Chapter 8

Reduced Basis Approximation and A Posteriori Error Estimation for Parametrized Parabolic PDEs: Application to Real-Time Bayesian Parameter Estimation

N.C. Nguyen, G. Rozza, D.B.P. Huynh and A.T. Patera

Massachusetts Institute of Technology, USA

8.1   Introduction

In this chapter we consider reduced basis (RB) approximation and a posteriori error estimation for linear functional outputs of affinely parametrized linear parabolic partial differential equations. The essential ingredients are Galerkin projection onto a low-dimensional space associated with a smooth parametrically induced manifold – dimension reduction; efficient and effective POD-Greedy$_{\text{RB}}$ sampling methods for identification of optimal and numerically stable approximations spaces – rapid convergence; rigorous and sharp a posteriori error bounds for the linear-functional outputs of interest – certainty; and Offline-Online computational decomposition strategies – minimum *marginal cost*. The RB approach is effective in the *real–time* context in which the expensive Offline stage is deemed unimportant – for example parameter estimation and control; the RB approach is also effective in the *many–query* context in which the expensive Offline stage is asymptotically negligible – for example design optimization, uncertainty quantification (Boyaval *et al.* 2008), and multi-scale analysis (Boyaval 2008; Nguyen 2008).

There are two parts to this chapter. In the first part we present a rather general RB formulation for linear parabolic equations: our development combines earlier work in the parabolic case (Grepl and Patera 2005; Haasdonk and Ohlberger 2008) and elliptic context (Prud'homme *et al.* 2002; Rozza *et al.* 2008) with new advances in sampling procedures and in particular the POD–Greedy$_{\text{RB}}$ approach. In the second part we develop an RB Bayesian framework for parameter estimation which exploits the rapid response and reliability of the certified reduced basis method; as an example we consider detection and characterization of a delamination crack by transient thermal analysis (Grepl 2005; Starnes 2002). In summary, the first part (Section 8.2) emphasizes the essential RB ingredients in a general linear context – the state of the art; the second part (Section 8.3) emphasizes the integration and impact of RB technology in applications – real–time and many–query applications; each section contains both background material and new contributions. Brief concluding remarks (Section 8.4) discuss the possibilities for the future.

8.2   Linear Parabolic Equations

8.2.1   Reduced Basis Approximation

We first introduce several notations required for the remainder of the chapter. Our parameter domain, a closed subset of $\mathbb{R}^P$, shall be denoted $\mathcal{D}$; a typical parameter value in $\mathcal{D}$ shall be denoted μ. Our time domain shall be denoted by $I = [0, t_f]$ with t_f the final time. Our physical domain in d space dimensions shall be denoted Ω with boundary $\partial\Omega$; a typical point in Ω shall be denoted $x = (x_1, \ldots, x_d)$. We can then define the function

space $X = X(\Omega)$ such that $(H_0^1(\Omega))^\mathcal{V} \subset X \subset (H^1(\Omega))^\mathcal{V}$; here $H^1(\Omega) = \{v|v \in L^2(\Omega), \nabla v \in (L^2(\Omega))^d\}$, $H_0^1(\Omega) = \{v \in H^1(\Omega)|v_{\partial\Omega} = 0\}$, $L^2(\Omega)$ is the space of square integrable functions over Ω, and $\mathcal{V} = 1$ (respectively, d) for scalar (respectively, vector) problems. We denote by $(\cdot, \cdot)_X$ the inner product associated with the Hilbert space X; this inner product in turn induces a norm $\|\cdot\|_X = \sqrt{(\cdot, \cdot)_X}$ equivalent to the usual $(H^1(\Omega))^\mathcal{V}$ norm. Similarly, we denote by $(\cdot, \cdot)$ and $\|\cdot\|$ the $L^2(\Omega)$ inner product and induced norm, respectively.

We consider a generalized convection-diffusion equation (expressed in weak form): Given $\mu \in \mathcal{D}$, we find $u(t; \mu)$ such that

$$m(u_t(t; \mu), v; \mu) + a(u(t; \mu), v; \mu) = f(v; t; \mu), \quad \forall v \in X, \quad \forall t \in I, \quad (8.1)$$

subject to initial condition $u(t = 0; \mu) = u_0 \in L^2(\Omega)$. We then evaluate our output as

$$s(t; \mu) = \ell(u(t; \mu); t; \mu), \quad \forall t \in I. \quad (8.2)$$

We shall assume that a – which represents convection and diffusion – is time–invariant, continuous over X, and coercive over X with coercivity constant

$$\alpha(\mu) = \inf_{v \in X} \frac{a(w, w; \mu)}{\|w\|_X^2}, \quad \forall \mu \in \mathcal{D};$$

we assume that m – which represents 'mass' or inertia – is time–invariant, symmetric, and continuous and coercive over $L^2(\Omega)$ with coercivity constant

$$\sigma(\mu) = \inf_{v \in X} \frac{m(w, w; \mu)}{\|w\|^2}, \quad \forall \mu \in \mathcal{D};$$

we assume that f and ℓ are linear continuous functionals over X and $L^2(\Omega)$, respectively.

Finally, to effect our Offline–Online decomposition we shall require that our bilinear and linear forms are 'affine in parameter' (more precisely, affine in functions of the parameter): for some finite Q_a and Q_m, a and m may be expressed as

$$a(w, v; \mu) = \sum_{q=1}^{Q_a} \Theta_a^q(\mu)a^q(w, v), \qquad m(w, v; \mu) = \sum_{q=1}^{Q_m} \Theta_m^q(\mu)m^q(w, v) \quad (8.3)$$

for given parameter-dependent functions $\Theta_a^q, 1 \le q \le Q_a$, $\Theta_m^q, 1 \le q \le Q_m$, and continuous parameter-independent bilinear forms $a^q, 1 \le q \le Q_a, m^q, 1 \le q \le Q_m$; furthermore, for some finite Q_f and Q_ℓ, f and ℓ may be expressed as

$$f(v; t; \mu) = \sum_{q=1}^{Q_f} \Theta_f^q(\mu)f^q(v; t), \qquad \ell(v; t; \mu) = \sum_{q=1}^{Q_\ell} \Theta_\ell^q(\mu)\ell^q(v; t), \quad (8.4)$$

for given time/parameter-dependent functions $\Theta_f^q, 1 \leq q \leq Q_f, \Theta_\ell^q, 1 \leq q \leq Q_\ell$, and continuous parameter–independent linear forms $f^q, 1 \leq q \leq Q_f, \ell^q, 1 \leq q \leq Q_\ell$.

We now describe a general class – through not *the* most general class – of problems which honors these hypotheses; for simplicity we consider a scalar field ($\mathcal{V} = 1$) in two space dimension ($d = 2$). We shall first define an 'original' problem (subscript o) for a field variable $u_o(t; \mu) \in X_o(\mu)$ over a *parameter-dependent* domain $\Omega_o(\mu) \subset \mathbb{R}^2$,

$$m_o(u_{o\,t}(t; \mu), v; \mu) + a_o(u_o(t; \mu), v; \mu) = f_o(v; t; \mu), \quad \forall v \in X_o(\mu), \quad (8.5)$$

$$s_o(t; \mu) = \ell_o(u_o(t; \mu); t; \mu); \quad (8.6)$$

we will then map $\Omega_o(\mu)$ to a *parameter-independent* reference domain $\Omega = \Omega(\mu_{\text{ref}}), \mu_{\text{ref}} \in \mathcal{D}$, to arrive at the 'transformed' problem (8.1), (8.2) – which is the point of departure of our reduced basis approach. It remains to place restrictions on both geometry ($\Omega_o(\mu)$) and operators (a_o, m_o, f_o, ℓ_o) such that (upon mapping) this transformed problem satisfies our hypotheses – in particular, the affine assumption (8.3), (8.4). Towards this end, a domain decomposition shall prove indispensable.

We first consider the class of admissible geometries. We may consider $\Omega_o(\mu)$ ($\Rightarrow \Omega = \Omega(\mu_{\text{ref}})$) of the form

$$\overline{\Omega}_o(\mu) = \bigcup_{j=1}^{J_{\text{dom}}} \overline{\Omega}_o^j(\mu) \quad \left(\Rightarrow \overline{\Omega} = \bigcup_{j=1}^{J_{\text{dom}}} \overline{\Omega}^j \right)$$

for which $\Omega_o^j(\mu) = \mathcal{T}^j(\Omega^j; \mu), 1 \leq j \leq J_{\text{dom}}, \forall \mu \in \mathcal{D}$; here the Ω_o^j (respectively, Ω^j), $1 \leq j \leq J_{\text{dom}}$, constitute a conforming triangulation of $\Omega_o(\mu)$ (respectively, Ω), and the $\mathcal{T}^j, 1 \leq j \leq J_{\text{dom}}$, are *affine* mappings. We next consider the class of admissible operators. We may consider

$$a_o(w, v; \mu) = \sum_{j=1}^{J_{\text{dom}}} \int_{\Omega_o^j(\mu)} \begin{bmatrix} w_{,1} & w_{,2} & w \end{bmatrix} \mathcal{K}_o^j(\mu) \begin{bmatrix} v_{,1} \\ v_{,2} \\ v \end{bmatrix},$$

$$m_o(w, v; \mu) = \sum_{j=1}^{J_{\text{dom}}} \int_{\Omega_o^j(\mu)} w \mathcal{M}_o^j(\mu) v, \quad (8.7)$$

$$f_o(v; t; \mu) = \sum_{j=1}^{J_{\text{dom}}} \int_{\Omega_o^j(\mu)} \mathcal{F}_o^j(t; \mu) v, \quad \ell_o(v; t; \mu) = \sum_{j=1}^{J_{\text{dom}}} \int_{\Omega_o^j(\mu)} \mathcal{L}_o^j(\mu) v; \quad (8.8)$$

here $w_{,i}$ refers to differentiation with respect to the ith spatial coordinate, and the $\mathcal{K}_o^j : \mathcal{D} \rightarrow \mathbb{R}^{3 \times 3}$, $\mathcal{M}_o^j : \mathcal{D} \rightarrow \mathbb{R}$, $\mathcal{F}_o^j : \mathcal{D} \rightarrow \mathbb{R}, \mathcal{L}_o^j : \mathcal{D} \rightarrow \mathbb{R}$, $1 \leq j \leq J_{\text{dom}}$, are prescribed coefficients. (There are additional standard restriction on $\mathcal{K}_o^j, \mathcal{M}_o^j, 1 \leq j \leq J_{\text{dom}}$, related to coercivity.)

The process by which we map this original problem to the transformed problem – in which the $\Theta_a^q, \Theta_m^q, \Theta_f^q, \Theta_\ell^q$ reflect the geometric and coefficient parametric dependence – can be largely automated (Huynh *et al.* 2007-2009; Rozza *et al.* 2008). There are many ways in which we can relax our assumptions and thus treat an even broader class of problems. For example, we may consider 'elliptical' or 'curvy' triangular subdomains (Rozza *et al.* 2008); we may consider a which satisfy only a weak coercivity (Garding) inequality (Knezevic and Patera 2009); we may consider non–time–invariant bilinear forms a and m; we may consider coefficient functions $\mathcal{K}, \mathcal{M}$ which are polynomial in the spatial coordinate (or more generally approximated by the Empirical Interpolation Method (Barrault *et al.* 2004; Grepl *et al.* 2007a)). These generalizations can be pursued, with no loss in rigor, by modification of the methods – the error estimators, the sampling procedures, and the Offline–Online decompositions – presented in this chapter. However, it is important to recognize that, in general, increased complexity in geometry and operator will result in more terms in our affine expansions – larger Q_a, Q_m, Q_f, Q_ℓ – with corresponding detriment to the reduced basis (Online) computational performance; we return to this point in the context of the operation counts provided in Section 8.3 and then again for our particular example of Section 8.3.

We next introduce the finite difference in time and finite element (FE) in space discretization of this parabolic problem (Quarteroni and Valli 1997). We first divide the time interval I into K subintervals of equal length $\Delta t = t_f/K$ and define $t^k \equiv k\Delta t, 0 \le k \le K$. We next define the finite element approximation space $X^\mathcal{N} \subset X$ of dimension $\mathcal{N}$. Then, given $\mu \in \mathcal{D}$, we look for $u^{\mathcal{N} k}(\mu) \in X^\mathcal{N}, 0 \le k \le K$, such that

$$\frac{1}{\Delta t}m(u^{\mathcal{N} k}(\mu) - u^{\mathcal{N} k-1}(\mu), v; \mu) + a(u^{\mathcal{N} k}(\mu), v; \mu) = f(v; t^k; \mu),$$
$$\forall v \in X^\mathcal{N}, 1 \le k \le K, \tag{8.9}$$

subject to initial condition $(u^{\mathcal{N} 0}, v) = (u_0, v), \forall v \in X^\mathcal{N}$. We then evaluate the output: for $0 \le k \le K$,

$$s^{\mathcal{N} k}(\mu) = \ell(u^{\mathcal{N} k}(\mu); t^k; \mu). \tag{8.10}$$

We shall sometimes denote $u^{\mathcal{N} k}(\mu)$ as $u^\mathcal{N}(t^k; \mu)$ and $s^{\mathcal{N} k}(\mu)$ as $s^\mathcal{N}(t^k; \mu)$ to more clearly identify the discrete time levels. Equation (8.9) – Euler-Backward Galerkin discretization of (8.1) – shall be our point of departure: we shall presume that Δt is sufficiently small and $\mathcal{N}$ is sufficiently large such that $u^\mathcal{N}(t^k; \mu)$ and $s^\mathcal{N}(t^k; \mu)$ are effectively indistinguishable from $u(t^k; \mu)$ and $s(t^k; \mu)$, respectively. (The development readily extends to Crank-Nicolson discretization; for purposes of exposition, we consider the simple Euler Backward approach.) Our goal is to accelerate the 'truth' discretization (8.9), (8.10), in the real-time and many-query contexts: we

shall build our RB approximation on this truth discretization; and we shall measure the RB computational performance and RB accuracy relative to this truth discretization.

We now introduce the reduced basis (RB) approximation (Almroth *et al.* 1978; Fink and Rheinboldt 1983; Noor and Peters 1980; Porsching 1985). Given a set of mutually $(\cdot, \cdot)_X$–orthogonal basis functions $\xi_n \in X^{\mathcal{N}}, 1 \leq n \leq N_{\max}$, the $N_{\max}$ hierarchical RB spaces are given by

$$X_N \equiv \operatorname{span}\{\xi_n, 1 \leq n \leq N\}, \quad 1 \leq N \leq N_{\max}. \tag{8.11}$$

In actual practice (see Section 8.2.3), the spaces $X_N \in X^{\mathcal{N}}$ will be generated by a POD–Greedy$_{\mathrm{RB}}$ sampling procedure which combines spatial snapshots in time and parameter – $u^{\mathcal{N}\,k}(\mu)$ – in an optimal fashion. Given $\mu \in \mathcal{D}$, we now look for $u_N^k(\mu) \in X_N, 0 \leq k \leq K$, such that

$$\frac{1}{\Delta t} m(u_N^k(\mu) - u_N^{k-1}(\mu), v; \mu) + a(u_N^k(\mu), v; \mu) = f(v; t^k; \mu),$$
$$\forall v \in X_N, 1 \leq k \leq K, \tag{8.12}$$

subject to $(u_N^0(\mu), v) = (u^{\mathcal{N}\,0}, v), \forall v \in X_N$. We then evaluate the associated output: for $0 \leq k \leq K$,

$$s_N^k(\mu) = \ell(u_N^k(\mu); t^k, \mu). \tag{8.13}$$

We shall sometimes denote $u_N^k(\mu)$ as $u_N(t^k; \mu)$ and $s_N^k(\mu)$ as $s_N(t^k; \mu)$ to more clearly identify the discrete time levels; note that the RB approximation inherits the timestep of the truth discretization – there is no reduction in the 'temporal' dimension. The RB quantities should in fact bear a $\mathcal{N}$–$X_N^{\mathcal{N}}, u_N^{\mathcal{N}\,k}(\mu), s_N^{\mathcal{N}\,k}(\mu)$ – since the RB approximation is defined in terms of a particular truth discretization: for clarity of exposition, we shall typically suppress the 'truth' superscript; we nevertheless insist on stability/uniformity as $\mathcal{N} \to \infty$ (and $\Delta t \to 0$).

In general we can choose our domain decomposition to ensure that the functions $\Theta_{a,m,f,\ell}^q$ are very smooth; it can then be demonstrated that the field variable $u(\mu)$ is very smooth with respect to the parameter. It is then plausible (and in certain cases may be proven theoretically (Maday *et al.* 2002; Rozza *et al.* 2008)) that a properly chosen RB approximation – a Galerkin – optimal linear combination of 'good' snapshots on a smooth manifold – will converge very rapidly (even exponentially) with increasing N. Numerical results for a large number of coercive elliptic and parabolic problems (Grepl and Patera 2005; Rozza *et al.* 2008) – including the example of this chapter – support this conjecture: typically $N \approx O(10 - 100)$ and hence $N \ll \mathcal{N}$; the 'sparse samples' identified by the POD–Greedy$_{\mathrm{RB}}$ of Section 2.3 play an important role. Of course performance will degrade as the number of parameters and the ranges of the parameters increase; we return to this point in our computational summary below.

8.2.2 A Posteriori Error Estimation

Rigorous, sharp, and inexpensive a posteriori error bounds are crucial for informed application of the reduced basis method: a posteriori error bounds confirm *in practice* the dimension N required for the desired accuracy. To construct the a posteriori error bounds for the RB approximation, we need two ingredients. The first ingredient is the dual norm of the residual

$$\varepsilon_N(t^k;\mu) = \sup_{v\in X^{\mathcal{N}}} \frac{r_N(v;t^k;\mu)}{\|v\|_X}, \quad 1\le k\le K, \tag{8.14}$$

where $r_N(v;t^k;\mu)$ is the residual associated with the RB approximation (8.12)

$$r_N(v;t^k;\mu) = f(v;t^k;\mu) - \frac{1}{\Delta t}m\big(u_N^k(\mu)-u_N^{k-1}(\mu),v;\mu\big)$$
$$-a\big(u_N^k(\mu),v;\mu\big), \quad \forall v\in X^{\mathcal{N}}, \quad 1\le k\le K. \tag{8.15}$$

The second ingredient is a lower bound $0 < \alpha_{\mathrm{LB}}^{\mathcal{N}}(\mu) \le \alpha^{\mathcal{N}}(\mu), \forall\mu\in\mathcal{D}$, for the coercivity constant $\alpha^{\mathcal{N}}(\mu)$ defined as

$$\alpha^{\mathcal{N}}(\mu) = \inf_{v\in X^{\mathcal{N}}} \frac{a(v,v;\mu)}{\|v\|_X^2}, \quad \forall\mu\in\mathcal{D}, \tag{8.16}$$

and a lower bound $0 < \sigma_{\mathrm{LB}}^{\mathcal{N}}(\mu) \le \sigma^{\mathcal{N}}(\mu), \forall\mu\in\mathcal{D}$, for the coercivity constant $\sigma^{\mathcal{N}}(\mu)$ defined as

$$\sigma^{\mathcal{N}}(\mu) = \inf_{v\in X^{\mathcal{N}}} \frac{m(v,v;\mu)}{\|v\|^2}, \quad \forall\mu\in\mathcal{D}. \tag{8.17}$$

Note that since $X^{\mathcal{N}}\subset X$, $\alpha^{\mathcal{N}}(\mu)\ge\alpha(\mu)>0$ and $\sigma^{\mathcal{N}}(\mu)\ge\sigma(\mu)>0$, $\forall\mu\in\mathcal{D}$; we shall ensure by construction that $\alpha_{\mathrm{LB}}^{\mathcal{N}}(\mu)>0$ and $\sigma_{\mathrm{LB}}^{\mathcal{N}}(\mu)>0$, $\forall\mu\in\mathcal{D}$.

We can now define our error bounds in terms of the dual norm of the residual and the lower bounds for the coercivity constants. In particular, it can readily be proven (Grepl and Patera 2005; Haasdonk and Ohlberger 2008; Nguyen *et al.* 2009) that for all $\mu\in\mathcal{D}$ and all N,

$$\|u^{\mathcal{N}\,k}(\mu) - u_N^k(\mu)\| \le \Delta_N^k(\mu),$$
$$|s^{\mathcal{N}\,k}(\mu) - s_N^k(\mu)| \le \Delta_N^{s\,k}(\mu), \quad 1\le k\le K, \tag{8.18}$$

where $\Delta_N^k(\mu) \equiv \Delta_N(t^k;\mu)$ (the L^2 error bound) and $\Delta_N^{s\,k}(\mu) \equiv \Delta_N^s(t^k;\mu)$ (the 'output error bound') are given by

$$\Delta_N^k(\mu) \equiv \sqrt{\frac{\Delta t}{\alpha_{\mathrm{LB}}^{\mathcal{N}}(\mu)\,\sigma_{\mathrm{LB}}^{\mathcal{N}}(\mu)} \sum_{k'=1}^k \varepsilon_N^2(t^{k'};\mu)},$$
$$\Delta_N^{s\,k}(\mu) \equiv \left(\sup_{v\in X^{\mathcal{N}}} \frac{\ell(v)}{\|v\|}\right) \Delta_N^k(\mu). \tag{8.19}$$

(We assume for simplicity that $u^{\mathcal{N} 0} \in X_N$; otherwise there will be an additional contribution to $\Delta_N^k(\mu)$.) Note again that the RB error is measured relative to the 'truth' discretization.

It should be clear that our error bound for the output is rather crude. We may pursue primal-dual RB approximations (Grepl and Patera 2005; Pierce and Giles 2000; Rozza *et al.* 2008) that provide both more rapid convergence of the output and also more robust (sharper) estimation of the output error. However, in cases in which *many* outputs are of interest, for example inverse problems, the primal-only approach described above can be more efficient and also more adaptive – efficiently expanded to include additional outputs.

8.2.3   Offline–Online Computational Approach

The arguments above indicate that a reduced basis space of greatly reduced dimension $N \ll \mathcal{N}$ may suffice for accurate approximation of the field and output. However, each member of this space – and in particular each basis function $\xi_n, 1 \leq n \leq N$ – will be represented as a vector in $\mathbb{R}^{\mathcal{N}}$ corresponding to (say) the finite element nodal values. It is thus not clear how we can efficiently compute our reduced basis solution and output without appeal to high – dimensional objects – more generally, how we can translate the reduced dimension into reduced computational effort. The error estimator is even more problematic: the dual norm of the residual requires the truth solution of a (Poisson–like) problem which will be almost as expensive as the original PDE. In both cases, the affine assumption is the crucial enabler that permits an efficient implementation in the real-time or many–query contexts: we can pre-compute components of the mass and stiffness matrices and residual dual norm. We now provide the details.

Construction-Evaluation Decomposition

The affine representation (8.3) permits a 'Construction-Evaluation' decomposition (Balmes 1996; Prud'homme *et al.* 2002) of computational effort that greatly reduces the marginal cost – relevant in the real-time and many-query contexts – of both the RB output evaluation, (8.13), *and* the associated error bounds, (8.19). The expensive Construction stage, performed once, provides the foundation for the subsequent very inexpensive Evaluation stage, performed many times for each new desired $\mu \in \mathcal{D}$. We first consider the Construction-Evaluation decomposition for the output and then address the error bounds. For simplicity, in this section we assume that $f(v; t^k; \mu) = g(t^k) f(v)$ for some control $g(t)$, and that $\ell(v; t^k; \mu) = \ell(v)$.

We represent $u_N^k(\mu)$ as $u_N^k(\mu) = \sum_{n=1}^{N} \omega_{N\,n}^k(\mu) \xi_n$, where we recall that the $\xi_n, 1 \leq n \leq N$, are the basis functions for our RB space X_N. We may then evaluate the RB output as

$$s_N^k(\mu) = L_N^T \omega_N^k(\mu), \quad 1 \leq k \leq K, \tag{8.20}$$

where $L_{N\,n} = \ell(\xi_n), 1 \leq n \leq N$. To find the $\omega_{N\,j}^k(\mu), 1 \leq j \leq N, 1 \leq k \leq K$, we insert $u_N^k(\mu) = \sum_{n=1}^{N} \omega_{N\,n}^k(\mu)\xi_n$, $u_N^{k-1}(\mu) = \sum_{n=1}^{N} \omega_{N\,n}^{k-1}(\mu)\xi_n$, and $v = \xi_m$ in (8.12) to obtain the discrete system

$$(M_N(\mu) + \Delta t A_N(\mu))\omega_N^k(\mu) = \Delta t g(t^k) F_N + M_N(\mu)\omega_N^{k-1}(\mu), \quad 1 \leq k \leq K \tag{8.21}$$

where $A_N(\mu) \in \mathbb{R}^{N \times N}, M_N(\mu) \in \mathbb{R}^{N \times N}$, and $F_N(\mu) \in \mathbb{R}^N$ are given by $A_{N\,m,n}(\mu) = a(\xi_n, \xi_m; \mu)$, $M_{N\,m,n}(\mu) = m(\xi_n, \xi_m; \mu)$, $1 \leq m, n \leq N$, and $F_{N\,n} = f(\xi_n), 1 \leq n \leq N$, respectively. We next note that $A_N(\mu)$ and $M_N(\mu)$ can be expressed, thanks to (8.3), as

$$A_N(\mu) = \sum_{q=1}^{Q_a} \Theta_a^q(\mu) A_N^q, \quad M_N(\mu) = \sum_{q=1}^{Q_m} \Theta_m^q(\mu) M_N^q \tag{8.22}$$

where the $A_{N\,m,n}^q \equiv a^q(\xi_n, \xi_m), 1 \leq m, n \leq N$, $1 \leq q \leq Q_a$, $M_{N\,m,n}^q \equiv m^q(\xi_n, \xi_m), 1 \leq m, n \leq N, 1 \leq q \leq Q_m$, are parameter-*independent*. We can now readily identify the Construction–Evaluation decomposition.

In the *Construction* stage we first form and store the time–independent and μ–independent matrices/vectors $A_{N_{\max}\,ij}^q, M_{N_{\max}\,ij}^{q'}, F_{N_{\max}\,i}$, and $L_{N_{\max}\,i}$, $1 \leq i, j \leq N_{\max}, 1 \leq q \leq Q_a, 1 \leq q' \leq Q_m$. The operation count in the Construction stage of course depends on $\mathcal{N}$ – even once the $\xi_i, 1 \leq i \leq N_{\max}$, are *known* (obtained by the sampling procedure of the next section), it remains to compute $O(N_{\max}^2)$ finite element quadratures over the $O(\mathcal{N})$ triangulation. Note that, thanks to the hierarchical nature of the RB spaces, the stiffness matrices/vectors $A_{N\,ij}^q, M_{N\,ij}^{q'}, F_{N\,i}$, and $L_{N\,i}, 1 \leq i, j \leq N, 1 \leq q \leq Q_a, 1 \leq q' \leq Q_m$, for any $N \leq N_{\max}$ can be extracted as principal subarrays of the corresponding $N_{\max}$ quantities. (For non-hierarchical RB spaces the storage requirements are much higher.)

In the *Evaluation* stage, we first form the left–hand side of (8.21) in $O((Q_a + Q_m)N^2)$ operations; we then invert the resulting $N \times N$ matrix in $O(N^3)$ operations (in general, we must anticipate that the RB matrices will be dense); finally, we compute $\omega_{N\,j}^k, 1 \leq j \leq N, 1 \leq k \leq K$, in $O(KN^2)$ operations – $O(KN^3)$ operations for non-LTI systems – by matrix-vector multiplication. Note that $g(t^k)$ need only be specified in the Online stage; we return to this point in our sampling strategy below. Once the $\omega_{N\,j}^k, 1 \leq j \leq N, 1 \leq k \leq K$, are obtained – $O((Q_a + Q_m + N + K)N^2)$ operations in total – we evaluate our output from (8.20) in $O(NK)$ operations. The storage and operation count in the Evaluation phase is clearly independent of $\mathcal{N}$, and we can thus anticipate – presuming $N \ll \mathcal{N}$ – very rapid RB response in the real-time and many-query contexts.

The Construction-Evaluation procedure for the output error bound is a bit more involved. There are three components to this bound: the dual norm of ℓ (readily computed, once, in the Construction phase); the lower bound

for the coercivity constants, $\alpha_{\mathrm{LB}}^{\mathcal{N}}(\mu)$ and $\sigma_{\mathrm{LB}}^{\mathcal{N}}(\mu)$, computed Offline–Online by the Successive Constraint Method (SCM) as described in detail in Huynh *et al.* (2007); Rozza *et al.* (2008), and not discussed further here; and the dual norm of the residual $\varepsilon_N(t^k;\mu)$. We consider here the Construction-Evaluation decomposition for the dual norm of the residual (Grepl and Patera 2005). We first note from duality arguments that $\varepsilon_N(t^k;\mu)$ can be expressed as

$$\varepsilon_N^2(t^k;\mu) = \|\hat{e}_N(t^k;\mu)\|_X^2, \quad 1 \le k \le K, \tag{8.23}$$

where $\hat{e}_N(t^k;\mu)$ is the Riesz representation of the residual,

$$(\hat{e}_N(t^k;\mu),v)_X = r_N(v;t^k;\mu), \quad \forall v \in X^{\mathcal{N}}. \tag{8.24}$$

Here $r_N(v;t^k;\mu)$ is the residual defined in (8.15) (with $f(v;t^k;\mu) = g(t^k)f(v)$), which we may further write – exploiting the reduced basis representation $u_N^k(\mu) = \sum_{n=1}^N \omega_{N\,n}^k(\mu)\xi_n$ and affine assumption (8.3) – as

$$r_N(v;t^k;\mu) = g(t^k)f(v) - \frac{1}{\Delta t}\sum_{q=1}^{Q_m}\sum_{j=1}^{N}\Theta_m^q(\mu)(\omega_{N\,j}^k(\mu) - \omega_{N\,j}^{k-1}(\mu))m^q(\xi_j,v)$$
$$ - \sum_{q=1}^{Q_a}\sum_{j=1}^{N}\Theta_a^q(\mu)\omega_{N\,j}^k(\mu)a^q(\xi_j,v), \tag{8.25}$$

for $1 \le k \le K$.

It now follows directly from (8.24) and (8.25) that

$$\hat{e}_N(t^k;\mu) = g(t^k)\Gamma_N + \frac{1}{\Delta t}\sum_{q=1}^{Q_m}\sum_{j=1}^{N}\Theta_m^q(\mu)(\omega_{N\,j}^k(\mu) - \omega_{N\,j}^{k-1}(\mu))\Lambda_N^{qj}$$
$$\phantom{\hat{e}_N(t^k;\mu) =} + \sum_{q=1}^{Q_a}\sum_{j=1}^{N}\Theta_a^q(\mu)\omega_{N\,j}^k(\mu)\Upsilon_N^{qj}, \quad 1 \le k \le K, \tag{8.26}$$

where

$$\begin{aligned}
(\Gamma_N,v)_X &= f(v), & \forall v \in X^{\mathcal{N}}, & \\
(\Lambda_N^{qj},v)_X &= -m^q(\xi_j,v), & \forall v \in X^{\mathcal{N}}, & \quad 1 \le q \le Q_m, 1 \le j \le N, \\
(\Upsilon_N^{qj},v)_X &= -a^q(\xi_j,v), & \forall v \in X^{\mathcal{N}}, & \quad 1 \le q \le Q_a, 1 \le j \le N.
\end{aligned} \tag{8.27}$$

It then follows from (8.28) that

$$\varepsilon_N^2(t^k;\mu)$$
$$= g(t^k)g(t^k)C_{N\,1}^{ff}(\mu) + \sum_{j=1}^{N}\sum_{j'=1}^{N}\omega_{N\,j}^k(\mu)\omega_{N\,j'}^k(\mu)C_{N\,jj'}^{aa}(\mu)$$

$$+\frac{1}{\Delta t^2}\sum_{j=1}^{N}\sum_{j'=1}^{N}(\omega_{N\,j}^{k}(\mu)-\omega_{N\,j}^{k-1}(\mu))(\omega_{N\,j'}^{k}(\mu)-\omega_{N\,j'}^{k-1}(\mu))C_{N\,jj'}^{mm}(\mu)$$

$$+2g(t^k)\sum_{j=1}^{N}\omega_{N\,j}^{k}(\mu)C_{N\,j}^{fa}(\mu)+\frac{2g(t^k)}{\Delta t}\sum_{j=1}^{N}(\omega_{N\,j}^{k}(\mu)-\omega_{N\,j}^{k-1}(\mu))C_{N\,j}^{fm}(\mu)$$

$$+\frac{2}{\Delta t}\sum_{j=1}^{N}\sum_{j'=1}^{N}(\omega_{N\,j}^{k}(\mu)-\omega_{N\,j}^{k-1}(\mu))\omega_{N\,j'}^{k}(\mu)C_{N\,jj'}^{am}(\mu),\quad 1\le k\le K,\ (8.28)$$

where, for $1\le j,j'\le N$,

$$
\begin{aligned}
C_{N\,1}^{ff}(\mu)&=(\Gamma_N,\Gamma_N)_X,\\
C_{N\,jj'}^{aa}(\mu)&=\sum_{q=1}^{Q_a}\sum_{q'=1}^{Q_a}\Theta_a^q(\mu)\Theta_a^{q'}(\mu)(\Upsilon_N^{qj},\Upsilon_N^{q'j'})_X,\\
C_{N\,jj'}^{mm}(\mu)&=\sum_{q=1}^{Q_m}\sum_{q'=1}^{Q_m}\Theta_m^q(\mu)\Theta_m^{q'}(\mu)(\Lambda_N^{qj},\Lambda_N^{q'j'})_X,\\
C_{N\,j}^{fa}(\mu)&=\sum_{q=1}^{Q_a}\Theta_a^q(\mu)(\Upsilon_N^{qj},\Gamma_N)_X,\\
C_{N\,j}^{fm}(\mu)&=\sum_{q=1}^{Q_m}\Theta_m^q(\mu)(\Lambda_N^{qj},\Gamma_N)_X,\\
C_{N\,jj'}^{am}(\mu)&=\sum_{q=1}^{Q_m}\sum_{q'=1}^{Q_a}\Theta_m^q(\mu)\Theta_a^{q'}(\mu)(\Lambda_N^{qj},\Upsilon_N^{q'j'})_X.
\end{aligned}
\qquad(8.29)
$$

The Construction–Evaluation decomposition is now clear. We emphasize that in infinite precision (8.28) and (8.23) are equivalent: (8.28) is a reformulation of (8.23) that admits an Offline-Online decomposition.[1]

In the *Construction* stage, we find the $\Gamma_{N_{\max}},\Lambda_{N_{\max}}^{qj},1\le q\le Q_m,1\le j\le N_{\max},\Upsilon_{N_{\max}}^{qj},1\le q\le Q_a,1\le j\le N_{\max}$, and form the inner products $(\Gamma_{N_{\max}},\ \Gamma_{N_{\max}})_X,\ (\Lambda_{N_{\max}}^{qj},\ \Lambda_{N_{\max}}^{q'j'})_X,\ 1\le q,q',\le Q_m,\ 1\le j,j'\le N_{\max},\ (\Upsilon_{N_{\max}}^{qj},\Upsilon_{N_{\max}}^{q'j'})_X,\ 1\le q,q',\le Q_a,1\le j,j'\le N_{\max},\ (\Lambda_{N_{\max}}^{qj},\Upsilon_{N_{\max}}^{q'j'})_X,1\le q\le Q_m,1\le q'\le Q_a,1\le j,j'\le N_{\max},\ (\Lambda_{N_{\max}}^{qj},\Gamma_{N_{\max}})_X,1\le q\le Q_m,1\le j\le N_{\max},(\Upsilon_{N_{\max}}^{qj},\Gamma_{N_{\max}})_X,1\le q\le Q_a,1\le j\le N_{\max}$. The operation count for the Construction stage clearly depends on $\mathcal{N}-1+(Q_a+Q_m)N$ finite element 'Poisson' problems (8.27) and $(1+(Q_a+Q_m)N)^2$ finite element quadratures over the triangulation. (The temporary storage associated with

---

[1] In finite precision (8.28) and (8.23) are not equivalent: $\varepsilon_N(t^k;\mu)$ computed from (8.28) will only be accurate to the square root of machine precision; $\varepsilon_N(t^k;\mu)$ computed from (8.23) will be accurate to machine precision. The former is rarely a limitation for actual error tolerances of interest.

the latter can be excessive for higher–dimensional problems: it is simple to develop procedures that balance temporary storage and re–computation.) Note that, thanks to the hierarchical nature of the reduced basis spaces, these inner products for any $N \leq N_{\max}$ can be directly extracted from the corresponding $N_{\max}$ quantities. (As already noted, for non-hierarchical reduced basis spaces the storage requirements will be considerably higher.)

In the *Evaluation* stage, given the reduced basis coefficients $\omega_{N\,j}(t^k; \mu)$, $1 \leq j \leq N, 1 \leq k \leq K$, and coefficient functions $\Theta_a^q(\mu), 1 \leq q \leq Q_a$, $\Theta_m^q(\mu)$, $1 \leq q \leq Q_m$: we can readily compute the coefficient functions (8.29) from the stored inner products in $O((Q_a + Q_m)^2 N^2)$ operations; we then simply perform the sum (8.28) in $O(N^2)$ operations per time step and hence $O(KN^2)$ operations in total. The operation count for the Evaluation stage is thus (roughly) $(K + (Q_a + Q_m)^2)N^2$; note that the operation count for the Evaluation stage is $O(K(Q_a + Q_m)^2 N^2)$ operations for *non*-LTI systems since the coefficient functions (8.29) must be evaluated for each timestep. The crucial point, again, is that the cost and storage in the Evaluation phase – the *marginal* cost for each new value of μ – is independent of $\mathcal{N}$: thus we can not only evaluate our output prediction but also our rigorous output error bound very rapidly in the parametrically interesting contexts of real-time or many-query investigation. In short, we inherit the high fidelity and certainty of the FE approximation but at the low cost of a reduced-order model.

This concludes the discussion of the Construction – Evaluation decomposition. The Construction stage is performed Offline; the Evaluation stage is invoked Online – for each new μ of interest in the real–time or many–query contexts. However, there is another component to the Offline stage: we must construct a good (rapidly convergent) reduced basis space and associated basis functions $\xi_i, 1 \leq i \leq N_{\max}$, by a POD-Greedy$_{\mathrm{RB}}$ procedure; this sampling process in fact relies on the Construction–Evaluation decomposition to greatly reduce the requisite number of (expensive) 'candidate' finite element calculations over an (extensive) Greedy$_{\mathrm{RB}}$ training sample, $\Xi_{\mathrm{train,RB}}$, as we now describe. (In actual practice there is also an Offline-Component to the SCM construction of $\alpha_{\mathrm{LB}}^{\mathcal{N}}(\mu)$ and $\sigma_{\mathrm{LB}}^{\mathcal{N}}(\mu)$ as reported in Huynh *et al.* (2007); Rozza *et al.* (2008).)

POD-Greedy$_{\mathrm{RB}}$ Procedure

We address here the generation of our reduced basis space X_N. Our sampling procedure combines, as first proposed in Haasdonk and Ohlberger (2008), the POD (Proper Orthogonal Decomposition) in t^k – to capture the causality associated with our evolution equation – with a Greedy procedure (Grepl and Patera 2005; Rozza *et al.* 2008; Veroy *et al.* 2003b;) in μ – to treat efficiently the higher dimensions and more extensive ranges of parameter variation. (For an alternative 'interpolation' approach to reduced order time-parameter spaces see Amsallem and Farhat (2008); Amsallem *et al.* (2009).)

To begin, we summarize the well-known optimality property of the POD (Kunisch and Volkwein 2002). Given J elements of $X^{\mathcal{N}}$, $w_j \in X^{\mathcal{N}}, 1 \leq j \leq J$, and any positive integer $M \leq J$, $\text{POD}(\{w_1, \ldots, w_J\}, M)$ returns M $(\cdot, \cdot)_X$-orthogonal functions $\{\chi_m, 1 \leq m \leq M\}$ such that the space $V_M = \text{span}\{\chi_m, 1 \leq m \leq M\}$ is optimal in the sense that

$$V_M = \arg \inf_{Y_M \subset \text{span}\{w_j, 1 \leq j \leq J\}} \left(\frac{1}{J} \sum_{j=1}^{J} \inf_{v \in Y_M} \|w_j - v\|_X^2 \right)^{1/2}$$

where Y_M denotes a M-dimensional linear space. We also recall that to find the χ_p we first form the correlation matrix C with entries $C_{ij} = (w_i, w_j)_X$, $1 \leq i, j \leq J$; we then find the largest M eigenvalues $\lambda^m, 1 \leq m \leq M$, and associated eigenvectors $v^m \in \mathbb{R}^J, 1 \leq m \leq M$, of the system $Cv^m = \lambda^m v^m$ with normalization $(v^m)^T v^m = 1$; finally we form $\chi_m = \sum_{j=1}^{J} v_j^m w_j, 1 \leq m \leq M$. Note that the χ_m thus satisfy the orthogonality condition $(\chi_m, \chi_n)_X = \lambda^m \delta_{mn}, 1 \leq m, n \leq M$.

To initiate the POD-Greedy$_{\text{RB}}$ sampling procedure we must specify a very large (exhaustive) 'training' sample of $n_{\text{train,RB}}$ points in $\mathcal{D}$, $\Xi_{\text{train,RB}}$, and an initial (say, random) RB parameter sample $S^* = \{\mu_0^*\}$. Typically we choose $\Xi_{\text{train,RB}}$ by Monte Carlo sampling over $\mathcal{D}$ with respect to a prescribed (usually uniform) density, however for P small (few parameters) often a uniform or log-uniform deterministic distribution is preferred. The algorithm is then given by

```
Set  𝒵 = ∅;
Set  μ* = μ₀*;
While   N ≤ N_max
    {χ_m, 1 ≤ m ≤ M₁} = POD({u^𝒩(t^k; μ*), 1 ≤ k ≤ K}, M₁);
    𝒵 ← {𝒵, {χ_m, 1 ≤ m ≤ M₁}};
    N ← N + M₂;
    {ξ_n, 1 ≤ n ≤ N} = POD(𝒵, N);
    X_N = span{ξ_n, 1 ≤ n ≤ N};
    μ* = arg max_{μ∈Ξ_train,RB} Δ_N(t^K = t_f; μ);
    S* ← {S*, μ*};
end.
Set  X_N = span{ξ_n, 1 ≤ n ≤ N}, 1 ≤ N ≤ N_max.
```

In actual practice, we typically exit the POD-Greedy sampling procedure at $N = N_{\text{max}} \leq N_{\text{max},0}$ for which a prescribed error tolerance is satisfied: to wit, we define

$$\varepsilon_{N,\text{max}}^* = \max_{\mu \in \Xi_{\text{train,RB}}} \frac{\Delta_N(t^K; \mu)}{\|u_N(t^K; \mu)\|},$$

and terminate when $\varepsilon_{N,\text{max}}^* \leq \varepsilon_{\text{tol}}$. Note, by virtue of the final re–definition, the POD-Greedy generates *hierarchical* spaces $X_N, 1 \leq N \leq N_{\text{max}}$, which is computationally very advantageous.

There are two 'tuning' variables in the POD-Greedy$_{\mathrm{RB}}$ procedure, M_1 and M_2. We choose M_1 to satisfy an internal POD error criterion based on the usual sum of eigenvalues; we choose $M_2 \leq M_1$ to minimize duplication in the reduced basis space – though typically we prefer $M_2 > 1$ in order to reduce the number of Greedy$_{\mathrm{RB}}$ iterations and hence Offline cost. We make two observations. First, the POD – Greedy$_{\mathrm{RB}}$ method readily accommodates a repeat μ^* in successive Greedy$_{\mathrm{RB}}$ cycles – new information will always be available and old information rejected; in contrast, a pure Greedy$_{\mathrm{RB}}$ approach in both t and μ (Grepl and Patera 2005), though often generating good spaces, can 'stall.' Second, thanks to the POD normalization $(\chi_m, \chi_n)_X = \lambda^m \delta_{mn}, 1 \leq m, n \leq M_1$, the modes generated in the first POD at any parameter value μ^* are automatically scaled by their respective importance in representing $u(t^k; \mu^*), 1 \leq k \leq K$; the inputs to the second POD (of $\mathcal{Z}$) are thus correctly weighted to accommodate modes from different parameter values. (An alternative single-stage POD-Greedy$_{\mathrm{RB}}$ procedure is proposed in Knezevic and Patera (2009).)

The procedure remains computationally feasible even for large parameter domains and very extensive training samples (and in particular in higher parameter dimensions $P > 1$): the POD is conducted in only one (time) dimension and the Greedy$_{\mathrm{RB}}$ addresses the remaining (parameter) dimensions. The crucial point to note is that the operation count for the POD-Greedy$_{\mathrm{RB}}$ algorithm is additive and not multiplicative in $n_{\mathrm{train,RB}}$ and $\mathcal{N}$: in searching for the next parameter value μ^*, we invoke the Construction–Evaluation decomposition to inexpensively calculate the a posteriori error bound at the $n_{\mathrm{train,RB}}$ candidate parameter values; $\Delta_N(t^k; \mu)$ over $\Xi_{\mathrm{train,RB}}$ is computed by first *constructing* the necessary parameter-independent inner products at cost which depends on $\mathcal{N}$ but not on $n_{\mathrm{train,RB}}$ and then *evaluating* the error over all training points at cost $n_{\mathrm{train,RB}}(K + (Q_a + Q_m)^2)N^2$ (independent of $\mathcal{N}$) – hence the additive and not multiplicative dependence on $\mathcal{N}$ and $n_{\mathrm{train,RB}}$. In contrast, in a pure POD approach, we would need to evaluate the finite element 'truth' solution at the $n_{\mathrm{train,RB}}$ candidate parameter values at cost $O(n_{\mathrm{train,RB}}\mathcal{N}^{\cdot})$. (Of course, much of the computational economies are due not to the Greedy$_{\mathrm{RB}}$ *per se*, but rather to the accommodation within the Greedy$_{\mathrm{RB}}$ of the inexpensive error bounds.) As a result, in the POD–Greedy$_{\mathrm{RB}}$ approach we can take $n_{\mathrm{train,RB}}$ relatively large: we can thus anticipate reduced basis spaces and approximations that provide rapid convergence *uniformly* over the entire parameter domain. (Note that more sophisticated and hence efficient search algorithms can be exploited in the Greedy$_{\mathrm{RB}}$ context, see for example Bui-Thanh *et al.* (2007).)

We pursue the POD-Greedy$_{\mathrm{RB}}$ sampling procedure – which involves both the Construction and Evaluation phases – in an Offline stage. Then, in the Online stage, we invoke only the very inexpensive Evaluation phase: $\mu \rightarrow s_N^k(\mu), \Delta_N^k(\mu), 1 \leq k \leq K$. Note that in the POD-Greedy$_{\mathrm{RB}}$ procedure we

choose for $g(t)$ the impulse function Grepl and Patera (2005); the resulting RB space will thus have good approximation properties for any $g(t)$, and hence $g(t)$ can be specified in the Online stage. (The latter property is of course lost for non-LTI problems.)

Summary

We briefly summarize here the various steps in the full algorithm. First a prerequisite: as described in Section 8.2.1, the original problem (potentially posed over a parameter–dependent domain) must be mapped to a transformed problem (over a reference domain); we must then confirm our hypotheses, including the affine assumption (8.3). Not all (original) problems will yield transformed problems that honor our hypotheses; and not all transformed problems can be efficiently treated by our RB approach – in particular if P or Q is too large, as discussed further below.

We then conduct the Offline stage. First, the (Offline component) of the SCM procedure is executed in order to provide the small database invoked by the Online component of the SCM lower bound for the coercivity constants, $\alpha_{\mathrm{LB}}^{\mathcal{N}}(\mu)$ of (8.16) and $\alpha_{\mathrm{LB}}^{\mathcal{N}}(\mu)$ of (8.17); we have chosen not to emphasize this algorithmic ingredient given extensive details in Huynh *et al.* (2007); Rozza *et al.* (2008). Second, the POD-Greedy$_{\mathrm{RB}}$ procedure is executed to provide the small database invoked by the Online component of the RB output and output error bound prediction; the database comprises $A_{N_{\max}}^q \in \mathbb{R}^{N_{\max} \times N_{\max}}, 1 \leq q \leq Q_a, M_{N_{\max}}^q \in \mathbb{R}^{N_{\max} \times N_{\max}},$ $1 \leq q \leq Q_m, F_{N_{\max}}^q \in \mathbb{R}^{N_{\max}}, \quad 1 \leq q \leq Q_f, L_{N_{\max}}^q \in \mathbb{R}^{N_{\max}}, \quad 1 \leq q \leq Q_\ell$ as defined in (8.22) for the output; the L^2 norm of ℓ and $(\Gamma_{N_{\max}}, \Gamma_{N_{\max}})_X,$ $(\Lambda_{N_{\max}}^{qj}, \Lambda_{N_{\max}}^{q'j'})_X, \ 1 \leq q, q', \leq Q_m, 1 \leq j, j' \leq N_{\max}, \ (\Upsilon_{N_{\max}}^{qj}, \Upsilon_{N_{\max}}^{q'j'})_X, \ 1 \leq$ $q, q', \leq Q_a, 1 \leq j, j' \leq N_{\max}, \ (\Lambda_{N_{\max}}^{qj}, \Upsilon_{N_{\max}}^{q'j'})_X, 1 \leq q \leq Q_m, 1 \leq q' \leq Q_a, 1 \leq$ $j, j' \leq N_{\max}, \ (\Lambda_{N_{\max}}^{qj}, \Gamma_{N_{\max}})_X, 1 \leq q \leq Q_m, 1 \leq j \leq N_{\max}, \ (\Upsilon_{N_{\max}}^{qj}, \Gamma_{N_{\max}})_X,$ $1 \leq q \leq Q_a, 1 \leq j \leq N_{\max}$, for the output error bound. (Recall that the necessary quantities for $N \leq N_{\max}$ can be extracted as subarrays of the corresponding $N_{\max}$ quantities.)

We may then exercise the Online stage. Given a parameter value $\mu \in \mathcal{D}$ of interest, we first calculate the output prediction: solution of (8.21) for the RB field coefficients followed by evaluation of the sum (8.20). We next calculate the output error bound: evaluation of the sums first of (8.29) and subsequently of (8.28) for the dual norm of the residual; computation of the SCM lower bounds $\alpha_{\mathrm{LB}}^{\mathcal{N}}, \sigma_{\mathrm{LB}}^{\mathcal{N}}$; and finally assembly of the final result, (8.19).

We shall focus here on the operation count for the Online stage. It is clear from our earlier discussions that to leading order the output and output error bound can be evaluated in $O(N^3 + (K + (Q_a + Q_m)^2)N^2)$ operations (the additional contribution of the SCM is typically negligible in the Online stage) – independent of $\mathcal{N}$. (The latter in turn implies that, at least as regards the Online stage, $\mathcal{N}$ may be chosen conservatively large.) We do

not have a strong a priori theory which permits us to forecast the requisite N as a function of desired accuracy. However, at least for parametrically smooth problems, the POD-Greedy$_{\mathrm{RB}}$ spaces and Galerkin projection should yield – and our posteriori error bounds will *confirm* – RB output predictions which are highly accurate for $N \ll \mathcal{N}$. Nevertheless, it is important to emphasize that the Online operation count will certainly increase with the difficulty of the problem: N will increase with the number of parameters P and the extent of variation in the parameters as reflected in $\mathcal{D}$; Q_a, Q_m will increase with the geometric and operator complexity of the problem. In actual practice, for coercive problems with as many as $O(10)$ parameters $N = O(100)$ suffices (see Sen *et al.* (2006) for an elliptic example); even more parameters may be considered if the parametric representation is effective – $\mathcal{D}$ of increasingly limited extent in the higher parameter dimensions (Boyaval *et al.* 2008). Our example of Section 8.3 shall serve as additional calibration of computational effort associated with both the Online and also Offline stages.

8.3   Bayesian Parameter Estimation

8.3.1   Bayesian Approach

In parameter estimation problems we would like to infer the unknown parameter $\mu_\star \in \mathcal{D} \subset \mathbb{R}^P$ from the measurements of outputs of interest, $s^{(m)}(t; \mu_\star), 1 \leq m \leq M_{\mathrm{out}}$, collected for $t = t^{k_j^{\mathrm{exp}}} = k_j^{\mathrm{exp}} \Delta t \in [0, t_f], 1 \leq j \leq J$; here M_{out} is the number of outputs and J is the number of measurements per output. (In actual practice, some of the P parameters – for example, measurement system design variables – may be specified (or optimized) rather than inferred.) In our case the outputs are expressed as functionals of the solution of the forward problem (8.1) – $s^{(m)}(t; \mu_\star) = \ell^{(m)}(u(t; \mu_\star))$ for $1 \leq m \leq M_{\mathrm{out}}$. In order to assess our approach to parameter estimation we create 'synthetic' data as

$$G_{mj}^{\mathrm{exp}}(\mu_\star; \varepsilon_{\mathrm{exp}}) = s^{(m)\mathcal{N}}(t^{k_j^{\mathrm{exp}}}; \mu_\star) + \varepsilon_{mj}^{\mathrm{exp}}, \quad 1 \leq m \leq M_{\mathrm{out}}, 1 \leq j \leq J,$$
(8.30)

where the $s^{(m)\mathcal{N}}(t^{k_j^{\mathrm{exp}}}; \mu_\star)$ are the finite element approximation to the exact output $s^{(m)}(t^{k_j^{\mathrm{exp}}}; \mu_\star)$ and the $\varepsilon_{mj}^{\mathrm{exp}}$ represent the 'experimental' error. We assume the $\varepsilon_{mj}^{\mathrm{exp}}$ to be independent identically distributed (i.i.d.) Gaussian random variables (hence white in time) with zero mean and known variance σ_{exp}^2; our formulation can in fact treat any desired probability distribution.

We apply the Bayesian approach to parameter estimation (Mosegaard and Tarantola 2002) to the truth discretization of the forward problem (8.1). The expected value[2] $\mathsf{E}^{\mathcal{N}}[\mu_\star | G^{\mathrm{exp}}]$ of the unknown parameter $\mu_\star$ conditional on

---

[2]For brevity we consider only the expectation; our methodology also applies to the variance and indeed the full empirical posterior distribution function.

the data G^{exp} is given by

$$\mathsf{E}^{\mathcal{N}}[\mu_\star|G^{\mathrm{exp}}] = \frac{\int_{\mathcal{D}} \mu \Pi^{\mathcal{N}}(G^{\mathrm{exp}}|\mu)\Pi_0(\mu)d\mu}{\int_{\mathcal{D}} \Pi^{\mathcal{N}}(G^{\mathrm{exp}}|\mu')\Pi_0(\mu')d\mu'}. \tag{8.31}$$

Here the likelihood function $\Pi^{\mathcal{N}}(G^{\mathrm{exp}}|\mu)$ is given by

$$\Pi^{\mathcal{N}}(G^{\mathrm{exp}}|\mu) = \left(\frac{1}{2\pi\sigma_{\mathrm{exp}}^2}\right)^{M_{\mathrm{out}}J/2} \exp\left(-\frac{(G^{\mathrm{exp}} - F^{\mathcal{N}}(\mu))^T(G^{\mathrm{exp}} - F^{\mathcal{N}}(\mu))}{2\sigma_{\mathrm{exp}}^2}\right), \tag{8.32}$$

where, for $1 \leq m \leq M_{\mathrm{out}}$ and $1 \leq j \leq J$, $F_{mj}^{\mathcal{N}} : \mu \in \mathcal{D} \to s^{(m)}\mathcal{N}(t^{k_j^{\mathrm{exp}}};\mu)$ denotes the finite element evaluation of the mth output at time $t^{k_j^{\mathrm{exp}}}$ at any given μ in our parameter domain $\mathcal{D}$. The prior distribution on the parameter μ, $\Pi_0(\mu)$, is also assumed Gaussian[3]

$$\Pi_0(\mu) = \left(\frac{1}{2\pi\sigma_0^2}\right)^{P/2} \exp\left(-\frac{(\mu - \mu_0)^T(\mu - \mu_0)}{2\sigma_0^2}\right), \tag{8.33}$$

where $\mu_0 \in \mathcal{D}$ is the prior mean and σ_0^2 is the associated variance (more generally a covariance); our approach is not limited to any particular prior. Note that $\mathsf{E}^{\mathcal{N}}[\mu_\star|G^{\mathrm{exp}}]$ in (8.31) is an expectation with respect to the 'random' parameter μ: for any given measurement, G^{exp}, $\mathsf{E}^{\mathcal{N}}[\mu_\star|G^{\mathrm{exp}}]$ is our estimator for $\mu_\star$; properly speaking, $\mathsf{E}^{\mathcal{N}}[\mu_\star|G^{\mathrm{exp}}]$ is a realization of a random variable – a function of G^{exp}. (To avoid cumbersome notation, G^{exp} refers both to the measurement random variable and to associated realizations.)

The expected value in (8.31) necessitates the computation of multidimensional integrals, which in turn require numerous evaluations of the truth outputs; in particular, in the remainder of this section we shall interpret

$$\int_{\mathcal{D}} \Phi(\mu) \equiv \sum_{i=1}^{n_{\mathrm{quad}}} w_i^{\mathrm{quad}}\Phi(\mu_i^{\mathrm{quad}}), \tag{8.34}$$

for $w_i^{\mathrm{quad}} \in \mathbb{R}_+$, $\mu_i^{\mathrm{quad}} \in \mathcal{D}$, $1 \leq i \leq n_{\mathrm{quad}}$.[4] As a consequence, the parameter estimation procedure can be very expensive. To reduce the computational cost of Bayesian inverse analysis Wang and Zabaras (2005) introduce POD–based model reduction. Our emphasis here is a posteriori error

---

[3]In theory, we must multiply (8.33) by a pre-factor reflecting the bounded $\mathcal{D}$. In practice, we shall consider small σ_0 and large μ_0 such that μ outside $\mathcal{D}$ are highly improbable – and hence $\mathcal{D}$ is effectively $\mathbb{R}^P$.

[4]In this chapter we consider an adaptive piecewise Gauss–Legendre technique: we first create a domain decomposition selectively refined near an approximate $\mu_\star$; we then apply standard tensor–product Gauss–Legendre quadrature within each subdomain. We denote by n_{quad} the total number of integrand evaluations required. For problems with more parameters, Monte Carlo techniques would be necessary.

estimation (absent in earlier Bayesian model reduction approaches): our error bounds shall ensure that our Bayesian inferences are (*i*) certifiably accurate (relative to the truth), and (*ii*) as efficient as possible – through optimal choice of N for a given error tolerance. In the subsequent subsection, we incorporate our a posteriori error bounds into the Bayesian approach to permit rapid and reliable parameter estimation. (See also Grepl (2005); Grepl *et al.* (2007b) for an alternative approach to RB inverse analysis which more explicitly characterizes parameter uncertainty.)

8.3.2   A Posteriori Bounds for the Expected Value

We develop here *inexpensive, rigorous* lower and upper bounds for the expected value (8.31) (with quadratures evaluated as (8.34)) based on the RB outputs and associated output error bounds. Toward this end, we first introduce $F_{N\,mj}(\mu) = s_N^{(m)}(t_j^{k^{\exp}};\mu)$ and $\Delta F_{N\,mj}(\mu) = \Delta_N^{s\,(m)}(t_j^{k^{\exp}};\mu)$ for $1 \le m \le M_{\mathrm{out}}$ and $1 \le j \le J$, and then $F_N^{\pm}(\mu) = F_N(\mu) \pm \Delta F_N(\mu)$; here $s_N^{(m)}(t^k;\mu)$ and $\Delta_N^{s\,(m)}(t^k;\mu)$ are the RB prediction and associated error bound for the mth output. We then define, for $1 \le m \le M_{\mathrm{out}}$ and $1 \le j \le J$,

$$B_{N\,mj}(\mu) = \max\{|G_{mj}^{\exp} - F_{N\,mj}^-(\mu)|, |G_{mj}^{\exp} - F_{N\,mj}^+(\mu)|\},$$

and

$$D_{N\,mj}(\mu) = \begin{cases} 0, & \text{if } G_{mj}^{\exp} \in [F_{N\,mj}^-(\mu), F_{N\,mj}^+(\mu)], \\ \min\{|G_{mj}^{\exp} - F_{N\,mj}^-(\mu)|, \\ \qquad |G_{mj}^{\exp} - F_{N\,mj}^+(\mu)|\}, & \text{otherwise .} \end{cases}$$
(8.35)

Note that $G^{\exp} \in \mathbb{R}^{M_{\mathrm{out}}J}, F_N^{\pm}(\mu) \in \mathbb{R}^{M_{\mathrm{out}}J}, D_N(\mu) \in \mathbb{R}^{M_{\mathrm{out}}J}$, and $B_N(\mu) \in \mathbb{R}^{M_{\mathrm{out}}J}$.

We now introduce two new likelihood functions

$$\Pi_N^a(G^{\exp}|\mu) = \left(\frac{1}{2\pi\sigma_{\exp}^2}\right)^{M_{\mathrm{out}}J/2} \exp\left(-\frac{D_N^T(\mu)D_N(\mu)}{2\sigma_{\exp}^2}\right),$$

$$\Pi_N^b(G^{\exp}|\mu) = \left(\frac{1}{2\pi\sigma_{\exp}^2}\right)^{M_{\mathrm{out}}J/2} \exp\left(-\frac{B_N^T(\mu)B_N(\mu)}{2\sigma_{\exp}^2}\right),$$
(8.36)

from which we may evaluate

$$\mathsf{E}_N^{\mathrm{LB}}[\mu_\star|G^{\exp}] = \frac{\int_{\mathcal{D}} \mu \Pi_N^b(G^{\exp}|\mu)\Pi_0(\mu)d\mu}{\int_{\mathcal{D}} \Pi_N^a(G^{\exp}|\mu')\Pi_0(\mu')d\mu'},$$

$$\mathsf{E}_N^{\mathrm{UB}}[\mu_\star|G^{\exp}] = \frac{\int_{\mathcal{D}} \mu \Pi_N^a(G^{\exp}|\mu)\Pi_0(\mu)d\mu}{\int_{\mathcal{D}} \Pi_N^b(G^{\exp}|\mu')\Pi_0(\mu')d\mu'}.$$
(8.37)

(If μ takes on negative values then (8.37) must be modified slightly.) We shall take

$$\mathsf{E}_N^{\mathrm{AV}}[\mu_\star|G^{\mathrm{exp}}] = \frac{1}{2}(\mathsf{E}_N^{\mathrm{LB}}[\mu_\star|G^{\mathrm{exp}}] + \mathsf{E}_N^{\mathrm{UB}}[\mu_\star|G^{\mathrm{exp}}])$$

as our RB approximation to $\mathsf{E}^{\mathcal{N}}[\mu_\star|G^{\mathrm{exp}}]$.

It can be shown that the expected values defined in (8.37) satisfy

$$\mathsf{E}_N^{\mathrm{LB}}[\mu_\star|G^{\mathrm{exp}}] \leq \mathsf{E}^{\mathcal{N}}[\mu_\star|G^{\mathrm{exp}}] \leq \mathsf{E}_N^{\mathrm{UB}}[\mu_\star|G^{\mathrm{exp}}], \tag{8.38}$$

and hence

$$|\mathsf{E}^{\mathcal{N}}[\mu_\star|G^{\mathrm{exp}}] - \mathsf{E}_N^{\mathrm{AV}}[\mu_\star|G^{\mathrm{exp}}]| \leq \frac{1}{2}\Delta\mathsf{E}_N[\mu_\star|G^{\mathrm{exp}}]$$
$$\equiv \frac{1}{2}(\mathsf{E}_N^{\mathrm{UB}}[\mu_\star|G^{\mathrm{exp}}] - \mathsf{E}_N^{\mathrm{LB}}[\mu_\star|G^{\mathrm{exp}}]).$$

We sketch the proof: we first note that, since $|s^{(m)\mathcal{N}}(t^{k_j^{\mathrm{exp}}};\mu) - s_N^{(m)}(t^{k_j^{\mathrm{exp}}};\mu)| \leq \Delta_N^{s^{(m)}}(t^{k_j^{\mathrm{exp}}};\mu)$,

$$F_N^-(\mu) \leq F^{\mathcal{N}}(\mu) \leq F_N^+(\mu), \quad \forall \mu \in \mathcal{D}; \tag{8.39}$$

it thus follows that

$$D_N(\mu)^T D_N(\mu) \leq (G^{\mathrm{exp}} - F^{\mathcal{N}}(\mu))^T(G^{\mathrm{exp}} - F^{\mathcal{N}}(\mu)) \leq B_N(\mu)^T B_N(\mu), \tag{8.40}$$

and hence

$$\Pi_N^b(G^{\mathrm{exp}}|\mu) \leq \Pi^{\mathcal{N}}(G^{\mathrm{exp}}|\mu) \leq \Pi_N^a(G^{\mathrm{exp}}|\mu). \tag{8.41}$$

The bound result (8.38) is a direct consequence of the definitions (8.37) and inequality (8.41), and the non-negativity of Π^a, Π^b, Π_0, (here) $\mu \in \mathcal{D}$, and finally the quadrature weights.

In words, Π_N^a is an upper bound for $\Pi^{\mathcal{N}}$ since we exploit the reduced basis error bounds to ensure that for each quadrature point the argument of the Π_N^a Gaussian is of smaller magnitude than the argument of the $\Pi^{\mathcal{N}}$ – we *underestimate* the difference between the experimental data and the model prediction; similar arguments demonstrate that Π_N^b constitutes a lower bound for $\Pi^{\mathcal{N}}$ – now we *overestimate* the difference between the experimental data and the model prediction. Thus (given our non-negativity hypotheses) we can selectively choose upper bounds or lower bounds for the numerator and denominator of (8.31) as provided in (8.37). Note that for probability distributions that do not decay monotonically away from the (assumed zero) mean the same procedure can be applied but D_N and B_N will now be slightly more complicated (though still inexpensive to evaluate). We also emphasize that our error estimator $\Delta\mathsf{E}_N[\mu_*|G^{\mathrm{exp}}]$ is a rigorous bound for the difference between the expectation as calculated from the truth and the expectation

as calculated from the reduced basis approximation *for the same quadrature formula* (8.34). Finally, we can not yet propose a similar error bound for the case of a Markov Chain Monte Carlo approach to the Bayesian estimation problem; we defer this topic to future work.

In the Offline stage the RB is constructed: the POD-Greedy$_{\mathrm{RB}}$ sampling procedure is invoked and all necessary Online quantities are computed and stored. Then, in the Online stage (which involves only the Evaluation phase), for each new identification ($\mu_\star$) – and hence for each new G^{exp} provided – we evaluate in 'real–time' the expectation lower and upper bounds (8.37). (Note for given G^{exp} the RB outputs and associated error bounds are computed (only once) and stored on the quadrature grid; we can then evaluate the several requisite integrals without further appeal to the RB approximation.) It is clear that the RB approach will be much faster than direct FE evaluation (of the requisite integrals) even for a single identification, and even more efficient for multiple identifications: in the limit that n_{quad} and/or the number of identifications tends to infinity, the RB Offline effort is negligible – only the very fast ($\mathcal{N}$–independent) RB Online evaluations are relevant. Equivalently, if our emphasis is on real–time identification, again only the very fast RB Online evaluations are important.

8.3.3   Numerical Example

We consider the application of transient thermal analysis to detection of flaws/defects in a Fiber-Reinforced Polymer (FRP) composite bonded to a concrete (C) slab (Grepl 2005; Starnes 2002). Since debonds or delaminations at the composite-concrete interface often occur (even at installation), effective and real-time quality control – providing reliable information about the thickness and fiber content of the composite, and the location and size of defects – is vital to safety.

We show the FRP-concrete system in Figure 8.1. The FRP layer is of thickness h_{FRP} and (truncated) lateral extent $10h_{\mathrm{FRP}}$; the concrete layer is of (truncated) depth and lateral extent $5h_{\mathrm{FRP}}$ and $10h_{\mathrm{FRP}}$, respectively. We presume that a delamination crack of *unknown* length w_{del} centered at $x_1 = 0$ is present at the FRP–concrete interface. The FRP thermal conductivity, specific heat, and density are given by $k, c,$ and ρ with subscripts FRP and C, respectively. We shall assume that the FRP and concrete share the same *known* values for both the density and specific heat. We assume that the FRP (respectively, concrete) conductivity is *unknown* (respectively, *known*); we denote the (*unknown*) conductivity ratio as $\kappa = k_{\mathrm{FRP}}/k_{\mathrm{C}}$. (In practice, the FRP conductivity depends on fiber orientation and content – and hence somewhat unpredictable.)

We nondimensionalize all lengths by $h_{\mathrm{FRP}}/2$ and all times by $h_{\mathrm{FRP}}^2 \rho_C c_C/4k_C$. The nondimensional temperature u is given by $(T - T_0)/(T_{\mathrm{FRP,max}} - T_0)$, where T is the dimensional temperature, T_0 is the initial
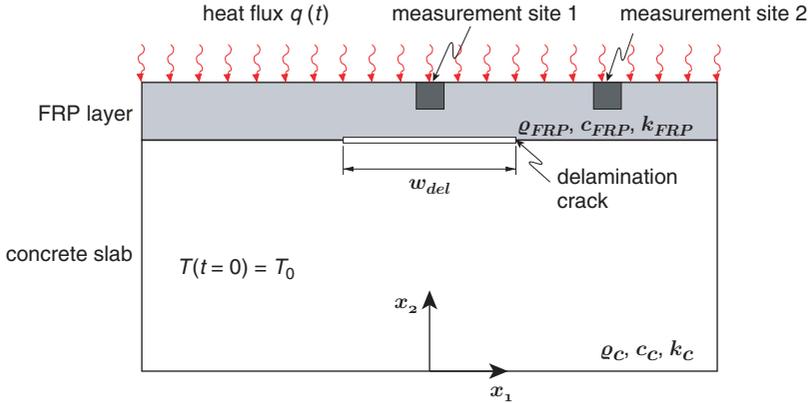
Figure 8.1   Delamination of a FRP layer bonded to a concrete slab.

temperature (uniform in both the FRP and concrete), and $T_{\mathrm{FRP,max}}$ is the maximum allowable FRP temperature. The nondimensional flux – imposed at the FRP exposed surface, as shown in Figure 8.1 – $g(t)$ is given by $q(t)h_{\mathrm{FRP}}/(2k_C(T_{\mathrm{FRP,max}} - T_0))$, where $q(t)$ is the dimensional flux. We presume that the nondimensional surface heat flux $g(t)$ – the stimulus – is unity for $0 \leq t \leq 5$ and zero for all $t > 5$. Henceforth, we refer only to non-dimensional quantities (and thus now w_{del} should be re-interpreted as the dimensional quantity normalized by $h_{\mathrm{FRP}}/2$.)

Upon application of our mapping procedures (to a reference domain with crack length $w_{\mathrm{del}} = 3$) as in Rozza *et al.* (2008) we arrive at the transformed problem statement (1) with affine expansions (2) for $Q_a = 15, Q_m = 2$. (In fact, due to symmetry, we consider only half the domain: $x_1 > 0$.) Our initial condition is $u = 0$; we integrate to a final time $t_f = 10.0$. Our $P = 2$ (both 'unknown') parameters are $\mu \equiv (\mu_1, \mu_2) \equiv (w_{\mathrm{del}}/2, \kappa)$ assumed to reside in the parameter domain $\mathcal{D} \equiv [1, 5] \times [0.5, 2]$. Finally, we introduce our truth discretization: we consider Euler backward discretization in time with $\Delta t = 0.05$ and hence $K = 200$ time levels $t^k = k\Delta t, 0 \leq k \leq K$; we consider a linear truth finite element approximation space $X^{\mathcal{N}}$ of dimension $\mathcal{N} = 3581$. (The triangulation provides high resolution in the vicinity of the surface and near the crack tip, the two regions which suffer sharp spatial gradients.) Finally, we consider $M_{\mathrm{out}} = 2$ outputs: as shown in Figure 8.1, each output functional corresponds to the average of the (temperature) field[5] over a 'small' square of side – length 1 (flush with the exposed FRP surface); the square for the first output is centered at (measurement site 1) $x_1 = 0$, while the square for the second output is centered at (measurement site 2) $x_1 = 6.5$.

---

[5]Note that we must consider a small area average (rather than pointwise measurement) to ensure that our output functionals remain bounded over $L^2(\Omega)$ (indeed, even over $H^1(\Omega)$); the $L^2(\Omega)$ norm of these 'area averaging' functionals increases as the inverse of the square root of the area.

We first briefly discuss the RB approximation and error bounds, and then turn to the inverse problem. This PDE is not too difficult: we need an RB space of dimension only $N = 50$ to ensure – based on $\Delta_N^{s\,(m)}(t^k, \mu), m = 1, 2$ – a 'certified' accuracy of roughly 0.5% in both outputs. In fact, the effectivity – the ratio of the output error bound to the true output error – is rather large, $O(100)$, and hence the *actual* accuracy for $N = 50$ is less than 10^{-4}; however, since in the Online stage our inferences are based on the (inexpensive) error bound, we must construct an RB approximation for which the *error bound* is sufficiently accurate. For $N = 50$ the Online RB calculation $\mu \rightarrow s_N^{(m)}(t^k; \mu), \Delta_N^{s\,(m)}(t^k; \mu), 0 \le k \le K$, is effected in 0.21 seconds; in contrast, direct truth evaluation requires 22 seconds. All computations in this section are carried out on a 1.73 GHz Pentium IV processor with 1GB memory.

We now turn to parameter estimation. We focus on the sensitivity of the parameter estimation procedure to the RB dimension N as (inexpensively but rigorously) quantified by our expectation error bounds. In this experiment, we set $\mu_\star = (\mu_{1\star}, \mu_{2\star}) = (w_{\text{del}\star}, \kappa_\star) = (2.8, 0.9)$ and $\sigma_{\text{exp}}^2 = 0.0025$; we choose for the prior mean and variance $\mu_0 = (3.3, 1.2)$ and $\sigma_0^2 = 0.04$, respectively. The synthetic experimental data (8.30) is generated by adding i.i.d. Gaussian random variables to our $M_{\text{out}} = 2$ outputs evaluated at $J = 20$ time levels $t^{k_j^{\text{exp}}}$, $k_j^{\text{exp}} = 10j, 1 \le j \le J$. We then apply our adaptive piecewise Gauss-Legendre quadrature algorithm with $n_{\text{quad}} = 10,000$ points.

We present in Table 8.1 the lower bound, $\mathsf{E}_N^{\text{LB}}[\mu_{p\star}]$, upper bound, $\mathsf{E}_N^{\text{UB}}[\mu_{p\star}]$, and bound gap $\Delta\mathsf{E}_N[\mu_{p\star}], p = 1, 2$, for the expected value of the unknown parameter $\mu_\star$; we consider a single realization G^{exp}. We observe that the bound gaps $\Delta\mathsf{E}_N[\mu_{p\star}] = \mathsf{E}_N^{\text{UB}}[\mu_{p\star}] - \mathsf{E}_N^{\text{LB}}[\mu_{p\star}], p = 1, 2$, decrease rapidly: as N increases, $\Delta_N^{s\,(m)}(t^k; \mu) \rightarrow 0$ and hence $D_N(\mu) \rightarrow B_N(\mu)$ rapidly. The parameter estimator is quite accurate: the expectation bounds for larger N are within the white noise (5.0%) of the true parameter value $\mu_\star = (2.8, 0.9)$, biased toward μ_0 as expected. The RB Online computation (for $N = 50$) of the lower and upper bounds for the expected value is completed in approximately

Table 8.1 Lower bound, upper bound, and bound gap for the expected value of the delamination half-width μ_1 and conductivity ratio μ_2 as a function of N. The true parameter value is $\mu_{1\star} = 2.8$ and $\mu_{2\star} = 0.9$.

	Delamination half-width			Conductivity ratio		
N	$\mathsf{E}_N^{\text{LB}}[\mu_{1\star}]$	$\mathsf{E}_N^{\text{UB}}[\mu_{1\star}]$	$\Delta\mathsf{E}_N[\mu_{1\star}]$	$\mathsf{E}_N^{\text{LB}}[\mu_{2\star}]$	$\mathsf{E}_N^{\text{UB}}[\mu_{2\star}]$	$\Delta\mathsf{E}_N[\mu_{2\star}]$
10	1.0527	7.5175	6.4648	0.3427	2.4468	2.1041
20	2.3896	3.3120	0.9224	0.7764	1.0759	0.2996
30	2.7417	2.8836	0.1419	0.8917	0.9378	0.0461
40	2.8008	2.8236	0.0228	0.9111	0.9185	0.0074
50	2.8096	2.8192	0.0096	0.9136	0.9171	0.0035

35 minutes – arguably 'real-time' for this particular application – as opposed to 61 hours for direct FE evaluation. The RB Offline time is roughly 2.8 hours, and hence even for one identification the RB approach 'pays off'; for several identifications, the RB Offline effort will be negligible. (If real-time response 'in the field' is imperative, then even for one identification the RB Offline effort is not important.) In short, we are guaranteed the fidelity of the truth FE approximation but at the cost of a low order model.

8.4   Concluding Remarks

In this chapter we have developed a framework for reduced basis approximation and a posteriori error estimation for parametrized linear parabolic partial differential equations. We have argued, and computationally confirmed, that the reduced basis approach can provide highly accurate, very inexpensive, rigorously certified predictions in the real–time and many–query contexts. We have further demonstrated that the certified reduced basis method can be integrated into a Bayesian framework to provide very rapid yet reliable parameter estimation procedures; similar advances should be possible for optimization and control applications as well as multi-scale analyses.

Certainly the most important outstanding issue (at least within the context of parabolic partial differential equations) is generality: given an 'original' problem of interest (i) is there an effective parametrization such that the resulting 'transformed' problem is amenable to efficient and rigorous reduced basis treatment, and (ii) can this effective parametrization be automatically deduced and subsequently implemented within a general framework? At present we know of large classes of linear problems and much smaller classes of nonlinear problems (Nguyen *et al.* 2009) which can be and have been successfully addressed by the certified reduced basis approach; future work must focus both on theoretical advances to identify important impediments and computational advances to address these restrictions.

Acknowledgement

References

Almroth, B.O., Stern, P. and Brogan, F.A. 1978 Automatic choice of global shape functions in structural analysis. *AIAA Journal* **16**, 525–528.

Amsallem, D. and Farhat, C. 2008 Interpolation method for adapting reduced-order models and application to aeroelasticity. *AIAA Journal* **46**(7), 1803–1813.

Amsallem, D., Cortial, J. and Farhat, C. 2009 On-demand CFD-based aeroelastic predictions using a database of reduced-order bases and models. *47th AIAA Aerospace Sciences Meeting, Including the New Horizons Forum and Aerospace Exposition, 5–8 January 2009, Orlando, Florida*. Paper 2009-800.

Balmes, E. 1996 Parametric families of reduced finite element models: Theory and applications. *Mechanical Systems and Signal Processing* **10**(4), 381–394.

Barrault, M., Maday, Y., Nguyen, N.C. and Patera, A.T. 2004 An 'empirical interpolation' method: Application to efficient reduced-basis discretization of partial differential equations. *C. R. Acad. Sci. Paris, Série I*. **339**, 667–672.

Boyaval, S. 2008 Reduced-basis approach for homogenization beyond the periodic setting. *SIAM Multiscale Modeling & Simulation* **7**(1), 466–494.

Boyaval, S., Le Bris, C., Nguyen, N.C. and Patera, A.T. 2008 A reduced-basis approach for variational problems with stochastic parameters: Application to heat conduction with variable robin coefficient. *Comp. Meth. Appl. Mech. and Eng*.

Bui-Thanh, T., Willcox, K. and Ghattas, O. 2007 Model reduction for large-scale systems with high-dimensional parametric input space (AIAA Paper 2007-2049) *Proceedings of the 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Material Conference*.

Cancès, E., Le Bris, C., Nguyen, N.C., Maday, Y., Patera, A.T. and Pau, G.S.H. 2007 Feasibility and competitiveness of a reduced basis approach for rapid electronic structure calculations in quantum chemistry. *Proceedings of the Workshop for High-dimensional Partial Differential Equations in Science and Engineering (Montreal)*.

Christensen, E.A., Brons, M. and Sorensen, J.N. 2000 Evaluation of pod-based decomposition techniques applied to parameter-dependent non-turbulent flows. *SIAM J. Sci. Comput*. **21**, 1419–1434.

Constantin, P. and Foias, C. 1988 *Navier-Stokes Equations*. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL.

Deane, A., Kevrekidis, I., Karniadakis, G. and Orszag, S. 1991 Low-dimensional models for complex geometry flows: Application to grooved channels and circular cylinders. *Phys. Fluids* **10**, 2337–2354.

Deparis, S. 2008 Reduced basis error bound computation of parameter-dependent Navier–Stokes equations by the natural norm approach. *SIAM Journal of Numerical Analysis* **46**, 2039–2067.

Fink, J.P. and Rheinboldt, W.C. 1983 On the error behavior of the reduced basis technique for nonlinear finite element approximations. *Z. Angew. Math. Mech*. **63**(1), 21–28.

Grepl, M. 2005 *Reduced-Basis Approximations and A Posteriori Error Estimation for Parabolic Partial Differential Equations*. PhD thesis. Massachusetts Institute of Technology.

Grepl, M.A. and Patera, A.T. 2005 *A posteriori* error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *M2AN (Math. Model. Numer. Anal.)* **39**(1), 157–181.

Grepl, M.A., Maday, Y., Nguyen, N.C. and Patera, A.T. 2007a Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations. *M2AN (Math. Model. Numer. Anal.)* **41**, 575–605.

Grepl, M.A., Nguyen, N.C., Veroy, K., Patera, A.T. and Liu, G.R. 2007b Certified rapid solution of partial differential equations for real-time parameter estimation and optimization. In *Proceedings of the 2nd Sandia Workshop of PDE-Constrained Optimization: Real-Time PDE-Constrained Optimization* (ed. Biegler, L.T., Ghattas, O., Heinkenschloss, M., Keyes, D. and van B. Wandeers, B.), pp. 197–216 SIAM Computational Science and Engineering Book Series.

Gunzburger, M.D. 1989 *Finite Element Methods for Viscous Incompressible Flows*. Academic Press.

Gunzburger, M.D., Peterson, J. and Shadid, J.N. 2007 Reduced-order modeling of time-dependent PDEs with multiple parameters in the boundary data. *Comp. Meth. Applied Mech.* **196**, 1030–1047.

Haasdonk, B. and Ohlberger, M. 2008 Reduced basis method for finite volume approximations of parametrized evolution equations. *M2AN (Math. Model. Numer. Anal.)* **42**(2), 277–302.

Hinze, M. and Volkwein, S. 2005 Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: error estimates and suboptimal control. *Lecture Notes in Computational Science and Engineering* vol. 45 Springer.

Huynh, D.B.P., Nguyen, N.C., Rozza, G. and Patera, A.T. 2007-2009 *rbMIT Software:* http://augustine.mit.edu/methodology/methodology_rbMIT_System.htm. Copyright MIT, Technology Licensing Office, case 12600, Cambridge, MA.

Huynh, D.B.P., Rozza, G., Sen, S. and Patera, A.T. 2007 A successive constraint linear optimization method for lower bounds of parametric coercivity and inf-sup stability constants. *C. R. Acad. Sci. Paris, Analyse Numérique* **345**(8), 473–478.

Ito, K. and Ravindran, S.S. 1998a A reduced basis method for control problems governed by PDEs. In *Control and Estimation of Distributed Parameter Systems* (ed. Desch, W., Kappel, F. and Kunisch, K.) Birkhäuser pp. 153–168.

Ito, K. and Ravindran, S.S. 1998b A reduced-order method for simulation and control of fluid flows. *Journal of Computational Physics* **143**(2), 403–425.

Ito, K. and Ravindran, S.S. 2001 Reduced basis method for optimal control of unsteady viscous flows. *International Journal of Computational Fluid Dynamics* **15**(2), 97–113.

Johansson, P.S., Andersson, H. and Rønquist, E. 2006 Reduced-basis modeling of turbulent plane channel flow. *Computers and Fluids* **35**(2), 189–207.

Johnson, C., Rannacher, R. and Boman, M. 1995 Numerical and hydrodynamic stability: Towards error control in computational fluid dynamics. *SIAM Journal of Numerical Analysis* **32**(4), 1058–1079.

Joseph, D. 1976 *Stability of fluid motions. I. & II*. vol. 27 & 28 of *Springer Tracts in Natural Philosophy*. Springer-Verlag, New York.

Karniadakis, G.E., Mikic, B.B. and Patera, A.T. 1988 Minimum-dissipation transport enhancement by flow destabilization – Reynolds' analogy revisited. *Journal of Fluid Mechanics* **192**, 365–391.

Knezevic, D. and Patera, A.T. 2009 A certified reduced basis method for fokker–planck equation of dilute polymeric fluids: Fene dumbbells in extensional flow. *SIAM J. Sci. Comput*. Submitted.

Kunisch, K. and Volkwein, S. 2002 Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. *SIAM J. Num. Analysis* **40**(2), 492–515.

Kunish, K. and Volkwein, S. 1999 Control of Burgers' equation by a reduced order approach using proper orthogonal decomposition. *J. Optimization Theory and Applications* **102**, 345–371.

Maday, Y., Patera, A. and Turinici, G. 2002 *A Priori* convergence theory for reduced-basis approximations of single-parameter elliptic partial differential equations. *Journal of Scientific Computing* **17**(1–4), 437–446.

Mosegaard, K. and Tarantola, A. 2002 Probabilistic approach to inverse problems. *International Handbook of Earthquake and Engineering Seismology, Part A* pp. 237–265.

Nguyen, N.C. 2008 Efficient and robust parameter estimation using the reduced basis method: Application to the inverse scattering problem. *SIAM J. Sci. Comput.*

Nguyen, N.C. 2008 A multiscale reduced-basis method for parametrized elliptic partial differential equations with multiple scales. *Journal of Computational Physics* **227**(23), 9807–9822.

Nguyen, N.C., Rozza, G. and Patera, A.T. 2009 Reduced basis approximation and a posteriori error estimation for the time-dependent viscous burgers equation. *Calcolo* **46**, 157–185.

Nguyen, N.C., Veroy, K. and Patera, A.T. 2005 Certified real-time solution of parametrized partial differential equations. In *Handbook of Materials Modeling* (ed. Yip, S.) Springer pp. 1523–1558.

Noor, A.K. and Peters, J.M. 1980 Reduced basis technique for nonlinear analysis of structures. *AIAA Journal* **18**(4), 455–462.

Pierce, N. and Giles, M.B. 2000 Adjoint recovery of superconvergent functionals from PDE approximations. *SIAM Review* **42**(2), 247–264.

Porsching, T.A. 1985 Estimation of the error in the reduced basis method solution of nonlinear equations. *Mathematics of Computation* **45**(172), 487–496.

Prud'homme, C., Rovas, D., Veroy, K., Maday, Y., Patera, A. and Turinici, G. 2002 Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bounds methods. *Journal of Fluids Engineering* **124**(1), 70–80.

Quarteroni, A. and Valli, A. 1997 *Numerical Approximation of Partial Differential Equations*, 2nd edn. Springer.

Rozza, G., Huynh, D.B.P. and Patera, A.T. 2008 Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations: Application to transport and continuum mechanics. *Archives Computational Methods in Engineering* **15**(3), 229–275.

Sen, S., Veroy, K., Huynh, D.B.P., Deparis, S., Nguyen, N.C. and Patera, A.T. 2006 'Natural norm' *a posteriori* error estimators for reduced basis approximations. *Journal of Computational Physics* **217**, 37–62.

Starnes, M. 2002 *Development of Technical Bases for Using Infrared Thermography for Nondestructive Evaluation of Fiber Reinforced Polymer Composites Bonded to Concrete* PhD thesis Massachusetts Institute of Technology.

Veroy, K. and Patera, A.T. 2005 Certified real-time solution of the parametrized steady incompressible Navier-Stokes equations; Rigorous reduced-basis *a posteriori* error bounds. *International Journal for Numerical Methods in Fluids* **47**, 773–788.

Veroy, K., Prud'homme, C. and Patera, A.T. 2003a Reduced-basis approximation of the viscous Burgers equation: Rigorous *a posteriori* error bounds. *C. R. Acad. Sci. Paris, Série I* **337**(9), 619–624.

Veroy, K., Prud'homme, C., Rovas, D.V. and Patera, A.T. 2003b *A Posteriori* error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations. *Proceedings of the 16th AIAA Computational Fluid Dynamics Conference*. Paper 2003-3847.

Wang, J. and Zabaras, N. 2005 Using Bayesian statistics in the estimation of heat source in radiation. *International Journal of Heat and Mass Transfer* **48**, 15–29.