted.

[20] Maday, Y., Patera, A. T., and Peraire, J., 1999, ''A general formulation for a posteriori bounds for output functionals of partial differential equations; Application to the eigenvalue problem,'' C. R. Acad. Sci. Paris, Série I, **328**, pp. 823–828.

[21] Machiels, L., Peraire, J., and Patera, A. T., 2001, ''A posteriori finite element output bounds for the incompressible Navier-Stokes equations; Application to a natural convection problem,'' J. Comput. Phys., **172**, pp. 401–425.

[22] Patera, A. T., and Rønquist, E. M., 2001, ''A general output bound result: Application to discretization and iteration error estimation and control,'' Math. Models Methods Appl. Sci., **11**, No. 4, pp. 685–712.

[23] Maday, Y., Patera, A. T., and Rovas, D. V., 2001, ''A blackbox reduced-basis output bound method for noncoercive linear problems,'' *Seminaive du College de France J. L. Lions, Series in Applied Mathematics*, eds. P. G. Ciarlet and P. L. Lions, Elsevier–Gauthier–Villars, Vol. 7, accepted, July 2002.

[24] Machiels, L., Maday, Y., Oliveira, I. B., Patera, A., and Rovas, D., 2000, ''Output bounds for reduced-basis approximations of symmetric positive definite eigenvalue problems,'' C. R. Acad. Sci. Paris, Série I, **331**, No. 2, pp. 153–158.

[25] Antoulas, A., and Sorensen, D., 2001 ''Approximation of large-scale dynamical systems: An overview,'' Technical report, Rice University.

[26] Sirovich, L., and Kirby, M., 1987, ''Low-dimensional procedure for the characterization of human faces,'' J. Opt. Soc. Am. A, **4**, No. 3, pp. 519–524.

[27] Machiels, L., Patera, A., and Rovas, D., 2001, ''Reduced basis output bound methods for parabolic problems,'' Comput. Methods Appl. Mech. Eng., Submitted.

[28] Machiels, L., Maday, Y., and Patera, A. T., 2000, ''A flux-free nodal Neumann subproblem approach to output bounds for partial differential equations,'' C. R. Acad. Sci. Paris, Série I, **330**, No. 3, pp. 249–254.