Reduced-Basis Methods Applied to Problems in Elasticity: Analysis and Applications

by

Karen Veroy

S.M. Civil Engineering (2000)
Massachusetts Institute of Technology

B.S. Physics (1996)
Ateneo de Manila University, Philippines

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2003

Author .....
Department of Civil and Environmental Engineering
April 15, 2003

Certified by .....
Anthony T. Patera
Professor of Mechanical Engineering
Thesis Supervisor

Certified by .....
Franz-Josef Ulm
Professor of Civil and Environmental Engineering
Chairman, Thesis Committee

Accepted by .....
Oral Buyukozturk
Professor of Civil and Environmental Engineering
Chairman, Department Committee on Graduate Students

Reduced-Basis Methods Applied to Problems in Elasticity: Analysis and Applications

by

Karen Veroy

Submitted to the Department of Civil and Environmental Engineering
on April 15, 2003, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Modern engineering problems require accurate, reliable, and efficient evaluation of quantities of interest, the computation of which often requires solution of a partial differential equation. We present a technique for the prediction of linear-functional outputs of elliptic partial differential equations with affine parameter dependence. The essential components are: (*i*) rapidly convergent global reduced-basis approximations — projection onto a space W_N spanned by solutions of the governing partial differential equation at N selected points in parameter space (*Accuracy*); (*ii*) *a posteriori* error estimation — relaxations of the error-residual equation that provide inexpensive bounds for the error in the outputs of interest (*Reliability*); and (*iii*) off-line/on-line computational procedures — methods which decouple the generation and projection stages of the approximation process (*Efficiency*). The operation count for the on-line stage depends only on N (typically very small) and the parametric complexity of the problem.

We present two general approaches for the construction of error bounds: Method I, *rigorous a posteriori* error estimation procedures which rely critically on the existence of a "bound conditioner" — in essence, an operator preconditioner that (*a*) satisfies an additional spectral "bound" requirement, and (*b*) admits the reduced-basis off-line/on-line computational stratagem; and Method II, *a posteriori* error estimation procedures which rely only on the rapid convergence of the reduced-basis approximation, and provide simple, inexpensive error bounds, albeit at the loss of complete certainty. We illustrate and compare these approaches for several simple test problems in heat conduction, linear elasticity, and (for Method II) elastic stability.

Finally, we apply our methods to the "static" (at conception) and "adaptive" (in operation) design of a multifunctional microtruss channel structure. We repeatedly and rapidly evaluate bounds for the average deflection, average stress, and buckling load for different parameter values to best achieve the design objectives subject to performance constraints. The output estimates are sharp — due to the rapid convergence of the reduced-basis approximation; the performance constraints are reliably satisfied — due to our *a posteriori* error estimation procedure; and the computation is essentially real-time — due to the off-line/on-line decomposition.

Thesis Supervisor: Anthony T. Patera
Title: Professor of Mechanical Engineering

Acknowledgments

I have had the most wonderful fortune of having Professor Anthony T. Patera as my advisor. For his support and guidance, his counsel and example, his humor and patience, I am truly grateful. I would also like to thank the members of my thesis committee, Professor Franz-Josef Ulm and Professor Jerome J. Connor, and also Professor John W. Hutchinson of Harvard University and Professor Anthony G. Evans of the University of California (Santa Barbara), for their comments, suggestions, and encouragement throughout my studies. I am very thankful to have had the opportunity to learn from and interact with them.

During my doctoral studies, I have had the good fortune to extensively collaborate with Ivan Oliveira, Christophe Prud'homme, Dimitrios Rovas, Shidrati Ali, and Thomas Leurent; this work would not have been possible without them. I would also like to thank Yuri Solodukhov, Ngoc Cuong Nguyen, Lorenzo Valdevit, and Gianluigi Rozza for many helpful and interesting discussions. Furthermore, I am most grateful to Debra Blanchard for her invaluable help and support, and for providing almost all of the wonderful artwork in this thesis.

Finally, I owe the deepest gratitude to my family — Raoul, Lizza, Butch, Chesca, Nicco, Iya, Ellen, Andring, and especially my parents, Bing and Renette — and to my best friend Martin, for their love and support without which I would not have been able to pursue my dreams.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

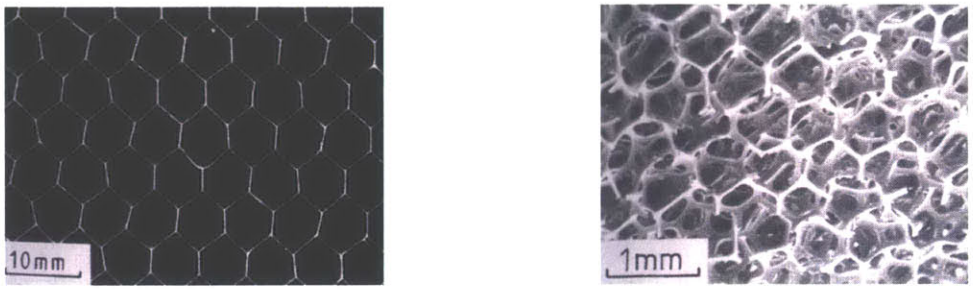
The optimization, control, and characterization of an engineering component or system requires the rapid (often real-time) evaluation of certain performance metrics, or *outputs*, such as deflections, maximum stresses, maximum temperatures, heat transfer rates, flowrates, or lifts and drags. These "quantities of interest" are typically functions of parameters which reflect variations in loading or boundary conditions, material properties, and geometry. The parameters, or *inputs*, thus serve to identify a particular "configuration" of the component. However, often implicit in these input-output relationships are underlying partial differential equations governing the behavior of the system, reliable solution of which demands great computational expense especially in the context of optimization, control, and characterization.

Our goal is the development of computational methods that permit *rapid* yet *accurate* 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. To further motivate our methods and illustrate the contexts in which we develop them, we consider the following example.

1.1 Motivation: A MicroTruss Example

Cellular solids consist of interconnected networks of solid struts or plates which form cells [16] and may, in general, be classified as either *periodic* (as in lattice and prismatic materials, shown in Figure 1-1(a)) or *stochastic* (as in sponges and foams, shown in Figure 1-1(b)) [11]. Numerous examples abound in nature — for instance, cork, wood, and bone — but recent materials and manufacturing advances have allowed synthetic cellular materials to be designed and fabricated for specific applications — for instance, lightweight structures, thermal insulation, energy absorption, and vibration control [16]. In particular, *open cellular metals* have received considerable attention largely due to their multifunctional capability: the metal struts offer relatively high structural load capacities at low weights, while the high thermal conductivities and open architecture allow for the efficient transfer of heat at low pumping costs.

We consider the example of a periodic open cellular structure (shown in Figure 1-2) simultaneously designed for both heat-transfer and structural capability; this structure could represent, for instance, a section of a combustion chamber wall in a reusable rocket engine [12, 19]. The prismatic microtruss consists of a frame (upper and lower faces) and a core of trusses. The structure conducts heat from a prescribed uniform flux source $\hat{q}''$ at the upper face to the coolant flowing through the open cells; the coolant enters the inlet at a temperature $\hat{T}_0$, and is forced through the cells by a pressure drop $\Delta\hat{P} = \hat{P}_{\text{high}} - \hat{P}_{\text{low}}$ from the inlet to the outlet. In addition, the microtruss transmits



(a) (b)

Figure 1-1: Examples of (a) periodic (honeycomb) and (b) stochastic (foam) cellular structures (Photographs taken from [16]).

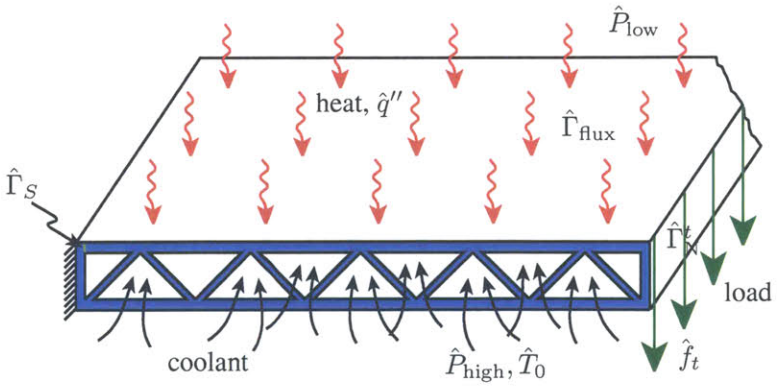


Figure 1-2: A multifunctional (thermo-structural) microtruss structure.

a force per unit depth $\hat{f}_t$ uniformly distributed over the tip $\hat{\Gamma}_{\rm N}$ through the truss system to the fixed left wall $\hat{\Gamma}_{\rm D}$. We assume that the associated Reynolds number is below the transition value and that the structure is sufficiently deep such that a physical model of fully-developed, laminar fluid flow and plane strain (two-dimensional) linear elasticity suffices.

1.1.1 Inputs

The structure, shown in Figure 1-3, is characterized by a seven-component nondimensional parameter vector or "input," $\mu = (\mu^1, \mu^2, \ldots, \mu^7)$, reflecting variations in geometry, material property, and loading or boundary conditions. Here,

$$\begin{array}{lcll} \mu^1 & = & \bar{t} & = \text{ thickness of the core trusses,} \\ \mu^2 & = & \bar{t}_{\rm t} & = \text{ thickness of the upper frame,} \\ \mu^3 & = & \bar{t}_{\rm b} & = \text{ thickness of the lower frame,} \\ \mu^4 & = & \bar{H} & = \text{ separation between the upper and lower frames,} \\ \mu^5 & = & \bar{\alpha} & = \text{ angle (in degrees) between the trusses and the frames,} \\ \mu^6 & = & \bar{k} & = \text{ thermal conductivity of the solid relative to the fluid, and} \\ \mu^7 & = & \bar{p} & = \text{ nondimensional pressure gradient;} \end{array}$$

furthermore, μ may take on any value in a specified design space, $\mathcal{D}^\mu \subset \mathbb{R}^7$, defined as

$$\mathcal{D}^\mu = [0.1, 2.0]^3 \times [6.0, 12.0] \times [35.0, 70.0] \times [5.0 \times 10^2, 1.0 \times 10^4] \times [1.0 \times 10^{-2}, 1.1 \times 10^2],$$

that is, $0.1 \le \bar{t}, \bar{t}_{\rm t}, \bar{t}_{\rm b} \le 2.0$, $6.0 \le \bar{H} \le 12.0$, $35.0 \le \bar{\alpha} \le 70.0$, $5.0 \times 10^2 \le \bar{k} \le 1.0 \times 10^4$, and $1.0 \times 10^{-2} \le \bar{p} \le 1.1 \times 10^2$. The thickness of the sides, $\bar{t}_{\rm s}$, is assumed to be equal to $t_{\rm b}$.

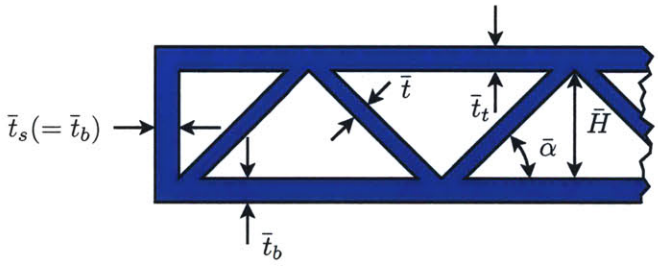


Figure 1-3: Geometric parameters for the microtruss structure.

1.1.2 Governing Partial Differential Equations

In this section, (and in much of this thesis) we shall omit the spatial dependence of the field variables. Furthermore, we shall use a bar to denote a general dependence on the parameter; for example, since the domain itself depends on the (geometric) parameters, we write $\bar{\Omega} \equiv \bar{\Omega}(\mu)$ to denote the domain, and $\bar{x}$ to denote any point in $\bar{\Omega}$. Also, we shall use repeated indices to signify summation.

Heat Transfer Model

The (nondimensionalized) temperature, $\bar{\vartheta}$, in the fluid satisfies the parametrized partial differential equation

$$-\bar{\nabla}^2 \bar{\vartheta} + \bar{p}\bar{V}_i \frac{\partial \bar{\vartheta}}{\partial \bar{x}_i} = 0, \quad \text{in } \bar{\Omega}_f \ , \tag{1.1}$$

with boundary conditions

$$\bar{\vartheta} = 0, \quad \text{on } \bar{\Gamma}_f^{\text{in}} \tag{1.2}$$

$$\frac{\partial \bar{\vartheta}}{\partial \bar{x}_i} \bar{\mathrm{e}}_i^n = 0, \quad \text{on } \bar{\Gamma}_f^{\text{out}} ; \tag{1.3}$$

note that the velocity field $\bar{V} = \begin{bmatrix} 0 & 0 & \bar{V}_3 \end{bmatrix}^T$, where $\bar{V}_3$ satisfies

$$\nabla^2 \bar{V}_3 = 1 . \tag{1.4}$$

Here, $\underline{\bar{\mathrm{e}}}^n$ denotes the unit outward normal, $\bar{\Omega}_f$ denotes the fluid domain, and $\bar{\Gamma}_f^{\text{in}}$ (respectively, $\bar{\Gamma}_f^{\text{out}}$) denotes the fluid inlet (respectively, outlet). The temperature in the solid is governed by

$$-\bar{k} \nabla^2 \bar{\vartheta} = 0, \quad \text{in } \bar{\Omega}_s , \tag{1.5}$$

with

$$\bar{k} \frac{\partial \bar{\vartheta}}{\partial \bar{x}_i} \bar{\mathrm{e}}_i^n = 1 \quad \text{on } \bar{\Gamma}_s^{\text{flux}} , \tag{1.6}$$

$$\frac{\partial \bar{\vartheta}}{\partial \bar{x}_i} \bar{\mathrm{e}}_i^n = 0 \quad \text{on } \bar{\Gamma}_s^{\text{ins}} , \tag{1.7}$$

reflecting the uniform flux and insulated boundary conditions, respectively. Continuity of temperature and heat flux along the interface $\bar{\Gamma}^{\text{int}}$ between the solid and fluid domains requires

$$\bar{\vartheta}\big|_s = \bar{\vartheta}\big|_f \quad \text{on } \bar{\Gamma}^{\text{int}} , \tag{1.8}$$

$$-\bar{k} \left. \frac{\partial \bar{\vartheta}}{\partial \bar{x}_i} \bar{\mathrm{e}}_i^n \right|_s = \left. \frac{\partial \bar{\vartheta}}{\partial \bar{x}_i} \bar{\mathrm{e}}_i^n \right|_f \quad \text{on } \bar{\Gamma}^{\text{int}}. \tag{1.9}$$

Structural (Solid Mechanics) Model

The (nondimensionalized) displacement, $\bar{u}_i, i = 1, 2$, satisfies the parametrized partial differential equation

$$\frac{\partial}{\partial \bar{x}_j} \left(\bar{C}_{ijkl} \frac{\partial \bar{u}_k}{\partial \bar{x}_l} \right) = 0, \qquad \text{in } \bar{\Omega}_s, \tag{1.10}$$

where $\bar{\Omega}_s$ denotes the truss domain, and the elasticity tensor $\bar{C}_{ijkl}$ is given by

$$\bar{C}_{ijkl} = \bar{c}_1 \left(\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk} \right) + \bar{c}_2 \delta_{ij} \delta_{kl}; \tag{1.11}$$

here, δ_{ij} is the Kronecker delta function, and $\bar{c}_1$, $\bar{c}_2$ are Lame's constants, related to Young's modulus, $\bar{E}$, and Poisson's ratio, $\bar{\nu}$, by

$$\bar{c}_1 = \frac{\bar{E}}{2(1+\bar{\nu})}, \tag{1.12}$$

$$\bar{c}_2 = \frac{\bar{E}\bar{\nu}}{(1+\bar{\nu})(1-2\bar{\nu})}. \tag{1.13}$$

The displacement and traction boundary conditions are

$$\bar{u}_i = 0 \qquad \text{on } \bar{\Gamma}_{\mathrm{D}}, \tag{1.14}$$

$$\bar{\sigma}_{ij}\bar{\mathrm{e}}_j^n = f^t\bar{\mathrm{e}}_i^t \qquad \text{on } \bar{\Gamma}_{\mathrm{N}}^t, \tag{1.15}$$

$$\bar{\sigma}_{ij}\bar{\mathrm{e}}_j^n = 0 \qquad \text{on } \bar{\Gamma}_s\backslash(\bar{\Gamma}_{\mathrm{D}} \cup \bar{\Gamma}_{\mathrm{N}}^t), \tag{1.16}$$

where the stresses, $\bar{\sigma}_{ij}$ are related to the displacements by

$$\bar{\sigma}_{ij} = \bar{C}_{ijkl}\frac{\partial \bar{u}_k}{\partial \bar{x}_l} \ . \tag{1.17}$$

Elastic Buckling

Furthermore, it can be shown [18] that the critical load parameter $\lambda^1 \in \mathbb{R}$, and associated buckling mode $\bar{\xi}^1 \in Y$, are solutions to the partial differential eigenproblem

$$\frac{\partial}{\partial \bar{x}_j}\left(\bar{C}_{ijkl}\frac{\partial \bar{\xi}_k}{\partial \bar{x}_l}\right) + \lambda\frac{\partial}{\partial \bar{x}_j}\left(\frac{\partial \bar{u}_m}{\partial \bar{x}_l}\bar{C}_{mlkj}\frac{\partial \bar{\xi}_i}{\partial \bar{x}_k}\right) = 0 \ , \tag{1.18}$$

with boundary conditions

$$\bar{C}_{ijkl}\frac{\partial \bar{\xi}_k}{\partial \bar{x}_l}\bar{\mathrm{e}}_j^n + \lambda\frac{\partial \bar{\xi}_i}{\partial \bar{x}_k}\bar{f}_n\bar{\mathrm{e}}_k^n = 0 \qquad \text{on } \bar{\Gamma}_N^n \tag{1.19}$$

$$\bar{C}_{ijkl}\frac{\partial \bar{\xi}_k}{\partial \bar{x}_l}\bar{\mathrm{e}}_j^n + \lambda\frac{\partial \bar{\xi}_i}{\partial \bar{x}_k}\bar{f}_t\bar{\mathrm{e}}_k^t = 0 \qquad \text{on } \bar{\Gamma}_N^t \ . \tag{1.20}$$

(The derivation of the equivalent weak form of (1.18)-(1.20) is presented in Chapter 6.)

1.1.3 Outputs

In the engineering context — i.e., in design, optimization, control, and characterization — the quantities of interest are often not the field variables themselves, but rather functionals of the field variables. For example, in our microtruss example the relevant "outputs" are neither the temperature field, $\bar{\vartheta}$, the fluid velocity field, $\bar{V}$, nor the displacement field, $\bar{u}$; rather, we may wish to evaluate as a function of the parameter μ the average temperature along $\bar{\Gamma}_s^{\text{flux}}$,

$$\vartheta_{\text{ave}}(\mu) = \frac{1}{|\bar{\Gamma}_s^{\text{flux}}|}\int_{\bar{\Gamma}_s^{\text{flux}}} \bar{\vartheta}\, d\bar{\Gamma} \ , \tag{1.21}$$

the flow rate,

$$\mathcal{Q}(\mu) = \int_{\bar{\Gamma}_f^{\text{out}}} \bar{V}_3\, d\bar{\Gamma} \ , \tag{1.22}$$

the average velocity at the outlet,

$$V_{\text{ave}} = \frac{1}{|\bar{\Gamma}_f^{\text{out}}|}\int_{\bar{\Gamma}_f^{\text{out}}} \bar{V}_3\, d\bar{\Gamma} \ , \tag{1.23}$$

the average deflection along $\bar{\Gamma}^t_{\rm N}$,

$$\delta_{\rm ave}(\mu) = -\frac{1}{|\bar{\Gamma}^t_{\rm N}|}\int_{\bar{\Gamma}^t_{\rm N}} \bar{u}_2 \, d\bar{\Gamma} \ , \tag{1.24}$$

and the normal stress near the support averaged along $\bar{\Gamma}_\sigma$,

$$\sigma_{\rm ave}(\mu) = -\frac{1}{|\bar{\Gamma}_\sigma|}\int_{\bar{\Gamma}_\sigma} \bar{\sigma}_{11}(\bar{u}). \tag{1.25}$$

In addition, one may also be interested in eigenvalues associated with the physical system; for example, in our microtruss example the buckling load

$$f^{\pm}_{\rm buckle}(\mu) = \lambda^{\pm}_1(\mu) f^t \ , \tag{1.26}$$

where λ^1 is the smallest eigenvalue of (1.18)-(1.20), is also of interest. Note that evaluation of these outputs requires solution of the governing partial differential equations of Section 1.1.2.

1.1.4 A Design-Optimize Problem

A particular microtruss design (corresponding to a particular value of μ) has associated with it (say) operational and material costs, as well as performance merits reflecting its ability to support the applied structural loads and efficiently transfer the applied heat to the fluid. Furthermore, a design must meet certain constraints reflecting, for example, safety and manufacturability considerations. The goal of the design process is then to minimize costs and optimize performance while ensuring that all design constraints are satisfied.

For example, we could define our *cost function,* $\mathcal{J}(\mu)$, as a weighted sum of the area of the structure, $\mathcal{A}(\mu)$, (reflecting material costs), and the power, $\mathcal{P}(\mu)$ required to pump the fluid through the structure (reflecting operational costs); that is,

$$\mathcal{J}(\mu) = a_{\mathcal{V}}\mathcal{V}(\mu) + a_{\mathcal{P}}\mathcal{P}(\mu), \tag{1.27}$$

where

$$\mathcal{V}(\mu) = 2\left[(t_t + t_b)\left(\frac{H}{\tan\alpha} + \frac{t}{\sin\alpha} + t_b\right) + H\left(\frac{t}{\sin\alpha} + t_b\right)\right], \tag{1.28}$$

$$\mathcal{P}(\mu) = \mathcal{Q}(\mu), \tag{1.29}$$

and $a_{\mathcal{V}}$, $a_{\mathcal{P}}$ are weights which measure the relative importance of material and operational costs in the design process.

Furthermore, we require that the average temperature be less than the melting temperature of the material:

$$\vartheta_{\rm ave}(\mu) \le \alpha_1 \vartheta_{\rm max}, \tag{1.30}$$

the deflection be less than a prescribed limit,

$$\delta_{\rm ave}(\mu) \le \alpha_2 \delta_{\rm max}, \tag{1.31}$$

the average normal stress near the support be less than the yield stress,

$$\sigma_{\rm ave}(\mu) \le \alpha_3 \sigma^T_Y, \tag{1.32}$$

and the magnitude of the applied load be less than the critical buckling load, so that

$$1 \leq \alpha_4 \lambda_1^{\pm}(\mu) \ . \tag{1.33}$$

We then define our *feasible set*, $\mathcal{F}$, as the set of all values of $\mu \in \mathcal{D}^{\mu}$ which satisfy the constraints (1.30)-(1.33).

Our *optimization problem* can now be stated as: find the optimal design, μ^*, which satisfies

$$\mu^* = \arg\min_{\mu \in \mathcal{F}} \mathcal{J}(\mu); \tag{1.34}$$

in other words, find that value of μ, μ^*, which minimizes the cost functional over all feasible designs $\mu \in \mathcal{F}$.

1.1.5 An Assess-Optimize Problem

The design of an engineering system, as illustrated in Section 1.1.4, involves the determination of the system configuration based on system requirements and environment considerations. During operation, however, the state of the system may be unknown or evolving, and the system may be subjected to dynamic system requirements, as well as changing environmental conditions. The system must therefore be adaptively designed and optimized, taking into consideration the uncertainty and variability of system state, requirements, and environmental conditions.

For example, we assume that extended deployment of our microtruss structure (for instance, as a component in an airplane wing) has led to the developement of defects (e.g., cracks) shown in Figure 1-4. The characteristics of the defects (e.g., crack lengths) are unknown, but we assume that we are privy to a set of experimental measurements which serve to assess the state of the structure. Clearly, the defects may cause the deflection to reach unacceptably high values; a shim is therefore introduced so as to stiffen the structure and maintain the deflection at the desired levels. However, this intervention leads to an increase in both material and operational costs. Our goal is to find, given the uncertainties in the crack lengths, the shim dimensions which minimize the weight while honoring our deflection constraint.

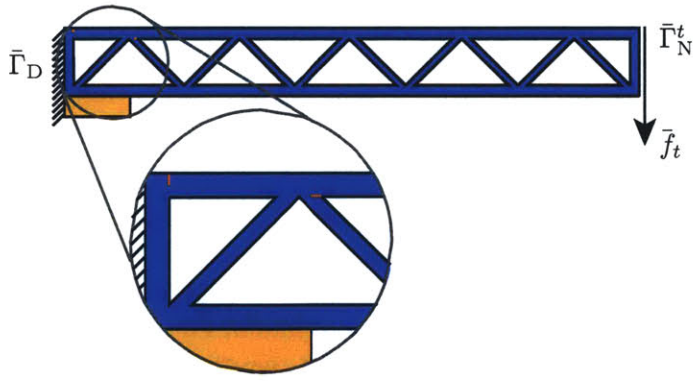


Figure 1-4: A "defective" microtruss structure. The insert highlights the defects (two cracks) and intervention (shim).

More precisely, we characterize our system with a multiparameter $\mu = (\mu_{\text{crack}}, \mu_{\text{shim}})$ where $\mu_{\text{crack}} = (\bar{L}_1, \bar{L}_2)$ and $\mu_{\text{shim}} = (\bar{L}_{\text{shim}}, \bar{t}_{\text{shim}})$. As shown in Figure 1-5, $\bar{L}_1$ and $\bar{L}_2$ are our "current guesses" for the relative lengths of the cracks on the upper frame and truss, respectively, while $\bar{t}_{\text{shim}}$

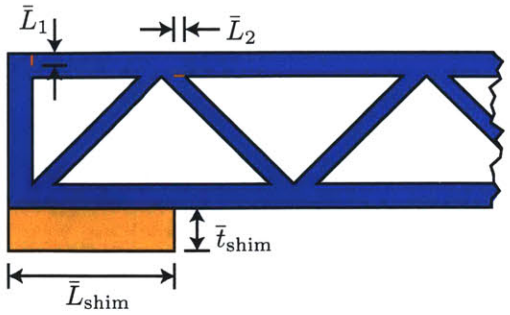


Figure 1-5: Parameters describing the defects, ($\bar{L}_1$ and $\bar{L}_2$), and the intervention, ($\bar{L}_{\text{shim}}$ and $\bar{t}_{\text{shim}}$).

and $\bar{L}_{\text{shim}}$ denote the thickness and length of the shim, respectively; we also denote by $\mu^*_{\text{crack}} = (\bar{L}^*_1, \bar{L}^*_2)$ the (real) *unknown* crack lengths. We further assume we are given M_{exp} intervals $[\delta^m_{LB}, \delta^m_{UB}]$, $m = 1, \ldots, M_{\text{exp}}$ representing experimental measurements of the deflection such that

$$\delta_{\text{ave}}(\mu^m_{\text{shim}}, \mu^*_{\text{crack}}) \in [\delta^m_{LB}, \delta^m_{UB}], \quad m = 1, \ldots, M_{\text{exp}} \; . \tag{1.35}$$

From these measurements, we then infer the existence of a set $\mathcal{L}^*$ such that $(\bar{L}^*_1, \bar{L}^*_2) \in \mathcal{L}^*$. We then wish to find the μ^*_{shim} which satisfies

$$\mu^*_{\text{shim}} = \arg \min_{\mu_{\text{shim}}} \mathcal{V}_{\text{shim}}(\mu_{\text{shim}}), \tag{1.36}$$

where $\mathcal{V}_{\text{shim}}(\mu_{\text{shim}}) = \bar{t}_{\text{shim}} \bar{L}_{\text{shim}}$ is simply the area of the shim, such that

$$\mu_{\text{shim}} \in \mathcal{D}^{\text{shim}} \tag{1.37}$$

$$\max_{\mu_{\text{crack}} \in \mathcal{L}^*} \delta_{\text{ave}}(\mu_{\text{shim}}, \mu_{\text{crack}}) \leq \delta_{\text{max}} \quad . \tag{1.38}$$

In words, we wish to find the shim dimensions which minimizes the area of the shim such that the maximum deflection (over all crack lengths consistent with the experiments) is less than the prescribed deflection limit δ_{max}.

1.2 Goals

Our first goal is the development of computational methods that permit *accurate*, *reliable*, and *rapid* evaluation of input-output relationships induced by partial differential equations *in the limit of many queries*. In particular, we seek to develop, especially for problems in elasticity, methods for (*i*) accurate approximation of the relevant outputs of interest; (*ii*) inexpensive and rigorous error bounds yielding upper and lower bounds for the error in the approximation; and (*iii*) a computational framework which allows rapid on-line calculation of the output approximation and associated error bounds.

Our second goal is the application of these computational methods to problems requiring repeated evaluation of these input-output relationships. In particular, we seek to use these computational methods to solve representative problems involving the design, optimization, and characterization of engineered systems.

1.3 Approach

1.3.1 Reduced-Basis Output Bounds

The difficulty is clear: evaluation of the output, $s(\mu)$, requires solution of the partial differential equation; the latter is computationally very expensive, too much so particularly in the limit of many queries. The "many queries" limit has certainly received considerable attention: from "fast loads" or multiple right-hand side notions (e.g., [10, 13]) to matrix perturbation theories (e.g., [1, 48]) to continuation methods (e.g., [3, 41]). Analytical "substructure" methods ([30]) which provide rapid on-line calculation of *exact* solutions to parametrized partial differential equations have also been developed; however, these techniques have limited applicability, relevant only to problems in which the parameters are the material properties, and are inapplicable even in the case of very simple geometric variations.

Our particular approach is based on the reduced–basis method, first introduced in the late 1970s for nonlinear structural analysis [4, 32], and subsequently developed more broadly in the 1980s and 1990s [7, 8, 14, 36, 37, 42]. 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 [7] 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.

We note that 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. Furthermore, theoretical and numerical results presented in Chapters 3 (for heat conduction), 5 (for linear elasticity), and 6 (for elastic stability) show that N may *indeed* be taken to be very small and that the computational economy is substantial. In addition, timing comparisons presented in Chapter 7 for the two-dimensional microtruss example show that a single on-line calculation of the output requires on the order of a few milliseconds, while conventional finite element calculation requires several seconds; the computational savings would be even larger in higher dimensions.

Furthermore, the *approximate* nature of reduced-basis solutions (as opposed to the *exact* solutions provided by [30]) do not pose a problem: our *a posteriori* error estimation procedures supply rigorous certificates of fidelity, thus providing the necessary accuracy assessment.

1.3.2 Real-Time (Reliable) Optimization

The numerical methods proposed are rather unique relative to more standard approaches to partial differential equations. Reduced–basis output bound methods are intended to render partial–differential-equation solutions truly useful: essentially real–time as regards operation count; "black-box" as regards reliability; and directly relevant as regards the (limited) input–output data required. But to be truly *useful*, these methods must directly enable solution of "real" optimization problems — rapidly, accurately, and reliably, even in the presence of uncertainty. Our work employs these reduced-basis output bounds in the context of "pre"-design — optimizing a system *at conception* with respect to prescribed objectives, constraints, and environmental conditions — and adaptive design — optimizing a system *in operation* subject to evolving system characteristics, dynamic

system requirements, and changing environmental conditions.

1.3.3 Architecture

Finally, to be truly *usable*, the entire methodology must reside within a special framework. This framework must permit the end-user to, *off-line*, (*i*) specify and define their problem in terms of high-level constructs; (*ii*) automatically and quickly generate the online simulation and optimization servers for their particular problem. Then, *on-line*, the user must be able to (*i*) specify the output and input values of interest; and to receive — quasi–instantaneously — the desired prediction and certificate of fidelity (error bound); and (*ii*) specify the objective, constraints, and relevant design variables; and to receive — real-time — the desired optimal system configuration.

1.4 Thesis Outline

In this thesis we focus on the development of reduced-basis output bound methods for problems in elasticity. In Chapter 2 we present an overview of reduced-basis methods, summarizing earlier work and focusing on the (new) critical ingredients. In Chapter 3 we describe, using the heat conduction problem for illustration, the reduced-basis approximation for coercive symmetric problems and "compliant" outputs; we present the associated *a posteriori* error estimation procedures in Chapter 4 . In Chapter 5 we develop the reduced-basis output bound method (approximation and *a posteriori* error estimation) to the linear elasticity problem and extend our methods to "noncompliant" outputs. In Chapter 6 we consider the nonlinear eigenvalue problem of elastic buckling; and in Chapter 7 we employ the reduced-basis methodology in the analysis, design (at conception and in operation) of our microtruss example. Finally, in Chapter 8 we conclude with some suggestions for future work, particularly in the area of *a posteriori* error estimation.

1.5 Thesis Contributions, Scope, and Limitations

In this thesis, we improve on earlier work on reduced-basis methods for linear elasticity in two ways: (*i*) we exploit the sparsity and symmetry of the elasticity tensor to substantially reduce both the off-line and on-line computational cost; and (*ii*) we extend the methodology to allow computation of more general outputs of interest such as average stresses and buckling loads.

Furthermore, we also introduce substantial advances to the general reduced-basis methodology. First, the challenges presented by the linear elasticity operator led us to achieve a better understanding of reduced-basis error estimation, and subsequently develop new techniques for constructing rigorous (Method I) bounds for the error in the approximation. Second, we also develop simple, inexpensive (Method II) error bounds for problems in which our rigorous error estimation methods are either inapplicable or too expensive computationally.

Finally, we apply our methods to design and optimization problems representative of applications requiring repeated and rapid evaluations of the outputs of interest. We illustrate how reduced-basis methods lend themselves naturally to existing solution methods (e.g., interior point methods for optimization), and how they allow the development of new methods (e.g. our assess-predict-optimize methodology) which would have been intractable with conventional methods.

We note that the goals presented in Section 1.3 are by no means trivial, and the variety of problems (i.e., partial differential equations) that must be addressed is extensive. Indeed, this work on linear elasticity is merely a small part of a much larger effort on developing reduced-basis output bound methods.

This thesis deals with methods that are generally applicable to linear coercive elliptic (second-order) partial differential equations with affine parameter dependence, and focuses on developing such methods for linear elasticity; in addition, this work also presents preliminary work on the (nonlinear) eigenvalue problem governing elastic stability, as well as some ideas for future work on the thermoelasticity and (noncoercive) Helmholtz problem. This thesis builds on earlier work on general coercive elliptic problems [43] and on linear elasticity [20].

However, reduced-basis methods have also been applied to parabolic problems [43], noncoercive problems [43], and problems which are (locally) non-affine in parameter [46]. These problems are not addressed in any great detail in this thesis, save for a short summary in Chapter 2 and a discussion of future work in Chapter 8.

Furthermore, in this thesis we assume that the mathematical model — particularly the parametrization of the partial differential equation — is exact. A discussion reduced-basis output bounds for approximately parametrized elliptic coercive partial differential operators may be found in [38]. Some preliminary ideas based on [38] for problems with approximately parametrized data or "loading" (as opposed to the partial differential operator) are presented in Chapter 8.

Next, the discussion in Chapter 7 on reduced-basis methods in the context of optimization problems is decidedly brief; a more detailed discussion may be found in [2, 34]. A related work [17] applies reduced-basis methods to problems in optimal control.

Finally, detailed discussions of the computational architecture in which the methodology resides may be found in [40].

Chapter 2

The Reduced Basis Approach: An Overview

2.1 Introduction

As earlier indicated, our goal is the development of computational methods that permit rapid and reliable evaluation of partial-differential-equation-induced input-output relationships in the limit of many queries. In Chapter 1 we present examples of these input-output relationships, as well as problems illustrating the "many-queries" context.

Our particular approach is based on the reduced-basis method, which 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 [39]. In this chapter, we provide an overview of the reduced-basis approach, introducing key ideas, surveying early work, highlighting more recent developments, and leaving more precise definitions and detailed development to later chapters. We focus particularly on the critical ingredients: (*i*) dimension reduction, effected by *global* approximation spaces; (*ii*) *a posteriori* error estimation, providing sharp and inexpensive upper and lower output bounds; and (*iii*) *off-line/on-line* computational decompositions, effecting (on-line) computational economies of several orders of magnitude. These three elements allow us, for the restricted but important class of "parameter-affine" problems, to compute output approximations rapidly, repeatedly, with certificates of fidelity.

2.2 Abstraction

Our model problem in Chapter 1 can be stated as: for any $\mu \in \mathcal{D}^\mu \subset \mathbb{R}^P$, find $s(\mu) \in \mathbb{R}$ given by

$$s(\mu) = \langle L(\mu), v \rangle, \tag{2.1}$$

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

$$\langle \mathcal{A}(\mu) u(\mu), v \rangle = \langle F(\mu), v \rangle \ , \quad \forall\ v \in Y. \tag{2.2}$$

Here, μ is a particular point in the parameter set, $\mathcal{D}^\mu$; Y is the infinite-dimensional space of admissible functions; $\mathcal{A}(\mu)$ is a symmetric, continuous and coercive distributional operator, and

the loading and output functionals, $F(\mu)$ and $L(\mu)$, respectively, are bounded linear forms.[1] In the language of Chapter 1, (2.2) is our parametrized partial differential equation (in weak form), μ is our input parameter, $u(\mu)$ is our field variable, and $s(\mu)$ is our output.

In actual practice, Y is replaced by an appropriate "truth" finite element approximation space $Y_{\mathcal{N}}$ of dimension $\mathcal{N}$ defined on a suitably fine truth mesh. We then approximate $u(\mu)$ and $s(\mu)$ by $u_{\mathcal{N}}(\mu)$ and $s_{\mathcal{N}}(\mu)$, respectively, and assume that $Y_{\mathcal{N}}$ is sufficiently rich such that $u_{\mathcal{N}}(\mu)$ and $s_{\mathcal{N}}(\mu)$ are indistinguishable from $u(\mu)$ and $s(\mu)$.

The difficulty is clear: evaluation of the output, $s(\mu)$, requires solution of the partial differential equation, (2.2); the latter is computationally very expensive, too much so particularly in the limit of many queries.

2.3 Dimension Reduction

In this section we give a brief overview of the essential ideas upon which the reduced basis approximation is based. For simplicity of exposition, we assume here and in Sections 2.4-2.5 that F and L do not depend on the parameter; in addition, we assume a "compliant" output:

$$\langle L, v\rangle \equiv \langle F, v\rangle \ , \quad \forall\, v \in Y. \tag{2.3}$$

2.3.1 Critical Observation

The difficulty in evaluating the output, $s(\mu)$, stems from the necessity of calculating the field variable, $u(\mu)$, which is a member of the infinite-dimensional solution space, Y, associated with the partial differential equation. However, we can intuit that the possible values of $u(\mu)$ do not "cover" the entire space, Y; if we imagine Y to be reduced to a three-dimensional space, then u — as a function of μ — can be conceived as lying on a curve or surface; this is depicted in Figure 2-1(a). For example, in our model problem of Chapter 1, we expect that the the displacement field which satisfies the governing equations ((1.10), (1.14)-(1.16)) does not vary randomly with the parameter μ (defined in Section 1.1.1), but in fact varies in a smooth fashion.

(a) (b)

Figure 2-1: (a) Low-dimensional manifold in which the field variable resides; and (b) approximation of the solution at μ^{new} by a linear combination of pre-computed solutions $u(\mu_i)$.

In other words, the field variable is not some arbitrary member of the high-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 [39]. This observation is fundamental to our approach, and is the basis for our approximation.

[1] See Chapters 3 and 5 for more precise definitions.

2.3.2 Reduced-Basis Approximation

By the preceding arguments, we see that to approximate $u(\mu)$, and hence $s(\mu)$, we need *not* represent *every possible function* in Y; instead, we need only approximate those functions in the low-dimensional manifold "spanned" by $u(\mu)$. We could, therefore, simply calculate the solution u at several points on the manifold corresponding to different values of μ; then, for any new parameter value, μ^{new}, we could "interpolate between the points," that is, approximate $u(\mu^{\text{new}})$ by some linear combination of the known solutions. This notion is illustrated in Figure 2-1(b).

More precisely, we introduce a sample in parameter space,

$$S_N^\mu = \{\mu_1, \ldots, \mu_N\} \tag{2.4}$$

where $\mu_n \in \mathcal{D}^\mu, n = 1, \ldots, N$. We then define our Lagrangian ([37]) reduced-basis approximation space as

$$W_N = \text{span}\{\zeta_n \equiv u(\mu_n), n = 1, \ldots, N\}, \tag{2.5}$$

where $u(\mu_n) \in Y$ is the solution to (2.2) for $\mu = \mu_n$. Our reduced-basis approximation is then: for any $\mu \in \mathcal{D}$, find

$$s_N(\mu) = \langle L, u_N(\mu) \rangle \tag{2.6}$$

where $u_N(\mu)$ is the Galerkin projection of $u(\mu)$ onto W_N.

$$\langle \mathcal{A}(\mu) u_N(\mu), v \rangle = \langle F, v \rangle, \quad \forall\, v \in W_N. \tag{2.7}$$

In other words, we express $u(\mu)$ as a linear combination of our basis functions,

$$u_N(\mu) = \sum_{j=1}^{N} u_{Nj}(\mu)\, \zeta_j \ ; \tag{2.8}$$

then, choosing the basis functions as test functions (i.e., setting $v = \zeta_n,\ n = 1, \ldots, N$, in (2.7)), we obtain

$$\underline{A}_N(\mu)\, \underline{u}_N(\mu) = \underline{F}_N, \tag{2.9}$$

where

$$(A_N)_{i\,j}(\mu) = \langle \mathcal{A}(\mu)\zeta_j, \zeta_i \rangle\ , \quad i, j = 1, \ldots, N, \tag{2.10}$$

$$(F_N)_i = \langle F, \zeta_i \rangle\ , i = 1, \ldots, N. \tag{2.11}$$

Our output approximation is then given by

$$s_N(\mu) = \underline{u}_N(\mu)^T \underline{L}_N, \tag{2.12}$$

where $\underline{L}_N = \underline{F}_N$ from (2.3).

2.3.3 *A Priori* Convergence Theory

We consider here the rate at which $u_N(\mu)$ and $s_N(\mu)$ converges to $u(\mu)$ and $s(\mu)$, respectively. To begin, it is standard to demonstrate the optimality of $u_N(\mu)$ in the sense that

$$||u(\mu) - u_N(\mu)||_Y \leq \sqrt{\frac{\gamma_{\mathcal{A}}^0}{\beta_{\mathcal{A}}^0}} \inf_{w_N \in W_N} ||u(\mu) - w_N||_Y \ . \tag{2.13}$$

where $||\cdot||_Y$ is the norm associated with Y, and $\gamma_{\mathcal{A}}^0$ and $\beta_{\mathcal{A}}^0$ are μ-independent constants associated with the operator $\mathcal{A}$. Furthermore, for our compliance output, it can be shown that

$$s(\mu) - s_N(\mu) = \gamma_{\mathcal{A}}(\mu)||u(\mu) - u_N(\mu)||_Y^2 \ , \tag{2.14}$$

It follows that $s_N(\mu)$ converges to $s(\mu)$ as the square of the error in $u_N(\mu)$.

At this point, we have not yet dealt with the question of how the sample points, μ_n, should be chosen. In particular, one may ask, is there an "optimal" choice for S_N? For certain simple problems, it can be shown [28] that, using a logarithmic point distribution, the error in the reduced-basis approximation is exponentially decreasing with N for N greater than some critical value N_{crit}. More generally, numerical tests (see Chapter 3) show that the logarithmic distribution performs considerably better than other more obvious candidates, in particular for large ranges of the parameter. A more detailed discussion of convergence and point distribution is presented in Chapter 3.

These theoretical considerations suggest that N may, indeed, be chosen very small. However, we note from (2.11) and (2.12) that $\underline{A}_N(\mu)$ (and therefore $\underline{u}_N(\mu)$ and $s_N(\mu)$) depend on the basis functions, ζ_n, and are therefore potentially computationally very expensive. We might then ask: can we calculate $s_N(\mu)$ *inexpensively*? We address this question in Section 2.4.

2.4 Off-Line/On-Line Computational Procedure

We recall that in this section, we assume that F and L are independent of the parameter, and that the output is compliant, $L = F$. Nevertheless, the development here can be easily extended (see Chapter 5) to the case of μ-dependent functionals L and F, and to noncompliance, $L \neq F$.

The output approximation, $s_N(\mu)$, will be inexpensive to evaluate *if* we make certain assumptions on the parametric dependence of $\mathcal{A}$; these assumptions will allow us to develop off-line/on-line computational procedures. (See [7] for an earlier application of this strategy within the reduced-basis context.) In particular, we shall suppose that

$$\langle \mathcal{A}(\mu)w, v\rangle = \sum_{q=1}^{Q_{\mathcal{A}}} \Theta^q(\mu)\, \langle \mathcal{A}^q w, v\rangle \ , \tag{2.15}$$

for some finite (preferably small) integer $Q_{\mathcal{A}}$. It follows that

$$\underline{A}_N(\mu) = \sum_{q=1}^{Q_{\mathcal{A}}} \Theta^q(\mu)\, \underline{A}_N^q(\mu) \ , \tag{2.16}$$

where

$$A_{N\,i,j}^q = \langle \mathcal{A}^q \zeta_j, \zeta_i\rangle, \quad 1 \leq i,j \leq N,\ 1 \leq q \leq Q \ . \tag{2.17}$$

Therefore, in the *off-line* stage, we compute the $u(\mu_n)$ and form the $\underline{A}_N^q$; this requires N (expensive) "$\mathcal{A}$" 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 $\underline{A}(\mu)$ from (2.16), then solve (2.9) for $\underline{u}_N(\mu)$, and finally evaluate $s_N(\mu) = \underline{u}_N(\mu)^T \underline{L}_N$; this requires $O(QN^2) + O(\frac{2}{3}N^3)$ 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 (we predict) N is small, thanks to the good convergence properties of W_N; and second, because (2.9) can be very rapidly assembled and inverted, thanks to the off-line/on-line decomposition.

These off-line/on-line computational procedures clearly exploit the dimension reduction of Section 2.3. However, apart from the discussion on the *a priori* convergence properties of our approximations, we have not presented any guidelines as to what value N must be taken. Furthermore, once $s_N(\mu)$ has been calculated, how does one know whether the approximation is accurate? We address these issues in Section 2.5.

2.5 A Posteriori Error Estimation

From Section 2.4 we know that, in theory, we can obtain $s_N(\mu)$ very inexpensively: the on–line computational effort scales as $O(\frac{2}{3}N^3) + O(QN^2)$; and N can, *in theory*, be chosen quite small. However, *in practice*, we do not know *how* small N can (nor how large it must) be chosen: this will depend on the desired accuracy, the selected output(s) of interest, and the particular problem in question. 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. We thus need *a posteriori* error estimators for $s_N(\mu)$. Surprisingly, 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 — *a posteriori* error estimation has received relatively little attention within the reduced–basis framework [32].

Efficiency and reliability of approximation are particularly important in the decision contexts in which reduced-basis methods typically serve. In many cases, we may wish to choose N *minimally* such that

$$|s(\mu) - s_N(\mu)| \leq \varepsilon_{\max}; \tag{2.18}$$

we therefore introduce an *error estimate:*

$$\Delta_N(\mu) \approx |s(\mu) - s_N(\mu)| \ , \tag{2.19}$$

which is *reliable, sharp,* and *inexpensive* to compute. Furthermore, we define the *effectivity* of our error estimate as

$$\eta_N(\mu) \equiv \frac{\Delta_N(\mu)}{|s(\mu) - s_N(\mu)|}, \tag{2.20}$$

and require that

$$1 \ \leq \ \eta_N(\mu) \ \leq \ \rho \ , \tag{2.21}$$

where $\rho \approx 1$. The left-hand inequality — which we denote the *lower effectivity inequality* — guarantees that $\Delta_N(\mu)$ is a rigorous upper bound for the error in the output of interest, while the right-hand inequality — which we denote the *upper effectivity inequality* — signifies that $\Delta_N(\mu)$ must be a *sharp* bound for the true error. The former relates to reliability; while the latter leads to efficiency. Our effectivity requirement, (2.21), then allows us to optimally select — on-line — N

such that

$$|s(\mu) - s_N(\mu)| \;\leq\; \Delta_N(\mu) \text{ “=” } \varepsilon_{\max}. \tag{2.22}$$

In addition to the constraint on the output approximation error (2.18), we may also prescribe contstraints on the output itself; that is, we may wish to ensure that

$$s_{\min} \leq s(\mu) \leq s_{\max}. \tag{2.23}$$

We therefore require not only an error bound, $\Delta_N(\mu)$, but also *lower* and *upper output bounds:*

$$s_N^-(\mu) \leq s(\mu) \leq s_N^+(\mu); \tag{2.24}$$

we likewise require that $s_N^-(\mu)$ and $s_N^+(\mu)$ be reliable, sharp, and inexpensive to compute. We then require that

$$s_{\min} \;\leq\; s_N^-(\mu)\,, \qquad s_N^+(\mu) \;\leq\; s_{\max}, \tag{2.25}$$

thereby ensuring that the constraint, (2.23), is met. In many applications, satisfaction of constraints such as (2.23) is critical to performance and, more importantly, safety; for example, the constraints on temperature, deflection, stress, and buckling in (1.30)-(1.33) must clearly be satisfied to ensure safe operation. Output bounds are therefore of great importance for the reduced-basis approximations to be of any practical use.

In this work we present rigorous (or *Method I*) as well as asymptotic (or *Method II*) *a posteriori* error estimation procedures; the former satisfy (2.21) and (2.24) for *all* N, while the latter only as $N \to \infty$. We briefly describe these error estimators in Sections 2.5.1 and 2.5.2. A more detailed discussion may be found in Chapters 4 and 5.

2.5.1 Method I

Method I error estimators are based on relaxations of the error-residual equation, and are derived from *bound conditioners* [21, 37, 47] — in essence, operator preconditioners $C(\mu)$ that satisfy (i) an additional spectral “bound” requirement:

$$1 \leq \frac{\langle \mathcal{A}(\mu)v, v\rangle}{\langle C(\mu)v, v\rangle} \leq \rho\,; \tag{2.26}$$

and (ii) a “computational invertibility” hypothesis:

$$C^{-1}(\mu) = \sum_{i\in\mathcal{I}(\mu)} \alpha_i(\mu) C_i^{-1} \tag{2.27}$$

so as to admit the reduced-basis off-line/on-line computational stratagem; here, $\mathcal{I}(\mu) \subset \{1, \ldots, I\}$ is a parameter-dependent set of indices, I is a finite (preferably small) integer, and the C_i are parameter-independent symmetric, coercive operators. In the compliance case ($L = F$), the error estimator is defined as

$$\Delta_N(\mu) = \langle R(\mu), C^{-1}(\mu) R(\mu)\rangle\,, \tag{2.28}$$

where the residual $R(\mu)$ is defined as

$$\begin{aligned} \langle R(\mu), v\rangle &= \langle F, v\rangle - \langle \mathcal{A}(\mu) u_N(\mu), v\rangle, && (2.29)\\ &= \langle \mathcal{A}(\mu)\left(u(\mu) - u_N(\mu)\right), v\rangle, \quad \forall\, v \in Y\ , && (2.30) \end{aligned}$$

the output bounds are then given by

$$\begin{aligned} s_N^-(\mu) &= s_N(\mu) && (2.31)\\ s_N^+(\mu) &= s_N(\mu) + \Delta_N(\mu). && (2.32) \end{aligned}$$

A more detailed discussion of the bounding properties of our Method I error estimators and the corresponding off-line/on-line computational procedure, as well as "recipes" for constructing the bound conditioner $\mathcal{C}(\mu)$, may be found in Chapter 4. The extension of the bound conditioner framework to the case of noncompliant outputs is addressed in Chapter 5. Numerical results are also presented in Chapters 4 and 5 for simple problems in heat conduction and linear elasticity, respectively.

The essential advantage of Method I error estimators is the guarantee of rigorous bounds. However, in some cases either the associated computational expense is much too high, or there is no self-evident good choice of bound conditioners. In Section 2.5.2 we briefly describe Method II error estimators which eliminate these problems, albeit at the loss of absolute certainty.

2.5.2 Method II

In cases for which there are no viable rigorous error estimation procedures, we may employ simple error estimates which replace the true output, $s(\mu)$, in (2.19) with a finer-approximation surrogate, $s_M(\mu)$. We thus set $M > N$ and compute

$$\Delta_{N,M}(\mu) = \frac{1}{\tau}\left(s_M(\mu) - s_N(\mu)\right)\ , \qquad (2.33)$$

for some $\tau \in (0,1)$. Here $s_M(\mu)$ is a reduced basis approximation to $s(\mu)$ based on a "richer" approximation space $W_M \supset W_N$. Since the Method II error bound is based entirely on evaluation of the output, the off-line/on-line procedure of Section 2.4 can be directly adapted; the on-line expense to calculate $\Delta_{N,M}(\mu)$ is therefore small. However, $\Delta_{N,M}(\mu)$ is no longer a strict upper bound for the true error: we can show [39] that $\Delta_{N,M}(\mu) > |s(\mu) - s_N(\mu)|$ only as $N \to \infty$. The usefulness of Method II error estimators — in spite of their asymptotic nature — is largely due to the rapid convergence of the reduced basis approximation.

In Chapter 4 we consider Method II error estimators in greater detail: the formulation and properties of the error estimators are discussed, and numerical effectivity results for the heat conduction problem are presented. Numerical results for the linear elasticity and elastic stability problems are also presented in Chapters 5 and 6, respectively. Method II estimators are also used for numerical tests in the specific application (microtruss optimization) problems of Chapter 7.

2.6 Extensions

2.6.1 Noncompliant Outputs

In Sections 2.3-2.5 we provide a brief overview of the reduced-basis method and associated error estimation procedure for the case of compliant outputs, $L = F$. In the case of more general linear

bounded output functionals ($L \neq F$), we introduce an adjoint or dual problem: for any $\mu \in \mathcal{D}^\mu$, find $\psi(\mu) \in Y$ such that

$$\langle A(\mu)v, \psi(\mu)\rangle = -\langle L, v\rangle \ , \qquad \forall\ v \in Y \ . \tag{2.34}$$

We then choose a sample set in parameter space,

$$S^\mu_{N/2} = \{\mu_1, \ldots, \mu_{N/2}\} \ , \tag{2.35}$$

where $\mu_i \in \mathcal{D}^\mu$, $i = 1, \ldots, N$ (N even) and define either an "integrated" reduced-basis approximation space

$$W_N = \text{span}\,\{u(\mu_n), \psi(\mu_n), n = 1, \ldots, N/2\} \tag{2.36}$$

for which the output approximation is given by

$$s_N(\mu) = \langle L, u_N(\mu)\rangle \tag{2.37}$$

where $u_N(\mu) \in W_N$ is the Galerkin projection of $u(\mu)$ onto W_N,

$$\langle A(\mu)u_N(\mu), v\rangle = \langle F, v\rangle \ , \qquad \forall\ v \in W_N \ ; \tag{2.38}$$

or a "nonintegrated" reduced-basis approximation space

$$W_N^{\text{pr}} = \text{span}\,\{\zeta_n^{\text{pr}} = u(\mu_n), n = 1, \ldots, N/2\} \tag{2.39}$$

$$W_N^{\text{du}} = \text{span}\left\{\zeta_n^{\text{du}} = \psi(\mu_n), n = 1, \ldots, N/2\right\} \tag{2.40}$$

for which the output approximation is given by

$$s_N(\mu) = \langle L(\mu), u_N(\mu)\rangle - (\langle F(\mu), \psi_N(\mu)\rangle - \langle \mathcal{A}(\mu)u_N(\mu), \psi_N(\mu)\rangle)\,, \tag{2.41}$$

where $u_N(\mu) \in W_N^{\text{pr}}$ and $\psi_N(\mu) \in W_N^{\text{du}}$ are the Galerkin projections of $u(\mu)$ and $\psi(\mu)$ onto W_N^{pr} and W_N^{du}, respectively, i.e.,

$$\langle \mathcal{A}(\mu)u_N(\mu), v\rangle = \langle F(\mu), v\rangle, \quad \forall\ v \in W_N^{\text{pr}} \ , \tag{2.42}$$

$$\langle \mathcal{A}(\mu)v, \psi_N(\mu)\rangle = -\langle L(\mu), v\rangle, \quad \forall\ v \in W_N^{\text{du}} \ . \tag{2.43}$$

As in the compliance case, the approximations $u_N(\mu)$ and $\psi_N(\mu)$ are optimal, and the "square" effect in the convergence rate of the output is recovered. Furthermore, both the integrated and nonintegrated approaches admit an off-line/on-line decomposition similar to that described in Section 2.4 for the compliant problem; as before, the on-line complexity and storage are independent of the dimension of the very fine ("truth") finite element approximation. The formulation and theory for noncompliant problems is discussed in greater detail in Chapter 5.

2.6.2 Nonsymmetric Operators

It is also possible to relax the assumption of symmetry in the operator $\mathcal{A}$, permitting treatment of a wider class of problems [43, 39] — a representative example is the convection-diffusion equation, in which the presence of the convective term renders the operator nonsymmetric. As in Section 2.3.2, we choose a sample set $S_N = \{\mu_1, \ldots, \mu_N\}$, and define the reduced-basis approximation space $W_N = \text{span}\{\zeta_n \equiv u(\mu_n), n = 1, \ldots, N\}$. The reduced-basis approximation is then $s_N(\mu) = \langle L, u_N(\mu)\rangle$

where $u_N(\mu) \in W_N$ satisfies $\langle \mathcal{A}(\mu)u_N(\mu), v\rangle = \langle F, v\rangle$, $\forall\ v \in W_N$, and $L = F$. Here, $u_N(\mu)$ is optimal in the sense that

$$||u(\mu) - u_N(\mu)||_Y \leq \left(1 + \frac{\gamma(\mu)}{\alpha(\mu)}\right) \inf_{w \in W_N} ||u(\mu) - w||_Y \ . \tag{2.44}$$

The off-line/on-line decomposition for the calculation of the output approximation is the same as for the symmetric case, with the exception that $\underline{A}_N$ and the $\underline{A}_N^q$ are no longer symmetric.

The *a posteriori* error estimation framework may also be extended to nonsymmetric problems. In the Method I approach, the bound conditioner $\mathcal{C}(\mu)$ must now satisfy

$$1 \leq \frac{\langle \mathcal{A}^S(\mu)v, v\rangle}{\langle \mathcal{C}(\mu)v, v\rangle} \leq \rho \tag{2.45}$$

where $\mathcal{A}^S(\mu) = \frac{1}{2}(\mathcal{A}(\mu) + \mathcal{A}^T(\mu))$ is the symmetric part of $\mathcal{A}(\mu)$; the procedure remains the same as for symmetric problems. In the Method II approach, since $s_N(\mu)$ is no longer a strict lower bound for $s(\mu)$, the error estimators and output bounds for noncompliant outputs (see Chapter 5) must be used.

2.6.3 Noncoercive Problems

There are many important problems for which the coercivity of $\mathcal{A}$ is lost — a representative example is the Helmholtz, or reduced–wave, equation. For noncoercive problems, well–posedness is now ensured only by the inf–sup condition: there exists a positive $\beta_{\mathcal{A}}^0$, $\beta_{\mathcal{A}}(\mu)$, such that

$$0 \ < \ \beta_{\mathcal{A}}^0 \ \leq \ \beta_{\mathcal{A}}(\mu) \ = \ \inf_{w \in Y} \sup_{v \in Y} \frac{\langle \mathcal{A}(\mu)w, v\rangle}{||w||_Y \ ||v||_Y}, \quad \forall\ \mu \in \mathcal{D}^\mu. \tag{2.46}$$

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. 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. However, it is possible to resolve both of these difficulties by considering projections other than standard Galerkin, and "enriched" approximation spaces [26, 43]. The second numerical difficulty is estimation of the inf–sup parameter, important for certain classes of Method I *a posteriori* error estimation techniques. In particular, $\beta_{\mathcal{A}}(\mu)$ can not typically be deduced analytically, and thus must be approximated. However, some ideas presented in Chapters 4 and 5 may be used to obtain the necessary approximation (more specifically, a lower bound) to the inf–sup parameter; this is discussed briefly in Chapter 8.

Method II techniques are also appropriate [39]: in particular, Method II approaches do not require accurate estimation of the inf-sup parameter, and thus one need be concerned only with stability in designing the reduced–basis spaces [39].

2.6.4 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 the 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.

We first consider the extension of our approach to (weakly nonlinear) symmetric positive definite eigenvalue problems of the form

$$\langle \mathcal{A}(\mu)\xi(\mu), v\rangle = \lambda(\mu)\langle \mathcal{M}(\mu)\xi(\mu), v\rangle \ , \forall\ v \in Y \ , \tag{2.47}$$

where $\mathcal{A}$ and $\mathcal{M}$ are symmetric, continuous, and coercive; these assumptions on the operators imply that the eigenvalues λ are real and positive. We suppose that the output of interest is the minimum eigenvalue $s(\mu) = \lambda_1(\mu)$.

One approach [23, 39, 43] utilizes a reduced-basis predictor and "Method I" error estimator similar to that for compliant symmetric problems except that (i) the reduced-basis space includes eigenfunctions associated with the first (smallest) and second eigenvalues $\lambda_1(\mu)$ and $\lambda_2(\mu)$, respectively, and (ii) the error estimator is only asymptotically valid. However, in practice, there is very little uncertainty in the asymptotic bounds such that the output bounds are valid even for $N = 1$. Furthermore, the effectivities are shown to be very good both in theory [23] and in numerical tests [39, 43].

A second approach [39] makes use of Method II error estimators, and no longer requires an estimate for the second eigenvalue. The reduced-basis space therefore includes eigenfunctions associated with only the smallest eigenvalue. The resulting bounds are asympotic, and the effectivity can be proven [39] to be bounded from above and in fact approach a constant as $N \to \infty$.

In Chapter 6, we consider a (strongly nonlinear) eigenvalue problem (in weak form)

$$\langle \mathcal{A}(\mu)\xi(\mu), v\rangle = \lambda(\mu)\langle \mathcal{B}(\mu; u(\mu))\xi(\mu), v\rangle \ , \forall\ v \in Y \ , \tag{2.48}$$

where $u(\mu)$ satisfies

$$\langle \mathcal{A}(\mu)u(\mu), v\rangle = \langle F, v\rangle \ , \forall\ v \in Y \ ; \tag{2.49}$$

(2.48)-(2.49) represent the equations governing the stability of elastic structures. Here, $\mathcal{A}(\mu)$ is symmetric, continuous, and coercive, while $\mathcal{B}(\mu; v)$ is linear in v for $v \in Y$, and is symmetric, continuous, positive semi-definite. The output of interest is the smallest eigenvalue $s(\mu) = \lambda_1(\mu)$ which corresponds to the critical buckling load; we formulate in Chapter 6 the reduced-basis approximation and Method II error estimation procedures for computing the output bounds.

2.6.5 Parabolic Problems

It is also possible to treat parabolic partial differential equations of the form $\langle \mathcal{M}(\mu)u_t(\mu), v\rangle = \langle \mathcal{A}(\mu)u(\mu), v\rangle$; 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 non-compliant nonsymmetric problems, except that now the time variable is included as an additional parameter. Thus, as in certain other time-domain model-order-reduction methods [5, 45], the basis functions are "snapshots" of the solution at selected time instants; however, in [43] an *ensemble* of such series is constructed 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 [43].

For the temporal discretization method, many possible choices are available. The most appropriate method — although not the only choice — is the discontinuous Galerkin method [23]. 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 Y.

2.6.6 Locally Nonaffine Problems

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 (2.2) for reasonably small $Q_{\mathcal{A}}$, 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 $\langle \mathcal{A}(\mu)\zeta_j, \zeta_i \rangle$. Unfortunately, the operation count (respectively, storage) for the on-line stage will now scale as $O(N^2 \dim(Y))$ (respectively, $O(N \dim(Y))$, where $\dim(Y)$ 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.

In [46], an approach is presented which addresses these difficulties and is slightly less general but potentially much more efficient. In particular, it is noted that in many cases — for example, boundary geometry modification — the nonaffine parametric dependence can be restricted to a small subdomain of Ω, Ω_{II}. The operator $\mathcal{A}$ can then be expressed as an affine/nonaffine sum,

$$\langle \mathcal{A}(\mu)w, v \rangle = \langle \mathcal{A}_I(\mu)w, v \rangle + \langle \mathcal{A}_{II}(\mu)w, v \rangle \ . \tag{2.50}$$

Here $\mathcal{A}_I$, defined over Ω_I — the majority of the domain — is affinely dependent on μ; and $\mathcal{A}_{II}$, 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 $\mathcal{A}_I$ and $\mathcal{A}_{II}$ respectively; that the stiffness matrix associated with $\mathcal{A}_I$ admits the usual on-line/off-line decomposition described in Section 2.4; and that the stiffness matrix associated with $\mathcal{A}_{II}$ 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 $O(N^2 \dim(Y|_{\Omega_{II}}))$, where $\dim(Y|_{\Omega_{II}})$ refers (in practice) to the number of finite-element nodes located within Ω_{II} — often extremely small. The method is therefore (almost) independent of $\dim(Y)$, though clearly the on-line code will be more complicated than in the purely affine case.

As regards *a posteriori* error estimation (see [46]), the nonaffine dependence of $\mathcal{A}$ (even locally) precludes the precomputation and linear superposition strategy required by Method I (unless domain decomposition concepts are exploited [22]); however, Method II directly extends to the locally nonaffine case.

Chapter 3

Reduced-Basis Output Approximation: A Heat Conduction Example

3.1 Introduction

In Chapter 2 we observed that the field variable is not some arbitrary member of the infinite-dimensional solution space, but rather resides on a much lower-dimensional manifold induced by the parametric dependence; this, we noted, is the essence of the reduced-basis approximation method. In this chapter, we present a more detailed discussion of the reduced-basis output approximation method for linear, coercive problems, illustrated in the context of steady-state heat conduction. We focus particularly on the *global* approximation spaces, *a priori* convergence theory, and the assumption of affine parameter dependence. Numerical results for several simple problems illustrate the rapid convergence of the reduced-basis approximation to the true output, as predicted by the *a priori* theory. For simplicity of exposition, we consider only the compliance case in which the inhomogeneity (loading) is the same as the output functional; noncompliant outputs are addressed in Chapter 5.

We begin by stating the most general problem (and all necessary hypotheses) to which the techniqes we develop will apply.

3.2 Abstraction

We consider a suitably regular (smooth) domain $\Omega \subset \mathbb{R}^d, d = 1, 2$, or 3, and associated function space $Y \subset (H^1(\Omega))^p$, where

$$H^1(\Omega) \equiv \left\{ v \mid v \in L^2(\Omega), \nabla v \in (L^2(\Omega))^d \right\}, \tag{3.1}$$

and

$$L^2(\Omega) \equiv \left\{ v \mid \int_\Omega v^2 < \infty \right\}. \tag{3.2}$$

The inner product and norm associated with Y are given by $(\cdot, \cdot)_Y$ and $||\cdot||_Y = (\cdot, \cdot)_Y^{1/2}$, respectively. The corresponding dual space of Y, Y' is then defined as the set of all functionals F such that the

dual norm of F, defined as,

$$||F||_{Y'} \equiv \sup_{v \in Y} \frac{\langle F, v \rangle}{||v||_Y}, \tag{3.3}$$

is bounded; here, $\langle \cdot, \cdot \rangle \equiv {}_{Y'}\langle \cdot, \cdot \rangle_Y$ is the associated duality pairing. As in the previous chapters, we also define a parameter set $\mathcal{D}^\mu \subset \mathbb{R}^P$, a particular point in which will be denoted μ. Note that Ω is a *reference domain*[1] and hence does *not* depend on the parameter.

We now introduce a distributional (second-order partial differential) operator $\mathcal{A}(\mu) : Y \to Y'$; we assume that $\mathcal{A}(\mu)$ is symmetric,

$$\langle \mathcal{A}(\mu) w, v \rangle = \langle \mathcal{A}(\mu) v, w \rangle, \quad \forall\, w, v \in Y,\ \forall\, \mu \in \mathcal{D}^\mu, \tag{3.4}$$

continuous,

$$\langle \mathcal{A}(\mu) w, v \rangle \le \gamma_{\mathcal{A}}(\mu) ||w||_Y ||v||_Y \le \gamma_{\mathcal{A}}^0 ||w||_Y ||v||_Y, \quad \forall\, w, v \in Y,\ \forall\, \mu \in \mathcal{D}^\mu, \tag{3.5}$$

and coercive,

$$0 < \beta_{\mathcal{A}}^0 \le \beta_{\mathcal{A}}(\mu) = \inf_{w \in Y} \frac{\langle \mathcal{A}(\mu) w, w \rangle}{||w||_Y^2}, \quad \forall\, w \in Y,\ \forall\, \mu \in \mathcal{D}^\mu. \tag{3.6}$$

We also introduce the bounded linear forms $F(\mu) \in Y'$ and $L(\mu) \in Y'$, and assume that they are "compliant,"

$$\langle L(\mu), v \rangle = \langle F(\mu), v \rangle, \quad \forall\, v \in Y; \tag{3.7}$$

noncompliant outputs are considered in Chapter 5.

We shall now make certain assumptions on the parametric dependence of $\mathcal{A}$. In particular, we shall suppose that, for some finite (preferably small) integer $Q_{\mathcal{A}}$, $\mathcal{A}$ may be expressed as

$$\mathcal{A}(\mu) = \sum_{q=1}^{Q_{\mathcal{A}}} \Theta^q(\mu)\, \mathcal{A}^q\ , \quad \forall\, \mu \in \mathcal{D}^\mu\ , \tag{3.8}$$

where $\Theta^q : \mathcal{D}^\mu \to \mathbb{R}$ and $\mathcal{A}^q : Y \to Y'$. As indicated in Section 2.4, this assumption of "separability" or affine parameter dependence is crucial to computational efficiency.

We shall also suppose that, for some finite (again, preferably small) integer Q_F, $F(\mu)$ may be expressed as

$$F(\mu) = \sum_{q=1}^{Q_F} \varphi_F^q(\mu)\, F^q\ , \quad \forall\, \mu \in \mathcal{D}^\mu\ , \tag{3.9}$$

where the $\varphi_F^q : \mathcal{D}^\mu \to \mathbb{R}$ and $F^q \in Y'$. However, for simplicity of exposition we shall assume in Sections 3.4 and 3.5 that $Q_F = 1$ and $\varphi_F^1(\mu) \equiv 1$.

For simplicity of exposition, we shall (at present) assume that F — and therefore L — is independent of μ; affine dependence of the linear forms is readily admitted, however, and is addressed in Chapter 5.

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

$$s(\mu) = \langle L, u(\mu) \rangle, \tag{3.10}$$

[1] The reference domain formulation is discussed in greater detail in Section 3.4.

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

$$\langle \mathcal{A}(\mu)u(\mu), v\rangle = \langle F, v\rangle, \quad \forall\, v \in Y. \tag{3.11}$$

In the language of Chapter 1, (3.11) is our partial differential equation (in weak form), μ is our parameter, $u(\mu)$ is our field variable, and $s(\mu)$ is our output.

In actual practice, Y is replaced by an appropriate "truth" finite element approximation space $Y_{\mathcal{N}}$ of dimension $\mathcal{N}$ defined on a suitably fine truth mesh. We then approximate $u(\mu)$ and $s(\mu)$ by $u_{\mathcal{N}}(\mu)$ and $s_{\mathcal{N}}(\mu)$, respectively, and assume that $Y_{\mathcal{N}}$ is sufficiently rich such that $u_{\mathcal{N}}(\mu)$ and $s_{\mathcal{N}}(\mu)$ are indistinguishable from $u(\mu)$ and $s(\mu)$.

3.3 Formulation of the Heat Conduction Problem

In this section we present the strong form of the equations governing heat conduction, from which we derive the weak statement; we then reformulate the problem in terms of a reference (parameter-independent) domain, thus recovering the abstract formulation of Section 3.2. In this and the following sections, repeated indices imply summation, and, unless otherwise indicated, indices take on the values 1 through d, where d is the dimensionality of the problem. Furthermore, in this section we use a bar to signify a general dependence on the parameter μ (e.g., $\bar{\Omega} \equiv \bar{\Omega}(\mu)$, or $\bar{t} \equiv \mu$) particularly when formulating the problem in a "non-reference" domain.

3.3.1 Governing Equations

We now present the governing equations and derive the weak statement for the case of a homogeneous body; we merely state the weak statement for the more general case of an inhomogeneous body.

Strong Form

We consider the transfer of heat in a homogeneous body $\bar{\Omega} \subset \mathbb{R}^d$ with boundary $\bar{\Gamma}$ and symmetric thermal diffusivity tensor $\bar{\kappa}_{ij}$. The field variable $\bar{u}$ — which here represents the temperature — satisfies the partial differential equation

$$\frac{\partial}{\partial \bar{x}_i}\left(\bar{\kappa}_{ij}\frac{\partial \bar{u}}{\partial \bar{x}_j}\right) + \bar{b} = 0 \quad \text{in } \bar{\Omega}\,, \tag{3.12}$$

with boundary conditions

$$\bar{u} = 0\,, \qquad \text{on } \bar{\Gamma}_{\mathrm{D}}\,, \tag{3.13}$$

$$\bar{\kappa}_{ij}\frac{\partial \bar{u}}{\partial \bar{x}_j}\bar{e}_i^n = \bar{f}\,, \qquad \text{on } \bar{\Gamma}_{\mathrm{N}}\,, \tag{3.14}$$

here, $\bar{b}$ is the rate of heat generated per unit volume, $\bar{f}$ is the prescribed heat flux input on the surface $\bar{\Gamma}_{\mathrm{N}}$, and $\bar{e}_i^n$ is the unit outward normal.

Weak Form

We now derive the weak form of the governing equations. To begin, we introduce the function space

$$\bar{Y} = \{\bar{v} \in (H^1(\bar{\Omega}))^{p=1} \mid \bar{v} = 0 \text{ on } \bar{\Gamma}_{\mathrm{D}}\} \ , \tag{3.15}$$

and associated norm

$$||\bar{v}||_{\bar{Y}} = \left(\int_{\bar{\Omega}} \sum_{i=1}^{d} \left(\frac{\partial \bar{v}}{\partial \bar{x}_i} \right)^2 d\bar{\Omega} \right)^{1/2} \ . \tag{3.16}$$

Multiplying (3.12) by a test function $\bar{v} \in \bar{Y}$ and integrating over $\bar{\Omega}$, we obtain

$$-\int_{\bar{\Omega}} \bar{v} \frac{\partial}{\partial \bar{x}_i} \left(\bar{\kappa}_{ij} \frac{\partial \bar{u}}{\partial \bar{x}_j} \right) d\bar{\Omega} = \int_{\bar{\Omega}} \bar{b}\, \bar{v}\, d\bar{\Omega} \ , \qquad \forall\, \bar{v} \in \bar{Y} \ . \tag{3.17}$$

Integrating by parts and applying the divergence theorem yields

$$-\int_{\bar{\Omega}} \bar{v} \frac{\partial}{\partial \bar{x}_i} \left(\bar{\kappa}_{ij} \frac{\partial \bar{u}}{\partial \bar{x}_j} \right) d\bar{\Omega} = -\int_{\bar{\Gamma}} \bar{v} \bar{\kappa}_{ij} \frac{\partial \bar{u}}{\partial \bar{x}_j} \bar{e}_i^n \, d\bar{\Gamma} + \int_{\bar{\Omega}} \frac{\partial \bar{v}}{\partial \bar{x}_i} \bar{\kappa}_{ij} \frac{\partial \bar{u}}{\partial \bar{x}_j} \, d\bar{\Omega} \ , \qquad \forall\, \bar{v} \in \bar{Y} \ . \tag{3.18}$$

Substituting (3.18) into (3.17), and using (3.13), (3.14), and the fact that $\bar{v} = 0$ on $\bar{\Gamma}_{\mathrm{D}}$, we obtain as our weak statement

$$\langle \bar{\mathcal{A}} \bar{u}, \bar{v} \rangle = \langle \bar{F}, \bar{v} \rangle \ , \qquad \forall\, \bar{v} \in \bar{Y} \ , \tag{3.19}$$

where

$$\langle \bar{\mathcal{A}} \bar{w}, \bar{v} \rangle = \int_{\bar{\Omega}} \frac{\partial \bar{v}}{\partial \bar{x}_i} \bar{\kappa}_{ij} \frac{\partial \bar{w}}{\partial \bar{x}_j} \tag{3.20}$$

$$\langle \bar{F}, \bar{v} \rangle = \langle \bar{F}_f, \bar{v} \rangle + \langle \bar{F}_b, \bar{v} \rangle \ ; \tag{3.21}$$

here,

$$\langle \bar{F}_f, \bar{v} \rangle = \int_{\bar{\Gamma}_{\mathrm{N}}^n} \bar{v}\, \bar{f}\, d\bar{\Gamma} \ , \qquad \langle \bar{F}_b, \bar{v} \rangle = \int_{\bar{\Omega}} \bar{v}\, \bar{b}\, d\bar{\Omega} \ . \tag{3.22}$$

We now generalize (3.19)-(3.22) to the case in which $\bar{\Omega}$ is inhomogeneous. Assuming that $\bar{\Omega}$ consists of $\bar{R}$ homogeneous subdomains $\bar{\Omega}^{\bar{r}}$ such that

$$\overline{\bar{\Omega}} = \bigcup_{\bar{r}=1}^{\bar{R}} \overline{\bar{\Omega}}^{\bar{r}} \tag{3.23}$$

(here, $\overline{\bar{\Omega}}$ denotes the closure of $\bar{\Omega}$) the weak statement takes the form of (3.19) where

$$\langle \bar{\mathcal{A}} \bar{w}, \bar{v} \rangle = \sum_{\bar{r}=1}^{\bar{R}} \int_{\bar{\Omega}^{\bar{r}}} \frac{\partial \bar{v}}{\partial \bar{x}_i} \bar{\kappa}_{ij}^{\bar{r}} \frac{\partial \bar{w}}{\partial \bar{x}_j} \, d\bar{\Omega} \tag{3.24}$$

$$\langle \bar{F}, \bar{v} \rangle = \langle \bar{F}_f, \bar{v} \rangle + \langle \bar{F}_b, \bar{v} \rangle \ , \tag{3.25}$$

and

$$\langle \bar{F}_f, \bar{v} \rangle = \sum_{\bar{r}=1}^{\bar{R}} \int_{\bar{\Gamma}_{\mathrm{N}}^{\bar{r}}} \bar{v}\, \bar{f}^{\bar{r}} \, d\bar{\Gamma} \ , \qquad \langle \bar{F}_b, \bar{v} \rangle = \sum_{r=1}^{\bar{R}} \int_{\bar{\Omega}^{\bar{r}}} \bar{v}\, \bar{b}^{\bar{r}} \, d\bar{\Omega} \ ; \tag{3.26}$$

here, $\bar{\kappa}_{ij}^{\bar{r}}$ is the thermal diffusivity tensor in $\bar{\Omega}^{\bar{r}}$, and $\bar{\Gamma}_{\rm N}^{\bar{r}}$ is the section of $\bar{\Gamma}_{\rm N}$ in $\overline{\bar{\Omega}^{\bar{r}}}$. The derivation of (3.24)-(3.26) is similar to that for the homogeneous case, but for the use of additional temperature and flux continuity conditions at the interfaces between the $\bar{\Omega}^{\bar{r}}$.

3.3.2 Reduction to Abstract Form

In this section, we reformulate the problem defined by (3.24)-(3.26) so as to recover the abstract formulation of Section 3.2.

Affine Geometric Mapping

To begin, we further partition the subdomains $\bar{\Omega}^{\bar{r}}$, $\bar{r} = 1, \ldots, \bar{R}$, into a total of R subdomains $\bar{\Omega}^r$, $r = 1, \ldots, R$ such that there exists a reference domain $\Omega = \bigcup_{r=1}^{R} \overline{\Omega}^r$ where, for any $\underline{\bar{x}} \in \bar{\Omega}^r$, $r = 1, \ldots, R$, its image $\underline{x} \in \Omega^r$ is given by

$$\underline{x} = \mathcal{G}^r(\mu; \underline{\bar{x}}) = \underline{G}^r(\mu)\underline{\bar{x}} + \underline{g}^r(\mu), \quad 1 \le r \le R \; ; \tag{3.27}$$

we thus write

$$\frac{\partial}{\partial \bar{x}_i} = \frac{\partial x_j}{\partial \bar{x}_i}\frac{\partial}{\partial x_j} = G_{ji}(\mu)\frac{\partial}{\partial x_j}, \tag{3.28}$$

and

$$\underline{x} = \mathcal{G}(\mu; \underline{\bar{x}}) = \underline{G}(\mu)\underline{\bar{x}} + \underline{g}(\mu) \; , \tag{3.29}$$

where $\underline{x} \in \Omega$, $\underline{\bar{x}} \in \bar{\Omega}$, $\underline{G}(\mu) \in \mathbb{R}^{d\times d}$ is a piecewise-constant matrix, $\underline{g}^r(\mu) \in \mathbb{R}^d$ is a piecewise-constant vector, and $\mathcal{G}(\mu)\colon \bar{\Omega} \to \Omega$ is a piecewise-affine geometric mapping. We then denote the boundary of Ω as Γ, where $\Gamma = \mathcal{G}(\mu; \bar{\Gamma})$.

Reference Domain Formulation

We now define the function space Y as $Y(\Omega) = \bar{Y}(\mathcal{G}^{-1}(\mu; \Omega)) = \bar{Y}(\bar{\Omega})$ such that

$$Y = \{v \in \left(H^1(\Omega)\right)^{p=1} \mid v = 0 \text{ on } \Gamma_{\rm D}\} \; , \tag{3.30}$$

and for any function $\bar{w} \in \bar{Y}$, we define $w \in Y$ such that $w(\underline{x}) = \bar{w}(\mathcal{G}^{-1}(\mu; \underline{x}))$. Furthermore, we have

$$d\bar{\Omega} = \det \underline{G}^{-1}(\mu)\, d\Omega \; , \tag{3.31}$$

$$d\bar{\Gamma} = \left|\underline{G}^{-1}(\mu)\, \underline{\rm e}^t\right| d\Gamma \; , \tag{3.32}$$

where $\underline{\rm e}^t$ is a unit vector tangent to the boundary Γ, and

$$\left|\underline{G}^{-1}(\mu)\, \underline{\rm e}^t\right| = \left(\sum_{i=1}^{d} \left(G_{ij}{\rm e}_j^t\right)^2\right)^{1/2} \; . \tag{3.33}$$

It then follows that $\langle \mathcal{A}(\mu)w, v\rangle = \langle \bar{\mathcal{A}}\bar{w}, \bar{v}\rangle$ for $\bar{\mathcal{A}}$ as in (3.24) and $\mathcal{A}(\mu)$ given by

$$\langle \mathcal{A}(\mu)w, v\rangle = \sum_{r=1}^{R} \int_{\Omega^r} \left(G^r_{ii'}(\mu) \frac{\partial w}{\partial x_i} \right) \bar{\kappa}^r_{i'j'} \left(G^r_{jj'}(\mu) \frac{\partial v}{\partial x_j} \right) \det(\underline{G}^r(\mu))^{-1} \, d\Omega \ , \tag{3.34}$$

$$= \sum_{r=1}^{R} \int_{\Omega^r} \frac{\partial w}{\partial x_i} \left(G^r_{ii'}(\mu) \bar{\kappa}^r_{i'j'} G^r_{jj'}(\mu) \det(\underline{G}(\mu))^{-1} \right) \frac{\partial v}{\partial x_j} \, d\Omega \quad \forall \ w, v \in Y \ , \tag{3.35}$$

and $\langle F(\mu)w, v\rangle = \langle \bar{F}\bar{w}, \bar{v}\rangle$ for $\bar{F}$ as in (3.25) and $F(\mu)$ given by

$$\langle F(\mu), v\rangle = \langle F_f, v\rangle + \langle F_b, v\rangle \tag{3.36}$$

where

$$\langle F_f, v\rangle = \sum_{r=1}^{R} \int_{\Gamma^r_{\mathrm{N}}} \left(\bar{f}^r \left| (\underline{G}^r(\mu))^{-1} \, \underline{\mathrm{e}}^t \right| \right) v \, d\Gamma \ , \quad \langle F_b, v\rangle = \sum_{r=1}^{R} \int_{\Omega^r} \left(\bar{b}^r \det\left(\underline{G}^r(\mu)\right)^{-1} \right) v \, d\Omega \ . \tag{3.37}$$

The abstract problem statement of Section 3.2 is then recovered for

$$\langle \mathcal{A}(\mu)w, v\rangle = \sum_{r=1}^{R} \int_{\Omega^r} \frac{\partial w}{\partial x_i} \kappa^r_{ij}(\mu) \frac{\partial v}{\partial x_j} \, d\Omega \qquad \forall \ w, v \in Y \ , \tag{3.38}$$

$$\langle F(\mu), v\rangle = \langle F_b(\mu), v\rangle + \langle F_f(\mu), v\rangle \tag{3.39}$$

where

$$\langle F_b(\mu), v\rangle = \sum_{r=1}^{R} \int_{\Omega^r} b^r(\mu) v \, d\Omega \tag{3.40}$$

$$\langle F_f(\mu), v\rangle = \sum_{r=1}^{R} \int_{\Gamma^r_{\mathrm{N}}} f^r(\mu) v \, d\Gamma \ ; \tag{3.41}$$

here $\kappa^r_{ij}(\mu)$ is given by

$$\kappa^r_{ij}(\mu) = G^r_{ii'}(\mu) \ \bar{\kappa}^r_{i'j'} \ G^r_{jj'}(\mu) \ \det(\underline{G}^r(\mu))^{-1} \ , \tag{3.42}$$

and $b^r(\mu)$, $f^r(\mu)$ are given by

$$b^r(\mu) = \bar{b}^r \det\left(\underline{G}^r(\mu)\right)^{-1} \ , \tag{3.43}$$

$$f^r(\mu) = \bar{f}^r \left| (\underline{G}^r(\mu))^{-1} \, \underline{\mathrm{e}}^t \right| \ . \tag{3.44}$$

Furthermore, clearly we may define

$$\Theta^{q(i,j,r)}(\mu) = \kappa^r_{ij}(\mu), \qquad \langle \mathcal{A}^{q(i,j,r)} w, v\rangle = \int_{\Omega^r} \frac{\partial v}{\partial x_i} \frac{\partial w}{\partial x_j} \, d\Omega \tag{3.45}$$

for $1 \le i,j \le d$, $1 \le r \le R$, and $q : \{1,\ldots,d\}^2 \times \{1,\ldots,R\} \to \{1,\ldots,Q_{\mathcal{A}}\}$; and

$$\varphi_F^{q'(r,\chi)} = \begin{cases} b^r(\mu) & \text{for } \chi = 1\ , \\ f^r(\mu) & \text{for } \chi = 2\ , \end{cases} \qquad F^{q'(r,\chi)} = \begin{cases} \int_{\Omega^r} v & \text{for } \chi = 1\ , \\ \int_{\Gamma_N^r} v & \text{for } \chi = 2\ , \end{cases} \tag{3.46}$$

for $1 \le r \le R$, and $q' : \{1,\ldots,R\} \times \{1,2\} \to \{1,\ldots,Q_F\}$. Note however that due to the symmetry of $\kappa_{ij}(\mu)$, $Q_{\mathcal{A}}$ can in fact be taken to be $d(d+1)R/2$ rather than d^2R; furthermore, $Q_{\mathcal{A}}$ and Q_F can be further reduced by eliminating elements which are identically zero.

3.4 Model Problems

We indicate here several examples which serve to illustrate our assumptions and methods. Examples 1 and 2 are simple one-dimensional instantiations of the abstract problem formulation of Section 3.2, while Examples 3 and 4 are two-dimensional instantiations of the heat conduction problem of Section 3.3.

3.4.1 Example 1

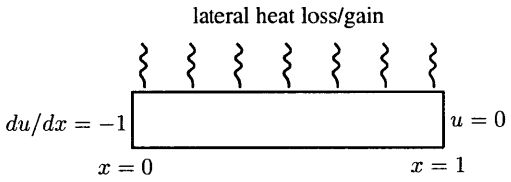


Figure 3-1: Example 1: Heat diffusion in a rod with lateral heat loss/gain.

We consider the flow of heat along the rod shown in Figure 3-1 with both diffusion d^2u/dx^2 along the rod and heat loss (or gain) across the lateral sides of the rod. A constant heat flux is applied at $x = 0$, and the temperature is zero at $x = 1$. The steady-state temperature distribution, u, satisfies

$$-\frac{d^2u}{dx^2} + \mu u = 0 \tag{3.47}$$

on the domain $\Omega \equiv (0,1)$ with Neumann and Dirichlet boundary conditions

$$\begin{aligned} \frac{du}{dx} &= -1, \qquad \text{at } x = 0, \\ u &= 0, \qquad \text{at } x = 1, \end{aligned} \tag{3.48}$$

respectively. Our output of interest is

$$s(\mu) = u(\mu)|_{x=0} \quad \text{for } \mu \in \mathcal{D}^\mu \equiv [0.01, 10^4]. \tag{3.49}$$

Our problem can then be formulated as: given a $\mu \in \mathcal{D}^\mu \subset \mathbb{R}^{P=1}$, find $s(\mu) = \langle L, u(\mu)\rangle$, where $u(\mu) \in Y = \{v \in H^1(\Omega) \mid v|_{x=1} = 0\}$ is the solution to (3.11); for this example, $\langle L, v\rangle = \langle F, v\rangle$ for

all $v \in Y$,

$$\langle \mathcal{A}(\mu)w, v\rangle = \int_0^1 \frac{dv}{dx}\,\frac{dw}{dx} + \mu \int_0^1 v\, w, \quad \forall\, w, v \in Y, \tag{3.50}$$

and

$$\langle F, v\rangle = v|_{x=0}, \quad \forall\, v \in Y. \tag{3.51}$$

The abstract problem statement of Section 3.2 is then recovered for $Q_{\mathcal{A}} = 2$, and

$$\Theta^1(\mu) = 1, \qquad \langle \mathcal{A}^1 w, v\rangle = \int_0^1 \frac{dv}{dx}\,\frac{dw}{dx}\,, \tag{3.52}$$

$$\Theta^2(\mu) = \mu, \qquad \langle \mathcal{A}^2 w, v\rangle = \int_0^1 v\, w. \tag{3.53}$$

3.4.2 Example 2

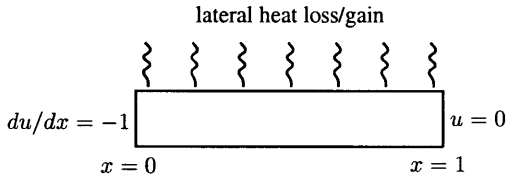


Figure 3-2: Example 2: Heat diffusion in a rod with lateral heat loss/gain and convective cooling.

We now consider a problem similar to Example 1 but with convective cooling, as shown in Figure 3-2, but with convective cooling at $x = 1$. The temperature distribution, u, satisfies

$$-\frac{d^2u}{dx^2} + \mu^1 u = 0 \tag{3.54}$$

on the domain $\Omega \equiv (0, 1)$ with Neumann and Robin boundary conditions

$$\begin{aligned} \frac{du}{dx} &= -1, \qquad \text{at } x = 0, \\ \frac{du}{dx} + \mu^2 u &= 0, \qquad \text{at } x = 1, \end{aligned} \tag{3.55}$$

respectively. Our output of interest is

$$s(\mu) = u(\mu)|_{x=0} \quad \text{for } \mu = (\mu^1, \mu^2) \in \mathcal{D}^\mu \equiv [1, 1000] \times [0.001, 0.1]. \tag{3.56}$$

Our problem can then be formulated as: given a $\mu \in \mathcal{D}^\mu \subset \mathbb{R}^{P=2}$, find $s(\mu) = \langle L, u(\mu)\rangle$, where $u(\mu) \in Y = H^1(\Omega)$ is the solution to (3.11); for this example, $\langle L, v\rangle = \langle F, v\rangle$ for all $v \in Y$,

$$\langle \mathcal{A}(\mu)w, v\rangle = \int_0^1 \frac{dv}{dx}\,\frac{dw}{dx} + \mu^1 \int_0^1 v\, w + \mu^2\, (vw)|_{x=1}\,, \quad \forall\, w, v \in Y, \tag{3.57}$$

and

$$\langle F, v\rangle = v|_{x=0}, \quad \forall\, v \in Y. \tag{3.58}$$

The abstract problem statement of Section 3.2 is then recovered for $Q = 3$, and

$$\Theta^1(\mu) = 1, \qquad \langle \mathcal{A}^1 w, v \rangle = \int_0^1 \frac{dv}{dx}\,\frac{dw}{dx}, \tag{3.59}$$

$$\Theta^2(\mu) = \mu^1, \qquad \langle \mathcal{A}^2 w, v \rangle = \int_0^1 v\, w, \tag{3.60}$$

$$\Theta^3(\mu) = \mu^2, \qquad \langle \mathcal{A}^3 w, v \rangle = (vw)|_{x=1}\,. \tag{3.61}$$

3.4.3 Example 3

Figure 3-3: Example 3: (a) Parameter-dependent rectangular domain with internal heat source; and (b) a reference domain.

We now consider the flow of heat in a rectangular region containing an internal heat source, shown in Figure 3-3(a). In particular, we consider the problem

$$-\bar{\nabla}^2 \bar{u} = \frac{1}{\bar{t}}, \qquad \bar{\nabla}^2 = \frac{\partial}{\partial \bar{x}_1^2} + \frac{\partial}{\partial \bar{x}_2^2} \tag{3.62}$$

in a domain $\bar{\Omega} \equiv (0,1) \times (0,\bar{t})$ with homogeneous Dirichlet conditions on the boundary,

$$\bar{u} = 0, \quad \text{on } \bar{\Gamma}_D. \tag{3.63}$$

Our output of interest is

$$s(\mu) = \frac{1}{\bar{t}} \int_{\bar{\Omega}} \bar{u}\, d\bar{\Omega}, \quad \text{for } \mu = \{\bar{t}\} \in \mathcal{D}^\mu \equiv [\bar{t}_{\min}, 1.0] = [0.1, 1.0]. \tag{3.64}$$

Our problem can then be formulated as: given a $\mu \in \mathcal{D}^\mu \subset \mathbb{R}^{P=1}$, find $s(\mu) = \langle \bar{L}, \bar{u} \rangle$ where $\bar{u} \in \bar{Y} = \{\bar{v} \in H^1(\bar{\Omega}) \mid \bar{v}|_{\bar{\Gamma}_D} = 0\}$ is the solution to

$$\langle \bar{\mathcal{A}} \bar{w}, \bar{v} \rangle = \langle \bar{F}, \bar{v} \rangle\,, \quad \forall\, \bar{v} \in \bar{Y}\,; \tag{3.65}$$

here, $\langle \bar{L}, \bar{v} \rangle = \langle \bar{F}, \bar{v} \rangle$ for all $\bar{v} \in \bar{Y}$,

$$\langle \bar{\mathcal{A}} \bar{w}, \bar{v} \rangle = \int_{\bar{\Omega}} \frac{\partial \bar{v}}{\partial \bar{x}_1}\,\frac{\partial \bar{w}}{\partial \bar{x}_1} + \frac{\partial \bar{v}}{\partial \bar{x}_2}\,\frac{\partial \bar{w}}{\partial \bar{x}_2}\, d\bar{\Omega}\,, \qquad \forall\, \bar{w}, \bar{v} \in \bar{Y}\,, \tag{3.66}$$

and

$$\langle \bar{F}, \bar{v} \rangle = \frac{1}{\bar{t}} \int_{\bar{\Omega}} \bar{v} \, d\bar{\Omega} \ , \qquad \forall \ \bar{v} \in \bar{Y} \ . \tag{3.67}$$

We now map our parameter-dependent domain $\bar{\Omega}$ onto a reference domain $\Omega = (0,1) \times (0,1)$, shown in Figure 3-3(b). The affine mapping $\mathcal{G}(\bar{x})(\mu) \colon \bar{\Omega} \rightarrow \Omega$ is then given by (3.27) for $R = 1$, $\underline{g}^1(\mu) = 0$, and

$$\underline{G}^1(\mu) = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\bar{t}} \end{bmatrix} \ . \tag{3.68}$$

Furthermore, we have

$$d\bar{\Omega} = \det \underline{G}^{-1}(\mu) \, d\Omega = \bar{t} \, d\Omega \ , \tag{3.69}$$

$$d\bar{\Gamma} = \left| \underline{G}^{-1}(\mu) \, \underline{e}^t \right| d\Gamma = \bar{t} \, d\Gamma \ . \tag{3.70}$$

We may now re-formulate our problem in terms of our reference domain. Our problem is then: find $s(\mu) = \langle L, u(\mu) \rangle$ where $u(\mu) \in Y = \{ v \in H^1(\Omega) \mid v|_{\Gamma_D} = 0 \}$ is the solution to (3.11); for this example, $\langle L, v \rangle = \langle F, v \rangle$ for all $v \in Y$,

$$\langle \mathcal{A}(\mu) w, v \rangle = \bar{t} \int_\Omega \frac{\partial v}{\partial x_1} \frac{\partial w}{\partial x_1} \, d\Omega + \frac{1}{\bar{t}} \int_\Omega \frac{\partial v}{\partial x_2} \frac{\partial w}{\partial x_2} \, d\Omega \ , \qquad \forall \ w, v \in Y \ , \tag{3.71}$$

and

$$\langle F, v \rangle = \int_\Omega v \, d\Omega \ , \qquad \forall \ v \in Y \ . \tag{3.72}$$

The abstract problem statement of Section 3.2 is then recovered for $P = 1$, $Q = 2$, and

$$\Theta^1(\mu) = \mu, \qquad \langle \mathcal{A}^1 w, v \rangle = \int_\Omega \frac{\partial v}{\partial x_1} \frac{\partial w}{\partial x_1}, \tag{3.73}$$

$$\Theta^2(\mu) = \frac{1}{\mu}, \qquad \langle \mathcal{A}^2 w, v \rangle = \int_\Omega \frac{\partial v}{\partial x_2} \frac{\partial w}{\partial x_2}. \tag{3.74}$$

3.4.4 Example 4

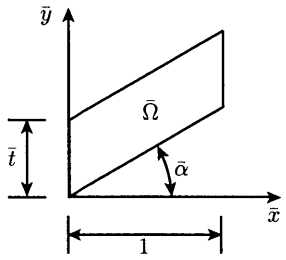


Figure 3-4: Example 4: Parameter-dependent domain undergoing both "stretch" and "shear."

In this example, we again consider the homogeneous rod and loading of Example 3, but in addition to the thickness $\bar{t}$, we allow the angle $\bar{\alpha}$ to vary, as shown in Figure 3-4. The output of

interest is again the average temperature in the domain $\bar{\Omega}$:

$$s(\mu) = -\frac{1}{\bar{t}} \int_{\bar{\Omega}} \bar{u}, \quad \text{for } \mu = \{\bar{t}, \bar{\alpha}\} \in \mathcal{D}^\mu = [0.1, 1.0] \times [0°, 45°] \ . \tag{3.75}$$

Our problem can then be formulated as: given a $\mu \in \mathcal{D}^\mu \subset \mathbb{R}^{P=2}$, find $s(\mu) = \langle \bar{L}, \bar{u} \rangle$, where $\bar{u} \in \bar{Y} = \{\bar{v} \in H^1(\bar{\Omega}) \mid \bar{v}|_{\bar{\Gamma}_D} = 0\}$ is the solution to $\langle \bar{\mathcal{A}} \bar{u}, \bar{v} \rangle = \langle \bar{F}, \bar{v} \rangle$, $\forall\ \bar{v} \in \bar{Y}$; here, $\langle \bar{L}, \bar{v} \rangle = \langle \bar{F}, \bar{v} \rangle$ for all $\bar{v} \in \bar{Y}$, and $\bar{\mathcal{A}}$, $\bar{F}$ are given by (3.66), (3.67), respectively.

We again map our parameter-dependent domain $\bar{\Omega}$ onto a reference domain $\Omega = (0,1) \times (0,1)$. The affine mapping $\mathcal{G}(\bar{x})(\mu) \colon \bar{\Omega} \to \Omega$ is then given by (3.27) for $R = 1$, $\underline{g}^1(\mu) = 0$, and

$$\underline{G}^1(\mu) = \begin{bmatrix} 1 & 0 \\ -\dfrac{\tan\bar{\alpha}}{\bar{t}} & \dfrac{1}{\bar{t}} \end{bmatrix} \ . \tag{3.76}$$

Furthermore, we have

$$d\bar{\Omega} = \det \underline{G}^{-1}(\mu)\ d\Omega = \bar{t}\, d\Omega \ , \tag{3.77}$$

$$d\bar{\Gamma} = \left| \underline{G}^{-1}(\mu)\ \underline{e}^t \right| d\Gamma = \bar{t}\, d\Gamma \ . \tag{3.78}$$

We may now re-formulate our problem in terms of the reference domain. Our problem is then: find $s(\mu) = \langle L, u(\mu) \rangle$ where $u(\mu) \in Y = H_0^1(\Omega)$ is the solution to (3.11); for this example, $\langle L, v \rangle = \langle F, v \rangle$ for all $v \in Y$,

$$\langle \mathcal{A}(\mu) w, v \rangle = \int_\Omega \frac{\partial v}{\partial x_i} \kappa_{ij}(\mu) \frac{\partial w}{\partial x_j}\, d\Omega, \quad \forall\ w, v \in Y \ , \tag{3.79}$$

and

$$\langle F, v \rangle = -\int_\Omega v\, d\Omega, \qquad \forall\ v \in Y \ , \tag{3.80}$$

and the effective diffusivity tensor $\kappa_{ij}(\mu) = G_{ii'}(\mu)\, \bar{\kappa}_{i'j'}\, G_{jj'}(\mu) \det \underline{G}^{-1}(\mu)$ is given by

$$\underline{\kappa} = \begin{bmatrix} \bar{t} & -\tan\bar{\alpha} \\ -\tan\bar{\alpha} & \dfrac{1 + \tan^2\bar{\alpha}}{\bar{t}} \end{bmatrix} \tag{3.81}$$

The abstract problem statement of Section 3.2 is then recovered for $P = 2$, $Q_\mathcal{A} = 3$,

$$\Theta^1(\mu) = \bar{t} \ , \tag{3.82}$$

$$\Theta^2(\mu) = \frac{1 + \tan^2\bar{\alpha}}{\bar{t}} \ , \tag{3.83}$$

$$\Theta^3(\mu) = -\tan\bar{\alpha} \ , \tag{3.84}$$

and

$$\langle \mathcal{A}^1 w, v\rangle = \int_\Omega \frac{\partial v}{\partial x_1}\frac{\partial w}{\partial x_1}\, d\Omega \tag{3.85}$$

$$\langle \mathcal{A}^2 w, v\rangle = \int_\Omega \frac{\partial v}{\partial x_2}\frac{\partial w}{\partial x_2}\, d\Omega\ , \tag{3.86}$$

$$\langle \mathcal{A}^3 w, v\rangle = \int_\Omega \left(\frac{\partial v}{\partial x_1}\frac{\partial w}{\partial x_2} + \frac{\partial v}{\partial x_2}\frac{\partial w}{\partial x_1}\right) d\Omega\ . \tag{3.87}$$

3.5 Reduced-Basis Output Approximation

We recall that in this chapter, we assume that $\mathcal{A}(\mu)$ is continuous, coercive, symmetric, and affine in μ. We also assume for simplicity that L and F are independent of parameter, and $L = F$.

3.5.1 Approximation Space

As indicated in Chapter 2, to approximate $u(\mu)$, and hence $s(\mu)$, we need *not* represent *every possible function* in Y; instead, we need only approximate those functions in the low-dimensional manifold "spanned" by $u(\mu)$. We therefore introduce a sample in parameter space,

$$S_N^\mu = \{\mu_1, \ldots, \mu_N\} \tag{3.88}$$

where $\mu_n \in \mathcal{D}^\mu \in \mathbb{R}^P, n = 1, \ldots, N$. We then define our Lagrangian [37] reduced-basis approximation space as

$$W_N = \text{span}\{\zeta_n \equiv u(\mu_n), n = 1, \ldots, N\}, \tag{3.89}$$

where $u(\mu_n) \in Y$ is the solution to (2.2) for $\mu = \mu_n$. In actual practice, Y is replaced by an appropriate "truth" space — a finite element approximation space defined on a suitably fine truth mesh — of dimension $\mathcal{N}$, where $\mathcal{N}$ is generally quite large.

Our reduced-basis approximation is then: for any $\mu \in \mathcal{D}^\mu$, find

$$s_N(\mu) = \langle L, u_N(\mu)\rangle, \tag{3.90}$$

where $u_N(\mu) \in W_N$ is the Galerkin projection of $u(\mu)$ onto W_N,

$$\langle \mathcal{A}(\mu) u_N(\mu), v\rangle = \langle F, v\rangle, \quad \forall\ v \in W_N. \tag{3.91}$$

3.5.2 A Priori Convergence Theory

Optimality

We consider here the rate at which $u_N(\mu)$ and $s_N(\mu)$ converges to $u(\mu)$ and $s(\mu)$, respectively. To begin, it is standard to demonstrate the optimality of $u_N(\mu)$ in the sense that

$$||u(\mu) - u_N(\mu)||_Y \leq \sqrt{\frac{\gamma_\mathcal{A}(\mu)}{\beta_\mathcal{A}(\mu)}} \inf_{w_N \in W_N} ||u(\mu) - w_N||_Y\ . \tag{3.92}$$

To prove (3.92), we first note that since $\langle \mathcal{A}u(\mu), v\rangle = \langle L, v\rangle, \forall v \in Y$ and $W_N \subset Y$, it follows that

$$\langle \mathcal{A}u(\mu), v\rangle = \langle L, v\rangle, \quad \forall v \in W_N. \tag{3.93}$$

Subtracting (3.91) from (3.93), we obtain

$$\langle \mathcal{A}(u(\mu) - u_N(\mu)), v\rangle = 0, \quad \forall\, v \in W_N, \tag{3.94}$$

which states that the error, $u(\mu) - u_N(\mu)$, is orthogonal to all members of W_N. We then note that for any $w_N = u_N + v_N \in W_N$ where $v_N \neq 0$,

$$\begin{aligned}
\langle \mathcal{A}(\mu)(u - w_N), (u - w_N)\rangle &= \langle \mathcal{A}(\mu)((u - u_N) - v_N), ((u - u_N) - v_N)\rangle \qquad (3.95)\\
&= \langle \mathcal{A}(\mu)(u - u_N), (u - u_N)\rangle - 2\langle \mathcal{A}(\mu)(u - u_N), v_N\rangle + \langle \mathcal{A}(\mu)v_N, v_N\rangle \\
&= \langle \mathcal{A}(\mu)(u - u_N), (u - u_N)\rangle + \langle \mathcal{A}(\mu)v_N, v_N\rangle \\
&> \langle \mathcal{A}(\mu)(u - u_N), (u - u_N)\rangle
\end{aligned}$$

from symmetry, Galerkin orthogonality, and coercivity; note that the μ-dependence has been omitted. It then follows that

$$\langle \mathcal{A}(\mu)(u(\mu) - u_N(\mu)), (u(\mu) - u_N(\mu))\rangle = \inf_{v_N \in W_N} \langle \mathcal{A}(\mu)(u(\mu) - v_N), (u(\mu) - v_N)\rangle \ . \tag{3.96}$$

Furthermore, from (3.5) we have

$$\langle \mathcal{A}(\mu)(u(\mu) - u_N(\mu)), (u(\mu) - u_N(\mu))\rangle^{1/2} \leq \sqrt{\gamma_{\mathcal{A}}(\mu)} \inf_{v_N \in W_N} ||u(\mu) - v_N||_Y \ ; \tag{3.97}$$

and from (3.6) we have

$$\sqrt{\beta_{\mathcal{A}}(\mu)}||u(\mu) - u_N(\mu)||_Y \leq \langle \mathcal{A}(\mu)(u(\mu) - u_N(\mu)), (u(\mu) - u_N(\mu))\rangle^{1/2} \ ; \tag{3.98}$$

optimality ((3.92)) then follows from (3.97) and (3.98). Furthermore, for our compliance output,

$$\begin{aligned}
s(\mu) - s_N(\mu) &= \langle L, u(\mu) - u_N(\mu)\rangle \qquad (3.99)\\
&= \langle \mathcal{A}(\mu)u(\mu), u(\mu) - u_N(\mu)\rangle \qquad (3.100)\\
&= \langle \mathcal{A}(\mu)(u(\mu) - u_N(\mu)), (u(\mu) - u_N(\mu))\rangle \qquad (3.101)\\
&\leq \gamma_{\mathcal{A}}(\mu)||u(\mu) - u_N(\mu)||_Y^2 \ , \qquad (3.102)
\end{aligned}$$

from (3.10), (3.11), symmetry, Galerkin orthogonality, and (3.5). Therefore, not only is $u_N(\mu)$ the best approximation among all members of W_N (in the sense of (3.92)), but the output approximation, $s_N(\mu)$, converges to $s(\mu)$ as the square of the error in $u_N(\mu)$.

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 [27, 28] is restricted to the case in which $P = 1$, $\mathcal{D}^\mu = [0, \mu_{\max}]$, and

$$\langle \mathcal{A}(\mu)w, v\rangle = \langle \mathcal{A}^0 w, v\rangle + \mu\langle \mathcal{A}^1 w, v\rangle \tag{3.103}$$

where $\mathcal{A}^0$ is continuous, coercive, and symmetric, and $\mathcal{A}^1$ is continuous, positive semi-definite ($\langle \mathcal{A}^1 w, w \rangle \geq 0$, $\forall\, w \in Y$), and symmetric. This model problem (3.103) is rather broadly relevant, for example to variable orthotropic conductivity, variable rectilinear geometry, variable piecewise-constant conductivity, and variable Robin boundary conditions.

Following [27, 28], we suppose that the μ_n, $n = 1, \ldots, N$, are logarithmically distributed in the sense that

$$\ln\left(\overline{\lambda}\,\mu_n + 1\right) = \frac{n-1}{N-1} \ln\left(\overline{\lambda}\,\mu_{\max} + 1\right), \quad n = 1, \ldots, N, \tag{3.104}$$

where $\overline{\lambda}$ is an upper bound for the maximum eigenvalue of $\mathcal{A}^1$ relative to $\mathcal{A}^0$. It can be shown [27] that, for $N > N_{\text{crit}} \equiv e \ln(\overline{\lambda}\,\mu_{\max} + 1)$,

$$\inf_{w_N \in W_N} ||u(\mu) - w_N||_Y \leq (1 + \mu_{\max}\,\overline{\lambda})\, ||u(0)||_Y \, \exp\left\{\frac{-(N-1)}{(N_{\text{crit}} - 1)}\right\}, \quad \forall\, \mu \in \mathcal{D}^\mu\,. \tag{3.105}$$

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 (3.105) are for the particular case in which $(\cdot,\cdot)_Y = \langle \mathcal{A}^0 \cdot, \cdot \rangle$.)

The proof [27, 28] exploits a parameter–space (non-polynomial) 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 (3.92) chooses to "illuminate" only certain points μ_n, automatically selecting the best approximation among all (combinatorially many — i.e., $N!$) possibilities, whereas conventional interpolation methods are "obligated" to use all of the points μ_n. 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 (3.104) implicated by our interpolant–based arguments is *not* simply an artifact of the proof. We present in Figure 3-4 the maximum relative error $\varepsilon_N(\mu)$ over $\mathcal{D}_i^\mu, i = 1, \ldots, 4$, for $s_N(\mu)$ calculated using the logarithmic distribution (3.104), a uniform distribution,

$$\mu_n = \frac{n-1}{N-1}\mu_{\max} \quad 1 \leq n \leq N\,, \tag{3.106}$$

and a Chebychev distribution,

$$\mu_n = \frac{\mu_{\max}}{2}\left(1 + \cos\left(\frac{2n-1}{N}\pi\right)\right)\,; \tag{3.107}$$

we observe that the logarithmic distribution performs considerably better than the other more obvious candidates.

The results presented in Figure 3-5 for Example 1 ($P = 1$) were obtained using "deterministic" grids; however, we expect that as P increases, tensor-product grids become prohibitively profligate. Fortunately, the convergence rate is not *too* sensitive to point selection: the theory only requires a log "on the average" distribution [27]. We present in Figure 3-6 the maximum relative error $\varepsilon_N(\mu)$ over $\mathcal{D}_i^\mu, i = 1, \ldots, 4$, for $s_N(\mu)$ calculated using the logarithmic grid (3.104), a *random* logarithmic distribution, and a *random* uniform distribution. We observe similar exponential behavior even for the random point distributions. We also observe that, for large ranges of the parameter, the "log-random" distribution generally performed better than the logarithmic grid; this will prove particularly significant in higher dimensions ($P > 1$).

The result (3.105) is certainly tied to the particular form (3.103) and associated regularity of

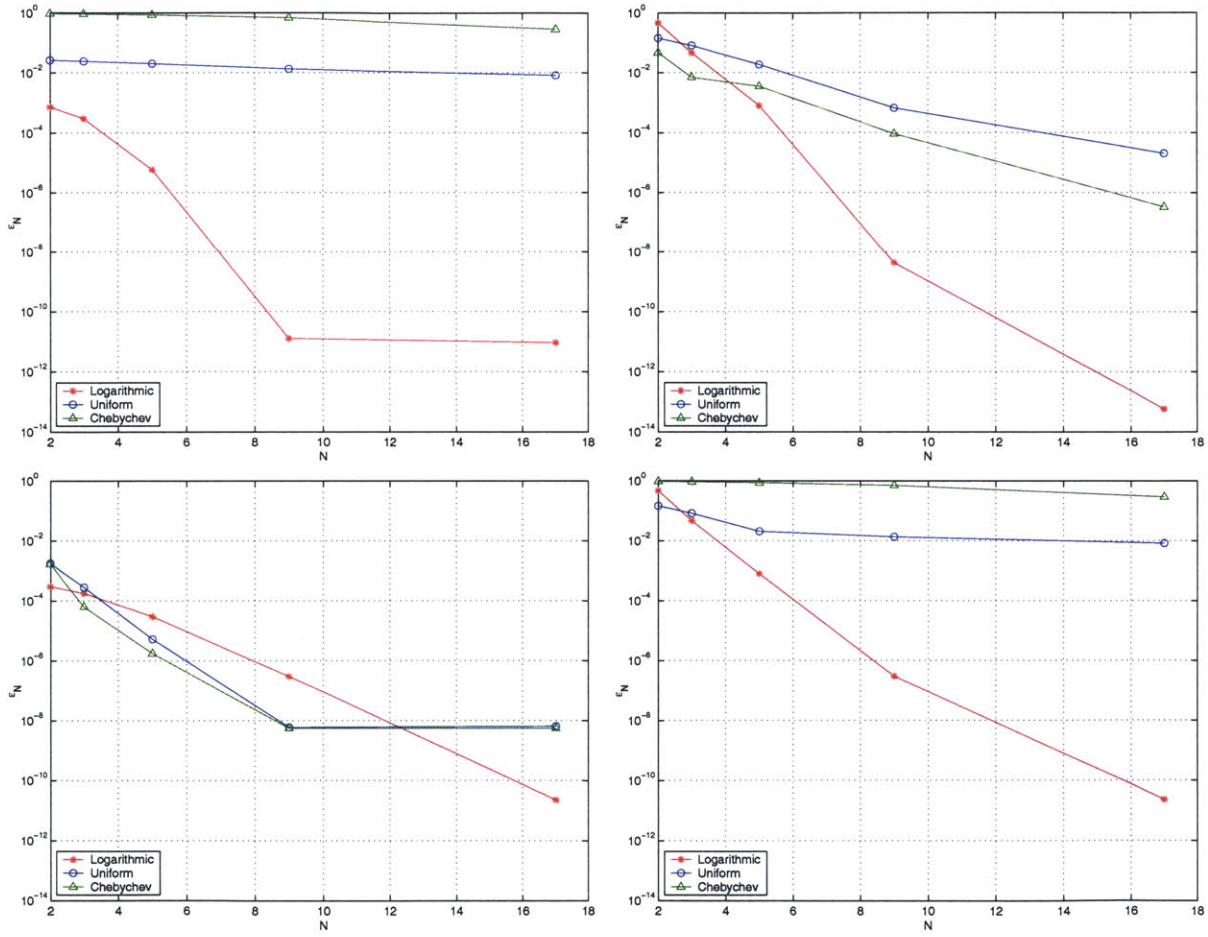


Figure 3-5: Logarithmic vs. other distributions (grid)

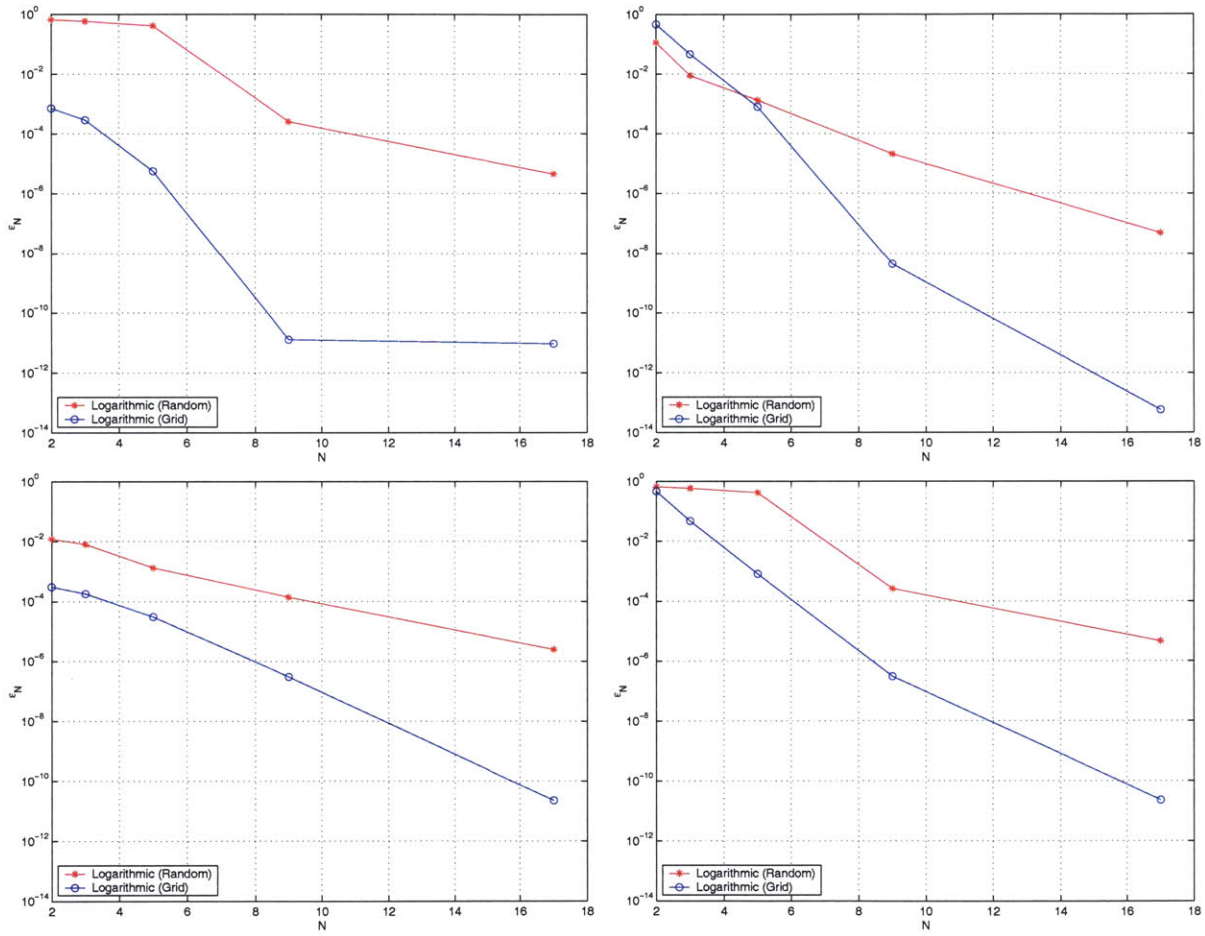


Figure 3-6: Logarithmic grid vs. logarithmic random vs. other random distributions.

$u(\mu)$. 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 Figures 3-7, 3-8, and 3-9 the maximum relative error $\varepsilon_N(\mu)$ as a function of N, at 100 randomly selected points μ in $\mathcal{D}$, for Examples 2, 3, and 4, respectively. In all three cases, the μ_n are chosen "log-randomly" over $\mathcal{D}$: we sample from a multivariate uniform probability density on $\log(\mu)$. This is particularly important for problems in which $P > 1$ (such as Examples 2 and 4) since tensor-product grids are prohibitively profligate as P increases. We observe that the error is remarkably small even for very small N; and that 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^μ and associated "good" weights so as to mitigate the curse of dimensionality as P increases. Furthermore, we observe that the log–random point distribution *is* important, as evidenced by the faster convergence (versus the *non–logarithmic* uniform random point distribution) shown in Figures 3-7, 3-8, and 3-9.

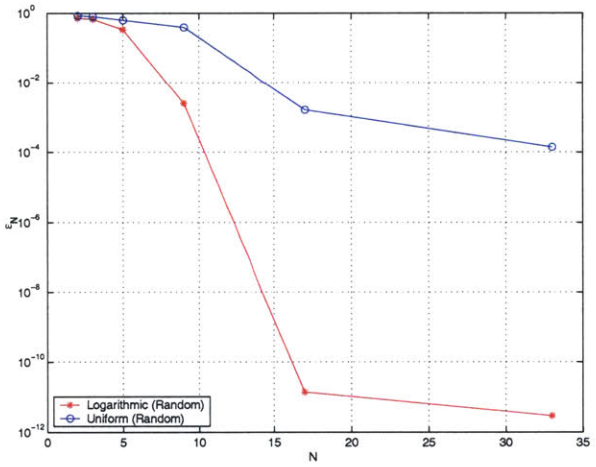


Figure 3-7: Convergence of the reduced-basis approximation for Example 2

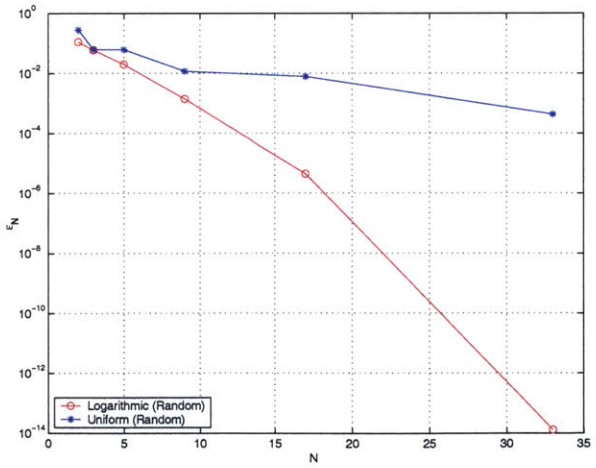


Figure 3-8: Convergence of the reduced-basis approximation for Example 3

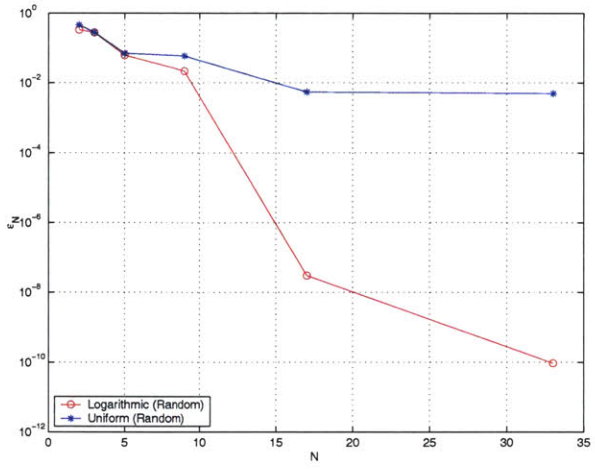


Figure 3-9: Convergence of the reduced-basis approximation for Example 4

3.5.3 Off-Line/On-Line Computational Procedure

The theoretical and empirical results of Sections 3.5.1 and 3.5.2 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 our approximation $u_N(\mu)$ to be a linear combination of the basis functions,

$$u_N(\mu) = \sum_{j=1}^{N} u_{N\,j}(\mu)\,\zeta_j \tag{3.108}$$

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 (3.91) yields the desired algebraic equations for $\underline{u}_N(\mu) \in \mathbb{R}^N$,

$$\underline{A}_N(\mu)\,\underline{u}_N(\mu) \;=\; \underline{F}_N \;, \tag{3.109}$$

in terms of which the output can then be evaluated as

$$s_N(\mu) = \underline{u}_N(\mu)^T \underline{L}_N \;. \tag{3.110}$$

Here $\underline{A}_N(\mu) \in \mathbb{R}^{N\times N}$ is the symmetric positive-definite (SPD) matrix with entries

$$A_{N\,i,j}(\mu) \equiv \langle \mathcal{A}(\mu)\zeta_j, \zeta_i\rangle, \quad 1 \le i,j \le N \;, \tag{3.111}$$

$\underline{F}_N \in \mathbb{R}^N$ is the "load" (and "output") vector with entries

$$F_{N\,i} \equiv \langle F, \zeta_i\rangle, \quad 1 \le i \le N \;, \tag{3.112}$$

and $\underline{L}_N = \underline{F}_N$. We now invoke (3.8) to write

$$A_{N\,i,j}(\mu) = \langle \mathcal{A}(\mu)\zeta_j, \zeta_i\rangle = \sum_{q=1}^{Q} \Theta^q(\mu)\,\langle \mathcal{A}^q\zeta_j, \zeta_i\rangle \;, \tag{3.113}$$

or

$$\underline{A}_N(\mu) = \sum_{q=1}^{Q} \Theta^q(\mu)\, \underline{A}_N^q \ , \tag{3.114}$$

where the $\underline{A}_N^q \in \mathbb{R}^{N\times N}$ are given by

$$A_{N\,i,j}^q = \langle \mathcal{A}^q \zeta_j, \zeta_i \rangle \ , \quad 1 \le i,j \le N,\ 1 \le q \le Q \ . \tag{3.115}$$

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) "$\mathcal{A}$" 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 $\underline{A}_N(\mu)$ from (3.113), then solve (3.109) for $\underline{u}_N(\mu)$, and finally evaluate $s_N(\mu) = \underline{u}_N(\mu)^T \underline{F}_N$: this requires $O(QN^2) + O(\frac{2}{3}N^3)$ 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 (3.109) can be very rapidly assembled and inverted — thanks to the off-line/on-line decomposition (see [7] for an earlier application of this strategy within the reduced–basis context). For the model problems discussed in this chapter, the resulting computational savings relative to standard (well-designed) finite-element approaches are significant.

Chapter 4

A Posteriori Error Estimation: A Heat Conduction Example

4.1 Introduction

In Chapter 3 we ascertained that $s_N(\mu)$ can, in theory, be obtained very inexpensively, and that N can, *in theory*, be chosen quite small. However, *in practice*, we do not know *how* small N can (nor how large it must) be chosen: this will depend on the desired accuracy, the selected output(s) of interest, and the particular problem in question. 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. We thus need *a posteriori* error estimators for $s_N(\mu)$.

In this chapter we develop methods for the *a posteriori* estimation of the error in reduced-basis output approximations. We present two general approaches for the construction of error bounds: Method I, *rigorous a posteriori* error estimation procedures which rely critically on the existence of a "bound conditioner" — in essence, an operator preconditioner that (*i*) satisfies an additional spectral "bound" requirement, and (*ii*) admits the reduced-basis off-line/on-line computational stratagem; and Method II, *a posteriori* error estimation procedures which rely only on the rapid convergence of the reduced-basis approximation, and provide simple, inexpensive error bounds, albeit at the loss of complete certainty. The general Method I approach and several "recipes" for the construction of bound conditioners are presented in Section 4.2. The Method II is presented in Section 4.5.

As in Chapter 2, we introduce an *error estimate*

$$\Delta_N(\mu) \approx |s(\mu) - s_N(\mu)| \ , \tag{4.1}$$

and define the *effectivity* of our error estimate as

$$\eta_N(\mu) \equiv \frac{\Delta_N(\mu)}{|s(\mu) - s_N(\mu)|} \ ; \tag{4.2}$$

we recall that for reliability and efficiency, the effectivity must satisfy

$$1 \ \leq \ \eta_N(\mu) \ \leq \ \rho \ , \tag{4.3}$$

where ρ is $O(1)$. In addition, we require *lower* and *upper output bounds*

$$s_N^-(\mu) \leq s(\mu) \leq s_N^+(\mu), \tag{4.4}$$

which are reliable, sharp, and inexpensive to compute. In this chapter, we continue to assume that $\mathcal{A}$ is continuous, coercive, symmetric, and affine in μ; and that F and L are μ-independent, and $\langle L, v\rangle = \langle F, v\rangle$ for all $v \in Y$.

We now present the general error estimation (and output bound) procedure for both Method I and Method II.

4.2 Method I: Uniform Error Bounds

The approach we describe here 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 [24, 25] and iterative solution error [35] both in the choice of (energy) relaxation and in the associated computational artifice; furthermore, the "bound conditioner" formulation presented here is a generalization of more recent work [39, 43] on *a posteriori* estimation for the reduced-basis method.

4.2.1 Bound Conditioner

To begin, we define the error $e(\mu) \in Y$ and the residual $\mathcal{R}(\mu) \in Y'$ as

$$e(\mu) \equiv u(\mu) - u_N(\mu) \tag{4.5}$$

$$\langle \mathcal{R}(\mu), v\rangle \equiv \langle F, v\rangle - \langle \mathcal{A}(\mu) u_N(\mu), v\rangle, \tag{4.6}$$

from which it follows that

$$\langle \mathcal{A}(\mu) e(\mu), v\rangle = \langle \mathcal{R}(\mu), v\rangle. \tag{4.7}$$

We then introduce a symmetric, continuous, and coercive *bound conditioner* [21, 37, 47] $\mathcal{C}(\mu)\colon Y \to Y'$ such that the minimum and maximum eigenvalues

$$\rho_{\min}(\mu) \equiv \min_{v\in Y} \frac{\langle \mathcal{A}(\mu)v, v\rangle}{\langle \mathcal{C}(\mu)v, v\rangle}, \tag{4.8}$$

$$\rho_{\max}(\mu) \equiv \max_{v\in Y} \frac{\langle \mathcal{A}(\mu)v, v\rangle}{\langle \mathcal{C}(\mu)v, v\rangle}, \tag{4.9}$$

satisfy

$$1 \leq \rho_{\min}(\mu), \qquad \rho_{\max}(\mu) \leq \rho, \tag{4.10}$$

for some (preferably small) constant $\rho \in \mathbb{R}$.

In addition to the spectral condition (4.10), we also require a "computational invertibility" hypothesis [38, 47]. In particular, we shall require that $\mathcal{C}^{-1}(\mu)$ be of the form

$$\mathcal{C}^{-1}(\mu) = \sum_{i\in\mathcal{I}(\mu)} \alpha_i(\mu) \mathcal{C}_i^{-1} \tag{4.11}$$

where $\mathcal{I}(\mu) \subset \{1, \ldots, I\}$ is a parameter-dependent set of indices, I is a finite (preferably small) integer, and the $\mathcal{C}_i\colon Y \to Y'$, are parameter-*independent* symmetric, coercive operators.

4.2.2 Error and Output Bounds

We now find $\hat{e}(\mu) \in Y$ such that

$$\langle \mathcal{C}(\mu)\hat{e}(\mu), v\rangle = \langle \mathcal{R}(\mu), v\rangle, \quad \forall\, v \in Y\ , \tag{4.12}$$

and define our lower and upper output bounds as

$$\begin{aligned} s_N^-(\mu) &= s_N(\mu)\ , & (4.13)\\ s_N^+(\mu) &= s_N(\mu) + \Delta_N(\mu)\ , & (4.14) \end{aligned}$$

where the error estimator is given by

$$\begin{aligned} \Delta_N(\mu) &= \langle \mathcal{C}(\mu)\hat{e}(\mu), \hat{e}(\mu)\rangle & (4.15)\\ &= \langle \mathcal{R}(\mu), \mathcal{C}^{-1}(\mu)\mathcal{R}(\mu)\rangle & (4.16)\\ &= \langle \mathcal{R}(\mu), \hat{e}(\mu)\rangle. & (4.17) \end{aligned}$$

It remains to demonstrate our claim that $s_N^-(\mu) \le s(\mu) \le s_N^+(\mu)$ for all $N \ge 1$, and to investigate the sharpness of our bounds.

4.2.3 Bounding Properties

To begin, we first note that for any $\hat{A}\colon Y \to Y'$ and any $\hat{C}\colon Y \to Y'$ such that

$$\hat{\rho}_{\min} \le \frac{\langle \hat{A}v, v\rangle}{\langle \hat{C}v, v\rangle} \le \hat{\rho}_{\max}, \quad \forall\, v \in Y\ , \tag{4.18}$$

and for any $U \in Y$, $W \in Y$ such that

$$\begin{aligned} \langle \hat{A}U, v\rangle &= \langle \hat{G}, v\rangle, \quad \forall\, v \in Y\ , & (4.19)\\ \langle \hat{C}W, v\rangle &= \langle \hat{G}, v\rangle, \quad \forall\, v \in Y\ , & (4.20) \end{aligned}$$

where $\hat{G} \in Y'$, we can show that

$$\hat{\rho}_{\min} \le \frac{\langle \hat{C}W, W\rangle}{\langle \hat{A}U, U\rangle} \le \hat{\rho}_{\max}\ . \tag{4.21}$$

To prove the lower inequality in (4.21) we note that $\langle \hat{A}U, v\rangle = \langle \hat{C}W, v\rangle$ for all $v \in Y$ and choose $v = U$ to obtain

$$\begin{aligned} \langle \hat{A}U, U\rangle &= \langle \hat{C}W, U\rangle\\ &\le \langle \hat{C}W, W\rangle^{1/2}\langle \hat{C}U, U\rangle^{1/2}\\ &\le \hat{\rho}_{\min}^{-1/2}\langle \hat{C}W, W\rangle^{1/2}\langle \hat{A}U, U\rangle^{1/2} & (4.22) \end{aligned}$$

from the Cauchy-Schwarz inequality and (4.18); thus $\hat{\rho}_{\min}\langle\hat{A}U,U\rangle \leq \langle\hat{C}W,W\rangle$ as desired. To prove the upper inequality in (4.21) we choose $v = W$ to obtain

$$\begin{aligned}\langle\hat{C}W,W\rangle &= \langle\hat{A}U,W\rangle \\ &\leq \langle\hat{A}U,U\rangle^{1/2}\langle\hat{A}W,W\rangle^{1/2} \\ &\leq \hat{\rho}_{\max}^{1/2}\langle\hat{C}W,W\rangle^{1/2}\langle\hat{A}U,U\rangle^{1/2} \ , \end{aligned} \tag{4.23}$$

from the Cauchy-Schwarz inequality and (4.18); thus $\langle\hat{\mathcal{C}}W,W\rangle \leq \hat{\rho}_{\max}\langle\hat{\mathcal{A}}U,U\rangle$ as desired.

We can now prove the bounding and sharpness properties of our bounds. We first prove that $s_N^-(\mu) \leq s(\mu)$. We note that

$$s(\mu) - s_N(\mu) = \langle\mathcal{A}(\mu)\left(u(\mu) - u_N(\mu)\right), u(\mu) - u_N(\mu)\rangle \tag{4.24}$$

$$\geq 0 \tag{4.25}$$

from (3.95) and the coercivity of $\mathcal{A}$, respectively. This lower bound proof is a standard result in variational approximation theory. We now turn to the less trivial upper bound.

To show that $s_N^+(\mu) \geq s(\mu)$, we note that

$$\eta_N(\mu) = \frac{\Delta_N(\mu)}{s(\mu) - s_N(\mu)} \tag{4.26}$$

$$= \frac{s_N^+(\mu) - s_N^-(\mu)}{s(\mu) - s_N^-(\mu)} \ ; \tag{4.27}$$

it thus only remains to prove that $\eta_N(\mu)$ is greater than unity. From (4.24), and the definitions of $e(\mu)$ and $\mathcal{R}(\mu)$, we obtain

$$s(\mu) - s_N(\mu) = \langle\mathcal{A}(\mu)e(\mu), e(\mu)\rangle \tag{4.28}$$

$$= \langle\mathcal{R}(\mu), \mathcal{A}^{-1}(\mu)\mathcal{R}(\mu)\rangle \tag{4.29}$$

$$= \langle\mathcal{R}(\mu), e(\mu)\rangle \ , \tag{4.30}$$

and therefore

$$\eta_N(\mu) = \frac{\langle\mathcal{C}(\mu)\hat{e}(\mu), \hat{e}(\mu)\rangle}{\langle\mathcal{A}(\mu)e(\mu), e(\mu)\rangle} \tag{4.31}$$

from (4.15) and (4.28). Taking $\hat{A} = \mathcal{A}(\mu)$, $\hat{C} = \mathcal{C}(\mu)$, and $\hat{G} = \mathcal{R}(\mu)$ in (4.18)-(4.21), it then follows that

$$\rho_{\min}(\mu) \leq \eta_N(\mu) \leq \rho_{\max}(\mu) \ , \tag{4.32}$$

from (4.7), (4.8), (4.9), and (4.12). Furthermore, by construction, $\rho_{\min}(\mu) \geq 1$ for all $\mu \in \mathcal{D}^\mu$ and therefore

$$\eta_N(\mu) \geq 1 \ , \tag{4.33}$$

and $s_N^+(\mu) \geq s(\mu)$, as required. Note that the result (4.32) also indicates the sharpness of our bounds: it follows from (4.10) that

$$\eta_N(\mu) \leq \rho \ . \tag{4.34}$$

The result (4.32) also provides insight as to the properties of a *good* bound conditioner. Clearly, we wish $\rho_{\max}(\mu)$ to be as close to unity, and hence as close to $\rho_{\min}(\mu)$, as possible. We thus see that good bound conditioners are similar to good (iterative) preconditioners — both satisfy

$\rho_{\max}(\mu)/\rho_{\min}(\mu) \cong 1$ — except that bound conditioners must satisfy the additional spectral requirement $\rho_{\min}(\mu) \geq 1$.

4.2.4 Off-line/On-line Computational Procedure

We indicate here the off-line/on-line calculation of $\Delta_N(\mu)$. We recall from (4.16) that $\Delta_N(\mu) = \langle \mathcal{R}(\mu), \mathcal{C}^{-1}(\mu)\mathcal{R}(\mu)\rangle$. From the definition of the residual (4.6), the separability assumption (3.8), and the expansion of $u_N(\mu)$ in terms of the basis functions (3.108), we have

$$\begin{aligned}\langle \mathcal{R}(\mu), v\rangle &= \langle F, v\rangle - \langle \mathcal{A}(\mu)u_N(\mu), v\rangle \\ &= \langle F, v\rangle - \sum_{q=1}^{Q_\mathcal{A}}\sum_{n=1}^{N} \Theta^q(\mu)\, u_{N\,n}(\mu)\, \langle \mathcal{A}^q\zeta_n, v\rangle, \qquad \forall\, v \in Y,\ \forall\, \mu \in \mathcal{D}^\mu\ . \end{aligned} \tag{4.35}$$

We now invoke the "computational invertibility" hypothesis on $\mathcal{C}(\mu)$, (4.11), to write

$$\Delta_N(\mu) = \sum_{i\in\mathcal{I}(\mu)} \alpha_i(\mu)\, \langle \mathcal{R}(\mu), \mathcal{C}_i^{-1}\mathcal{R}(\mu)\rangle \tag{4.36}$$

$$= \sum_{i\in\mathcal{I}(\mu)} \alpha_i(\mu) \left\langle F - \sum_{q=1}^{Q_\mathcal{A}}\sum_{n=1}^{N} \Theta^q(\mu)\, u_{N\,n}(\mu)\, \mathcal{A}^q\zeta_n,\ \mathcal{C}_i^{-1}\left(F - \sum_{q'=1}^{Q_\mathcal{A}}\sum_{n'=1}^{N} \Theta^{q'}(\mu)\, u_{N\,n'}(\mu)\, \mathcal{A}^{q'}\zeta_{n'}\right)\right\rangle \tag{4.37}$$

$$\begin{aligned} = \sum_{i\in\mathcal{I}(\mu)} \alpha_i(\mu) \Bigg[& \langle F, \mathcal{C}_i^{-1}F\rangle \\ & - \sum_{q=1}^{Q_\mathcal{A}}\sum_{n=1}^{N} \Theta^q(\mu)\, u_{N\,n}(\mu)\, \left(\langle F, \mathcal{C}_i^{-1}\mathcal{A}^q\zeta_n\rangle + \langle \mathcal{A}^q\zeta_n, \mathcal{C}_i^{-1}F\rangle\right) \\ & + \sum_{q=1}^{Q_\mathcal{A}}\sum_{q'=1}^{Q_\mathcal{A}}\sum_{n=1}^{N}\sum_{n'=1}^{N} \Theta^q(\mu)\,\Theta^{q'}(\mu)\, u_{N\,n}(\mu)\, u_{N\,n'}(\mu)\, \langle \mathcal{A}^q\zeta_n, \mathcal{C}_i^{-1}\mathcal{A}^{q'}\zeta_{n'}\rangle \Bigg]\ . \end{aligned} \tag{4.38}$$

Thus, in the *off-line* stage, we compute the μ-independent inner products

$$c_i = \langle F, \mathcal{C}_i^{-1}F\rangle\ , \tag{4.39}$$

$$\Lambda_{qn}^i = -\langle F, \mathcal{C}_i^{-1}\mathcal{A}^q\zeta_n\rangle = -\langle \mathcal{A}^q\zeta_n, \mathcal{C}_i^{-1}F\rangle \tag{4.40}$$

$$\Gamma_{qq'nn'}^i = \langle \mathcal{A}^q\zeta_n, \mathcal{C}_i^{-1}\mathcal{A}^{q'}\zeta_{n'}\rangle\ , \tag{4.41}$$

$1 \leq i \leq I$, $1 \leq q, q' \leq Q_\mathcal{A}$, and $1 \leq n, n' \leq N$; note that (4.40) follows from the symmetry of $\mathcal{C}_i$. This requires QN $\mathcal{A}^q\zeta_n$ multiplications, $I(1+QN)$ $\mathcal{C}$-solves, and $I(Q^2N^2+QN+1)$ inner products.

In the *on-line* stage, given any new value of μ, we simply perform the sum

$$\Delta_N(\mu) = \sum_{i\in\mathcal{I}(\mu)} \alpha_i(\mu) \left[c_i + 2\sum_{q=1}^{Q_\mathcal{A}}\sum_{n=1}^{N} \Theta^q(\mu)\, u_{N\,n}(\mu)\, \Lambda^i_{qn} + \sum_{q=1}^{Q_\mathcal{A}}\sum_{q'=1}^{Q_\mathcal{A}}\sum_{n=1}^{N}\sum_{n'=1}^{N} \Theta^q(\mu)\, \Theta^{q'}(\mu)\, u_{N\,n}(\mu)\, u_{N\,n'}(\mu)\, \Gamma^i_{qq'nn'} \right] . \tag{4.42}$$

The on-line complexity is thus, to leading order, $O(|\mathcal{I}(\mu)|Q^2N^2)$ — and hence, independent of $\mathcal{N}$ (the dimension of Y).

Apart from the spectral condition and invertibility requirement of Section 4.2, we have not yet specifically defined, or provided methods for developing, the bound conditioner $\mathcal{C}(\mu)$; we shall address this issue in Sections 4.3 and 4.4.

4.3 Bound Conditioner Constructions — Type I

We recall from Section 4.2 that a bound conditioner is defined as a symmetric, continuous, and coercive distributional operator $\mathcal{C}(\mu)\colon Y \to Y'$ which satisfies (i) a spectral condition —

$$\rho_{\min}(\mu) \equiv \min_{v\in Y} \frac{\langle \mathcal{A}(\mu)v, v\rangle}{\langle \mathcal{C}(\mu)v, v\rangle} \geq 1 , \qquad \rho_{\max}(\mu) \equiv \max_{v\in Y} \frac{\langle \mathcal{A}(\mu)v, v\rangle}{\langle \mathcal{C}(\mu)v, v\rangle} \leq \rho , \tag{4.43}$$

for some (preferably small) constant $\rho \in \mathbb{R}$; and (ii) a computational invertibility condition —

$$\mathcal{C}^{-1}(\mu) = \sum_{i\in\mathcal{I}(\mu)} \alpha_i(\mu)\mathcal{C}_i^{-1} , \tag{4.44}$$

where $\mathcal{I}(\mu) \subset \{1,\ldots,I\}$ is a parameter-dependent set of indices, I is a finite (preferably small) integer, and the $\mathcal{C}_i\colon Y \to Y', i = 1,\ldots,I$, are parameter-independent symmetric, coercive operators.

In this section, we consider several classes of bound conditioners for which $|\mathcal{I}(\mu)| = 1$, such that (4.44) takes the simpler form

$$\mathcal{C}^{-1}(\mu) = \alpha_{\mathcal{I}(\mu)}(\mu)\, \mathcal{C}^{-1}_{\mathcal{I}(\mu)} ; \tag{4.45}$$

here, $\mathcal{I}(\mu) \in \{1,\ldots,I\}$ is a parameter-dependent indicator function. In this case, the problem of constructing good bound conditioners then reduces to finding μ-independent operators $\mathcal{C}_i$, $i = 1,\ldots,I$, and associated μ-dependent coefficients $\alpha_i(\mu)$ such that the spectral condition (4.43) is satisfied.

We note from (4.43) that the $\mathcal{C}_i$ must be "equivalent" to $\mathcal{A}$ in the sense that the norms $\|\cdot\|_{\mathcal{C}_i} = \langle \mathcal{C}_i\cdot,\cdot\rangle$ and $\|\cdot\|_\mathcal{A} = \langle \mathcal{A}\cdot,\cdot\rangle$ are equivalent — there exists positive constants a and b such that [31]

$$a\|v\|_{\mathcal{C}_i} \leq \|v\|_\mathcal{A} \leq b\|v\|_{\mathcal{C}_i} , \quad \forall\, v \in Y . \tag{4.46}$$

Furthermore, if we define $\rho^i_{\min}(\mu)$ and $\rho^i_{\max}(\mu)$ as

$$\rho^i_{\min}(\mu) \equiv \min_{v\in Y} \frac{\langle \mathcal{A}(\mu)v, v\rangle}{\langle \mathcal{C}_i\, v, v\rangle} , \qquad \rho^i_{\max}(\mu) \equiv \max_{v\in Y} \frac{\langle \mathcal{A}(\mu)v, v\rangle}{\langle \mathcal{C}_i\, v, v\rangle} , \tag{4.47}$$

then for $C(\mu) = \alpha_i(\mu)^{-1}\mathcal{C}_i$, $\rho_{\min}$ and $\rho_{\max}$ are given by

$$\rho_{\min}(\mu) = \rho^i_{\min}(\mu)\alpha_i(\mu) \ , \qquad \rho_{\max}(\mu) = \rho^i_{\max}(\mu)\alpha_i(\mu) \tag{4.48}$$

Therefore, given any "$\mathcal{A}$-equivalent," μ-independent $\mathcal{C}_i$, $\alpha_i(\mu)^{-1}\mathcal{C}_i$ is a bound conditioner if and only if $\alpha_i(\mu)^{-1}$ is a lower bound to the minimum eigenvalue:

$$\alpha_i(\mu)^{-1} \leq \rho^i_{\min}(\mu) \ , \quad \forall\ \mu \in \mathcal{D}^\mu \ ; \tag{4.49}$$

furthermore, since the effectivities are bounded from above and below by $\rho_{\max}(\mu)$ and $\rho_{\min}(\mu)$, respectively, $\alpha_i(\mu)^{-1}$ must be a sharp lower bound to $\rho^i_{\min}(\mu)$, and $\rho^i_{\max}(\mu)$ must be as close to $\rho^i_{\min}(\mu)$ as possible.

We now consider several methods for choosing the $\mathcal{C}_i$ and the associated $\alpha_i(\mu)$.

4.3.1 Minimum Coefficient Bound Conditioner

To begin, we recall our separability assumption on $\mathcal{A}(\mu)$:

$$\mathcal{A}(\mu) = \sum_{q=1}^{Q_\mathcal{A}} \Theta^q(\mu)\, \mathcal{A}^q \ , \quad \forall\ \mu \in \mathcal{D}^\mu \ , \tag{4.50}$$

where the $\Theta^q(\mu) : \mathcal{D}^\mu \to \mathbb{R}$ and the $\mathcal{A}^q \colon Y \to Y'$. We now define

$$A(\theta) \equiv \sum_{q=1}^{Q_\mathcal{A}} \theta^q\, A^q \tag{4.51}$$

where $\theta \in \mathbb{R}^{Q_\mathcal{A}}$, $A(\theta) \colon Y \to Y'$, and $A^q \equiv \mathcal{A}^q$. We may then write

$$A(\Theta(\mu)) = \mathcal{A}(\mu) \ , \tag{4.52}$$

where $\Theta \colon \mathcal{D}^\mu \to \mathcal{D}^\theta$, and $\mathcal{D}^\theta \equiv \text{Range}(\Theta) \subset \mathbb{R}^{Q_\mathcal{A}}$. We now assume that $A(\theta)$ is coercive for any $\theta \in \mathcal{D}^\theta$, and the A^q (and therefore the $\mathcal{A}^q$) are symmetric, positive semi-definite, i.e.,

$$\langle A^q v, v\rangle \geq 0 \ , \quad \forall\ v \in Y \ , \ 1 \leq q \leq Q_\mathcal{A} \ ; \tag{4.53}$$

and $\mathcal{D}^\theta \subset \mathbb{R}^{Q_\mathcal{A}}_+$, where $\mathbb{R}_+$ refers to the positive real numbers.

Bound Conditioner

We then choose I points $\bar{\theta}_i \in \mathcal{D}^\theta$, $i = 1, \ldots, I$, and define $\alpha_i(\mu)$, $\mathcal{C}_i$, and $C(\mu)$ as

$$\alpha_i(\mu) = \left(\min_{1\leq q\leq Q_\mathcal{A}} \left(\frac{\Theta^q(\mu)}{\bar{\theta}^q_i} \right) \right)^{-1} \ , \qquad \mathcal{C}_i = \sum_{q=1}^{Q_\mathcal{A}} \bar{\theta}^q_i A^q \ , \tag{4.54}$$

and $C(\mu) = \alpha_{\mathcal{I}(\mu)}(\mu)^{-1} C_{\mathcal{I}(\mu)}$, respectively, where $\mathcal{I}(\mu)$ selects the "best" among all the points $\bar{\theta}_i$. The effectivity then satisfies

$$1 \leq \eta_N(\mu) \leq \max_{1 \leq m \leq I} \left(\frac{\max\limits_{1 \leq q \leq Q_A} \left(\frac{\Theta^q(\mu)}{\bar{\theta}_i^q} \right)}{\min\limits_{1 \leq q \leq Q_A} \left(\frac{\Theta^q(\mu)}{\bar{\theta}_i^q} \right)} \right) , \quad \forall\, v \in Y \,, \forall\, \mu \in \mathcal{D}^\mu \,. \tag{4.55}$$

We note that if $\Theta(\mu) = \bar{\theta}_i$, then $\alpha_i(\mu) = 1$, $\langle C_i v, v \rangle = \langle \mathcal{A}(\mu) v, v \rangle$, and therefore $\eta_N(\mu) = 1$. We also note that given any new $\mu \in \mathcal{D}^\mu$, there are, in fact, I possible choices for $\mathcal{I}(\mu)$; we can either establish a definition of closeness and choose $\mathcal{I}(\mu)$ such that $\bar{\theta}_{\mathcal{I}(\mu)}$ is closest to $\Theta(\mu)$; or, alternatively, consider all possible candidates and select that which yields the best effectivity.

Proof of Bounding Properties

To prove (4.55), we note that for any $\mu \in \mathcal{D}^\mu$, $i \in \{1, \ldots, I\}$,

$$\begin{aligned}
\langle \mathcal{A}(\mu) v, v \rangle &= \langle A(\Theta(\mu)) v, v \rangle && (4.56)\\
&= \left\langle \sum_{q=1}^{Q_A} \Theta^q(\mu) A^q v, v \right\rangle && (4.57)\\
&= \left\langle \sum_{q=1}^{Q_A} \left(\frac{\Theta^q(\mu)}{\bar{\theta}_i^q} \right) \bar{\theta}_i^q A^q v, v \right\rangle && (4.58)\\
&\geq \left\langle \left(\min_{1 \leq q \leq Q_A} \left(\frac{\Theta^q(\mu)}{\bar{\theta}_i^q} \right) \right) \left(\sum_{q=1}^{Q_A} \bar{\theta}_i^q A^q \right) v, v \right\rangle && (4.59)\\
&= \langle \alpha_i(\mu)^{-1} C_i v, v \rangle, \quad \forall\, v \in Y \,. && (4.60)
\end{aligned}$$

Furthermore, for any $\mu \in \mathcal{D}^\mu$, $i \in \{1, \ldots, I\}$, we have

$$\begin{aligned}
\langle \mathcal{A}(\mu) v, v \rangle &\leq \left\langle \left(\max_{1 \leq q \leq Q_A} \left(\frac{\Theta^q(\mu)}{\bar{\theta}_i^q} \right) \right) \left(\sum_{q=1}^{Q_A} \bar{\theta}_i^q A^q \right) v, v \right\rangle \,, && (4.61)\\
&= \frac{\max\limits_{1 \leq q \leq Q_A} \left(\frac{\Theta^q(\mu)}{\bar{\theta}_i^q} \right)}{\min\limits_{1 \leq q \leq Q_A} \left(\frac{\Theta^q(\mu)}{\bar{\theta}_i^q} \right)} \langle \alpha_i(\mu)^{-1} C_i v, v \rangle \,, \quad \forall\, v \in Y \,. && (4.62)
\end{aligned}$$

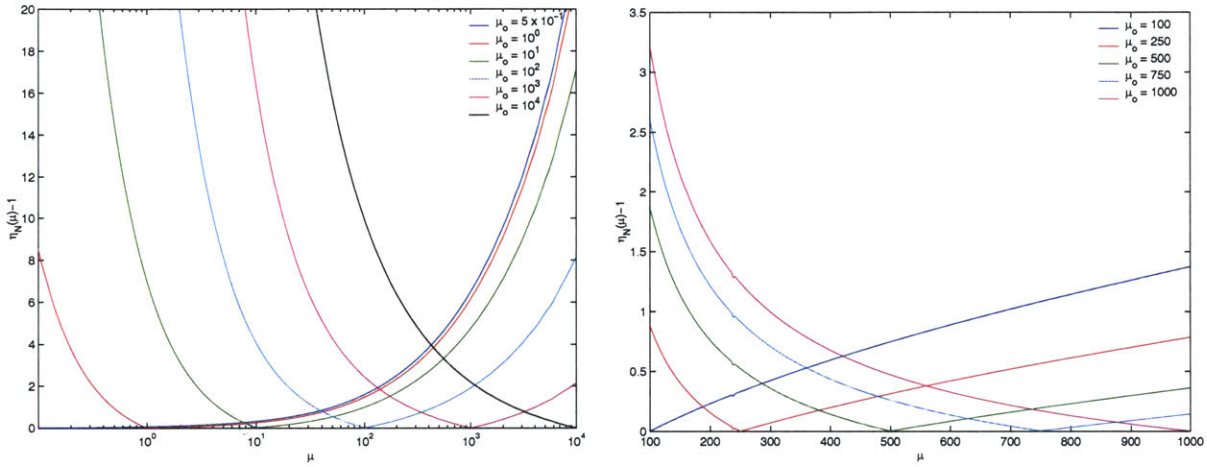
The result (4.55) directly follows from (4.59) and (4.62).

Remarks

We present in Figure 4-1, Table 4.1, and Figure 4-2 the effectivities for Examples 1-3, respectively, and $\bar{\theta} = \mu_0$ for different values of μ_0. Note that $\eta_N(\mu) - 1 > 0$ in all cases, and, as illustrated in Figures 4-1 and 4-2, $\eta_N(\mu_0) = 1$.

Numerical tests also confirm that the sharpness of the resulting bounds does depend on the choice of $\bar{\theta}$. In particular, we present in Figure 4-1(a) the effectivities for Example 1, calculated

as a function of μ for the entire range $\mathcal{D}^\mu$. For this example we choose $\bar{\theta} = \Theta(\mu_0)$ for different values of μ_0. We present similar data in Figure 4-1(b) also for Example 1 but for a smaller range ($\mu \in [10^2, 10^3]$), and in Figure 4-2 for Example 3. We observe that in all three cases the effectivities increase as $|\mu - \mu_0| \to \infty$, and the rate of increase is greater for $\mu < \mu_0$. Nevertheless, the results show that even for $I = 1$ — *if* μ_0 is chosen "optimally" — the resulting effectivities are remarkably low over large ranges of μ ($\eta_N(\mu) - 1 \leq 20 \; \forall \; \mu \in \mathcal{D}^\mu$ for $\mu_0 \sim 1.0$). The existence of an "optimal" choice of μ_0 is made evident in (*i*) Figure 4-1(b) — for Example 1 and $\mu \in [10^2, 10^3]$, $\mu_0 \sim 2.5 \times 10^2$ yields the best effectivities; (*ii*) Table 4.1 — for Example 2 and $\mu = \{\mu_1, \mu_2\} \in [\mu^1_{\min}, \mu^1_{\max}] \times [\mu^2_{\min}, \mu^2_{\max}]$, $\mu_0 \sim \{\mu^1_{\min}, \mu^2_{\min}\}$ yields the best effectivities; and (*iii*) Figure 4-2 — for Example 3 and $\mu \in [0.1, 1.0]$, $\mu_0 \sim 0.1$ yields the best effectivities.



(a) (b)

Figure 4-1: Effectivity as a function of μ for Example 1 and (a) $\mu \in [0.01, 10^4]$, and (b) $\mu \in [10^2, 10^3]$, calculated using the minimum coefficient bound conditioner with $\bar{\theta} = \Theta(\mu_0)$ for different choices of μ_0.

$\mu_0 = \{\mu^1_0, \mu^2_0\}$	$\min_\mu (\eta_N(\mu) - 1)$	$\max_\mu (\eta_N(\mu) - 1)$	$\text{ave}_\mu (\eta_N(\mu) - 1)$
$\{\mu^1_{\min}, \mu^2_{\min}\}$	0.04	4.19	0.65
$\{\mu^1_{\min}, \mu^2_{\max}\}$	0.24	259.87	25.37
$\{\mu^1_{\max}, \mu^2_{\min}\}$	0.75	333.96	40.34
$\{\mu^1_{\max}, \mu^2_{\max}\}$	1.58	325.49	45.87
$\left\{\frac{\mu^1_{\min} + \mu^1_{\max}}{2}, \frac{\mu^2_{\min} + \mu^2_{\max}}{2}\right\}$	0.79	210.66	30.24

Table 4.1: Minimum and maximum effectivity for Example 2 calculated using the minimum coefficient bound conditioner with $\bar{\theta} = \Theta(\mu_0)$ for different choices of μ_0.

The numerical results presented for Examples 1-3 illustrate that the minimum coefficient bound conditioner guarantees rigorous bounds, and yields good effectivities. There is, however, one disad-

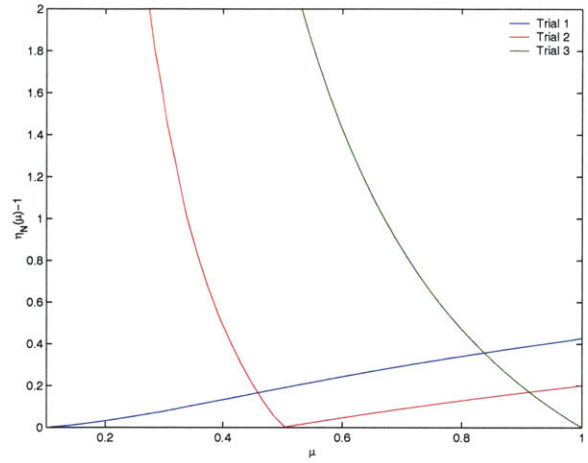


Figure 4-2: Effectivity as a function of μ for Example 3 and $\mu \in [0.1, 1.0]$ calculated using the minimum coefficient bound conditioner with $\bar{\theta} = \Theta(\mu_0)$ for different choices of μ_0.

vantage: the method is restricted to problems for which the A^q are positive semidefinite, as stated in (4.53). Although many property, boundary condition, load, and some geometry variations yield A^q's with the required properties, there are many problems which do not admit such a representation. For example, while the "stretch" geometry variation of Example 3 yields A^q's which satisfy (4.53), the addition of a "shear" parameter in Example 4 does not. We therefore need to consider other bound conditioner constructions — particularly for more general geometry variations.

4.3.2 Eigenvalue Interpolation: Quasi-Concavity in μ

In this section, we investigate the possibility of constructing bound conditioners based on observations on the behavior of the relevant eigenvalues with respect to the parameter. In particular, given some $\mathcal{C}_1$ (preferably chosen "optimally," such that $\rho^i_{\max}/\rho^i_{\min}$ is small), we wish to find a lower bound $\alpha_1(\mu)^{-1}$ to $\rho^i_{\min}$ based on the behavior of $\rho^i_{\min}$ as a function of μ.

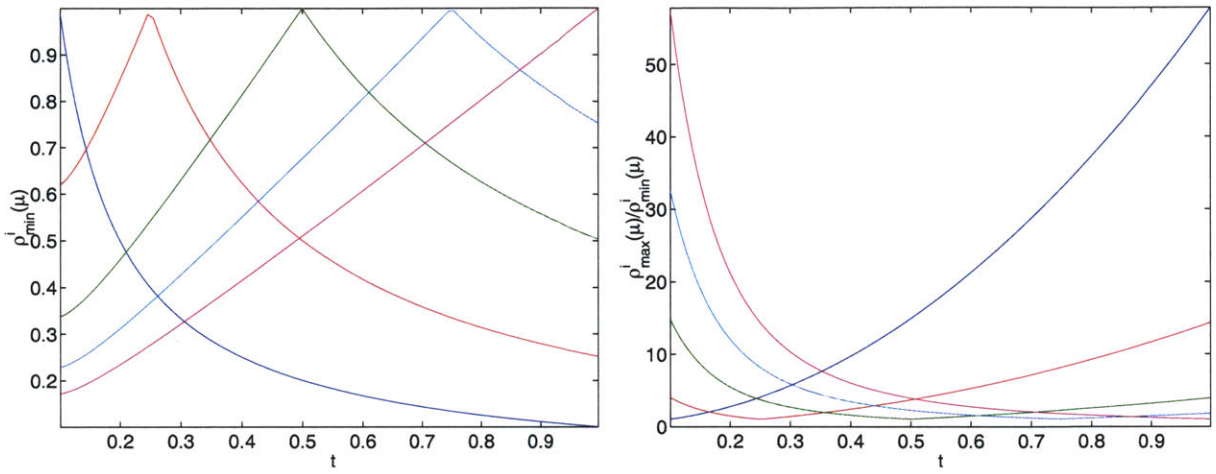
We first consider Example 3, and plot in Figure 4-3(a) the eigenvalues $\rho^i_{\min}$ as a function of the stretch parameter $\bar{t}$, calculated using $\mathcal{C}_1 = \mathcal{A}(\mu_0)$ for several values of $\mu_0 = \{t_0\}$. We note that for this example, and for all choices of $\mathcal{C}_1$, the eigenvalues appear to be *quasi-concave* in μ — a function $\hat{f}(\hat{x})$ is quasi-concave in $\hat{x} \in \mathcal{D}^{\hat{x}} \subset \mathbb{R}^1$ if, for any $\hat{x}_1 \in \mathcal{D}^{\hat{x}}$ and $\hat{x}_2 \in \mathcal{D}^{\hat{x}}$ such that $\hat{x}_1 < \hat{x}_2$,

$$\hat{f}(\hat{x}) \geq \min[\hat{f}(\hat{x}_1), \hat{f}(\hat{x}_2)] \ , \qquad \forall\ \hat{x} \in [\hat{x}_1, \hat{x}_2] \ . \tag{4.63}$$

Furthermore, the plot of $\rho^i_{\max}/\rho^i_{\min}$ Figure 4-3(b) again illustrates the existence of an "optimal" μ_0: the ratio of the maximum and minimum eigenvalues is smallest for $\bar{t}_0 \sim 0.25$.

We present similar results in Figures 4-4 and 4-5 for Example 4 for which we take $\mathcal{C}_1 = \mathcal{A}(\mu_0)$ with $\mu_0 = \{\bar{t}_0, \bar{\alpha}_0\} = \{0.25, 0.0\}$ (as suggested by the results for Example 3). We observe that $\rho^i_{\min}$ appears to be quasi-concave with respect to $\bar{t}$ *and* $\bar{\alpha}$. Furthermore, the ratio $\rho^i_{\max}/\rho^i_{\min}$ as plotted in Figures 4-4 and 4-5 are not too large, indicating that our choice of $\mathcal{C}_1$ is relatively good.

We now construct a bound conditioner based on these observations.



(a) (b)

Figure 4-3: Plot of (a) the eigenvalues $\rho^i_{\min}$ and (b) the ratio $\rho^i_{\max}/\rho^i_{\min}$ as a function of $\bar{t}$ for Example 3, with $\mathcal{C}_1 = \mathcal{A}(\mu_0)$, $\mu_0 = \{\bar{t}_0\}$, and $\bar{t}_0 = 0.10, 0.25, 0.50, 0.75$, and 1.00.

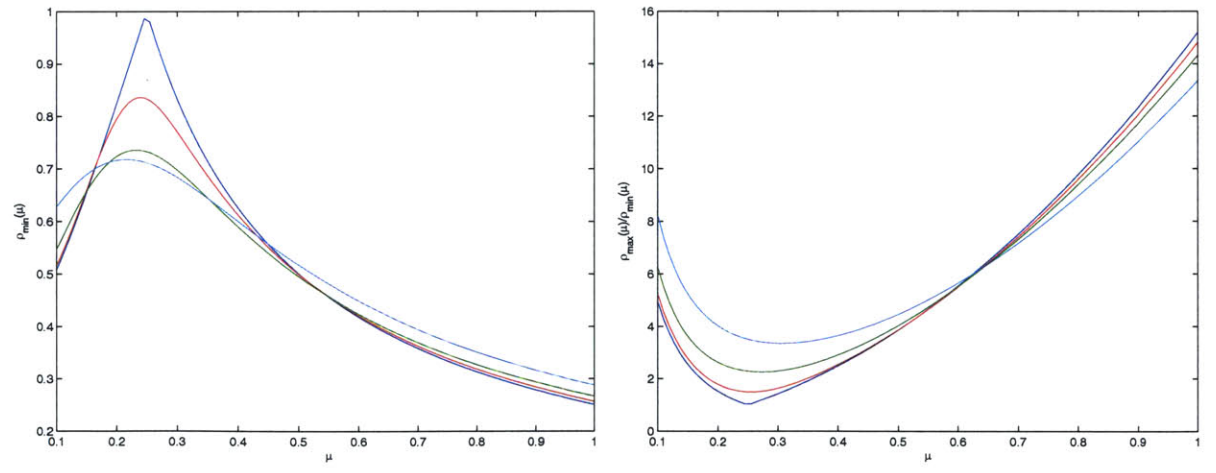


Figure 4-4: Contours of the eigenvalues $\rho^i_{\min}$ as a function of $\bar{t}$ for Example 4, with $\mathcal{C}_1 = \mathcal{A}(\mu_0)$, $\mu_0 = \{\bar{t}_0, \bar{\alpha}_0\} = \{0.25, 0.0\}$. The contours are calculated at constant $\bar{\alpha}$, for $\bar{\alpha} = 0°$, 15°, 30°, and 45°.

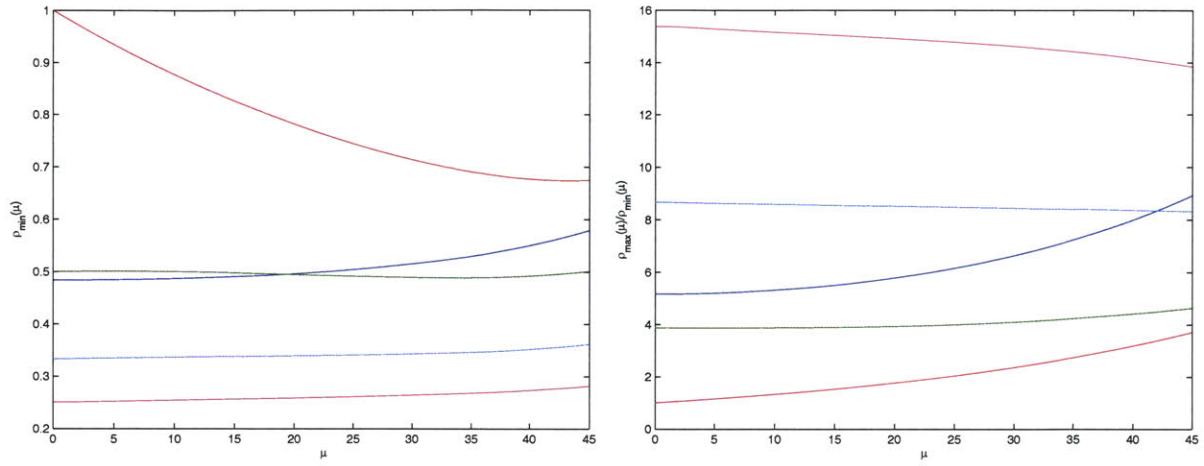


Figure 4-5: Contours of the eigenvalues $\rho^i_{\min}$ as a function of $\bar{\alpha}$ for Example 4, with $\mathcal{C}_1 = \mathcal{A}(\mu_0)$, $\mu_0 = \{\bar{t}_0, \bar{\alpha}_0\} = \{0.25, 0.0\}$. The contours are calculated at constant $\bar{t}$, for $\bar{t} = 0.10$, 0.25, 0.50, 0.75, and 1.0.

Bound Conditioner

To begin, we introduce a sample $S^\mu_T = \{\mu_1, \cdots, \mu_T\}$, where $\mu_m \in \mathcal{D}^\mu$, $m = 1, \ldots, T$. We also assume that, given any $\mu \in \mathcal{D}^\mu$, we can find a simplex $\mathcal{S}(\mu)$ such that $\mu \in \mathcal{S}(\mu)$, where

$$\mathcal{S}(\mu) \equiv \left\{ \hat{\mu} \mid \hat{\mu} = \sum_{m \in \mathcal{T}(\mu)} \hat{a}_m \mu_m \right\} \; ; \tag{4.64}$$

here, the $\hat{a}_m$ are *weights* —

$$0 \leq \hat{a}_m \leq 1 \; , \quad \forall \; m \in \mathcal{T}(\mu) \; , \tag{4.65}$$

and

$$\sum_{m \in \mathcal{T}(\mu)} \hat{a}_m = 1 \; . \tag{4.66}$$

Given any μ-independent symmetric, continuous, coercive norm-equivalent operator $\mathcal{C}_1$, we then define the minimum eigenvalues $\rho^{1\,m}_{\min}$, $m = 1, \ldots, T$, as

$$\rho^{1\,m}_{\min} = \min_{v \in Y} \frac{\langle \mathcal{A}(\mu_m) v, v \rangle}{\langle \mathcal{C}_1 v, v \rangle} \; . \tag{4.67}$$

We now make the hypothesis that $\rho^1_{\min}(\mu)$ is quasi-concave with respect to μ for $\mu \in \mathcal{D}^\mu \subset \mathbb{R}^P$ — that is,

$$\rho^1_{\min}(\mu) \geq \min_{m \in \mathcal{T}(\mu)} \rho^{1\,m}_{\min} \; ; \tag{4.68}$$

our empirical results indicate that this hypothesis is certainly plausible for the case of heat conduction; however, we have no theoretical proof at present.

To each $\mu \in \mathcal{D}^\mu$, we then define $\alpha_1(\mu)$ and $\mathcal{C}(\mu)$ as

$$\alpha_1(\mu) = \left(\min_{m \in \mathcal{T}(\mu)} \rho_{\min}^{1\,m} \right)^{-1} , \tag{4.69}$$

and $\mathcal{C}(\mu) = \alpha_1(\mu)^{-1}\mathcal{C}_1$, respectively. Note that if (4.68) is true, it directly follows that $\alpha_1(\mu)^{-1}$ is a lower bound to $\rho_{\min}^1(\mu)$, as required.

Remarks

We present in Figure 4-6 the effectivities for Example 3 obtained using a uniform sample S_T^μ with $T = 2$ and 4; we present similar results in Table 4.2 for Example 4 obtained using a uniform sample S_T^μ with $T = 4$, 9, and 25. We note that we obtain good effectivities (and therefore sharp bounds) even for small T, and, as expected, the effectivities improve as the sample size T is increased.

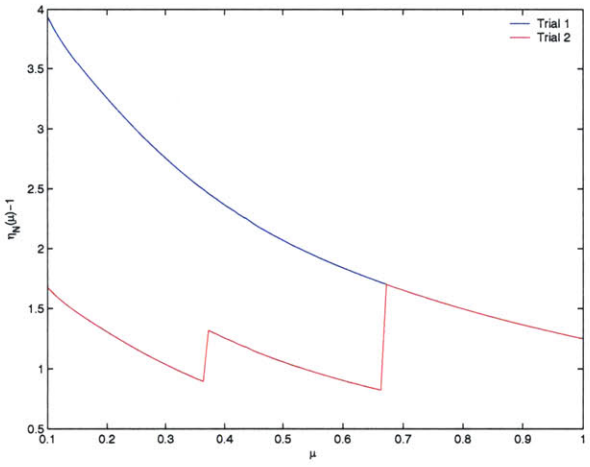


Figure 4-6: Effectivity as a function of μ for Example 3, obtained using the quasi-concave eigenvalue bound conditioner with $\mathcal{C}_1 = \mathcal{A}(\mu_0)$, $\mu_0 = \bar{t}_0 = 0.25$, and a uniform sample S_T^μ with $T = 4$ (Trial 1) and $T = 4$ (Trial 2).

T	$\min_\mu \eta_N(\mu)$	$\max_\mu \eta_N(\mu)$	ave $\eta_N(\mu)$
4	0.6075	13.0946	2.4175
9	0.2483	7.5051	1.3841
25	0.2357	6.8562	1.2175

Table 4.2: Minimum, maximum, and average effectivity over $\mu \in \mathcal{D}^\mu$ for Example 4, obtained using the quasi-concave eigenvalue bound conditioner with $\mathcal{C}_1 = \mathcal{A}(\mu_0)$, $\mu_0 = \{\bar{t}_0, \bar{\alpha}_0\} = \{0.25, 0.0\}$, and a uniform sample S_T^μ.

The main disadvantage of this method is in the (current) lack of rigor: the bound conditioner is constructed based on empirical observations of the behavior of the relevant eigenvalues. We thus turn to more rigorous approaches.

4.3.3 Eigenvalue Interpolation: Concavity in θ

To begin, we define $A(\theta)$ for any $\theta \in \mathbb{R}^{Q_{\mathcal{A}}}$ as

$$A(\theta) = \sum_{q=1}^{Q_{\mathcal{A}}} \theta^q A^q \tag{4.70}$$

where $A(\theta)\colon Y \rightarrow Y'$, and $A^q \equiv \mathcal{A}^q$, such that $A(\Theta(\mu)) = \mathcal{A}(\mu)$ for $\Theta\colon \mathcal{D}^\mu \subset \mathbb{R}^P \rightarrow \mathcal{D}^\theta \subset \mathbb{R}^{Q_{\mathcal{A}}}$. We shall also define $\theta_{\min}(\geq 0)$, $\theta_{\max}$ (assumed finite), and $\mathcal{D}^\theta_{\text{box}} \subset \mathbb{R}^{Q_{\mathcal{A}}}$ as

$$\theta^q_{\min} \equiv \sup \hat{\vartheta}_{\{\hat{\vartheta}\in\mathbb{R} \mid \Theta^q(\mu)\geq\hat{\vartheta},\ \forall\ \mu\in\mathcal{D}^\mu\}}\ , \quad q = 1,\ldots,Q_{\mathcal{A}} \tag{4.71}$$

$$\theta^q_{\max} \equiv \inf \hat{\vartheta}_{\{\hat{\vartheta}\in\mathbb{R} \mid \Theta^q(\mu)\leq\hat{\vartheta},\ \forall\ \mu\in\mathcal{D}^\mu\}}\ , \quad q = 1,\ldots,Q_{\mathcal{A}}\ , \tag{4.72}$$

and

$$\mathcal{D}^\theta_{\text{box}} \equiv \prod_{q=1}^{Q_{\mathcal{A}}} [\theta^q_{\min}, \theta^q_{\max}]\ , \tag{4.73}$$

respectively. We assume that $A(\Theta(\mu))$ is continuous, symmetric, and coercive for all $\mu \in \mathcal{D}^\mu$. We further assume that the A^q, $q = 1,\ldots,Q_{\mathcal{A}}$, are symmetric and continuous (but not necessarily coercive); note that $A(\theta)$ will be coercive in some neighborhood of $\mathcal{D}^\theta$, and symmetric for all $\theta \in \mathbb{R}^{Q_{\mathcal{A}}}$.

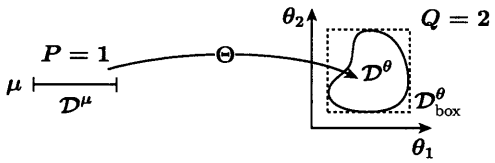


Figure 4-7: Mapping (Θ) between $\mathcal{D}^\mu \in \mathbb{R}^P$ to $\mathcal{D}^\theta \in \mathbb{R}^{Q_{\mathcal{A}}}$, for $P = 1$ and $Q = 2$.

We now consider a class of bound conditioners based on direct approximation to the minimum eigenvalue. For simplicity of exposition, we shall consider $I = 1$ and $\mathcal{I}(\mu) = \{1\}$. The extension to $I > 1$ is straightforward.

Bound Conditioner

We now introduce a "θ" sample $S^\theta_T = \{\theta_1,\ldots,\theta_T\}$, where $\theta_m \in \mathcal{D}^\theta_{\text{box}}$, $m = 1,\ldots,T$, and $T \geq Q+1$. We assume that, given any $\theta \in \mathcal{D}^\theta_{\text{box}}$, we can find a simplex $\mathcal{S}(\theta)$ such that $\theta \in \mathcal{S}(\theta)$, where

$$\mathcal{S}(\theta) \equiv \left\{ \hat{\vartheta} \mid \hat{\vartheta} = \sum_{m\in\mathcal{T}(\theta)} \hat{a}_m\, \theta_m \right\}\ ; \tag{4.74}$$

here, the $\hat{a}_m$ are *weights* —

$$\hat{a}_m \geq 0\ , \quad \forall\ m \in \mathcal{T}(\theta)\ , \tag{4.75}$$

and

$$\sum_{m \in \mathcal{T}(\theta)} \hat{a}_m = 1 \ . \tag{4.76}$$

We implicitly assume that S_T^θ is chosen such that, for all $\theta \in \mathcal{D}_{\text{box}}^\theta$, such a construction is possible; a deficient sample S_T^θ can always be rendered compliant simply by replacing one point with $\theta_{\min}$.

Given any μ-independent symmetric, continuous, coercive norm-equivalent operator $\mathcal{C}_1$, we define the minimum eigenvalues $\rho_{\min}^{1\,m}$, $i = 1, \ldots, T$, as

$$\rho_{\min}^{1\,m} = \min_{v \in Y} \frac{\langle A(\theta_m)v, v\rangle}{\langle \mathcal{C}_1 v, v\rangle} \ . \tag{4.77}$$

To each $\mu \in \mathcal{D}^\mu$, we then define $\alpha_1(\mu)$ and $\mathcal{C}(\mu)$ as

$$\alpha_1(\mu) = \left(\sum_{m \in \mathcal{T}(\Theta(\mu))} a_m(\Theta(\mu)) \rho_{\min}^{1\,m} \right)^{-1} \ , \tag{4.78}$$

and $\mathcal{C}(\mu) = \alpha_1(\mu)^{-1}\mathcal{C}_1$, respectively.

Proof of Bounding Properties

To prove that our bound conditioner satisfies (4.43), we need only show that

$$\alpha_1(\mu)^{-1} \leq \rho_{\min}^1(\Theta(\mu)) \ , \quad \forall\ \mu \in \mathcal{D}^\mu \ , \tag{4.79}$$

where we recall $\rho_{\min}^1(\theta)$ is defined (in (4.47)) as the minimum eigenvalue

$$\rho_{\min}^1(\theta) = \min_{v \in Y} \frac{\langle A(\theta)v, v\rangle}{\langle \mathcal{C}_1 v, v\rangle} \ . \tag{4.80}$$

To begin, we note that if a function $\hat{f}(\theta) \in \mathbb{R}$ is *concave* in the set $\hat{\mathcal{D}}_\theta$, then for any $\hat{\theta}_1, \ldots, \hat{\theta}_{\hat{T}} \in \hat{\mathcal{D}}_\theta$,

$$\hat{f}(\hat{a}_1\hat{\theta}_1 + \cdots + \hat{a}_{\hat{T}}\hat{\theta}_{\hat{T}}) \geq \hat{a}_1\hat{f}(\hat{\theta}_1) + \cdots + \hat{a}_{\hat{T}}\hat{f}(\hat{\theta}_{\hat{T}}) \ , \tag{4.81}$$

where the weights $\hat{a}_m$, $1 \leq m \leq \hat{T}$ satisfy (4.158) and (4.159) [6]. We now claim that the eigenvalue $\rho_{\min}^1(\theta)$ given by (4.80) is concave for all $\theta \in \mathbb{R}^{Q_\mathcal{A}}$; (4.79) then follows from the choice of the a_m and the definition of $\alpha_1(\mu)$. It thus remains only to demonstrate the concavity of $\rho_{\min}^1(\theta)$.

We first define $A^{\text{seg}}(\tau; \theta_1, \theta_2) \colon Y \to Y'$ for any two points $\theta_1 \in \mathbb{R}^{Q_\mathcal{A}}$, $\theta_2 \in \mathbb{R}^{Q_\mathcal{A}}$ as

$$\begin{aligned} A^{\text{seg}}(\tau; \theta_1, \theta_2) &= A(\theta_1 + \tau(\theta_2 - \theta_1)) && (4.82) \\ &= \sum_{q=1}^{Q_\mathcal{A}} \theta_1^q A^q + \tau \sum_{q=1}^{Q_\mathcal{A}} (\theta_2^q - \theta_1^q) A^q \ , && (4.83) \\ &\equiv A_0^{\text{seg}}(\theta_1, \theta_2) + \tau A_1^{\text{seg}}(\theta_1, \theta_2) \ , && (4.84) \end{aligned}$$

for $\tau \in [0, 1]$. We then define $\rho^{\text{seg}}(\tau; \theta_1, \theta_2)$ as the eigenvalue and $\xi^{\text{seg}}(\tau; \theta_1, \theta_2)$ as the corresponding eigenvector of

$$\langle A^{\text{seg}}(\tau; \theta_1, \theta_2)\xi^{\text{seg}}(\tau; \theta_1, \theta_2), v\rangle = \lambda^{\text{seg}}(\tau; \theta_1, \theta_2)\langle \mathcal{C}_1\, \xi^{\text{seg}}(\tau; \theta_1, \theta_2), v\rangle \ , \quad \forall\ v \in Y \ , \tag{4.85}$$

such that the eigenvectors are normalized with respect to $\mathcal{C}_1$:

$$\langle \mathcal{C}_1 \xi^{\text{seg}}(\tau;\theta_1,\theta_2), \xi^{\text{seg}}(\tau;\theta_1,\theta_2)\rangle = 1 \ . \tag{4.86}$$

Defining $\rho^{\text{seg}}_{\min}$ to be the minimum eigenvalue of (4.85), we can then write

$$\rho^{\text{seg}}_{\min}(\tau;\theta_1,\theta_2) = \rho^1_{\min}(\theta_1 + \tau(\theta_2-\theta_1)) \ . \tag{4.87}$$

We would therefore like to show that for any $\theta_1 \in \mathbb{R}^{Q_\mathcal{A}}$, $\theta_2 \in \mathbb{R}^{Q_\mathcal{A}}$,

$$\frac{d^2}{d\tau^2}\rho^{\text{seg}}_{\min}(\tau;\theta_1,\theta_2) \le 0 \ , \quad \forall\, \tau \in [0,1] \ , \tag{4.88}$$

from which it follows that $\rho^{\text{seg}}_{\min}(\theta)$ is concave for all $\theta \in \mathbb{R}^{Q_\mathcal{A}}$.

For brevity, we shall henceforth omit the dependence on (θ_1,θ_2).

We first differentiate the normalization (4.86) with respect to τ to obtain

$$\left\langle \mathcal{C}_1 \frac{d}{d\tau}\xi^{\text{seg}}(\tau), \xi^{\text{seg}}(\tau)\right\rangle = 0 \ , \tag{4.89}$$

since $\mathcal{C}_1$ is symmetric and independent of τ. We then differentiate (4.85) with respect to τ to obtain

$$\left\langle A^{\text{seg}}(\tau)\frac{d}{d\tau}\xi^{\text{seg}}(\tau), v\right\rangle - \rho^{\text{seg}}(\tau)\left\langle \mathcal{C}_1\frac{d}{d\tau}\xi^{\text{seg}}(\tau), v\right\rangle + \langle A^1_{\text{seg}}(\tau)\xi^{\text{seg}}(\tau), v\rangle - \frac{d}{d\tau}\rho^{\text{seg}}(\tau)\langle \mathcal{C}_1\xi^{\text{seg}}(\tau), v\rangle = 0 \ , \quad \forall\, v \in Y \ . \tag{4.90}$$

We now choose $v = \xi^{\text{seg}}(\tau)$ in (4.90) and exploit symmetry, (4.85), and (4.86), to obtain

$$\frac{d}{d\tau}\rho^{\text{seg}}(\tau) = \langle A^1_{\text{seg}}(\tau)\xi^{\text{seg}}(\tau), \xi^{\text{seg}}(\tau)\rangle \tag{4.91}$$

We then choose $v = d\xi^{\text{seg}}(\tau)/d\tau \in Y$ in (4.90), and exploit symmetry and (4.89) to obtain

$$\left\langle A^{\text{seg}}(\tau)\bar{\xi}'(\tau), \frac{d}{d\tau}\bar{\xi}(\tau)\right\rangle - \rho^{\text{seg}}(\tau)\left\langle \mathcal{C}_1\frac{d}{d\tau}\bar{\xi}(\tau), \frac{d}{d\tau}\bar{\xi}(\tau)\right\rangle + \left\langle A^1_{\text{seg}}\frac{d}{d\tau}\bar{\xi}(\tau), \xi^{\text{seg}}(\tau)\right\rangle = 0 \ . \tag{4.92}$$

We now differentiate (4.91) with respect to τ to obtain

$$\frac{d^2}{d\tau^2}\rho^{\text{seg}}(\tau) = 2\left\langle A^1_{\text{seg}}\frac{d}{d\tau}\bar{\xi}(\tau), \xi^{\text{seg}}(\tau)\right\rangle \tag{4.93}$$

since A^{seg}_1 is symmetric and independent of τ. We then obtain from (4.92)

$$\frac{d^2}{d\tau^2}\rho^{\text{seg}}(\tau) \ = \ 2\left(\rho^{\text{seg}}(\tau)\left\langle \mathcal{C}_1\frac{d}{d\tau}\bar{\xi}(\tau), \frac{d}{d\tau}\bar{\xi}(\tau)\right\rangle - \left\langle A^{\text{seg}}(\tau)\frac{d}{d\tau}\bar{\xi}(\tau), \frac{d}{d\tau}\bar{\xi}(\tau)\right\rangle\right) \ . \tag{4.94}$$

Finally, we note that

$$\rho^{\text{seg}}_{\min}(\tau) \le \frac{\langle A^{\text{seg}}(\tau)v, v\rangle}{\langle \mathcal{C}_1 v, v\rangle} \ , \quad \forall\, v \in Y \ , \tag{4.95}$$

and therefore

$$\frac{d^2}{d\tau^2}\rho_{\min}^{\text{seg}}(\tau) = 2\left(\rho_{\min}^{\text{seg}}(\tau)\left\langle \mathcal{C}_1\frac{d}{d\tau}\bar{\xi}(\tau), \frac{d}{d\tau}\bar{\xi}(\tau)\right\rangle - \left\langle A^{\text{seg}}(\tau)\frac{d}{d\tau}\bar{\xi}(\tau), \frac{d}{d\tau}\bar{\xi}(\tau)\right\rangle\right) \quad (4.96)$$

$$\leq 0\,. \quad (4.97)$$

Remarks

We present in Figure 4-8, Table 4.3, and Figure 4-10 the effectivities for Examples 1-3, respectively, with $\mathcal{C}_1 = A(\theta_0)$ for different values of θ_0. Note that in all cases, $\eta_N(\mu) - 1 > 0$ — we obtain rigorous bounds.

However, as in the minimum coefficient bound conditioner, the numerical tests also show that the sharpness of the resulting bounds depends on the choice of θ_0. In particular, we observe in Figure 4-8 (Example 1) that taking $\theta_0 = \mu_0 \leq 10^2$ yields $\eta_N(\mu) < 20$, but taking $\theta_0 = \mu_0 \geq 10^3$ yields $\eta \sim 100$ — in the latter, the error is overestimated by two orders of magnitude. In Table 4.3, a good choice of θ_0 yields $\eta_N(\mu) \leq 5$, while a bad choice yields effectivities as high as 200.

In Figure 4-9 we show the points θ_m which form the convex hull $\mathcal{S}(\theta)$, and in Figure 4-10 we plot the effectivities as a function of μ calculated using $\mathcal{C}_1 = A(\theta_0)$ for different choices of θ_0 $(= \mu_0)$: $\theta_0 = \{\mu_{\min}, 1/\mu_{\min}\}$ for Trial 1, $\theta_0 = \{\mu_{\min}, 1/\mu_{\max}\}$ for Trial 2, $\theta_0 = \{\mu_{\max}, 1/\mu_{\max}\}$ for Trial 3, and $\theta_0 = \{\mu_{\max}/2, 2/\mu_{\max}\}$ for Trial 4. Better samples for the convex hull can certainly be used (e.g., using the tangent to the $\Theta_1 - \Theta - 2$ curve), however the resulting effectivities with our rather simple choice are exceptionally good for all the trials.

The concave eigenvalue bound conditioner shows great promise — the method is general, and we obtain sharp bounds even for $I = 1$. While the numerical results show that $\mathcal{C}_1$ must be chosen well, the resulting effectivities seem less sensitive to the choice of $\mathcal{C}_1$ than in the minimum coefficient bound conditioner. Perhaps the biggest disadvantage of the approach is that it requires, *on-line* (for any new μ), the search for a simplex $\mathcal{S}(\Theta(\mu))$ that contains $\Theta(\mu)$. However, the problem of finding the *optimal* simplex — that yielding the largest lower bound to the minimum eigenvalue — may be stated as a linear programming problem:

$$\alpha_1(\mu) = \max_{a_m}\left(\sum_{m=1}^{T} a_m \rho_{\min}^{1\,m}\right)^{-1} \quad (4.98)$$

$$\text{s.t.} \quad a_m \geq 0 \quad (4.99)$$

$$\sum_{m=1}^{T} a_m = 1\,. \quad (4.100)$$

Note that feasibility of (4.98) requires that T be *at least* $2^{Q_{\mathcal{A}}}$; however there is otherwise little *a priori* guidance for the choice of T or of S_T^θ. Furthermore, the off-line stage may be expensive and complicated, particularly for large $Q_{\mathcal{A}}$.

4.3.4 Effective Property Bound Conditioners

We now consider a problem which, after affine mapping, is described by

$$\langle \mathcal{A}(\mu)w, v\rangle = \sum_{r=1}^{R}\int_{\Omega^r}\frac{\partial w}{\partial x_i}\kappa_{ij}^r(\mu)\frac{\partial v}{\partial x_j} = \sum_{r=1}^{R}\int_{\Omega^r}\underline{D}[w]\,\underline{\kappa}^r(\mu)\,\underline{D}[v] \quad (4.101)$$

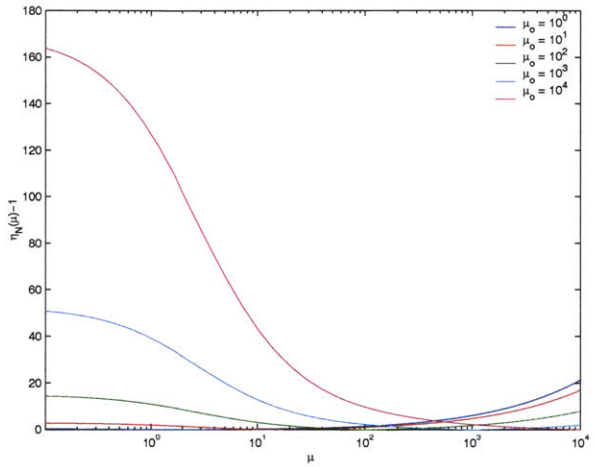


Figure 4-8: Effectivity as a function of μ for Example 1, calculated using the concave eigenvalue bound conditioner with $C_1 = A(\theta_0)$ for different choices of θ_0 $(= \mu_0)$.

$\theta_0 \equiv= \{\mu_0^1, \mu_0^2\}$	$\min_\mu (\eta_N(\mu) - 1)$	$\max_\mu (\eta_N(\mu) - 1)$	$\text{ave}_\mu (\eta_N(\mu) - 1)$
$\{\mu_{\min}^1, \mu_{\min}^2\}$	0.0379	4.1912	0.6447
$\{\mu_{\min}^1, \mu_{\max}^2\}$	0.7467	236.6623	34.4212
$\{\mu_{\max}^1, \mu_{\min}^2\}$	0.7859	230.7950	33.7431
$\{\mu_{\max}^1, \mu_{\max}^2\}$	0.0279	3.8670	0.5852
$\left\{\frac{\mu_{\min}^1 + \mu_{\max}^1}{2}, \frac{\mu_{\min}^2 + \mu_{\max}^2}{2}\right\}$	0.9724	149.2957	22.3509

Table 4.3: Minimum, maximum, and average effectivity for Example 2, calculated using the concave eigenvalue bound conditioner with $T = 16$, a uniform sample S_T^θ, and $C_1 = A(\theta_0)$ for different choices of θ_0.

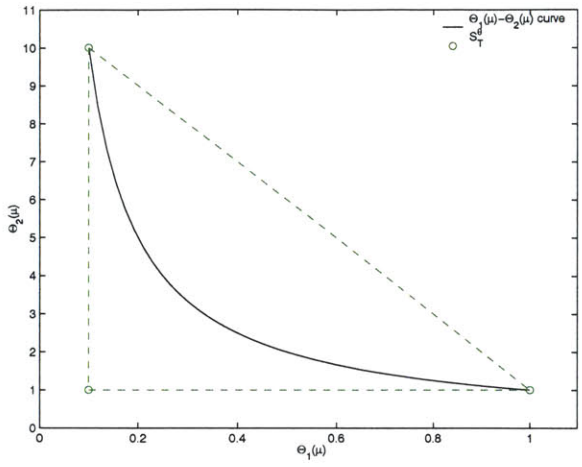


Figure 4-9: Sample S_T^θ and the convex hull, $\mathcal{S}(\theta)$ for all $\theta \in \mathcal{D}^\theta$.

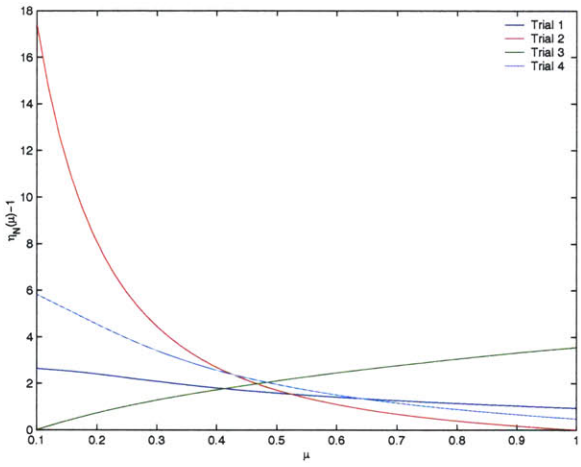


Figure 4-10: Effectivity as a function of μ for Example 3, calculated using the concave eigenvalue bound conditioner with $\mathcal{C}_1 = A(\theta_0)$ for different choices of θ_0.

where $\overline{\Omega} = \cup_{r=1}^R \overline{\Omega}^r$ is the domain ($\Omega^i \cap \Omega^j = 0$, and $\overline{\Omega}$ denotes the closure of Ω), $\underline{\kappa}^r(\mu) \in \mathbb{R}^{d\times d}$, $d = 1, 2$, or 3, is the symmetric, positive definite effective diffusivity tensor (e.g., due to geometric mapping or non-isotropic properties), and

$$D_i[w] = \frac{\partial w}{\partial x_i} , \quad 1 \le i \le d . \tag{4.102}$$

We now present a class of bound conditioners based on approximations to the diffusivity tensor. For simplicity of exposition, we shall again consider $I = 1$ and $\mathcal{I}(\mu) = \{1\}$. The extension to $I > 1$ is straightforward.

Bound Conditioner

We assume that we can find a set of μ-independent, symmetric, positive-definite matrices $\hat{\underline{\kappa}}^r$, $r = 1, \ldots, R$, such that such that

$$\frac{\underline{v}^T \underline{\kappa}^r(\mu)\, \underline{v}}{\underline{v}^T \hat{\underline{\kappa}}^r\, \underline{v}} > 0 , \quad \forall\, \underline{v} \in \mathbb{R}^2 , \forall\, \mu \in \mathcal{D}^\mu , \; 1 \le r \le R . \tag{4.103}$$

We then choose $\alpha_1(\mu)$, $\mathcal{C}_1$, and $\mathcal{C}(\mu)$ such that

$$1 \le \alpha_1(\mu) \frac{\underline{v}^T \underline{\kappa}^r(\mu)\, \underline{v}}{\underline{v}^T \hat{\underline{\kappa}}^r\, \underline{v}} \le \rho , \quad \forall\, \underline{v} \in \mathbb{R}^2 , \forall\, \mu \in \mathcal{D}^\mu , \; 1 \le r \le R , \tag{4.104}$$

$$\langle \mathcal{C}_1 w, v\rangle = \sum_{r=1}^R \int_{\Omega^r} \underline{D}^T[w] \hat{\underline{\kappa}}^r \underline{D}[v] , \tag{4.105}$$

and $\mathcal{C}(\mu) = \alpha_1(\mu)^{-1}\mathcal{C}_1$. It then follows that

$$1 \leq \eta_N(\mu) \leq \rho \; . \tag{4.106}$$

Proof of Bounding Properties

We note that, by construction,

$$\underline{D}^T[v]\,\underline{\kappa}^r(\mu)\,\underline{D}[v] \;\geq\; \alpha_1(\mu)^{-1}\,\underline{D}^T[v]\,\underline{\hat{\kappa}}^r\,\underline{D}[v]\;, \quad \forall\, v \in Y\;, \forall\, \mu \in \mathcal{D}^\mu\;,\; 1 \leq r \leq R\;; \tag{4.107}$$

it therefore follows that

$$\frac{\langle \mathcal{A}(\mu)v, v\rangle}{\langle \mathcal{C}(\mu)v, v\rangle} \;=\; \frac{\displaystyle\sum_{r=1}^{R}\int_{\Omega^r} \underline{D}^T[v]\,\underline{\kappa}^r(\mu)\,\underline{D}[v]}{\displaystyle\sum_{r=1}^{R}\int_{\Omega^r} \alpha_1(\mu)^{-1}\underline{D}^T[v]\,\underline{\hat{\kappa}}^r\,\underline{D}[v]} \tag{4.108}$$

$$\geq\; 1\;, \quad \forall\, v \in Y\;. \tag{4.109}$$

Furthermore, since

$$\underline{D}^T[v]\,\underline{\kappa}^r(\mu)\,\underline{D}[v] \leq \rho\left(\alpha_1(\mu)^{-1}\underline{D}^T[v]\,\underline{\hat{\kappa}}^r\,\underline{D}[v]\right)\;, \tag{4.110}$$

it follows from (4.108) that

$$\frac{\langle A(\mu)v, v\rangle}{\langle \mathcal{C}(\mu)v, v\rangle} \;\leq\; \frac{\displaystyle\sum_{r=1}^{R}\rho\,\alpha_1(\mu)^{-1}\int_{\Omega^r} \underline{D}^T[v]\underline{\hat{\kappa}}^r\underline{D}^T[v]}{\displaystyle\sum_{r=1}^{R}\alpha_1(\mu)^{-1}\int_{\Omega^r} \underline{D}^T[v]\underline{\hat{\kappa}}^r\underline{D}[v]} \tag{4.111}$$

$$\leq\; \rho\;. \tag{4.112}$$

The result (4.106) directly follows from (4.108) and (4.112).

Finding the $\underline{\hat{\kappa}}^r$ and associated $\alpha_1(\mu)$ that satisfy (4.104) and minimize ρ is not so trivial, but it is not difficult to find suboptimal solutions. We shall illustrate this with two general examples.

We first consider the case $R = 1$ and

$$\underline{\kappa}^1(\mu) = \begin{bmatrix} a(\mu) & 0 \\ 0 & b(\mu) \end{bmatrix} \tag{4.113}$$

with $a_{\max} \geq a(\mu) \geq a_{\min} \geq 0$, and $b_{\max} \geq b(\mu) \geq b_{\min} \geq 0$; these conditions ensure that the $\underline{\kappa}^1(\mu)$ is symmetric, positive-definite. We then define $\alpha_1(\mu) = 1$, and choose

$$\underline{\hat{\kappa}}^1 = \begin{bmatrix} a_{\min} & 0 \\ 0 & b_{\min} \end{bmatrix}\;. \tag{4.114}$$

Clearly, our choices satisfy (4.104) since

$$\min_{\underline{v}\in\mathbb{R}^2} \alpha_1(\mu)\frac{\underline{v}^T\underline{\kappa}^1(\mu)\underline{v}}{\underline{v}^T\underline{\hat{\kappa}}^1\underline{v}} \;=\; \min\left[\frac{a(\mu)}{a_{\min}}, \frac{b(\mu)}{b_{\min}}\right] \tag{4.115}$$

$$\geq\; 1\;, \tag{4.116}$$

and

$$\max_{\underline{v}\in\mathbb{R}^2} \alpha_1(\mu)\frac{\underline{v}^T\underline{\kappa}^1(\mu)\underline{v}}{\underline{v}^T\hat{\underline{\kappa}}^1\underline{v}} = \max\left[\frac{a(\mu)}{a_{\min}}, \frac{b(\mu)}{b_{\min}}\right] \tag{4.117}$$

$$\leq \max_{\mu\in\mathcal{D}^\mu}\left(\max\left[\frac{a(\mu)}{a_{\min}}, \frac{b(\mu)}{b_{\min}}\right]\right) . \tag{4.118}$$

We now consider the case $R = 1$ and

$$\underline{\kappa}^1(\mu) = \begin{bmatrix} a(\mu) & c(\mu) \\ c(\mu) & b(\mu) \end{bmatrix} \tag{4.119}$$

with $a_{\max} \geq a(\mu) \geq a_{\min} > 0$, $b_{\max} \geq b(\mu) \geq b_{\min} > 0$, and $a(\mu)b(\mu) - c(\mu)^2 > \beta_\kappa^0 > 0$ for all μ in $\mathcal{D}^\mu$; these three conditions ensure that the $\underline{\kappa}^1(\mu)$ is symmetric, positive-definite. We then define

$$\alpha_1(\mu) = \frac{a(\mu)b(\mu)}{a(\mu)b(\mu) - c(\mu)^2} \tag{4.120}$$

and choose

$$\hat{\underline{\kappa}}^1 = \frac{1}{2}\begin{bmatrix} a_{\min} & 0 \\ 0 & b_{\min} \end{bmatrix} . \tag{4.121}$$

To verify that our choices satisfy (4.104), we let $\ell_{\min}$ and $\ell_{\max}$ be the minimum and maximum eigenvalues given by

$$\ell_{\min}(\mu) = \min_{\underline{v}\in\mathbb{R}^2} \alpha_1(\mu)\frac{\underline{v}^T\underline{\kappa}^1(\mu)\underline{v}}{\underline{v}^T\hat{\underline{\kappa}}\,\underline{v}} , \qquad \ell_{\max}(\mu) = \max_{\underline{v}\in\mathbb{R}^2} \alpha_1(\mu)\frac{\underline{v}^T\underline{\kappa}^1(\mu)\underline{v}}{\underline{v}^T\hat{\underline{\kappa}}\,\underline{v}} . \tag{4.122}$$

The characteristic equation for the eigenvalues $\ell(\mu)$ is given by

$$\left(a(\mu) - \frac{a_{\min}}{2\alpha_1(\mu)}\ell(\mu)\right)\left(b(\mu) - \frac{b_{\min}}{2\alpha_1(\mu)}\ell(\mu)\right) - c(\mu)^2 = 0 \tag{4.123}$$

or

$$\ell(\mu)^2 - \ell(\mu)\left(2\alpha_1(\mu)\left(\frac{a(\mu)}{a_{\min}} + \frac{b(\mu)}{b_{\min}}\right)\right) + 4\alpha_1(\mu)^2\frac{a(\mu)b(\mu) - c(\mu)^2}{a_{\min}b_{\min}} = 0. \tag{4.124}$$

It then follows that

$$\ell_{\min}(\mu) + \ell_{\max}(\mu) = 2\alpha_1(\mu)\left(\frac{a(\mu)}{a_{\min}} + \frac{b(\mu)}{b_{\min}}\right) \tag{4.125}$$

$$\ell_{\min}(\mu)\ell_{\max}(\mu) = 4\alpha_1(\mu)^2\frac{a(\mu)b(\mu) - c(\mu)^2}{a_{\min}b_{\min}}. \tag{4.126}$$

To prove the left-hand inequality in (4.104), we note that

$$\ell_{\min}(\mu)\ell_{\max}(\mu) \leq \ell_{\min}(\mu)(\ell_{\min}(\mu) + \ell_{\max}(\mu)), \quad \forall\ \mu \in \mathcal{D}^\mu , \tag{4.127}$$

so that

$$\begin{aligned}
\ell_{\min}(\mu) &> \frac{\ell_{\min}(\mu)\ell_{\max}(\mu)}{\ell_{\min}(\mu)+\ell_{\max}(\mu)} && (4.128)\\
&\geq 2\alpha_1(\mu)\left(\frac{a(\mu)b(\mu)-c(\mu)^2}{a(\mu)b_{\min}+a_{\min}b(\mu)}\right) && (4.129)\\
&= 2\alpha_1(\mu)\left(\frac{a(\mu)b(\mu)-c(\mu)^2}{a(\mu)b(\mu)}\right)\left(\frac{a(\mu)b(\mu)}{a(\mu)b_{\min}+a_{\min}b(\mu)}\right) && (4.130)\\
&= 2\left(\frac{b_{\min}}{b(\mu)}+\frac{a_{\min}}{a(\mu)}\right)^{-1} && (4.131)\\
&\geq 1\,. && (4.132)
\end{aligned}$$

Also, we have

$$\begin{aligned}
\ell_{\max}(\mu) &\leq \ell_{\min}(\mu)+\ell_{\max}(\mu) && (4.133)\\
&= 2\alpha_1(\mu)\left(\frac{a(\mu)}{a_{\min}}+\frac{b(\mu)}{b_{\min}}\right) && (4.134)\\
&\leq 2\left(\frac{a_{\max}}{a_{\min}}+\frac{b_{\max}}{b_{\min}}\right)\max_{\mu\in\mathcal{D}^\mu}\left(\frac{a(\mu)b(\mu)}{a(\mu)b(\mu)-c(\mu)^2}\right) && (4.135)\\
&\leq 2\left(\frac{a_{\max}}{a_{\min}}+\frac{b_{\max}}{b_{\min}}\right)\left(\frac{a_{\max}b_{\max}}{\beta_\kappa^0}\right)\,, \quad \forall\ \mu\in\mathcal{D}^\mu\,. && (4.136)
\end{aligned}$$

We shall now apply our Effective Diffusivity Bound Conditioner to Examples 3 and 4.

Example 3

As a first example we consider the case of Example 3 of Section 3.4 in which $R=1$, $P=1$, $\mu=\bar{t}$, $\mathcal{D}^\mu=[0.1,1.0]$, and

$$\underline{\kappa}^1(\mu)=\begin{pmatrix}\bar{t} & 0\\ 0 & \frac{1}{\bar{t}}\end{pmatrix}. \tag{4.137}$$

We therefore choose $\alpha_1(\mu)=1$ and

$$\underline{\hat{\kappa}}^1=\begin{bmatrix}\bar{t}_{\min} & 0\\ 0 & \dfrac{1}{\bar{t}_{\max}}\end{bmatrix}. \tag{4.138}$$

The effectivity then satisfies

$$1\leq\eta_N(\mu)\leq\frac{t_{\max}}{t_{\min}}\,, \qquad \forall\ \mu\in\mathcal{D}^\mu\ ; \tag{4.139}$$

this is confirmed in Figure 4-11, in which we plot the resulting effecivity as a function of μ.

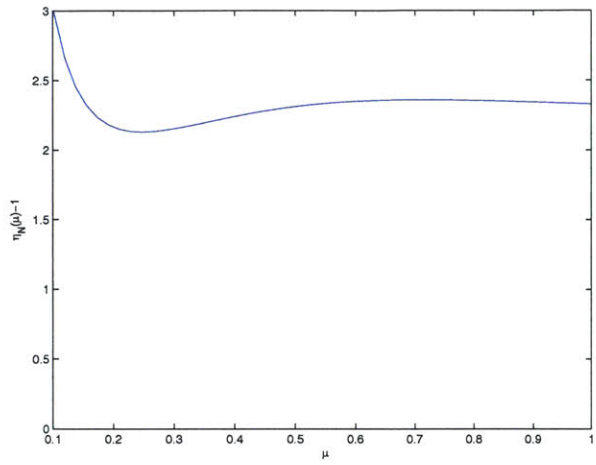


Figure 4-11: Effectivity as a function of μ for Example 3 obtained using the effective property bound conditioner.

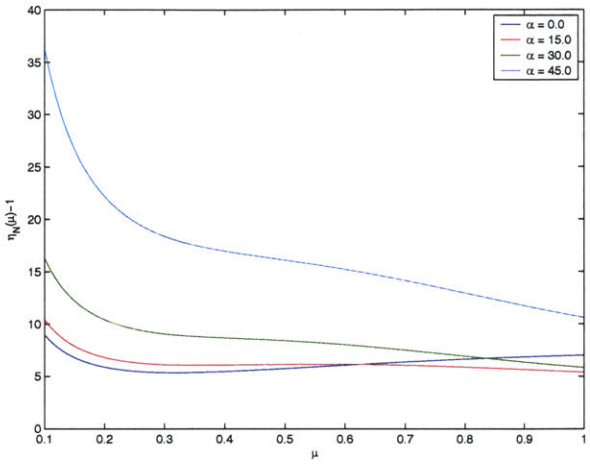


Figure 4-12: Effectivity as a function of μ for Example 4 obtained using the effective property bound conditioner.

Example 4

As a second example, we consider Example 4 of Section 3.4 in which $R = 1$, $P = 2$, $\mu = (\bar{t}, \bar{\alpha})$, $\mathcal{D}^\mu = [0.1, 1.0] \times [0°, 45°]$, and

$$\underline{\kappa}^1(\mu) = \begin{bmatrix} \bar{t} & -\tan\bar{\alpha} \\ -\tan\bar{\alpha} & \dfrac{1+\tan^2\bar{\alpha}}{\bar{t}} \end{bmatrix} \tag{4.140}$$

We therefore choose

$$\alpha_1(\mu) = 1 + \tan^2\bar{\alpha} \tag{4.141}$$

and

$$\underline{\hat{\kappa}}^1 = \frac{1}{2}\begin{bmatrix} \bar{t}_{\min} & 0 \\ 0 & \dfrac{1}{\bar{t}_{\max}} \end{bmatrix} \tag{4.142}$$

The effectivity then satisfies

$$1 \le \eta_N(\mu) \le 4\frac{\bar{t}_{\max}}{\bar{t}_{\min}}, \qquad \forall\, \mu \in \mathcal{D}^\mu\ ; \tag{4.143}$$

this is also confirmed in Figure 4-12, in which we plot the resulting effecivity as a function of μ.

Remarks

The effective property bound conditioner presents several advantages: numerical tests show that the effective property conditioner yields good effectivities and, unlike the concave eigenvalue bound conditioner, the computational cost is independent of $Q_{\mathcal{A}}$. Furthermore, although the method is restricted to problems of the form (4.101), unlike the minimum coefficient conditioner it is applicable even for general geometry variations. However, in Chapter 5 we shall see that the requirement that $\underline{\kappa}$ be positive-definite proves very restrictive — the relevant material property tensor is no longer

positive-definite.

4.4 Bound Conditioner Constructions — Type II

We recall from Section 4.2 our "computational invertibility" assumption:

$$\mathcal{C}^{-1}(\mu) = \sum_{i \in \mathcal{I}(\mu)} \alpha_i(\mu) \mathcal{C}_i^{-1} \tag{4.144}$$

where $\mathcal{I}(\mu) \subset \{1, \ldots, I\}$ is a parameter-dependent set of indices, I is a finite (preferably small) integer, and the $\mathcal{C}_i : Y \to Y'$, are parameter-*independent* symmetric, coercive operators.

We now consider a particular class of bound conditioners in which $\mathcal{C}^{-1}(\mu)$ is constructed as a direct approximation to $\mathcal{A}^{-1}(\mu)$.

4.4.1 Convex Inverse Bound Conditioner

To begin, we recall our separability assumption (3.8) on $\mathcal{A}(\mu)$, and assume that $\mathcal{A}(\mu)$ takes the particular (slightly different) form

$$\mathcal{A}(\mu) = \mathcal{A}^0 + \sum_{q=1}^{Q_\mathcal{A}} \Theta^q(\mu)\, \mathcal{A}^q \ , \quad \forall\ \mu \in \mathcal{D}^\mu \ , \tag{4.145}$$

where the $\Theta^q(\mu) : \mathcal{D}^\mu \to \mathbb{R}_{+,0}$, $q = 1, \ldots, Q_\mathcal{A}$; here, $\mathbb{R}_{+,0}$ refers to the non-negative real numbers. We also assume that $\mathcal{A}^0 : Y \to Y'$ is symmetric, continuous, and coercive, and that the $\mathcal{A}^q : Y \to Y'$, $q = 1, \ldots, Q_\mathcal{A}$ are symmetric, continuous, and positive-semidefinite, i.e.,

$$\langle \mathcal{A}^0 w, w \rangle \ > \ 0 \ , \quad \forall\ w \in Y \ , \tag{4.146}$$

$$\langle \mathcal{A}^q w, w \rangle \ \geq \ 0 \ , \quad \forall\ w \in Y \ , \quad 1 \leq q \leq Q_\mathcal{A} \ ; \tag{4.147}$$

it follows that $\mathcal{A}(\mu)$ is symmetric, continuous, and coercive for all $\mu \in \mathcal{D}^\mu$.

For any $\theta \in \mathbb{R}^{Q_\mathcal{A}}_{+,0}$, we now define $A(\theta) : Y \to Y'$ as

$$A(\theta) = A^0 + \sum_{q=1}^{Q_\mathcal{A}} \theta^q\, A^q \ , \tag{4.148}$$

where $A^q \equiv \mathcal{A}^q$, $0 \leq q \leq Q_\mathcal{A}$. We may then write

$$A(\Theta(\mu)) = \mathcal{A}(\mu) \ , \tag{4.149}$$

where $\Theta : \mathcal{D}^\mu \subset \mathbb{R}^P \to \mathcal{D}^\theta \subset \mathbb{R}^{Q_\mathcal{A}}$, and $\mathcal{D}^\theta$ is the range of Θ. We shall also define $\theta_{\min} (\geq 0)$, $\theta_{\max}$ (assumed finite), and $\mathcal{D}^\theta_{\text{box}} \subset \mathbb{R}^{Q_\mathcal{A}}_{+,0}$ as

$$\theta^q_{\min} \ \equiv \ \sup \ \hat{\vartheta}_{\{\hat{\vartheta} \in \mathbb{R}_+ \ | \ \Theta^q(\mu) \geq \hat{\vartheta},\ \forall\ \mu \in \mathcal{D}^\mu\}} \ , \quad q = 1, \ldots, Q_\mathcal{A} \tag{4.150}$$

$$\theta^q_{\max} \ \equiv \ \inf \ \hat{\vartheta}_{\{\hat{\vartheta} \in \mathbb{R}_+ \ | \ \Theta^q(\mu) \leq \hat{\vartheta},\ \forall\ \mu \in \mathcal{D}^\mu\}} \ , \quad q = 1, \ldots, Q_\mathcal{A} \ , \tag{4.151}$$

and

$$\mathcal{D}^{\theta}_{\text{box}} \equiv \prod_{q=1}^{Q_{\mathcal{A}}} \left[\theta^q_{\min}, \theta^q_{\max}\right] \ , \tag{4.152}$$

respectively. It follows that $A(\theta)$ is symmetric, continuous, and coercive for all $\theta \in \mathcal{D}^{\theta}_{\text{box}}$.

Finally, we note that our compliance output can be expressed as an energy:

$$\begin{aligned} s(\mu) &= \langle L, u(\mu)\rangle & (4.153)\\ &= \langle F, u(\mu)\rangle & (4.154)\\ &= \langle A(\Theta(\mu))u(\mu), u(\mu)\rangle \ , & (4.155) \end{aligned}$$

or, equivalently,

$$s(\mu) = \langle F, A^{-1}(\Theta(\mu))F\rangle \ , \tag{4.156}$$

where $A^{-1}(\theta)$ is the (symmetric, coercive) inverse of $A(\theta)$.

Bound Conditioner Formulation

We now introduce a "θ" sample $S^{\theta}_T = \{\theta_1, \ldots, \theta_T\}$, where $\theta_i \in \mathcal{D}^{\theta}_{\text{box}}$, $i = 1, \ldots, T$. We assume that, given any $\theta \in \mathcal{D}^{\theta}_{\text{box}}$, we can find a simplex $\mathcal{S}(\theta)$ such that $\theta \in \mathcal{S}(\theta)$, where

$$\mathcal{S}(\theta) \equiv \left\{ \hat{\vartheta} \mid \hat{\vartheta} = \sum_{i \in \mathcal{T}(\theta)} \hat{a}_i \, \theta_i \right\} \ ; \tag{4.157}$$

here, $\mathcal{T}(\theta) \subset \{1, \ldots, T\}$ is an index set of cardinality $Q_{\mathcal{A}}$, and the $\hat{a}_i$ are *weights* —

$$0 \le \hat{a}_i \le 1 \ , \quad \forall \ r \in \mathcal{T}(\theta) \ , \tag{4.158}$$

and

$$\sum_{i \in \mathcal{T}(\theta)} \hat{a}_i = 1 \ . \tag{4.159}$$

In words, the simplex $\mathcal{S}(\theta)$ is a polyhedral set, defined by $|\mathcal{T}(\theta)|$ vertices $\mathcal{T}(\theta)$ chosen such that $\mathcal{S}(\theta)$ contains θ. We implicitly assume that S^{θ}_T is chosen such that, for all $\theta \in \mathcal{D}^{\theta}_{\text{box}}$, such a construction is possible; a deficient sample S^{θ}_T can always be rendered compliant simply by replacing one point with $\theta_{\min}$.

To each μ in $\mathcal{D}^{\mu}$ we then associate (i) a point $\bar{\theta}(\mu) \in \mathcal{D}^{\theta}_{\text{box}}$, such that

$$\bar{\theta}^q(\mu) \le \Theta^q(\mu) \ , \quad 1 \le q \le Q_{\mathcal{A}} \ , \tag{4.160}$$

(ii) an index set $\mathcal{I}(\mu)$ given by

$$\mathcal{I}(\mu) = \mathcal{T}(\bar{\theta}(\mu)) \ , \tag{4.161}$$

and (iii) a set of weights $a_i(\bar{\theta}(\mu))$, such that

$$\bar{\theta}(\mu) = \sum_{i \in \mathcal{I}(\mu)} a_i(\bar{\theta}(\mu)) \, \theta_i \ . \tag{4.162}$$

We then choose the $\alpha_i(\mu)$ and $\mathcal{C}_m$ as

$$\alpha_i(\mu) = a_i(\bar{\theta}(\mu)) \tag{4.163}$$

$$\mathcal{C}_i = A^{-1}(\theta_i) \tag{4.164}$$

so that $C(\mu)$ is given by

$$C(\mu) = \sum_{i \in \mathcal{I}(\mu)} \alpha_i(\mu) \mathcal{C}_i^{-1} = \left(\sum_{i \in \mathcal{I}(\mu)} \alpha_i(\mu)\, A^{-1}(\theta_i) \right)^{-1} . \tag{4.165}$$

(Clearly, $\mathcal{C}^{-1}(\mu)$ (and therefore $\mathcal{C}(\mu)$) is symmetric positive-definite.) In words, $\mathcal{C}^{-1}(\mu)$ is an approximation to $A^{-1}(\Theta(\mu))$ constructed as a convex combination of A^{-1} at "neighboring" θ.

Proof of Bounding Properties

To prove that the bound conditioner (4.163) provides uniform bounds, we shall need to demonstrate certain properties of the quadratic form associated with $A^{-1}(\theta)$. To begin, we define $\hat{\mathcal{J}}: \mathcal{D}^{\theta}_{\text{box}} \times Y' \rightarrow \mathbb{R}$ as

$$\hat{\mathcal{J}}(\theta, \hat{G}) = \langle \hat{G}, A^{-1}(\theta)\, \hat{G} \rangle , \tag{4.166}$$

Also, given $\theta_1 \in \mathcal{D}^{\theta}_{\text{box}}$, $\theta_2 \in \mathcal{D}^{\theta}_{\text{box}}$, and $\tau \in [0,1]$, we define

$$\hat{\mathcal{J}}^{\text{seg}}(\tau; \theta_1, \theta_2; \hat{G}) = \hat{\mathcal{J}}(\theta_1 + \tau(\theta_2 - \theta_1), \hat{G}) \tag{4.167}$$

We also define $A^{\text{seg}}(\tau; \theta_1, \theta_2) = A(\theta_1 + \tau(\theta_2 - \theta_1))$. We can then write

$$A^{\text{seg}}(\tau; \theta_1, \theta_2) = A_1 + \sum_{q=1}^{Q_{\mathcal{A}}} \theta_1^q\, A^q + \tau \sum_{q=1}^{Q_{\mathcal{A}}} (\theta_2^q - \theta_1^q)\, A^q . \tag{4.168}$$

and

$$\hat{\mathcal{J}}^{\text{seg}}(\tau; \theta_1, \theta_2; \hat{G}) = \langle \hat{G}, (A^{\text{seg}}(\tau; \theta_1, \theta_2))^{-1}\, \hat{G} \rangle . \tag{4.169}$$

We now consider the monotonicity and convexity of $\hat{\mathcal{J}}(\theta)$ in the parameter θ; in particular, we can prove [47] that:

(A) $\hat{\mathcal{J}}(\theta, \hat{G})$ is a non-increasing function: for any $\theta_1 \in \mathcal{D}^{\theta}_{\text{box}}$, $\theta_2 \in \mathcal{D}^{\theta}_{\text{box}}$, such that $\theta_2 \geq \theta_1$ (i.e., $\theta_2^q \geq \theta_1^q$, $q = 1, \ldots, Q_{\mathcal{A}}$),

$$\hat{\mathcal{J}}(\theta_2, \hat{G}) \leq \hat{\mathcal{J}}(\theta_1, \hat{G}) , \quad \forall\, \hat{G} \in Y' ; \tag{4.170}$$

(B) $\hat{\mathcal{J}}(\theta, \hat{G})$ is a convex function of θ: for any $\theta_1 \in \mathcal{D}^{\theta}_{\text{box}}$, $\theta_2 \in \mathcal{D}^{\theta}_{\text{box}}$, and for all $\tau \in [0,1]$,

$$\hat{\mathcal{J}}(\theta_1 + \tau(\theta_2 - \theta_1), \hat{G}) \leq (1-\tau)\, \hat{\mathcal{J}}(\theta_1, \hat{G}) + \tau \hat{\mathcal{J}}(\theta_2, \hat{G}) , \quad \forall\, \hat{G} \in Y' . \tag{4.171}$$

We shall refer to (4.170) and (4.171) as Propositions (A) and (B), respectively.

Assuming that Propositions (A) and (B) are indeed true (we shall present the proofs subsequently), we can prove that our bound conditioner satisfies the spectral condition. We recall from

Section 4.2 that the effectivity may be written as

$$\eta_N(\mu) = \frac{\langle \mathcal{R}(\mu), \mathcal{C}^{-1}(\mu)\mathcal{R}(\mu)\rangle}{\langle \mathcal{R}(\mu), \mathcal{A}^{-1}(\mu)\mathcal{R}(\mu)\rangle} \ . \tag{4.172}$$

From the definitions (4.163) and (4.145) we immediately note that

$$\eta_N(\mu) = \frac{\sum\limits_{i\in\mathcal{I}(\mu)} \alpha_i(\mu)\, \hat{\mathcal{J}}(\theta_i, \mathcal{R}(\mu))}{\hat{\mathcal{J}}(\Theta(\mu), \mathcal{R}(\mu))} \ . \tag{4.173}$$

But from the construction of the $\alpha_j(\mu)$, the choice of $\mathcal{I}(\mu)$, Proposition (B), classical results in convex analysis, and Proposition (A), it directly follows that, for any $\hat{G} \in Y'$ (and therefore for $\hat{G} = \mathcal{R}(\mu)$),

$$\sum_{i\in\mathcal{I}(\mu)} \alpha_i(\mu)\, \hat{\mathcal{J}}(\theta_i, \hat{G}) \geq \hat{\mathcal{J}}(\overline{\theta}(\mu), \hat{G}) \geq \hat{\mathcal{J}}(\Theta(\mu), \hat{G}) \ , \tag{4.174}$$

which concludes the proof. It now remains only to prove Propositions (A) and (B).

Proof of Proposition (A)

To prove Proposition (A), we need only demonstrate that, for any (fixed) θ_1, θ_2 such that $\theta_2 \geq \theta_1$, and any (fixed) $\hat{G} \in Y'$,

$$\frac{d}{d\tau}\hat{\mathcal{J}}^{\text{seg}}(\tau; \theta_1, \theta_2; \hat{G}) \leq 0, \quad \forall\ \tau \in [0,1] \ . \tag{4.175}$$

To evaluate $d\hat{\mathcal{J}}^{\text{seg}}/d\tau$, we note that

$$\frac{d}{d\tau}\hat{\mathcal{J}}^{\text{seg}}(\tau; \theta_1, \theta_2; \hat{G}) = \left\langle \hat{G}, \frac{d}{d\tau}\left((A^{\text{seg}})^{-1}(\tau; \theta_1, \theta_2)\right), \hat{G} \right\rangle \ ; \tag{4.176}$$

it thus remains only to show that $d(A^{\text{seg}}(\tau; \theta_1, \theta_2))^{-1}/d\tau$ is symmetric negative-semidefinite.

To this end, we note that $(A^{\text{seg}}(\tau; \theta_1, \theta_2))^{-1}\, A^{\text{seg}}(\tau; \theta_1, \theta_2) = I\!I$ (the identity), and thus

$$\frac{d}{d\tau}\left(A^{\text{seg}}(\tau; \theta_1, \theta_2)\right)^{-1} A^{\text{seg}}(\tau; \theta_1, \theta_2) + \left(A^{\text{seg}}(\tau; \theta_1, \theta_2)\right)^{-1} \frac{d}{d\tau}\left(A^{\text{seg}}(\tau; \theta_1, \theta_2)\right) = 0 \ . \tag{4.177}$$

Application of (4.168) then yields

$$\frac{d}{d\tau}\left(A^{\text{seg}}(\tau; \theta_1, \theta_2)\right)^{-1} = -\left(A^{\text{seg}}(\tau; \theta_1, \theta_2)\right)^{-1} \left(\sum_{q=1}^{Q_{\mathcal{A}}} (\theta_2^q - \theta_1^q)\, A_q\right) \left(A^{\text{seg}}(\tau; \theta_1, \theta_2)\right)^{-1} \ ; \tag{4.178}$$

the desired result then directly follows, since $\theta_2^q \geq \theta_1^q$, and the A_q are symmetric positive-semidefinite.

Proof of Proposition (B)

To prove Proposition (B), we need only demonstrate that, for any $\theta_1 \in \mathcal{D}^{\theta}_{\text{box}}$, $\theta_2 \in \mathcal{D}^{\theta}_{\text{box}}$, and $\tau \in [0,1]$,

$$\hat{\mathcal{J}}^{\text{seg}}(\tau; \theta_1, \theta_2, \hat{G}) \leq (1-\tau)\hat{\mathcal{J}}_{\text{seg}}(0; \theta_1, \theta_2; \hat{G}) + \tau\hat{\mathcal{J}}^{\text{seg}}(1; \theta_1, \theta_2; \hat{G}) \ , \quad \forall\ \hat{G} \in Y' \ . \tag{4.179}$$

From standard results in convex analysis [6] it suffices to show that, for any (fixed) $\hat{G} \in Y'$,

$$\frac{d^2}{d\tau^2}\mathcal{J}^{\text{seg}}(\tau;\theta_1,\theta_2;\hat{G}) \geq 0, \quad \forall\, \tau \in [0,1]\ . \tag{4.180}$$

From the definition of $\mathcal{J}^{\text{seg}}(\tau;\theta_1,\theta_2;\hat{G})$, it thus remains only to show that $d^2((A^{\text{seg}})^{-1}(\tau;\theta_1,\theta_2))/d\tau^2$ is symmetric positive-semidefinite.

To this end, we continue the differentiation of (4.178) to obtain

$$\begin{aligned}
\frac{d^2}{d\tau^2}(A^{\text{seg}}(\tau;\theta_1,\theta_2))^{-1} &= -\frac{d}{d\tau}(A^{\text{seg}}(\tau;\theta_1,\theta_2))^{-1}\left(\sum_{q=1}^{Q_{\mathcal{A}}}(\theta_2^q-\theta_1^q)A_q\right)(A^{\text{seg}}(\tau;\theta_1,\theta_2))^{-1} \\
&\quad -(A^{\text{seg}}(\tau;\theta_1,\theta_2))^{-1}\left(\sum_{q=1}^{Q_{\mathcal{A}}}(\theta_2^q-\theta_1^q)\,A_q\right)\frac{d}{d\tau}(A^{\text{seg}}(\tau;\theta_1,\theta_2))^{-1} \\
&= 2(A^{\text{seg}}(\tau;\theta_1,\theta_2))^{-1}\left(\sum_{q=1}^{Q_{\mathcal{A}}}(\theta_2^q-\theta_1^q)\,A_q\right)(A^{\text{seg}}(\tau;\theta_1,\theta_2))^{-1} \\
&\quad \times\left(\sum_{q=1}^{Q_{\mathcal{A}}}(\theta_2^q-\theta_1^q)\,A_q\right)(A^{\text{seg}}(\tau;\theta_1,\theta_2))^{-1}\ .
\end{aligned} \tag{4.181}$$

The desired result then directly follows since $(A^{\text{seg}}(\tau;\theta_1,\theta_2))^{-1}$ is symmetric positive-definite.

Remarks

We shall consider three different bound conditioners.

The first is a single-point conditioner, and will be labelled SP. Here we set $T=1$, $S_T^\theta=\{\theta_{\min}\}$, $|\mathcal{I}(\mu)|=1$, $\mathcal{I}(\mu)=\{1\}$, and $\overline{\theta}(\mu)=\theta_{\min}$. This bound conditioner is a special case of our earlier Type I bound conditioner formulation. In particular, SP is equivalent to the minimum coefficient conditioner (with $I=1$ and $\bar{\theta}_1=\theta_{\min}$).

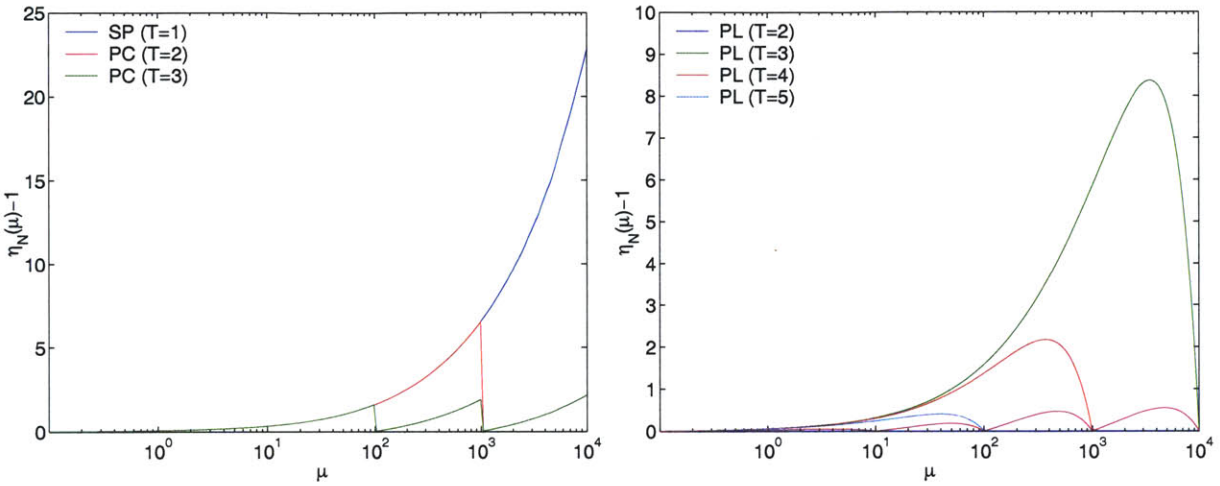
The second bound conditioner we develop here is piecewise-constant, and will be labelled PC. Now we set $T\geq 1$, $S_T^\theta=\{\theta_1=\theta_{\min},\theta_2,\ldots,\theta_i\}$, and $|\mathcal{I}(\mu)|=1$, and choose $\mathcal{I}(\mu)=\{i_1(\mu)\}$ such that $\overline{\theta}(\mu)\equiv\theta_{i_1(\mu)}\leq\Theta(\mu)$. There will often be many possible choices for $i_1(\mu)$; we can either establish a definition of closeness, or alternatively consider all possible candidates and select the best (in the sense of yielding the lowest effectivities).

The third bound conditioner we develop here is piecewise-linear, and will be labelled PL. Now we set $T\geq Q_{\mathcal{A}}+1$, $S_T^\theta=\{\theta_1,\ldots,\theta_T\}$, $|\mathcal{I}(\mu)|=Q_{\mathcal{A}}+1$, and choose $\mathcal{I}(\mu)$ such that the θ_i, $i\in\mathcal{I}(\mu)$, form a $(Q_{\mathcal{A}}+1)$-simplex containing $\overline{\theta}(\mu)\equiv\Theta(\mu)$. Again, there will often be several choices for the index set $\mathcal{I}(\mu)$ and associated simplex; we can either establish an *a priori* criterion for goodness (e.g., related to simplex size), or instead evaluate all candidates and select that yielding the best effectivities. Note that, for μ for which S_T^θ contains no $\Theta(\mu)$-containing $(Q_{\mathcal{A}}+1)$-simplex, we must accept a lower-order simplex and $\overline{\theta}(\mu)<\Theta(\mu)$ (e.g., in the worst case, we revert to PC).

We present in Figure 4-13 the effectivity as a function of μ for Example 1 obtained using the SP, PC, and PL bound conditioners; similar results are present in Figure 4-14 for Example 3. We also present in Tables 4.4 and 4.5 the minimum, maximum, and average effectivity over $\mu\in\mathcal{D}^\mu$ for Example 2. In all three cases, the PC conditioner performs considerably better than SP; and the PL conditioner is even better than PC. Of course, the purpose of these higher-order bound conditioners is to achieve some fixed accuracy (measured by $\Delta_N(\mu)$) at lower computational effort;

considering only the on-line complexity, no doubt PC requires the lowest cost at fixed accuracy. However, in practice, we must also consider the off-line complexity and on-line storage.

Our numerical tests show that higher-order constructions do yield very good effectivities. However, an important restriction of these methods is the assumption that the operator $\mathcal{A}$ must be of the form (4.145), and, in addition, the $\mathcal{A}^q$ must be positive semi-definite. There are many problems which cannot be expressed in the required form (for instance, Example 4), or for which the $\mathcal{A}^q$ are indefinite (for instance, the elasticity operator).



(a) (b)

Figure 4-13: Effectivity as a function of μ for Example 1 obtained using (a) SP ($T = 1$), PC ($T = 2, 3$), and (b) PL ($T = 2, 3, 4, 5$).

T	$\min_{\mu}(\eta_N(\mu) - 1)$	$\max_{\mu}(\eta_N(\mu) - 1)$	$\text{ave}_{\mu}(\eta_N(\mu) - 1)$
1	0.0379	4.1932	0.6450
9	0.0109	2.3677	0.3345

Table 4.4: Minimum, maximum, and average effectivities for Example 2 obtained using the SP and PC conditioners.

T	$\min_{\mu}(\eta_N(\mu) - 1)$	$\max_{\mu}(\eta_N(\mu) - 1)$	$\text{ave}_{\mu}(\eta_N(\mu) - 1)$
4	0.0133	2.4546	0.3788
16	0.0027	1.6437	0.1892

Table 4.5: Minimum, maximum, and average effectivities for Example 2 obtained using the PL conditioner.

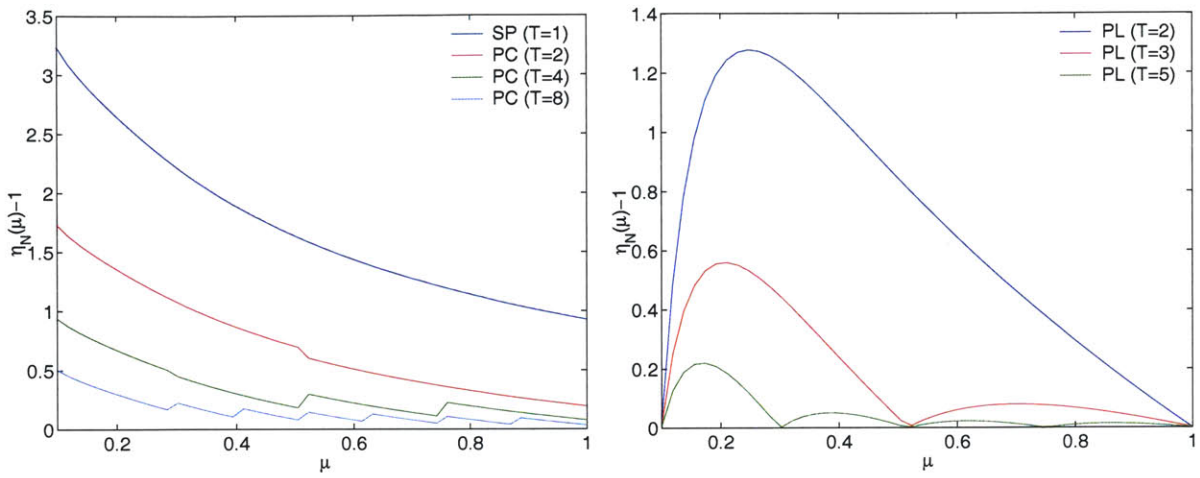


Figure 4-14: Effectivity as a function of μ for Example 3 obtained using (a) SP ($T = 1$), PC ($T = 2, 4, 8$), and (b) PL ($T = 2, 3, 5$).

4.5 Method II: Asymptotic Error Bounds

The essential point is that Method I error estimators guarantee rigorous bounds: in some cases, the resulting error estimates are quite sharp — the effectivity is close to (but always greater than) unity. However, in other cases, either the necessary bound conditioners can not be found or yield unacceptably large effectivities; or the associated computational expense is much too high due to the $O(Q^2)$ on-line scaling induced by (4.42). These disadvantages are eliminated in Method II to be discussed in the next section; however, Method II provides only *asymptotic* bounds as $N \to \infty$. The choice thus depends on the relative importance of absolute certainty and computational efficiency.

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

4.5.1 Error and Output Bounds

To begin, we set $M > N$, and introduce a parameter sample

$$S_M^\mu = \{\mu_1, \ldots, \mu_M\} \tag{4.182}$$

and associated reduced–basis approximation space

$$W_M = \text{span}\,\{\zeta_i \equiv u(\mu_i),\ m = 1, \ldots, M\}\ ; \tag{4.183}$$

for both theoretical and practical reasons we require $S_N^\mu \subset S_M^\mu$ and therefore $W_N \subset W_M$. The procedure is very simple: we first find $u_M(\mu) \in W_M$ such that

$$\langle \mathcal{A}(\mu) u_M(\mu), v\rangle = \langle F, v\rangle, \quad \forall\ v \in W_M\ ; \tag{4.184}$$

we then evaluate

$$s_M(\mu) = \langle L, u_M(\mu)\rangle; \tag{4.185}$$

finally, we compute our upper and lower output estimators as

$$s^-_{N,M}(\mu) = s_N(\mu) , \tag{4.186}$$
$$s^+_{N,M}(\mu) = s_N(\mu) + \Delta_{N,M}(\mu), \tag{4.187}$$

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

$$\Delta_{N,M}(\mu) = \frac{1}{\tau}\left(s_M(\mu) - s_N(\mu)\right) \tag{4.188}$$

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

$$\eta_{N,M}(\mu) = \frac{\Delta_{N,M}(\mu)}{s(\mu) - s_N(\mu)} . \tag{4.189}$$

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

4.5.2 Bounding Properties

As in Section 4.2, we would like to be able to prove the effectivity inequality $1 \leq \eta_{N,2N}(\mu) \leq \rho$, for all $N \geq 1$. However, we can only demonstrate an *asymptotic* form of this inequality [39].

As in [39], we shall require the hypothesis that

$$\mathcal{E}_{N,2N}(\mu) \equiv \frac{s(\mu) - s_{2N}(\mu)}{s(\mu) - s_N(\mu)} \rightarrow 0, \quad \text{as } N \rightarrow \infty. \tag{4.190}$$

We note that the assumption (4.190) is certainly plausible: if the (*a priori*) exponential convergence of (3.105) in fact reflects general asymptotic behavior of the reduced basis approximation, then

$$s(\mu) - s_N(\mu) \sim c_1 e^{-c_2 N}, \tag{4.191}$$
$$s(\mu) - s_{2N}(\mu) \sim c_1 e^{-2c_2 N}, \tag{4.192}$$

and hence

$$\mathcal{E}_{N,2N}(\mu) \sim e^{-c_2 N}, \tag{4.193}$$

as desired.

We first prove the lower effectivity inequality, $\eta_{N,M}(\mu) \geq 1$, which is equivalent to

$$s^-_{N,2N}(\mu) \leq s(\mu) \leq s^+_{N,2N}(\mu), \quad N \rightarrow \infty. \tag{4.194}$$

The lower bound property $s^-_{N,2N}(\mu) \leq s(\mu)$ directly follows from (4.25); indeed, this result still obtains for *all* N. To demonstrate the upper bound, we write

$$s^+_{N,2N}(\mu) = s + \left(\frac{1}{\tau} - 1\right)\left(s(\mu) - s_N(\mu)\right) - \frac{1}{\tau}(s(\mu) - s_{2N}(\mu)) \tag{4.195}$$
$$= s + \left(\frac{1}{\tau}\left(1 - \mathcal{E}_{N,2N}(\mu)\right) - 1\right)\left(s(\mu) - s_N(\mu)\right). \tag{4.196}$$

Since $s(\mu) - s_N(\mu) \geq 0$, and $0 < \tau < 1$, it follows from (4.190) and (4.196) that there exists a finite $N^*(\tau)$ such that

$$1 - \mathcal{E}_{N,2N}(\mu) \geq \tau , \quad \forall\ N > N^*; \tag{4.197}$$

it follows that

$$s^{+}_{N,2N}(\mu) \geq s(\mu) \ , \quad \forall \ N > N^* \ . \tag{4.198}$$

We therefore obtain *asymptotic* bounds.

We now investigate the *sharpness* of our bounds by proving the upper effectivity inequality, $\eta_{N,2N}(\mu) \leq \rho$. From the definitions of $\eta_{N,2N}(\mu)$, $\Delta_{N,2N}(\mu)$, and $\mathcal{E}_{N,2N}(\mu)$, we directly obtain

$$\eta_{N,2N}(\mu) = \frac{1}{\tau}\frac{s_{2N}(\mu) - s_N(\mu)}{s(\mu) - s_N(\mu)} \tag{4.199}$$

$$= \frac{1}{\tau}\frac{(s_{2N}(\mu) - s(\mu)) - (s_N(\mu) - s(\mu))}{(s(\mu) - s_N(\mu))} \tag{4.200}$$

$$= \frac{1}{\tau}(1 - \mathcal{E}_{N,2N}(\mu)). \tag{4.201}$$

We know from (4.25) that $\mathcal{E}_{N,2N}(\mu)$ is strictly non-negative; it therefore follows that

$$\eta_{N,2N}(\mu) \leq \frac{1}{\tau} \ , \quad \forall \ N. \tag{4.202}$$

Also, since $W_N \subset W_{2N}$, it follows from optimality (3.92) and (4.31) that

$$s(\mu) \geq s_{2N}(\mu) \geq s_N(\mu), \tag{4.203}$$

and hence $\mathcal{E}_{N,2N}(\mu) \leq 1$. It therefore follows that

$$\eta_{N,2N}(\mu) \geq 0 \ , \quad \forall \ N. \tag{4.204}$$

Furthermore, from our hypothesis (4.190), we know that

$$\eta_{N,2N}(\mu) \rightarrow \frac{1}{\tau} \ , \quad N \rightarrow \infty. \tag{4.205}$$

4.5.3 Off-line/On-line Computational Procedure

Since the error bounds are based entirely on the evaluation of the output, we can directly adapt the off-line/on-line procedure of Section 3.5.3. 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) "$\mathcal{A}$" finite element solutions and $O(4Q_{\mathcal{A}}N^2)$ 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 $O(4Q_{\mathcal{A}}N^2) + O(\frac{16}{3}N^3)$ operations and $O(4Q_{\mathcal{A}}N^2)$ 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 3.5.

Method II is in some sense very naïve: we simply replace the true output $s(\mu)$ with a finer–approximation surrogate $s_{2N}(\mu)$. (Note there are other ways to describe the method in terms of a reduced–basis approximation for the error [20].) The essential computation enabler is again exponential convergence, which permits us to choose $M = 2N$ — hence controlling the addi-

tional computational effort attributable to error estimation — while simultaneously ensuring that $\mathcal{E}_{N,2N}(\mu)$ 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.

4.5.4 Numerical Results

From our discussion in Section 4.5.2, we observe that 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. These claims are demonstrated in Tables 4.6 and 4.7; the results tabulated correspond to the choice $\tau = 1/2$. In all cases, we clearly obtain bounds even for very small N; and we observe that $\eta_{N,2N}(\mu)$ does, indeed, rather quickly approach $1/\tau$.

N	$\min_\mu \eta_N(\mu)$	$\max_\mu \eta_{N,M}(\mu)$
1	0.28	1.43
2	1.44	2.00
3	1.99	2.00
4	1.99	2.00
5	1.99	2.00

N	$\min_\mu \eta_N(\mu)$	$\max_\mu \eta_{N,M}(\mu)$
1	1.26	1.99
2	1.84	2.00
3	1.95	2.00
4	2.00	2.00
5	2.00	2.00

Table 4.6: Minimum and maximum effectivities for the Method II error estimators for Examples 1 and 2.

N	$\min_\mu \eta_N(\mu)$	$\max_\mu \eta_{N,M}(\mu)$
1	0.81	2.00
2	1.55	2.00
3	1.68	2.00
4	1.82	2.00

N	$\min_\mu \eta_N(\mu)$	$\max_\mu \eta_{N,M}(\mu)$
1	0.01	2.00
2	0.24	2.00
3	1.21	2.00
4	1.71	2.00

Table 4.7: Minimum and maximum effectivities for the Method II error estimators for Examples 3 and 4.

The choice between Method I and Method II error estimators depends on the relative importance of absolute certainty and computational efficiency. There are certainly cases in which the loss of complete certainty is acceptable; however, there are many cases — for instance, when the output $s(\mu)$ must satisfy constraints critical to performance and safety — in which certainty in the approximation is crucial, so that Method I error estimators are strongly preferred. However, the ability to calculate rigorous uniform bounds hinges on the availability of a bound conditioner which satisfies the spectral condition (4.10) and the computational invertibility hypothesis (4.11); in some cases either the necessary bound conditioners cannot be found, or the computational and algorithmic complexity is much too high. In these cases, Method II error estimators provide a simple and effective alternative.

Chapter 5

Reduced-Basis Output Bounds for Linear Elasticity and Noncompliant Outputs

5.1 Introduction

In Chapter 3 we present the reduced basis method for approximating compliant linear functional outputs associated with elliptic partial differential equations. In Chapter 4, we develop two general approaches for *a posteriori* error estimation: Method I, which provides rigorous uniform error bounds obtained from relaxations of the error residual equation; and Method II, which provides asymptotic error bounds obtained from "finer" reduced basis approximations. In both chapters, we utilize several simple problems in steady-state heat conduction to illustrate our methods and numerically verify our claims.

In this chapter, we apply reduced basis approximation and error estimation methods to linear elasticity. The problem of linear elasticity brings about some challenging but important issues: first, a decidedly more complex parametric dependence — related to the affine decomposition of the operator — causes an increase in the required computational expense and storage; and second, singularities in the elasticity tensor associated with pure-rotation modes — related to the coercivity constant of the operator — causes the effectivity of the error estimators in certain cases to be unacceptably high.

Furthermore, in Chapters 3 and 4 the methods are restricted to compliant parameter-independent outputs, $L = F$. We formulate here the reduced-basis method and associated error estimation procedures for more general linear bounded output functionals.

5.2 Abstraction

As in Chapter 3, we consider a suitably regular (smooth) parameter-independent domain $\Omega \subset \mathbb{R}^d, d = 1, 2$, or 3, and associated function space $Y \subset (H^1(\Omega))^p$ with inner product $(\cdot, \cdot)_Y$, norm $\|\cdot\|_Y = (\cdot, \cdot)_Y^{1/2}$, dual space Y', and duality pairing $\langle \cdot, \cdot \rangle = {}_{Y'}\langle \cdot, \cdot \rangle_Y$; as before, we define a parameter set $\mathcal{D}^\mu \subset \mathbb{R}^P$, a particular point in which will be denoted μ.

We now consider the symmetric, continuous, and coercive distributional (second-order partial differential) operator $\mathcal{A}(\mu) : Y \to Y'$, and introduce the bounded linear forms $F(\mu) \in Y'$ and $L(\mu) \in Y'$; note we no longer assume that $L = F$. We shall again make certain assumptions on

the parametric dependence of $\mathcal{A}$, F, and L. In particular, we shall suppose that, for some finite (preferably small) integers $Q_\mathcal{A}$, Q_F, and Q_L, $\mathcal{A}(\mu)$, $F(\mu)$, and $L(\mu)$ may be expressed as

$$\mathcal{A}(\mu) = \sum_{q=1}^{Q_\mathcal{A}} \Theta^q(\mu)\, \mathcal{A}^q , \quad \forall\, \mu \in \mathcal{D}^\mu , \tag{5.1}$$

$$F(\mu) = \sum_{q=1}^{Q_F} \varphi_F^q(\mu)\, F^q , \quad \forall\, \mu \in \mathcal{D}^\mu , \tag{5.2}$$

$$L(\mu) = \sum_{q=1}^{Q_L} \varphi_L^q(\mu)\, L^q , \quad \forall\, \mu \in \mathcal{D}^\mu , \tag{5.3}$$

where the $\Theta^q(\mu)\colon \mathcal{D}^\mu \to \mathbb{R}$, $\varphi_F^q(\mu)\colon \mathcal{D}^\mu \to \mathbb{R}$, $\varphi_L^q(\mu)\colon \mathcal{D}^\mu \to \mathbb{R}$, $\mathcal{A}^q\colon Y \to Y'$, $F^q \in Y'$, and $L^q \in Y'$. As earlier indicated, this assumption of "separability" or affine parameter dependence is crucial to computational efficiency.

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

$$s(\mu) = \langle L(\mu), u(\mu) \rangle, \tag{5.4}$$

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

$$\langle \mathcal{A}(\mu) u(\mu), v \rangle = \langle F(\mu), v \rangle, \quad \forall\, v \in Y. \tag{5.5}$$

Thus, (5.5) is our partial differential equation (in weak form), μ is our parameter, $u(\mu)$ is our field variable, and $s(\mu)$ is our output.

In actual practice, Y is replaced by an appropriate "truth" finite element approximation space $Y_\mathcal{N}$ of dimension $\mathcal{N}$ defined on a suitably fine truth mesh. We then approximate $u(\mu)$ and $s(\mu)$ by $u_\mathcal{N}(\mu)$ and $s_\mathcal{N}(\mu)$, respectively, and assume that $Y_\mathcal{N}$ is sufficiently rich such that $u_\mathcal{N}(\mu)$ and $s_\mathcal{N}(\mu)$ are indistinguishable from $u(\mu)$ and $s(\mu)$.

5.3 Formulation of the Linear Elasticity Problem

In this section we present the strong form of the equations of linear elasticity, from which we derive the weak statement; we then reformulate the problem in terms of a reference (parameter-independent) domain, thus recovering the abstract formulation of Section 5.2. In this and the following sections, repeated indices imply summation, and, unless otherwise indicated, indices take on the values 1 through d, where d is the dimensionality of the problem. Furthermore, in this section we use a bar to signify a general dependence on the parameter μ (e.g., $\bar{\Omega} \equiv \bar{\Omega}(\mu)$, or $\bar{E} \equiv \mu$) particularly when formulating the problem in a "non-reference" domain.

5.3.1 Governing Equations

For simplicity of exposition, we present the governing equations for the case of a homogeneous body, and merely state the weak statement for the more general case of an inhomogeneous body.

Strong Formulation

We consider the deformation of a homogeneous body $\bar{\Omega} \subset \mathbb{R}^d$ with boundary $\bar{\Gamma}$ subjected to external body forces, surface tractions and homogeneous displacement boundary conditions. We assume that the displacement gradients are sufficiently small that a linear elasticity model adequately describes the deformation; in particular, we assume that the equations of equilibrium are satisfied in the undeformed configuration

$$\frac{\partial \bar{\sigma}_{ij}}{\partial \bar{x}_j} + \bar{b}_i = 0 \quad \text{in } \bar{\Omega} \ , \tag{5.6}$$

and the stresses $\bar{\sigma}_{ij}$ are related to the linearized strains $\bar{\varepsilon}_{ij}$ by the constitutive equations

$$\bar{\sigma}_{ij} = \bar{C}_{ijkl}\bar{\varepsilon}_{kl} \ , \tag{5.7}$$

where the linearized strains are related to the displacements $\bar{u}$ by

$$\bar{\varepsilon}_{kl} = \frac{1}{2}\left(\frac{\partial \bar{u}_k}{\partial \bar{x}_l} + \frac{\partial \bar{u}_l}{\partial \bar{x}_k}\right). \tag{5.8}$$

Assuming the material is isotropic[1], the elasticity tensor has the form

$$\bar{C}_{ijkl} = \bar{c}_1 \delta_{ij}\delta_{kl} + \bar{c}_2 \left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) \ ; \tag{5.9}$$

we note that from the symmetry of $\bar{\sigma}_{ij}$ and $\bar{\varepsilon}_{kl}$, and from isotropy, the elasticity tensor satisfies

$$\bar{C}_{ijkl} = \bar{C}_{jikl} = \bar{C}_{ijlk} = \bar{C}_{klij} \ , \tag{5.10}$$

from which it follows that

$$\bar{\sigma}_{ij} = \bar{C}_{ijkl}\frac{\partial \bar{u}_k}{\partial \bar{x}_l} \ , \tag{5.11}$$

Here, $\bar{c}_1$ and $\bar{c}_2$ are Lame's constants, related to Young's modulus, $\bar{E}$, and Poisson's ratio, $\bar{\nu}$, by

$$\bar{c}_1 = \frac{\bar{E}\bar{\nu}}{(1+\bar{\nu})(1-2\bar{\nu})} \ , \tag{5.12}$$

$$\bar{c}_2 = \frac{\bar{E}}{2(1+\bar{\nu})} \ . \tag{5.13}$$

The displacement and traction boundary conditions are given by

$$\bar{u}_i \bar{e}_i^n = 0 \ , \qquad \text{on } \bar{\Gamma}_{\mathrm{D}}^n \ , \tag{5.14}$$

$$\bar{u}_i \bar{e}_i^t = 0 \ , \qquad \text{on } \bar{\Gamma}_{\mathrm{D}}^t \ , \tag{5.15}$$

[1]Note the assumption of isotropy is introduced only for simplicity — the methods we develop are similarly applicable to general anisotropic materials so long as the spatial and parametric dependence of the elasticity tensor are separable, i.e., $\bar{C}_{ijkl}(\bar{x};\mu) = \bar{\bar{f}}(\bar{x})\bar{\bar{C}}_{ijkl}(\mu)$.

and

$$\bar{\sigma}_{ij}\bar{e}_j^n = \begin{cases} \bar{f}_n\bar{e}_i^n & \text{on } \bar{\Gamma}_N^n \\ 0 & \text{on } \bar{\Gamma}\backslash(\bar{\Gamma}_D^n \cup \bar{\Gamma}_N^n) \end{cases} \tag{5.16}$$

$$\bar{\sigma}_{ij}\bar{e}_j^n = \begin{cases} \bar{f}_t\bar{e}_i^t & \text{on } \bar{\Gamma}_N^t \\ 0 & \text{on } \bar{\Gamma}\backslash(\bar{\Gamma}_D^t \cup \bar{\Gamma}_N^t) \ , \end{cases} \tag{5.17}$$

respectively, where $\bar{e}_i^n$ and $\bar{e}_i^t$ are the unit normal and tangent vectors on $\bar{\Gamma}$.

Weak Formulation

We now derive the weak form of the governing equations. To begin, we introduce the function space

$$\bar{Y} = \{\bar{v} \in \left(H^1(\bar{\Omega})\right)^{p=d} \mid \bar{v}_i\bar{e}_i^n = 0 \text{ on } \bar{\Gamma}_D^n,\ \bar{v}_i\bar{e}_i^t = 0 \text{ on } \bar{\Gamma}_D^t\} \ , \tag{5.18}$$

and associated norm

$$||\bar{v}||_{\bar{Y}} = \left(\sum_{i=1}^{d} ||\bar{v}_i||^2_{H_1(\bar{\Omega})}\right)^{1/2} = \left(\int_{\bar{\Omega}} \sum_{i,j=1}^{d} \left(\frac{\partial \bar{v}_i}{\partial \bar{x}_j}\right)^2 d\bar{\Omega}\right)^{1/2} . \tag{5.19}$$

Multiplying (5.6) by a test function $\bar{v} \in \bar{Y}$ and integrating over $\bar{\Omega}$, we obtain

$$-\int_{\bar{\Omega}} \bar{v}_i \frac{\partial \bar{\sigma}_{ij}}{\partial \bar{x}_j}\, d\bar{\Omega} = \int_{\bar{\Omega}} \bar{b}_i\bar{v}_i\, d\bar{\Omega} \ , \qquad \forall\ \bar{v} \in \bar{Y} \ . \tag{5.20}$$

Integrating by parts and applying the divergence theorem yields

$$-\int_{\bar{\Omega}} \bar{v}_i \frac{\partial \bar{\sigma}_{ij}}{\partial \bar{x}_j}\, d\bar{\Omega} = -\int_{\bar{\Gamma}} \bar{v}_i\bar{\sigma}_{ij}\bar{e}_j^n\, d\bar{\Gamma} + \int_{\bar{\Omega}} \frac{\partial \bar{v}_i}{\partial \bar{x}_j}\bar{\sigma}_{ij}\, d\bar{\Omega} \ , \qquad \forall\ \bar{v} \in \bar{Y} \ . \tag{5.21}$$

Substituting (5.21) into (5.20), and using (5.14), (5.16), and the fact that $\bar{v} \in \bar{Y}$, we find that

$$\int_{\bar{\Omega}} \frac{\partial \bar{v}_i}{\partial \bar{x}_j}\bar{\sigma}_{ij}\, d\bar{\Omega} = \int_{\bar{\Omega}} \bar{v}_i\bar{b}_i\, d\bar{\Omega} + \int_{\bar{\Gamma}_N^n} \bar{v}_i\bar{f}_n\bar{e}_i^n\, d\bar{\Gamma} + \int_{\bar{\Gamma}_N^t} \bar{v}_i\bar{f}_t\bar{e}_i^t\, d\bar{\Gamma} \ . \tag{5.22}$$

Finally, using (5.11), we obtain as our weak statement

$$\langle \bar{\mathcal{A}}\bar{u}, \bar{v}\rangle = \langle \bar{F}, \bar{v}\rangle \ , \qquad \forall\ \bar{v} \in \bar{Y} \ , \tag{5.23}$$

where

$$\langle \bar{\mathcal{A}}\bar{w}, \bar{v}\rangle = \int_{\bar{\Omega}} \frac{\partial \bar{v}_i}{\partial \bar{x}_j}\bar{C}_{ijkl}\frac{\partial \bar{w}_k}{\partial \bar{x}_l} \tag{5.24}$$

$$\langle \bar{F}, \bar{v}\rangle = \langle \bar{F}_f, \bar{v}\rangle + \langle \bar{F}_b, \bar{v}\rangle \ ; \tag{5.25}$$

here,

$$\langle \bar{F}_f, \bar{v}\rangle = \int_{\bar{\Gamma}_N^n} \bar{v}_i\bar{f}_n\bar{e}_i^n\, d\bar{\Gamma} + \int_{\bar{\Gamma}_N^t} \bar{v}_i\bar{f}_t\bar{e}_i^t\, d\bar{\Gamma} \ , \qquad \langle \bar{F}_b, \bar{v}\rangle = \int_{\bar{\Omega}} \bar{v}_i\bar{b}_i\, d\bar{\Omega} \ . \tag{5.26}$$

We now generalize (5.23)-(5.26) to the case in which $\bar{\Omega}$ is inhomogeneous. Assuming that $\bar{\Omega}$

consists of $\bar{R}$ homogeneous subdomains $\bar{\Omega}^{\bar{r}}$ such that

$$\overline{\bar{\Omega}} = \bigcup_{\bar{r}=1}^{\bar{R}} \overline{\bar{\Omega}}^{\bar{r}} , \tag{5.27}$$

the weak statement takes the form of (5.23) where

$$\langle \bar{\mathcal{A}}\bar{w}, \bar{v} \rangle = \sum_{\bar{r}=1}^{\bar{R}} \int_{\bar{\Omega}^{\bar{r}}} \frac{\partial \bar{v}_i}{\partial \bar{x}_j} \bar{C}^{\bar{r}}_{ijkl} \frac{\partial \bar{w}_k}{\partial \bar{x}_l} \, d\bar{\Omega} \tag{5.28}$$

$$\langle \bar{F}, \bar{v} \rangle = \langle \bar{F}_f, \bar{v} \rangle + \langle \bar{F}_b, \bar{v} \rangle , \tag{5.29}$$

and

$$\langle \bar{F}_b, \bar{v} \rangle = \sum_{r=1}^{\bar{R}} \int_{\bar{\Omega}^{\bar{r}}} \bar{v}_i \bar{b}^{\bar{r}}_i \, d\bar{\Omega} , \qquad \langle \bar{F}_f, \bar{v} \rangle = \sum_{\bar{r}=1}^{\bar{R}} \left(\int_{\bar{\Gamma}^{n\,\bar{r}}_{\mathrm{N}}} \bar{v}_i \bar{f}^{\bar{r}}_n \bar{\mathrm{e}}^n_i \, d\bar{\Gamma} + \int_{\bar{\Gamma}^{t\,\bar{r}}_{\mathrm{N}}} \bar{v}_i \bar{f}^{\bar{r}}_t \bar{\mathrm{e}}^t_i \, d\bar{\Gamma} \right) ; \tag{5.30}$$

here, $\overline{\bar{\Omega}}$ is the closure of $\bar{\Omega}$, $\bar{C}^{\bar{r}}_{ijkl}$ is the elasticity tensor in $\bar{\Omega}^{\bar{r}}$, and $\bar{\Gamma}^{\bar{r}}_{\mathrm{N}}$ is the section of $\bar{\Gamma}_{\mathrm{N}}$ in $\overline{\bar{\Omega}}^{\bar{r}}$. The derivation of (5.28)-(5.30) is similar to that for the homogeneous case, but for the use of additional displacement and traction continuity conditions at the interfaces between the $\bar{\Omega}^{\bar{r}}$.

5.3.2 Reduction to Abstract Form

In this section, we reformulate the problem defined by (5.28)-(5.30) so as to recover the abstract formulation of Section 5.2.

Affine Geometric Mapping

As before, we further partition the subdomains $\bar{\Omega}^{\bar{r}}$, $\bar{r} = 1, \ldots, \bar{R}$, into a total of R subdomains $\bar{\Omega}^r$, $r = 1, \ldots, R$ such that there exists a reference domain $\overline{\bar{\Omega}} = \bigcup_{r=1}^{R} \overline{\bar{\Omega}}^r$ where, for any $\underline{\bar{x}} \in \bar{\Omega}^r$, $r = 1, \ldots, R$, its image $\underline{x} \in \Omega^r$ is given by

$$\underline{x} = \mathcal{G}^r(\mu; \underline{\bar{x}}) = \underline{G}^r(\mu)\underline{\bar{x}} + \underline{g}^r(\mu); \tag{5.31}$$

we write

$$\frac{\partial}{\partial \bar{x}_i} = \frac{\partial x_j}{\partial \bar{x}_i} \frac{\partial}{\partial x_j} = G_{ji}(\mu) \frac{\partial}{\partial x_j}, \tag{5.32}$$

and

$$\underline{x} = \mathcal{G}(\mu; \underline{\bar{x}}) = \underline{G}(\mu)\underline{\bar{x}} + \underline{g}(\mu) , \tag{5.33}$$

where $\underline{x} \in \Omega$, $\underline{\bar{x}} \in \bar{\Omega}$, $\underline{G}(\mu) \in \mathbb{R}^{d \times d}$ is a piecewise-constant matrix, $\underline{g}^r(\mu) \in \mathbb{R}^d$ is a piecewise-constant vector, and $\mathcal{G}(\mu) \colon \bar{\Omega} \to \Omega$ is a piecewise-affine geometric mapping.

Reference Domain Formulation

We now define the function space Y as $Y(\Omega) = \bar{Y}(\mathcal{G}^{-1}(\mu; \Omega)) = \bar{Y}(\bar{\Omega})$, i.e.,

$$Y = \{ v \in \left(H^1(\Omega)\right)^{p=d} \mid v_i \mathrm{e}^n_i = 0 \text{ on } \Gamma^n_{\mathrm{D}}, \; v_i \mathrm{e}^t_i = 0 \text{ on } \Gamma^t_{\mathrm{D}} \} , \tag{5.34}$$

and for any function $\bar{w} \in \bar{Y}$, we define $w \in Y$ such that $w(\underline{x}) = \bar{w}(\mathcal{G}^{-1}(\mu; \underline{x}))$. Furthermore, we have

$$d\bar{\Omega} = \det \underline{G}^{-1}(\mu)\, d\Omega \,, \tag{5.35}$$

$$d\bar{\Gamma} = \left|\underline{G}^{-1}(\mu)\, \underline{e}^t\right| d\Gamma \,, \tag{5.36}$$

where $\underline{e}^t$ is a unit vector tangent to the boundary Γ, and

$$\left|\underline{G}^{-1}(\mu)\, \underline{e}^t\right| = \left(\sum_{i=1}^{d} \left(G_{ij} e_j^t\right)^2\right)^{1/2} . \tag{5.37}$$

It then follows that $\langle \mathcal{A}(\mu)w, v\rangle = \langle \bar{\mathcal{A}}\bar{w}, \bar{v}\rangle$ for $\bar{\mathcal{A}}$ as in (5.28) and $\mathcal{A}(\mu)$ given by

$$\langle \mathcal{A}(\mu)w, v\rangle = \sum_{r=1}^{R} \int_{\Omega^r} \left(G^r_{jj'}(\mu)\frac{\partial w_i}{\partial x_j}\right) \bar{E}^r_{ij'kl'} \left(G^r_{ll'}(\mu)\frac{\partial v_k}{\partial x_l}\right) \det(\underline{G}^r(\mu))^{-1}\, d\Omega \,, \tag{5.38}$$

$$= \sum_{r=1}^{R} \int_{\Omega^r} \frac{\partial w_i}{\partial x_j} \left(G^r_{jj'}(\mu) \bar{E}^r_{ij'kl'} G^r_{ll'}(\mu) \det(\underline{G}(\mu))^{-1}\right) \frac{\partial v_k}{\partial x_l}\, d\Omega \quad \forall\, w, v \in Y \,, \tag{5.39}$$

and $\langle F(\mu)w, v\rangle = \langle \bar{F}\bar{w}, \bar{v}\rangle$ for $\bar{F}$ as in (5.29) and $F(\mu)$ given by

$$\langle F(\mu), v\rangle = \sum_{r=1}^{R} \left(\int_{\Omega^r} \left(\bar{b}^r_i \det\left(\underline{G}^r(\mu)\right)^{-1}\right) v_i\, d\Omega + \int_{\Gamma_N^{n\,r}} \left(\bar{f}^r_n \bar{e}^n_i \left|(\underline{G}^r(\mu))^{-1}\underline{e}^t\right|\right) v_i\, d\Gamma \right.$$
$$\left. + \int_{\Gamma_N^{t\,r}} \left(\bar{f}^r_t \bar{e}^t_i \left|(\underline{G}^r(\mu))^{-1}\underline{e}^t\right|\right) v_i\, d\Gamma \right) . \tag{5.40}$$

The abstract problem statement of Section 5.2 is then recovered for

$$\langle \mathcal{A}(\mu)w, v\rangle = \sum_{r=1}^{R} \int_{\Omega^r} \frac{\partial w_i}{\partial x_j} C^r_{ijkl}(\mu) \frac{\partial v_k}{\partial x_l} \qquad \forall\, w, v \in Y \,, \tag{5.41}$$

$$\langle F(\mu), v\rangle = \left(\langle F_b(\mu), v\rangle + \langle F_f(\mu), v\rangle\right) \tag{5.42}$$

where

$$\langle F_b(\mu), v\rangle = \sum_{r=1}^{R} \int_{\Omega^r} b^r_i(\mu) v_i \tag{5.43}$$

$$\langle F_f(\mu), v\rangle = \sum_{r=1}^{R} \left(\int_{\Gamma_N^{n\,r}} f_i^{n\,r}(\mu) v_i\, d\Gamma + \int_{\Gamma_N^{t\,r}} f_i^{t\,r}(\mu) v_i\, d\Gamma \right) \,; \tag{5.44}$$

here $C^r_{ijkl}(\mu)$ is given by

$$C^r_{ijkl}(\mu) = G^r_{jj'}(\mu)\ \bar{E}^r_{ij'kl'}\ G^r_{ll'}(\mu)\ \det(\underline{G}^r(\mu))^{-1} \,, \tag{5.45}$$

and $b_i^r(\mu)$, $f_i^{n\,r}(\mu)$, and $f_i^{t\,r}(\mu)$ are given by

$$b_i^r(\mu) = \bar{b}_i^r \det\left(\underline{G}^r(\mu)\right)^{-1} , \tag{5.46}$$

$$f_i^{n\,r}(\mu) = \bar{f}_n^r \bar{e}_i^n \left|(\underline{G}^r(\mu))^{-1}\underline{e}^t\right| , \tag{5.47}$$

$$f_i^{t\,r}(\mu) = \bar{f}_t^r \bar{e}_i^t \left|(\underline{G}^r(\mu))^{-1}\underline{e}^t\right| . \tag{5.48}$$

Furthermore, we define

$$\Theta^{q(i,j,k,l,r)}(\mu) = C_{ijkl}^r(\mu), \qquad \langle \mathcal{A}^{q(i,j,k,l,r)} w, v\rangle = \int_\Omega \frac{\partial v_i}{\partial x_j}\frac{\partial w_k}{\partial x_l} \tag{5.49}$$

for $i,j,k,l \in \{1,\ldots,d\}$, $1 \le r \le R$, and q: $\{1,\ldots,d\}^4 \times \{1,\ldots,R\} \to \{1,\ldots,Q_{\mathcal{A}}\}$; and

$$\varphi_F^{q'(i,r,\chi)} = \begin{cases} b_i^r(\mu) & \text{for } \chi = 1 , \\ f_i^{n\,r}(\mu) & \text{for } \chi = 2 , \\ f_i^{t\,r}(\mu) & \text{for } \chi = 3 , \end{cases} \qquad F^{q(i,r,\chi)} = \begin{cases} \int_{\Omega^r} v_i & \text{for } \chi = 1 , \\ \int_{\Gamma_{\rm N}^{n\,r}} v_i & \text{for } \chi = 2 , \\ \int_{\Gamma_{\rm N}^{t\,r}} v_i & \text{for } \chi = 3 , \end{cases} \tag{5.50}$$

for $1 \le i \le d$, $1 \le r \le R$, and q': $\{1,\ldots,d\} \times \{1,\ldots,R\} \to \{1,\ldots,Q_F\}$. Note however that due to the symmetry of $C_{ijkl}(\mu)$, $Q_{\mathcal{A}}$ can in fact be taken to be $d^2(d^2+1)R/2$ rather than d^4R; furthermore, $Q_{\mathcal{A}}$ and Q_F can be further reduced by eliminating the elements of $C_{ijkl}(\mu)$ which are identically zero. This is illustrated in the model problems to be discussed in the following section.

5.4 Model Problems

We indicate here several simple instantiations of the linear elasticity problem; these examples will serve to illustrate our assumptions and methods, and show ways in which the affine parameter dependence may be further simplified.

5.4.1 Example 5

We consider the compression of the prismatic bi-material rod shown in Figure 5-1; the physical model is simple plane-strain linear elasticity. The rod has a Poisson's ratio of $\nu = 0.25$, and a Young's modulus of $E^1 = 1$ in Ω^1, and $E^2 = \mu$ in Ω^2, where $\Omega^1 = (0,2)\times(0.0,0.5)$ and $\Omega^2 = (0,2)\times(0.5,1.0)$). The structure is clamped ($u = 0$) along $\Gamma_{\rm D} = \Gamma_{\rm D}^n = \Gamma_{\rm D}^t$, and subjected to a compressive unit load (per unit length) uniformly distributed along $\Gamma_{\rm N}^n$; the remainder of the boundary is traction-free. The output of interest is the displacement in the negative x_1-direction averaged along $\Gamma_{\rm N}^n$:

$$s(\mu) = -\int_{\Gamma_{\rm N}^n} u_1(\mu) \quad \text{for } \mu \in \mathcal{D}^\mu \equiv [0.1,1]. \tag{5.51}$$

Our problem can then be formulated as: given any $\mu \in \mathcal{D}^\mu \subset \mathbb{R}^{P=1}$, find $s(\mu) = \langle L, u(\mu)\rangle$, where $u(\mu) \in Y = (H_0^1(\Omega))^2$ is the solution to $\langle \mathcal{A}(\mu)u(\mu), v\rangle = \langle F, v\rangle$, $\forall\ v \in Y$, $Y = \{v \in$

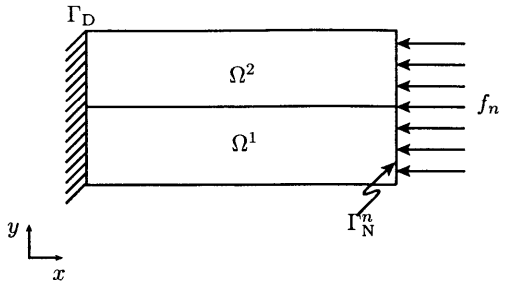


Figure 5-1: Bi-material rectangular rod under compressive loading.

$(H^1(\Omega))^2 \mid v|_{\Gamma_D} = 0\}$. For this example, $\langle L, v\rangle = \langle F, v\rangle$ for all $v \in Y$,

$$\langle \mathcal{A}(\mu)w, v\rangle = \int_{\Omega^1} \frac{\partial v_i}{\partial x_j} C^1_{ijkl} \frac{\partial w_k}{\partial x_l}\, d\Omega + \mu \int_{\Omega^2} \frac{\partial v_i}{\partial x_j} C^1_{ijkl} \frac{\partial w_k}{\partial x_l}\, d\Omega\ , \quad \forall\ w, v \in Y, \tag{5.52}$$

$$\langle F, v\rangle = -\int_{\Gamma_N^n} v_1\, d\Gamma\ , \quad \forall\ v \in Y\ , \tag{5.53}$$

and,

$$C^1_{ijkl} = c^1_1 \delta_{ij}\delta_{kl} + c^1_2 \left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right)\ , \tag{5.54}$$

where

$$c^1_1 = \frac{E^1 \nu}{(1+\nu)(1-2\nu)}\ , \tag{5.55}$$

$$c^1_2 = \frac{E^1}{2(1+\nu)}\ , \tag{5.56}$$

for $E^1 = 1$ and $\nu = 0.25$. The abstract problem statement of Section 5.2 is then recovered for $Q_{\mathcal{A}} = 2$, and

$$\Theta^1(\mu) = 1, \qquad \langle \mathcal{A}^1 w, v\rangle = \int_{\Omega^1} \frac{\partial v_i}{\partial x_j} C^1_{ijkl} \frac{\partial w_k}{\partial x_l}\, d\Omega\ , \tag{5.57}$$

$$\Theta^2(\mu) = \mu, \qquad \langle \mathcal{A}^2 w, v\rangle = \int_{\Omega^2} \frac{\partial v_i}{\partial x_j} C^1_{ijkl} \frac{\partial w_k}{\partial x_l}\, d\Omega\ . \tag{5.58}$$

Note that F and L are compliant ($L = F$) and independent of μ.

5.4.2 Example 6

We now consider a homogeneous rod of variable thickness, as shown in Figure 5-2. The rod has a Young's modulus of $E = 1$ and Poisson's ratio $\nu = 0.25$ in the domain $\bar{\Omega} = (0,1) \times (0,t)$. The structure is clamped along $\bar{\Gamma}_D (= \bar{\Gamma}^n_D = \bar{\Gamma}^t_D)$, and subjected to a compressive load (per unit length) with magnitude

$$\bar{f}_n = \frac{1}{t}\ , \tag{5.59}$$

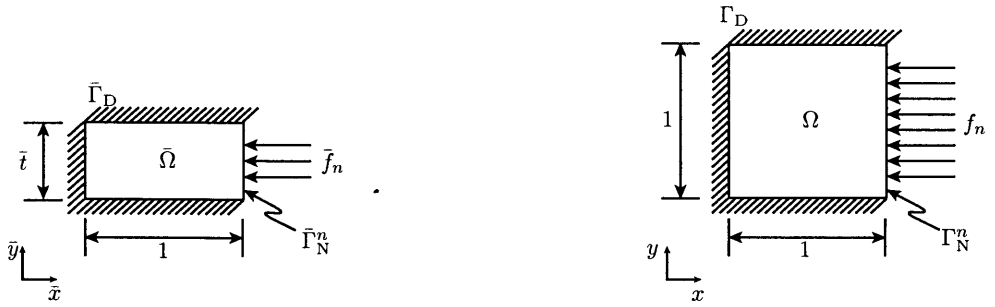


Figure 5-2: (a) Homogeneous rectangular rod with variable thickness subjected to a compressive load; and (b) parameter-independent reference domain.

uniformly distributed along $\bar{\Gamma}_N^n$. The output of interest is the displacement in the negative $\bar{x}_1$-direction averaged along $\bar{\Gamma}_N^n$:

$$s(\mu) = -\frac{1}{\mu}\int_{\bar{\Gamma}_N^n} \bar{u}_1 \, d\bar{\Gamma} \quad \text{for } \mu = \{t\} \in \mathcal{D}^\mu \equiv [0.03125, 2.0] \; . \tag{5.60}$$

The problem can then be formulated as: given a $\mu \in \mathcal{D}^\mu \subset \mathbb{R}^{P=1}$, find $s(\mu) = \langle \bar{L}, \bar{u} \rangle$, where $\bar{u} \in \bar{Y} = (H_0^1(\bar{\Omega}))^2$ is the solution to $\langle \bar{\mathcal{A}}(\mu)\bar{u}, \bar{v} \rangle = \langle \bar{F}, \bar{v} \rangle$, $\forall\ \bar{v} \in \bar{Y}$; here, $\langle \bar{L}, \bar{v} \rangle = \langle \bar{F}, \bar{v} \rangle$ for all $\bar{v} \in \bar{Y}$,

$$\langle \bar{\mathcal{A}}\bar{w}, \bar{v} \rangle = \int_{\bar{\Omega}} \frac{\partial \bar{v}_i}{\partial \bar{x}_j} \bar{C}_{ijkl} \frac{\partial \bar{w}_k}{\partial \bar{x}_l} \, d\bar{\Omega} \; , \quad \forall\ \bar{w}, \bar{v} \in \bar{Y} \; , \tag{5.61}$$

$$\langle \bar{F}, \bar{v} \rangle = -\frac{1}{\mu}\int_{\bar{\Gamma}_N(\mu)} \bar{v}_1 \, d\bar{\Gamma}, \quad \forall\ \bar{v} \in \bar{Y} \; , \tag{5.62}$$

and $\bar{C}_{ijkl}$ is given by (5.9).

We now map our parameter-dependent domain $\bar{\Omega}$ onto a reference domain Ω. The affine mapping $\mathcal{G}(\bar{x})(\mu)\colon \bar{\Omega}(\mu) \to \Omega$ is given by (5.31) where $\underline{g} = 0$ and

$$\underline{G}(\mu) = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{t} \end{bmatrix} \; . \tag{5.63}$$

Furthermore, we have

$$d\bar{\Omega} = \det \underline{G}^{-1}(\mu) d\Omega = t \, d\Omega \tag{5.64}$$

$$d\bar{\Gamma} = \left| \underline{G}^{-1}(\mu) \, \underline{e}^t \right| d\Gamma = t d\Gamma \; . \tag{5.65}$$

We may now re-formulate our problem in terms of the reference domain. Our problem is then: find $s(\mu) = \langle L, u(\mu) \rangle$ where $u(\mu) \in Y = (H_0^1(\Omega))^2$ is the solution to $\langle \mathcal{A}(\mu)u(\mu), v \rangle = \langle F, v \rangle\ \forall\ v \in Y$; for this example, $\langle L, v \rangle = \langle F, v \rangle$ for all $v \in Y$,

$$\langle \mathcal{A}(\mu)w, v \rangle = \int_{\Omega} \frac{\partial v_i}{\partial x_j} C_{ijkl}(\mu) \frac{\partial w_k}{\partial x_l} \, d\Omega \; , \quad \forall\ w, v \in Y \; , \tag{5.66}$$

and

$$\langle F, v\rangle = -\int_{\Gamma_N} v_1 \, d\Gamma \;, \qquad \forall\, v \in Y \;; \tag{5.67}$$

the effective elasticity tensor $C_{ijkl}(\mu) = G_{jj'}(\mu) C_{ij'kl'} G_{ll'}(\mu) \det \underline{G}^{-1}(\mu)$ is given in Table 5.1, where c_1 and c_2 are given by (5.55) for $E = 1$, $\nu = 0.25$. The abstract problem statement of Section 5.2

$ij\backslash kl$	11	12	21	22
11	$t\,(c_1 + 2c_2)$	0	0	c_1
12	·	$\frac{1}{t}(c_2)$	c_2	0
21	·	·	$t(c_2)$	0
22	·	·	·	$\frac{1}{t}\,(c_1 + 2c_2)$

Table 5.1: Elements of the effective elasticity tensor.

is then recovered for $P = 1$, $Q_{\mathcal{A}} = 3$,

$$\Theta^1(\mu) = 1, \qquad \Theta^2(\mu) = t, \qquad \Theta^3(\mu) = \frac{1}{t}, \tag{5.68}$$

and

$$\langle \mathcal{A}^1 w, v\rangle = c_1 \int_\Omega \left(\frac{\partial v_1}{\partial x_1}\frac{\partial w_2}{\partial x_2} + \frac{\partial v_2}{\partial x_2}\frac{\partial w_1}{\partial x_1} \right) d\Omega + c_2 \int_\Omega \left(\frac{\partial v_1}{\partial x_2}\frac{\partial w_2}{\partial x_1} + \frac{\partial v_2}{\partial x_1}\frac{\partial w_1}{\partial x_2} \right) d\Omega \;, \tag{5.69}$$

$$\langle \mathcal{A}^2 w, v\rangle = (c_1 + 2c_2) \int_\Omega \frac{\partial v_1}{\partial x_1}\frac{\partial w_1}{\partial x_1}\, d\Omega + c_2 \int_\Omega \frac{\partial v_2}{\partial x_1}\frac{\partial w_2}{\partial x_1}\, d\Omega \;, \tag{5.70}$$

$$\langle \mathcal{A}^3 w, v\rangle = (c_1 + 2c_2) \int_\Omega \frac{\partial v_2}{\partial x_2}\frac{\partial w_2}{\partial x_2}\, d\Omega + c_2 \int_\Omega \frac{\partial v_1}{\partial x_2}\frac{\partial w_1}{\partial x_2}\, d\Omega \;. \tag{5.71}$$

5.4.3 Example 7

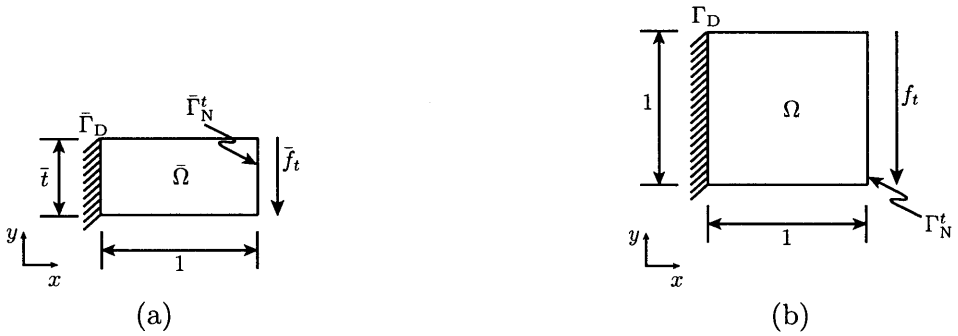


Figure 5-3: (a) Homogeneous rectangular rod with variable thickness subjected to a shear load; and (b) parameter-independent reference domain.

We again consider the homogeneous rod of Example 6, but with $\bar{\Gamma}_D$ as illustrated in Figure 5-3, and subjected to a uniform shear load per unit length with magnitude

$$\bar{f}_t = \frac{1}{t} \tag{5.72}$$

along $\bar{\Gamma}_{\rm N}$. The output of interest is the displacement in the negative $\bar{x}_2$-direction averaged along $\bar{\Gamma}_{\rm N}^t$:

$$s(\mu) = -\frac{1}{\mu}\int_{\bar{\Gamma}_{\rm N}} \bar{u}_2 \quad \text{for } \mu = \{t\} \in \mathcal{D}^\mu \equiv [0.03125, 2.0] \ . \tag{5.73}$$

The problem can then be formulated as: given a $\mu \in \mathcal{D}^\mu \subset \mathbb{R}^{P=1}$, find $s(\mu) = \langle \bar{L}, \bar{u}\rangle$, where $\bar{u} \in \bar{Y} = (H_0^1(\bar{\Omega}))^2$ is the solution to $\langle \bar{\mathcal{A}}(\mu)\bar{u}, \bar{v}\rangle = \langle \bar{F}, \bar{v}\rangle$, $\forall\ \bar{v} \in \bar{Y}$; here, $\langle \bar{L}, \bar{v}\rangle = \langle \bar{F}, \bar{v}\rangle$ for all $\bar{v} \in \bar{Y}$,

$$\langle \bar{\mathcal{A}}\bar{w}, \bar{v}\rangle = \int_{\bar{\Omega}} \frac{\partial \bar{v}_i}{\partial \bar{x}_j} \bar{C}_{ijkl} \frac{\partial \bar{w}_k}{\partial \bar{x}_l} \ , \quad \forall\ \bar{w}, \bar{v} \in \bar{Y} \ , \tag{5.74}$$

$$\langle \bar{F}, \bar{v}\rangle = -\frac{1}{\mu}\int_{\bar{\Gamma}_{\rm N}} \bar{v}_2, \quad \forall\ \bar{v} \in \bar{Y} \ , \tag{5.75}$$

and $\bar{C}_{ijkl}$ is given by (5.9).

Upon application of the affine mapping of Example 6, the problem may then be re-formulated as: find $s(\mu) = \langle L, u(\mu)\rangle$ where $u(\mu) \in Y = (H_0^1(\Omega))^2$ is the solution to $\langle \mathcal{A}(\mu)\bar{u}, \bar{v}\rangle = \langle \bar{F}, \bar{v}\rangle$, $\forall\ \bar{v} \in \bar{Y}$; for this example, $\langle L, v\rangle = \langle F, v\rangle$ for all $v \in Y$,

$$\langle \mathcal{A}(\mu)w, v\rangle = \int_{\Omega} \frac{\partial v_i}{\partial x_j} C_{ijkl}(\mu) \frac{\partial w_k}{\partial x_l}, \quad \forall\ w, v \in Y \ , \tag{5.76}$$

and

$$\langle F, v\rangle = -\int_{\Gamma_{\rm N}} v_2, \qquad \forall\ v \in Y \ ; \tag{5.77}$$

the effective elasticity tensor $C_{ijkl}(\mu)$ is given in Table 5.1. The abstract problem statement of Section 5.2 is then recovered for $Q_{\mathcal{A}} = 3$, where the $\Theta^q(\mu)$ and $\mathcal{A}^q$ are given by (5.68)-(5.71).

5.4.4 Example 8

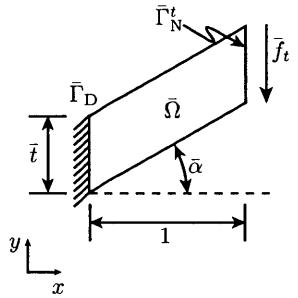


Figure 5-4: Homogeneous rectangular rod with variable thickness and angle subjected to a uniform shear load.

In this example, we again consider the homogeneous rod and loading of Example 7, but in addition to the thickness t, we allow the angle α to vary, as shown in Figure 5-4. The output of interest is again the displacement in the negative x_2-direction averaged along $\bar{\Gamma}_{\rm N}^x t$:

$$s(\mu) = -\frac{1}{t}\int_{\bar{\Gamma}_{\rm N}^t} \bar{u}_2, \quad \text{for } \mu = \{t, \alpha\} \in \mathcal{D}^\mu = [0.03125, 1.0] \times [0°, 45°] \ . \tag{5.78}$$

Our problem can then be formulated as: given a $\mu \in \mathcal{D}^\mu \subset \mathbb{R}^{P=2}$, find $s(\mu) = \langle \bar{L}, \bar{u}\rangle$, where

$\bar{u} \in \bar{Y} = (H_0^1(\bar{\Omega}))^2$ is the solution to $\langle \bar{\mathcal{A}}(\mu)\bar{u}, \bar{v}\rangle = \langle \bar{F}, \bar{v}\rangle$, $\forall\ \bar{v} \in \bar{Y}$; here, $\langle \bar{L}, \bar{v}\rangle = \langle \bar{F}, \bar{v}\rangle$ for all $\bar{v} \in \bar{Y}$, and $\bar{\mathcal{A}}$, $\bar{F}$ are given by (5.74), (5.75), respectively.

We again map our parameter-dependent domain $\bar{\Omega}$ onto a reference domain $\Omega =]0,1[\times]0,1[$, shown in Figure 5-3(b). The affine mapping $\mathcal{G}(\bar{x})(\mu)\colon \bar{\Omega}(\mu) \to \Omega$ is given by (5.31) where $\underline{g} = 0$ and

$$\underline{G}(\mu) = \begin{bmatrix} 1 & \tan\alpha \\ 0 & \frac{1}{t} \end{bmatrix} . \tag{5.79}$$

Furthermore, we have

$$d\bar{\Omega} = \det \underline{G}^{-1}(\mu)\ d\Omega = t\ d\Omega\ , \tag{5.80}$$

$$d\bar{\Gamma} = \left|\underline{G}^{-1}(\mu)\ \underline{e}^t\right| d\Gamma = t\ d\Gamma\ . \tag{5.81}$$

We may now re-formulate our problem in terms of the reference domain. Our problem is then: find $s(\mu) = \langle L, u(\mu)\rangle$ where $u(\mu) \in Y = (H_0^1(\Omega))^2$ is the solution to (5.5); for this example, $\langle L, v\rangle = \langle F, v\rangle$ for all $v \in Y$,

$$\langle \mathcal{A}(\mu)w, v\rangle = \int_\Omega \frac{\partial v_i}{\partial x_j} C_{ijkl}(\mu) \frac{\partial w_k}{\partial x_l}, \quad \forall\ w, v \in Y\ , \tag{5.82}$$

and

$$\langle F, v\rangle = -\int_{\Gamma_N} v_1, \qquad \forall\ v \in Y\ ; \tag{5.83}$$

the effective elasticity tensor $C_{ijkl}(\mu) = G_{jj'}(\mu) C_{ij'kl'} G_{ll'}(\mu) \det \underline{G}^{-1}(\mu)$ is given in Table 5.2, where c_1 and c_2 are given by (5.55) for $E = 1$, $\nu = 0.25$.

$ij\backslash kl$	11	12	21	22
11	$t\left(c_1 + 2c_2 + \tan^2\alpha c_2\right)$	$(\tan\alpha)(c_2)$	$(t\tan\alpha)(c_1 + c_2)$	c_1
12	·	$\frac{1}{t}(c_2)$	c_2	0
21	·	·	$t(c_2 + \tan^2\alpha(c_1 + 2c_2))$	$(\tan\alpha)(c_1 + 2c_2)$
22	·	·	·	$\frac{1}{t}(c_1 + 2c_2)$

Table 5.2: Components of the effective elasticity tensor of Example 7.

The abstract problem statement of Section 5.2 is then recovered for $P = 2$, and $Q_{\mathcal{A}} = 6$, where the $\Theta^q(\mu)$ are the unique elements of the elasticity tensor in Table5.2, and the $\mathcal{A}^q$ are the associated operators.

5.4.5 Example 9

In this example, we again consider the rod of Example 6, but remove Dirichlet conditions and instead apply homogeneous traction conditions on the top surface. We then take as our output of interest

$$s(\mu) = \int_{\bar{\Gamma}_D} (1 - x)\bar{\sigma}_{1j}\bar{e}_j^n\ d\bar{\Gamma} \quad \text{for } \mu = \{t, \alpha\} \in \mathcal{D}^\mu \equiv [0.03125, 1.0] \times [0°, 45°]\ . \tag{5.84}$$

Our problem can then be formulated as: given a $\mu \in \mathcal{D}^\mu \subset \mathbb{R}^{P=2}$, find $s(\mu) = \langle \bar{L}(\mu), \bar{u} \rangle$, where $\bar{u} \in \bar{Y} = (H_0^1(\bar{\Omega}))^2$ is the solution to $\langle \bar{A}(\mu)\bar{u}, \bar{v} \rangle = \langle \bar{F}, \bar{v} \rangle$, $\forall\ \bar{v} \in \bar{Y}$ for this example, $\bar{A}$, $\bar{F}$ are given by (5.74), (5.75), respectively, and

$$\begin{aligned} \langle \bar{L}(\mu), v \rangle &= \langle \bar{A}(\mu)\bar{\chi}, \bar{v} \rangle \\ &= \int_{\bar{\Omega}} \frac{\partial \bar{v}_i}{\partial \bar{x}_j} \bar{C}_{ijkl} \frac{\partial \bar{\chi}_k}{\partial x_l}\, d\bar{\Omega} \ . \end{aligned} \tag{5.85}$$

where $\bar{\chi} \in \bar{Y}$, $\bar{\chi}_1(x) = 1 - x$, and $\bar{\chi}_2(x) = 0$. We note that $s(\mu) = \frac{1}{t}\langle \bar{L}(\mu), \bar{u} \rangle$ since

$$\begin{aligned} \langle \bar{L}(\mu), \bar{u} \rangle &= \int_{\bar{\Omega}} \frac{\partial \bar{u}_i}{\partial \bar{x}_j} \bar{C}_{ijkl} \frac{\partial \bar{\chi}_k}{\partial x_l}\, d\bar{\Omega} \\ &= \int_{\bar{\Omega}} \frac{\partial \bar{\chi}_i}{\partial \bar{x}_j} \bar{C}_{ijkl} \frac{\partial \bar{u}_k}{\partial x_l}\, d\bar{\Omega} \\ &= \int_{\bar{\Omega}} \frac{\partial}{\partial \bar{x}_j} \left(\bar{\chi}_i \bar{C}_{ijkl} \frac{\partial \bar{u}_k}{\partial x_l} \right) d\bar{\Omega} - \int_{\bar{\Omega}} \bar{\chi}_i \frac{\partial}{\partial \bar{x}_j} \left(\bar{C}_{ijkl} \frac{\partial \bar{u}_k}{\partial x_l} \right) d\bar{\Omega} \qquad (5.86) \\ &= \int_{\bar{\Gamma}_{\mathrm{D}}} \bar{\chi}_i \bar{\sigma}_{ij} \bar{\mathrm{e}}_j^n\, d\bar{\Gamma} + \int_{\bar{\Gamma}_{\mathrm{N}}^t} \bar{\chi}_i \bar{\sigma}_{ij} \bar{\mathrm{e}}_j^n\, d\bar{\Gamma} \qquad (5.87) \\ &= \int_{\bar{\Gamma}_{\mathrm{D}}} (1 - x) \bar{\sigma}_{1j} \bar{\mathrm{e}}_j^n\, d\bar{\Gamma} + \frac{1}{t} \int_{\bar{\Gamma}_{\mathrm{N}}^t} \bar{\chi}_2\, d\bar{\Gamma} \qquad (5.88) \\ &= \int_{\bar{\Gamma}_{\mathrm{D}}} (1 - x) \bar{\sigma}_{1j} \bar{\mathrm{e}}_j^n\, d\bar{\Gamma} \ . \qquad (5.89) \end{aligned}$$

We also note that (5.85) is bounded for all $\bar{v} \in \bar{Y}$ while the more obvious alternative (5.84) is not.

5.5 Reduced-Basis Output Approximation: Compliance

The theoretical basis and general computational procedure presented in Chapter 3 for the reduced-basis approximation of compliant linear functional outputs of interest applies to all problems described by the abstract formulation of Section 5.2 (or Section 3.2) for the case $L = F$. We review the computational procedure here — this time allowing F to be μ-dependent — and present numerical results for the linear elasticity model problems of Section 5.4.

5.5.1 Approximation Space

We again introduce a sample in parameter space,

$$S_N^\mu = \{\mu_1, \ldots, \mu_N\} \tag{5.90}$$

where $\mu_n \in \mathcal{D}^\mu \subset \mathbb{R}^P, n = 1, \ldots, N$, and define our Lagrangian ([37]) reduced-basis approximation space as

$$W_N = \mathrm{span}\{\zeta_n \equiv u(\mu_n), n = 1, \ldots, N\}, \tag{5.91}$$

where $u(\mu_n) \in Y$ is the solution to (5.5) for $\mu = \mu_n$.

Our reduced-basis approximation is then: for any $\mu \in \mathcal{D}^\mu$, find

$$s_N(\mu) = \langle L(\mu), u_N(\mu) \rangle, \tag{5.92}$$

where $u_N(\mu) \in W_N$ is the Galerkin projection of $u(\mu)$ onto W_N,

$$\langle \mathcal{A}(\mu) u_N(\mu), v \rangle = \langle F(\mu), v \rangle, \quad \forall\, v \in W_N. \tag{5.93}$$

5.5.2 Off-line/On-line Computational Procedure

We express our approximation $u_N(\mu) \approx u(\mu)$ to be a linear combination of the basis functions,

$$u_N(\mu) = \sum_{j=1}^{N} u_{N\,j}(\mu)\, \zeta_j \tag{5.94}$$

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 (5.93) yields the desired algebraic equations for $\underline{u}_N(\mu) \in \mathbb{R}^N$,

$$\underline{A}_N(\mu)\, \underline{u}_N(\mu) \;=\; \underline{F}_N(\mu)\;, \tag{5.95}$$

in terms of which the output can then be evaluated as

$$s_N(\mu) = \underline{u}_N(\mu)^T \underline{L}_N(\mu)\;. \tag{5.96}$$

Here $\underline{A}_N(\mu) \in \mathbb{R}^{N\times N}$ is the symmetric positive-definite (SPD) matrix with entries

$$A_{N\,i,j}(\mu) \equiv \langle \mathcal{A}(\mu)\zeta_j, \zeta_i \rangle, \quad 1 \le i, j \le N\;, \tag{5.97}$$

$\underline{F}_N(\mu) \in \mathbb{R}^N$ is the "load" vector with entries

$$F_{N\,i}(\mu) \equiv \langle F(\mu), \zeta_i \rangle, \quad 1 \le i \le N\;, \tag{5.98}$$

and $\underline{L}_N(\mu) = \underline{F}_N(\mu)$. We now invoke (5.1) and (5.2) to write

$$A_{N\,i,j}(\mu) \;=\; \sum_{q=1}^{Q_\mathcal{A}} \Theta^q(\mu)\, \langle \mathcal{A}^q \zeta_j, \zeta_i \rangle\;, \tag{5.99}$$

$$F_{N\,i}(\mu) \;=\; \sum_{q=1}^{Q_F} \varphi_F^q(\mu)\, \langle F^q, \zeta_i \rangle\;, \tag{5.100}$$

or

$$\underline{A}_N(\mu) \;=\; \sum_{q=1}^{Q_\mathcal{A}} \Theta^q(\mu)\, \underline{A}_N^q\;, \tag{5.101}$$

$$\underline{F}_N(\mu) \;=\; \sum_{q=1}^{Q_F} \varphi_F^q(\mu)\, \underline{F}_N^q\;, \tag{5.102}$$

where the $\underline{A}_N^q \in \mathbb{R}^{N\times N}$ and $\underline{F}_N^q$ are given by

$$A_{N\,i,j}^q \;=\; \langle \mathcal{A}^q \zeta_j, \zeta_i \rangle\;, \quad 1 \le i, j \le N,\; 1 \le q \le Q_\mathcal{A}\;, \tag{5.103}$$

$$F_{N\,i}^q \;=\; \langle F^q, \zeta_i \rangle\;, \quad 1 \le i \le N,\; 1 \le q \le Q_F\;. \tag{5.104}$$

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 $\underline{F}_N^q$: this requires N (expensive) "$\mathcal{A}$" finite element solutions, and $O(Q_\mathcal{A}N^2) + O(Q_F N)$ finite-element-vector ($O(\mathcal{N})$) inner products. In the *on-line* stage, for any given new μ, we first form $\underline{A}_N(\mu)$ and $F_N(\mu)$ from (5.101) and (5.102), then solve (5.99) for $\underline{u}_N(\mu)$, and finally evaluate $s_N(\mu) = \underline{u}_N(\mu)^T \underline{F}_N(\mu)$: this requires $O(Q_\mathcal{A}N^2 + Q_F N) + O(\frac{2}{3}N^3)$ operations and $O(Q_\mathcal{A}N^2 + Q_F N)$ storage.

Thus, as required, the incremental, or marginal, cost to evaluate $s_N(\mu)$ for any given new μ is very small: first, because N is very small — thanks to the good convergence properties of W_N (see Chapter 3); and second, because (5.99) can be very rapidly assembled and inverted — thanks to the off-line/on-line decomposition. We shall now demonstrate the former by applying our methods to the compliant model problems of Section 5.4.

5.5.3 Numerical Results

We present in Figures 5-6 to 5-8 the maximum relative error $\varepsilon_N(\mu)$ as a function of N for $\mu \in \mathcal{D}^\mu$, for Examples 5 to 8, respectively. In all four cases, the μ_n are chosen "log-randomly" over $\mathcal{D}^\mu$: we sample from a multivariate uniform probability density on $\log(\mu)$, as described in Chapter 3. We note that the error is monotonically decreasing with N (as predicted by our *a priori* theory), and even for cases (such as Example 8) for which $P > 1$, the error is remarkably small even for relatively small N. Furthermore, we again observe that the log–random point distribution *is* important, as evidenced by generally faster convergence (versus the *non–logarithmic* uniform random point distribution), particularly for large ranges of parameter.

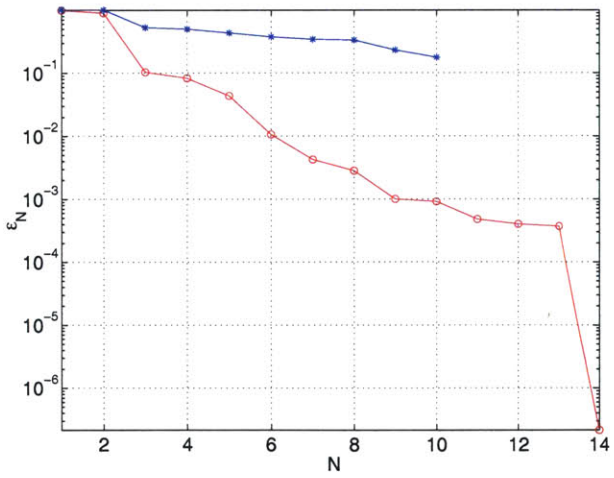


Figure 5-5: Convergence of the reduced-basis approximation for Example 5.

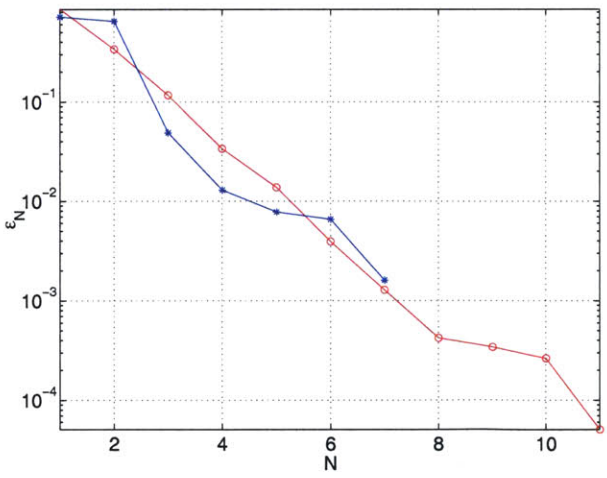


Figure 5-6: Convergence of the reduced-basis approximation for Example 6.

5.6 Reduced-Basis Output Approximation: Noncompliance

In Section 5.5, we formulate the reduced-basis method for the case of compliant outputs, $L(\mu) = F(\mu)$. We briefly summarize here the formulation and theory for more general linear bounded output functionals.

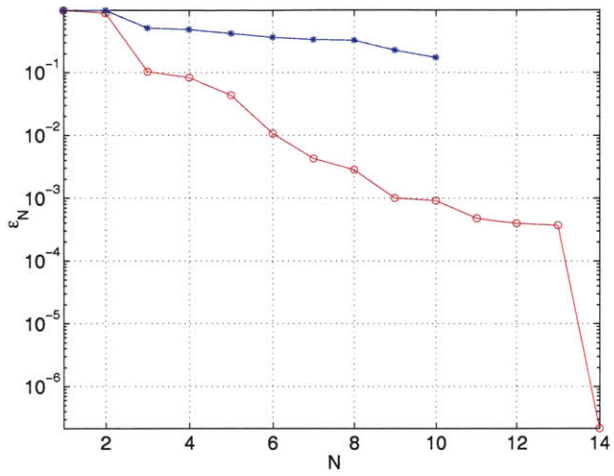


Figure 5-7: Convergence of the reduced-basis approximation for Example 7.

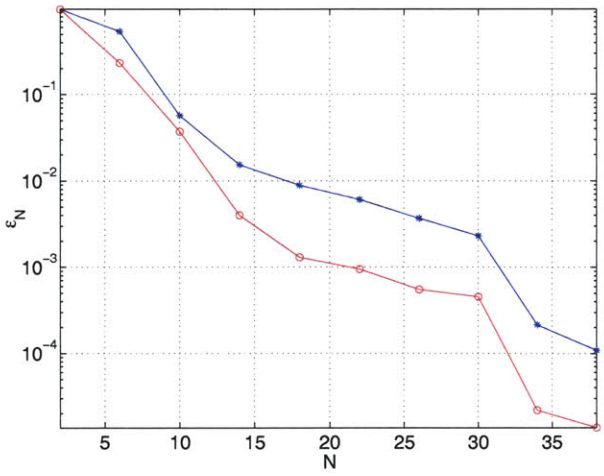


Figure 5-8: Convergence of the reduced-basis approximation for Example 8.

As before, we define the "primal" problem as in (5.5); however, we also introduce an associated adjoint or "dual" problem: for any $\mu \in \mathcal{D}^\mu$, find $\psi(\mu) \in Y$ such that

$$\langle \mathcal{A}(\mu)v, \psi(\mu)\rangle = -\langle L(\mu), v\rangle \ , \qquad \forall\ v \in Y \ ; \tag{5.105}$$

recall that $L(\mu)$ is our output functional.

5.6.1 Approximation Space

To develop the reduced-basis space, we first choose, randomly or log-randomly as described in Chapter 3, a sample set in parameter space,

$$S^\mu_{N/2} = \{\mu_1, \ldots, \mu_{N/2}\} \tag{5.106}$$

where $\mu_n \in \mathcal{D}^\mu \subset \mathbb{R}^P, n = 1, \ldots, N/2$; we now consider two approaches — "nonsegregated" and "segregated" — for defining our Lagrangian ([37]) reduced-basis approximation.

Nonsegregated Approach

We define the *nonsegregated* reduced-basis approximation space as

$$W_N = \text{span}\,\{u(\mu_n), \psi(\mu_n), n = 1, \ldots, N/2\} \tag{5.107}$$
$$\equiv \text{span}\,\{\zeta_n, n = 1, \ldots, N\} \tag{5.108}$$

where $u(\mu_n) \in Y$ and $\psi(\mu_n) \in Y$ are the solutions to (5.5) and (5.105), respectively, for $\mu = \mu_n$.

Our reduced-basis approximation is then: for any $\mu \in \mathcal{D}^\mu$, find

$$s_N(\mu) = \langle L(\mu), u_N(\mu)\rangle, \tag{5.109}$$

where $u_N(\mu) \in W_N$ and $\psi_N(\mu) \in W_N$ are the Galerkin projections of $u(\mu)$ and $\psi(\mu)$, respectively,

onto W_N,

$$\langle \mathcal{A}(\mu) u_N(\mu), v \rangle = \langle F(\mu), v \rangle, \quad \forall\, v \in W_N \,, \tag{5.110}$$

$$\langle \mathcal{A}(\mu) \psi_N(\mu), v \rangle = -\langle L(\mu), v \rangle, \quad \forall\, v \in W_N \,. \tag{5.111}$$

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

$$||u(\mu) - u_N(\mu)||_Y \leq \sqrt{\frac{\gamma_{\mathcal{A}}(\mu)}{\alpha_{\mathcal{A}}(\mu)}} \inf_{w_N \in W_N} ||u(\mu) - w_N||_Y \,, \tag{5.112}$$

$$||\psi(\mu) - \psi_N(\mu)||_Y \leq \sqrt{\frac{\gamma_{\mathcal{A}}(\mu)}{\alpha_{\mathcal{A}}(\mu)}} \inf_{w_N \in W_N} ||\psi(\mu) - w_N||_Y \,. \tag{5.113}$$

The proof of (5.112) and the best approximation analysis is similar to that presented in Section 3.5.2. As regards our output, we now have

$$|s(\mu) - s_N(\mu)| = |\langle L(\mu), u(\mu) \rangle - \langle L(\mu), u_N(\mu) \rangle| \tag{5.114}$$

$$= |\langle \mathcal{A}(\mu)(u(\mu) - u_N(\mu)), \psi(\mu) \rangle| \tag{5.115}$$

$$= |\langle \mathcal{A}(\mu)(u(\mu) - u_N(\mu)), \psi(\mu) - \psi_N(\mu) \rangle| \tag{5.116}$$

$$\leq \gamma_{\mathcal{A}}^0 ||u(\mu) - u_N(\mu)||_Y \, ||\psi(\mu) - \psi_N(\mu)||_Y \tag{5.117}$$

from Galerkin orthogonality, the definition of the primal and adjoint problems, and (3.5). We now understand why we include the $\psi(\mu_n)$ in W_N: to ensure that $||\psi(\mu) - \psi_N(\mu)||_Y$ is small. We thus recover the "square" effect in the convergence rate of the output, albeit 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. However, there are significant computational and conditioning advantages associated with a "segregated" approach; this approach (and the corresponding advantages) is discussed in the next section.

Segregated Approach

In the *segregated* approach, we introduce separate primal and dual approximation spaces for $u(\mu)$ and $\psi(\mu)$, respectively. In particular, we define

$$W_N^{\text{pr}} = \text{span}\,\{\zeta_n^{\text{pr}} = u(\mu_n), n = 1, \ldots, N/2\} \tag{5.118}$$

$$W_N^{\text{du}} = \text{span}\left\{\zeta_n^{\text{du}} = \psi(\mu_n), n = 1, \ldots, N/2\right\} \tag{5.119}$$

Our reduced-basis approximation is then: for any $\mu \in \mathcal{D}^\mu$, find

$$s_N(\mu) = \langle L(\mu), u_N(\mu) \rangle - (\langle F(\mu), \psi_N(\mu) \rangle - \langle \mathcal{A}(\mu) u_N(\mu), \psi_N(\mu) \rangle) \,, \tag{5.120}$$

where $u_N(\mu) \in W_N^{\rm pr}$ and $\psi_N(\mu) \in W_N^{\rm du}$ are the Galerkin projections of $u(\mu)$ and $\psi(\mu)$ onto $W_N^{\rm pr}$ and $W_N^{\rm du}$, respectively,

$$\langle \mathcal{A}(\mu) u_N(\mu), v\rangle = \langle F(\mu), v\rangle, \quad \forall\, v \in W_N^{\rm pr}\ , \tag{5.121}$$

$$\langle \mathcal{A}(\mu) \psi_N(\mu), v\rangle = -\langle L(\mu), v\rangle, \quad \forall\, v \in W_N^{\rm du}\ . \tag{5.122}$$

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

$$||u(\mu) - u_N(\mu)||_Y \leq \sqrt{\frac{\gamma_{\mathcal{A}}(\mu)}{\alpha_{\mathcal{A}}(\mu)}} \inf_{w_N \in W_N^{\rm pr}} ||u(\mu) - w_N||_Y\ , \tag{5.123}$$

$$||\psi(\mu) - \psi_N(\mu)||_Y \leq \sqrt{\frac{\gamma_{\mathcal{A}}(\mu)}{\alpha_{\mathcal{A}}(\mu)}} \inf_{w_N \in W_N^{\rm du}} ||\psi(\mu) - w_N||_Y\ . \tag{5.124}$$

The proof of (5.123) and (5.124), and the best approximation analysis is similar to that presented in Section 3.5.2. As regards our output, we now have

$$|s(\mu) - s_N(\mu)| = |\langle L(\mu), u(\mu)\rangle - \langle L(\mu), u_N(\mu)\rangle + (\langle F(\mu), \psi_N(\mu)\rangle - \langle \mathcal{A}(\mu) u_N(\mu), \psi_N(\mu)\rangle)| \tag{5.125}$$

$$= |\langle \mathcal{A}(\mu)(u(\mu) - u_N(\mu)), \psi(\mu)\rangle - (\langle \mathcal{A}(\mu)(u(\mu) - u_N(\mu)), \psi_N(\mu)\rangle)\,| \tag{5.126}$$

$$= |\langle \mathcal{A}(\mu)(u(\mu) - u_N(\mu)), \psi(\mu) - \psi_N(\mu)\rangle| \tag{5.127}$$

$$\leq \gamma_{\mathcal{A}}^0 ||u(\mu) - u_N(\mu)||_Y\, ||\psi(\mu) - \psi_N(\mu)||_Y \tag{5.128}$$

from (5.120), the definition of the primal and dual problems, and (3.5). Note in this 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 $(\langle F(\mu), \psi_N(\mu)\rangle - \langle \mathcal{A}(\mu) u_N(\mu), \psi_N(\mu)\rangle)$ [26].

5.6.2 Off-line/On-line Computational Decomposition

Both the nonsegregated and segregated approaches admit an off-line/on-line decomposition similar to that described in Section 5.5.2 for the compliant problem; as before, the on-line complexity and storage are independent of the dimension of the very fine ("truth") finite element approximation.

Nonsegregated Approach

The off-line/on-line decomposition for the nonsegregated approach is, for the most part, similar to that described in Section 5.5.2. As before, the output is given by $s_N(\mu) = \underline{u}_N(\mu)^T \underline{L}_N(\mu)$; however, in this case the output vector $\underline{L}_N(\mu)$ is given by

$$L_{N\,i}(\mu) \equiv \langle L(\mu), \zeta_i\rangle, \quad 1 \leq i \leq N\ . \tag{5.129}$$

We then invoke (5.4) to write

$$L_{N\,i}(\mu) = \sum_{q=1}^{Q_L} \varphi_L^q(\mu)\, \langle L^q, \zeta_i\rangle\ , \tag{5.130}$$

or

$$\underline{L}_N(\mu) = \sum_{q=1}^{Q_L} \varphi_L^q(\mu)\, \underline{L}_N^q \ , \tag{5.131}$$

where the $\underline{L}_N^q \in \mathbb{R}^N$ is given by

$$L_{N\,i}^q = \langle L^q, \zeta_i \rangle \ , \quad 1 \le i \le N,\ 1 \le q \le Q_L \ . \tag{5.132}$$

Thus, in the *off-line* stage, we compute $u(\mu_n)$ and $\psi(\mu_n)$ for $n = 1, \ldots, N/2$ and form $\underline{A}_N^q$, $\underline{F}_N^q$, and $\underline{L}_N^q$: this requires N (expensive) "$\mathcal{A}$" finite element solutions, and $O(Q_{\mathcal{A}}N^2) + O(Q_L N + Q_F N)$ finite-element-vector ($O(\mathcal{N})$) inner products. In the *on-line* stage, for any given new μ, we first form $\underline{A}_N(\mu)$, $\underline{F}_N(\mu)$, and $\underline{L}_N(\mu)$ from (5.101),(5.102) and (5.130), solve (5.99) for $\underline{u}_N(\mu)$, and finally evaluate $s_N(\mu) = \underline{u}_N(\mu)^T \underline{L}_N(\mu)$: this requires $O(Q_{\mathcal{A}}N^2 + Q_F N + Q_L N) + O(\frac{2}{3}N^3)$ operations and $O(Q_{\mathcal{A}}N^2 + Q_F N + Q_L N)$ storage.

Segregated Approach

In the segregated approach, we express $u_N(\mu)$ and $\psi_N(\mu)$ to be linear combinations of the appropriate basis functions:

$$u_N(\mu) = \sum_{j=1}^{N/2} u_{N\,j}(\mu)\, \zeta_j^{\rm pr} \tag{5.133}$$

$$\psi_N(\mu) = \sum_{j=1}^{N/2} \psi_{N\,j}(\mu)\, \zeta_j^{\rm du} \ . \tag{5.134}$$

Then, choosing for test functions $v = \zeta_i^{\rm pr}$ in (5.110) and $v = \zeta_i^{\rm du}$ in (5.111) yields the desired algebraic equations for $\underline{u}_N(\mu) \in \mathbb{R}^N$ and $\underline{\psi}_N(\mu) \in \mathbb{R}^N$,

$$\underline{A}_N^{\rm pr}(\mu)\, \underline{u}_N(\mu) = \underline{F}_N(\mu) \ , \tag{5.135}$$

$$\underline{A}_N^{\rm du}(\mu)\, \underline{\psi}_N(\mu) = -\underline{L}_N(\mu) \ , \tag{5.136}$$

in terms of which the output can then be evaluated as

$$s_N(\mu) = \underline{u}_N^T(\mu)\underline{L}_N^{\rm pr}(\mu) + \underline{\psi}_N^T(\mu)\underline{F}_N^{\rm du}(\mu) - \underline{\psi}_N^T(\mu)\underline{A}_N^{\rm pr,du}\underline{u}_N(\mu) \tag{5.137}$$

Here $\underline{A}_N^{\rm pr}(\mu) \in \mathbb{R}^{N\times N}$, $\underline{A}_N^{\rm du}(\mu) \in \mathbb{R}^{N\times N}$, and $\underline{A}_N^{\rm pr,du}(\mu) \in \mathbb{R}^{N\times N}$ are the symmetric positive-definite matrices with entries

$$A_{N\,i,j}^{\rm pr}(\mu) = \langle \mathcal{A}(\mu)\zeta_j^{\rm pr}, \zeta_i^{\rm pr} \rangle, \quad 1 \le i,j \le N/2 \ , \tag{5.138}$$

$$A_{N\,i,j}^{\rm du}(\mu) = \langle \mathcal{A}(\mu)\zeta_j^{\rm du}, \zeta_i^{\rm du} \rangle, \quad 1 \le i,j \le N/2 \ , \tag{5.139}$$

$$A_{N\,i,j}^{\rm pr,du}(\mu) = \langle \mathcal{A}(\mu)\zeta_j^{\rm pr}, \zeta_i^{\rm du} \rangle, \quad 1 \le i,j \le N/2 \ , \tag{5.140}$$

$\underline{F}_N^{\rm pr}(\mu) \in \mathbb{R}^N$ and $\underline{F}_N^{\rm du}(\mu) \in \mathbb{R}^N$ are the "load" vectors with entries

$$F_{N\,i}^{\rm pr}(\mu) = \langle F(\mu), \zeta_i^{\rm pr} \rangle, \quad 1 \le i \le N/2 \ , \tag{5.141}$$

$$F_{N\,i}^{\rm du}(\mu) = \langle F(\mu), \zeta_i^{\rm du} \rangle, \quad 1 \le i \le N/2 \ , \tag{5.142}$$

and $\underline{L}_N^{\rm pr}(\mu)$ and $\underline{L}_N^{\rm du}(\mu)$ are the vectors with entries

$$L_{N\,i}^{\rm pr}(\mu) = \langle L(\mu), \zeta_i^{\rm pr} \rangle, \quad 1 \le i \le N/2 \ , \tag{5.143}$$

$$L_{N\,i}^{\rm du}(\mu) = \langle L(\mu), \zeta_i^{\rm du} \rangle, \quad 1 \le i \le N/2 \ . \tag{5.144}$$

We again invoke (5.1)-(5.4) to write

$$A_{N\,i,j}^{\rm pr}(\mu) = \sum_{q=1}^{Q_{\mathcal{A}}} \Theta^q(\mu) \, \langle \mathcal{A}^q \zeta_j^{\rm pr}, \zeta_i^{\rm pr} \rangle \ , \tag{5.145}$$

$$A_{N\,i,j}^{\rm du}(\mu) = \sum_{q=1}^{Q_{\mathcal{A}}} \Theta^q(\mu) \, \langle \mathcal{A}^q \zeta_j^{\rm du}, \zeta_i^{\rm du} \rangle \ , \tag{5.146}$$

$$A_{N\,i,j}^{\rm pr,du}(\mu) = \sum_{q=1}^{Q_{\mathcal{A}}} \Theta^q(\mu) \, \langle \mathcal{A}^q \zeta_j^{\rm pr}, \zeta_i^{\rm du} \rangle \ , \tag{5.147}$$

$$F_{N\,i}^{\rm pr}(\mu) = \sum_{q=1}^{Q_F} \varphi_F^q(\mu) \, \langle F^q, \zeta_i^{\rm pr} \rangle \ , \tag{5.148}$$

$$F_{N\,i}^{\rm du}(\mu) = \sum_{q=1}^{Q_F} \varphi_F^q(\mu) \, \langle F^q, \zeta_i^{\rm du} \rangle \ , \tag{5.149}$$

$$L_{N\,i}^{\rm pr}(\mu) = \sum_{q=1}^{Q_L} \varphi_L^q(\mu) \, \langle L^q, \zeta_i^{\rm pr} \rangle \ , \tag{5.150}$$

$$L_{N\,i}^{\rm du}(\mu) = \sum_{q=1}^{Q_L} \varphi_L^q(\mu) \, \langle L^q, \zeta_i^{\rm du} \rangle \ , \tag{5.151}$$

or

$$\underline{A}_N^{\text{pr}}(\mu) = \sum_{q=1}^{Q_\mathcal{A}} \Theta^q(\mu)\, \underline{A}_N^{q\,\text{pr}} , \tag{5.152}$$

$$\underline{A}_N^{\text{du}}(\mu) = \sum_{q=1}^{Q_\mathcal{A}} \Theta^q(\mu)\, \underline{A}_N^{q\,\text{du}} , \tag{5.153}$$

$$\underline{A}_N^{\text{pr,du}}(\mu) = \sum_{q=1}^{Q_\mathcal{A}} \Theta^q(\mu)\, \underline{A}_N^{q\,\text{pr,du}} , \tag{5.154}$$

$$\underline{F}_N^{\text{pr}}(\mu) = \sum_{q=1}^{Q_F} \varphi_F^q(\mu)\, \underline{F}_N^{q\,\text{pr}} , \tag{5.155}$$

$$\underline{F}_N^{\text{du}}(\mu) = \sum_{q=1}^{Q_F} \varphi_F^q(\mu)\, \underline{F}_N^{q\,\text{du}} , \tag{5.156}$$

$$\underline{L}_N^{\text{pr}}(\mu) = \sum_{q=1}^{Q_L} \varphi_L^q(\mu)\, \underline{L}_N^{q\,\text{pr}} , \tag{5.157}$$

$$\underline{L}_N^{\text{du}}(\mu) = \sum_{q=1}^{Q_L} \varphi_L^q(\mu)\, \underline{L}_N^{q\,\text{du}} , \tag{5.158}$$

where the $\underline{A}_N^{q\,\text{pr}}$, $\underline{A}_N^{q\,\text{du}}$, $\underline{A}_N^{q\,\text{pr,du}}$, $\underline{F}_N^{q\,\text{pr}}$, $\underline{F}_N^{q\,\text{du}}$, $\underline{L}_N^{q\,\text{pr}}$ and $\underline{L}_N^{q\,\text{du}}$ are given by

$$A_{N\,i,j}^{q\,\text{pr}} = \langle \mathcal{A}^q \zeta_j^{\text{pr}}, \zeta_i^{\text{pr}} \rangle , \quad 1 \le i,j \le N/2 \; 1 \le q \le Q_\mathcal{A} , \tag{5.159}$$

$$A_{N\,i,j}^{q\,\text{du}} = \langle \mathcal{A}^q \zeta_j^{\text{du}}, \zeta_i^{\text{du}} \rangle , \quad 1 \le i,j \le N/2 \; 1 \le q \le Q_\mathcal{A} , \tag{5.160}$$

$$A_{N\,i,j}^{q\,\text{pr,du}} = \langle \mathcal{A}^q \zeta_j^{\text{pr}}, \zeta_i^{\text{du}} \rangle , \quad 1 \le i,j \le N/2 \; 1 \le q \le Q_\mathcal{A} , \tag{5.161}$$

$$F_{N\,i}^{q\,\text{pr}} = \langle F^q, \zeta_i^{\text{pr}} \rangle , \quad 1 \le i \le N/2 \; 1 \le q \le Q_F , \tag{5.162}$$

$$F_{N\,i}^{q\,\text{du}} = \langle F^q, \zeta_i^{\text{du}} \rangle , \quad 1 \le i \le N/2 \; 1 \le q \le Q_F , \tag{5.163}$$

$$L_{N\,i}^{q\,\text{pr}} = \langle L^q, \zeta_i^{\text{pr}} \rangle , \quad 1 \le i \le N/2 \; 1 \le q \le Q_L , \tag{5.164}$$

$$L_{N\,i}^{q\,\text{du}} = \langle L^q, \zeta_i^{\text{du}} \rangle , \quad 1 \le i \le N/2 \; 1 \le q \le Q_L . \tag{5.165}$$

Thus, in the *off-line* stage, we compute the $u(\mu_n)$, $\psi(\mu_n)$, and form the $\underline{A}_N^{q\,\text{pr}}$, $\underline{A}_N^{q\,\text{du}}$, $\underline{A}_N^{q\,\text{pr,du}}$, $\underline{F}_N^{q\,\text{pr}}$, $\underline{F}_N^{q\,\text{du}}$, $\underline{L}_N^{q\,\text{pr}}$, and $\underline{L}_N^{q\,\text{du}}$: this requires N (expensive) "$\mathcal{A}$" finite element solutions, and $O(\frac{3}{4}QN^2) + O(Q_F N + Q_L N)$ finite-element-vector ($O(\mathcal{N})$) inner products. In the *on-line* stage, for any given new μ, we first form $\underline{A}_N^{\text{pr}}(\mu)$, $\underline{A}_N^{\text{du}}(\mu)$, $\underline{A}_N^{\text{pr,du}}(\mu)$, $\underline{F}_N^{\text{pr}}(\mu)$, $\underline{F}_N^{\text{du}}(\mu)$, $\underline{L}_N^{\text{pr}}(\mu)$, and $\underline{L}_N^{\text{du}}(\mu)$ from (5.152)-(5.158), then solve (5.135) and (5.136) for $\underline{u}_N(\mu)$ and $\underline{\psi}_N(\mu)$, respectively, and finally evaluate $s_N(\mu)$: this requires $O(\frac{3}{4}Q_\mathcal{A}N^2 + Q_F N + Q_L N) + O(\frac{4}{3}N^3)$ operations and $O(\frac{3}{4}Q_\mathcal{A}N^2 + Q_F N + Q_L N)$ storage.

5.6.3 Numerical Results

We present in Figure 5-9 the maximum relative error $\varepsilon_N(\mu)$ as a function of N for $\mu \in \mathcal{D}^\mu$ for Example 9. For this example, reduced-basis space is nonsegregated, and the μ_n are chosen "log-

randomly" over $\mathcal{D}^\mu$. We note that, unlike in the compliance case, the error no longer decreases monotonically with N; nevertheless, the approximation converges relatively rapidly. Furthermore, the error is remarkably small even for relatively small N — we obtain approximations which are accurate to within 1% for $N \sim 25$. Furthermore, we again observe that the log–random point distribution *is* important, as evidenced by generally faster convergence (versus the *non–logarithmic* uniform random point distribution).

We again note that although our reduced-basis approximations are accurate and inexpensive to calculate, it is important to *assess* the accuracy of our predictions. In Sections 5.7 and 5.8 we extend the *a posteriori* Method I and Method II error estimation methods developed in Chapter 4 to the more general case of noncompliant problems.

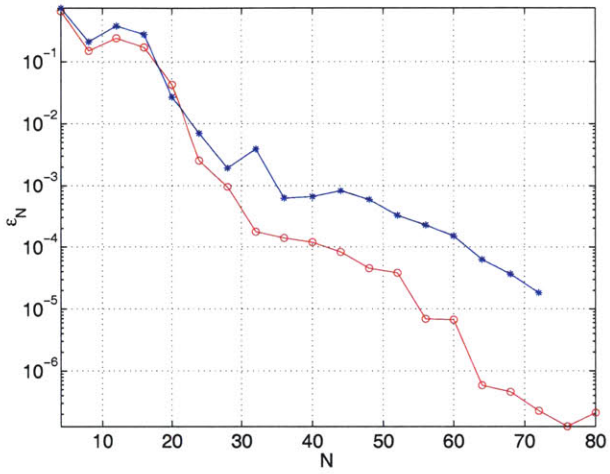


Figure 5-9: Convergence of the (noncompliant) reduced-basis approximation for Example 9.

5.7 *A Posteriori* Error Estimation: Method I

The methods presented in Section 4.2 still apply to the compliance case, and the extension to μ-dependent linear functionals ($F(\mu) = L(\mu)$) is straightforward. In this section, we extend the bound conditioner constructions of Sections 4.3 and 4.4 to more general linear functionals ($F(\mu) \neq L(\mu)$). We then apply our methods (for compliance and noncompliance) to the linear elasticiy model problems of Section 5.4; we conclude with numerical results.

To begin, we define the primal and dual error, $e^{\rm pr}(\mu)$, $e^{\rm du}(\mu) \in Y$, as well as the primal and dual residual $\mathcal{R}^{\rm pr}(\mu)$, $\mathcal{R}^{\rm du}(\mu) \in Y$ as

$$e^{\rm pr}(\mu) \equiv u(\mu) - u_N(\mu) \ , \qquad \langle \mathcal{R}^{\rm pr}(\mu), v\rangle \equiv \ \langle F(\mu), v\rangle - \langle \mathcal{A}(\mu) u_N(\mu), v\rangle \ , \tag{5.166}$$

$$e^{\rm du}(\mu) \equiv \psi(\mu) - \psi_N(\mu) \ , \qquad \langle \mathcal{R}^{\rm du}(\mu), v\rangle \equiv -\langle L(\mu), v\rangle - \langle \mathcal{A}(\mu) v, \psi_N(\mu)\rangle \ , \tag{5.167}$$

from which it follows that

$$\langle \mathcal{A}(\mu) e^{\rm pr}(\mu), v\rangle = \langle \mathcal{R}^{\rm pr}(\mu), v\rangle \ , \tag{5.168}$$

$$\langle \mathcal{A}(\mu) v, e^{\rm du}(\mu)\rangle = \langle \mathcal{R}^{\rm du}(\mu), v\rangle \ . \tag{5.169}$$

5.7.1 Bound Conditioner

We again introduce a symmetric, continuous, and coercive *bound conditioner* [21, 37, 47] $\mathcal{C}(\mu)$: $Y \rightarrow Y'$ such that the eigenvalues

$$\rho_{\min}(\mu) = \min_{v \in Y} \frac{\langle \mathcal{A}(\mu)v, v\rangle}{\langle \mathcal{C}(\mu)v, v\rangle} , \tag{5.170}$$

$$\rho_{\max}(\mu) = \max_{v \in Y} \frac{\langle \mathcal{A}(\mu)v, v\rangle}{\langle \mathcal{C}(\mu)v, v\rangle} , \tag{5.171}$$

satisfy

$$1 \leq \rho_{\min}(\mu), \qquad \rho_{\max}(\mu) \leq \rho, \tag{5.172}$$

for some (preferably small) constant $\rho \in \mathbb{R}$. As before, we also require an additional "computational invertibility" hypothesis [38, 47]: we require that $\mathcal{C}^{-1}(\mu)$ be of the form

$$\mathcal{C}^{-1}(\mu) = \sum_{i \in \mathcal{I}(\mu)} \alpha_i(\mu) \mathcal{C}_i^{-1} \tag{5.173}$$

where $\mathcal{I}(\mu) \subset \{1, \ldots, I\}$ is a parameter-dependent set of indices, I is a finite (preferably small) integer, and the $\mathcal{C}_i$: $Y \rightarrow Y', i = 1, \ldots, I$, are parameter-*independent* symmetric, coercive operators.

5.7.2 Error and Output Bounds

We now find $\hat{e}^{\rm pr}(\mu) \in Y$ and $\hat{e}^{\rm du}(\mu) \in Y$ such that

$$\langle \mathcal{C}(\mu)\hat{e}^{\rm pr}(\mu), v\rangle = \langle \mathcal{R}^{\rm pr}(\mu), v\rangle, \quad \forall\, v \in Y , \tag{5.174}$$

$$\langle \mathcal{C}(\mu)\hat{e}^{\rm du}(\mu), v\rangle = \langle \mathcal{R}^{\rm du}(\mu), v\rangle, \quad \forall\, v \in Y . \tag{5.175}$$

We then define our lower and upper output bounds as

$$s_N^-(\mu) = s_N(\mu) - \Delta_N(\mu) \tag{5.176}$$

$$s_N^+(\mu) = s_N(\mu) + \Delta_N(\mu) \tag{5.177}$$

where $\Delta_N(\mu)$ is given by

$$\Delta_N(\mu) = \langle \mathcal{C}(\mu)\hat{e}^{\rm pr}(\mu), \hat{e}^{\rm pr}(\mu)\rangle^{1/2} \langle \mathcal{C}(\mu)\hat{e}^{\rm du}(\mu), \hat{e}^{\rm du}(\mu)\rangle^{1/2} \tag{5.178}$$

$$= \langle \mathcal{R}^{\rm pr}(\mu), \mathcal{C}^{-1}(\mu)\mathcal{R}^{\rm pr}(\mu)\rangle^{1/2} \langle \mathcal{R}^{\rm du}(\mu), \mathcal{C}^{-1}(\mu)\mathcal{R}^{\rm du}(\mu)\rangle^{1/2} \tag{5.179}$$

$$= \langle \mathcal{R}^{\rm pr}(\mu), \hat{e}^{\rm pr}(\mu)\rangle^{1/2} \langle \mathcal{R}^{\rm du}(\mu), \hat{e}^{\rm du}(\mu)\rangle^{1/2} . \tag{5.180}$$

5.7.3 Bounding Properties

It remains to demonstrate that

$$\eta_N(\mu) = \frac{\Delta_N(\mu)}{|s(\mu) - s_N(\mu)|} \geq 1 , \qquad \forall\, N \geq 1 , \tag{5.181}$$

and to investigate the sharpness of our bounds; note that from (5.176), (5.177), and (5.181) it directly follows that that $s_N^-(\mu) \leq s(\mu) \leq s_N^+(\mu)$ for all $N \geq 1$.

To prove (5.181), we first note from (5.117) for the nonsegregated case, and (5.127) for the

segregated case, that

$$|s(\mu) - s_N(\mu)| = |\langle \mathcal{A}(\mu) e^{\rm pr}(\mu), e^{\rm du}(\mu)\rangle|. \tag{5.182}$$

It then follows that

$$\begin{aligned}
\eta_N(\mu) &= \frac{\langle \mathcal{C}(\mu)\hat{e}^{\rm pr}(\mu), \hat{e}^{\rm pr}(\mu)\rangle^{1/2}\langle \mathcal{C}(\mu)\hat{e}^{\rm du}(\mu), \hat{e}^{\rm du}(\mu)\rangle^{1/2}}{|\langle \mathcal{A}(\mu) e^{\rm pr}(\mu), e^{\rm du}(\mu)\rangle|} && (5.183)\\
&\geq \frac{\langle \mathcal{C}(\mu)\hat{e}^{\rm pr}(\mu), \hat{e}^{\rm pr}(\mu)\rangle^{1/2}\langle \mathcal{C}(\mu)\hat{e}^{\rm du}(\mu), \hat{e}^{\rm du}(\mu)\rangle^{1/2}}{\langle \mathcal{A}(\mu) e^{\rm pr}(\mu), e^{\rm pr}(\mu)\rangle^{1/2}\langle \mathcal{A}(\mu) e^{\rm du}(\mu), e^{\rm du}(\mu)\rangle^{1/2}} && (5.184)\\
&\geq \rho_{\rm min}(\mu) && (5.185)\\
&\geq 1\ , && (5.186)
\end{aligned}$$

from the triangle inequality and (4.21).

We now turn to the upper effectivity inequality (sharpness property). We note from (5.183) that if the primal and dual errors are $\mathcal{A}$-orthogonal, or become increasingly orthogonal as N increases, such that $|\langle \mathcal{A}(\mu) e^{\rm pr}(\mu), e^{\rm du}(\mu)\rangle| \to 0$, then the effectivity will not be bounded as $N \to \infty$. However, if we make the (plausible) hypothesis [39] that

$$|s(\mu) - s_N(\mu)| = \left|\langle \mathcal{A}(\mu) e^{\rm pr}(\mu), e^{\rm du}(\mu)\rangle\right| \geq \hat{c}^0_{\mathcal{A}} ||e^{\rm pr}(\mu)||_Y ||e^{\rm du}(\mu)||_Y\ , \tag{5.187}$$

then it follows that

$$\begin{aligned}
\eta_N(\mu) &= \frac{\langle \mathcal{C}(\mu)\hat{e}^{\rm pr}(\mu), \hat{e}^{\rm pr}(\mu)\rangle^{1/2}\langle \mathcal{C}(\mu)\hat{e}^{\rm du}(\mu), \hat{e}^{\rm du}(\mu)\rangle^{1/2}}{|\langle \mathcal{A}(\mu) e^{\rm pr}(\mu), e^{\rm du}(\mu)\rangle|} && (5.188)\\
&\leq \frac{\langle \mathcal{C}(\mu)\hat{e}^{\rm pr}(\mu), \hat{e}^{\rm pr}(\mu)\rangle^{1/2}\langle \mathcal{C}(\mu)\hat{e}^{\rm du}(\mu), \hat{e}^{\rm du}(\mu)\rangle^{1/2}}{\hat{c}^0_{\mathcal{A}} ||e^{\rm pr}(\mu)||_Y ||e^{\rm du}(\mu)||_Y} && (5.189)\\
&\leq \frac{\gamma^0_{\mathcal{A}}}{\hat{c}^0_{\mathcal{A}}} \frac{\langle \mathcal{C}(\mu)\hat{e}^{\rm pr}(\mu), \hat{e}^{\rm pr}(\mu)\rangle^{1/2}\langle \mathcal{C}(\mu)\hat{e}^{\rm du}(\mu), \hat{e}^{\rm du}(\mu)\rangle^{1/2}}{\langle \mathcal{A}(\mu) e^{\rm pr}(\mu), e^{\rm pr}(\mu)\rangle^{1/2}\langle \mathcal{A}(\mu) e^{\rm du}(\mu), e^{\rm du}(\mu)\rangle^{1/2}} && (5.190)\\
&\leq \frac{\gamma^0_{\mathcal{A}}}{\hat{c}^0_{\mathcal{A}}} \rho_{\rm max}(\mu) && (5.191)\\
&\leq \frac{\gamma^0_{\mathcal{A}}}{\hat{c}^0_{\mathcal{A}}} \rho\ . && (5.192)
\end{aligned}$$

5.7.4 Off-line/On-line Computational Procedure

We indicate here the off-line/on-line calculation of $\Delta_N(\mu)$ for the nonsegregated cases; the segregated approach admit a similar computational decomposition.

Nonsegregated Case

We recall from (5.179) that

$$\Delta_N(\mu) = \langle \mathcal{R}^{\rm pr}(\mu), \mathcal{C}^{-1}(\mu)\mathcal{R}^{\rm pr}(\mu)\rangle^{1/2}\langle \mathcal{R}^{\rm du}(\mu), \mathcal{C}^{-1}(\mu)\mathcal{R}^{\rm du}(\mu)\rangle^{1/2}\ . \tag{5.193}$$

From the definition of the residual (5.166)-(5.167), the separability assumption (5.1), and the expansion of $u_N(\mu)$ and $\psi_N(\mu)$ in terms of the basis functions, we have

$$\begin{aligned}\langle \mathcal{R}^{\rm pr}(\mu), v\rangle &= \langle F(\mu), v\rangle - \langle \mathcal{A}(\mu) u_N(\mu), v\rangle \\ &= \sum_{q=1}^{Q_F} \varphi_F^q(\mu)\langle F^q, v\rangle - \sum_{q=1}^{Q_\mathcal{A}}\sum_{n=1}^{N} \Theta^q(\mu)\, u_{N\,n}(\mu)\, \langle \mathcal{A}^q\zeta_n, v\rangle, \quad \forall\, v\in Y,\ \forall\, \mu\in\mathcal{D}^\mu\ ,\end{aligned}$$

and

$$\begin{aligned}\langle \mathcal{R}^{\rm du}(\mu), v\rangle &= -\langle L(\mu), v\rangle - \langle \mathcal{A}(\mu)\psi_N(\mu), v\rangle \\ &= -\sum_{q=1}^{Q_L} \varphi_L^q(\mu)\langle L^q, v\rangle - \sum_{q=1}^{Q_\mathcal{A}}\sum_{n=1}^{N} \Theta^q(\mu)\, \psi_{N\,n}(\mu)\, \langle \mathcal{A}^q\zeta_n, v\rangle, \quad \forall\, v\in Y,\ \forall\, \mu\in\mathcal{D}^\mu\ .\end{aligned}$$

We now invoke the "computational invertibility" hypothesis on $\mathcal{C}(\mu)$, (5.173), to write

$$\begin{aligned}\langle \mathcal{R}^{\rm pr}(\mu), \mathcal{C}^{-1}(\mu)\mathcal{R}^{\rm pr}(\mu)\rangle &= \sum_{i\in\mathcal{I}(\mu)} \alpha_i(\mu)\, \langle \mathcal{R}^{\rm pr}(\mu), \mathcal{C}_i^{-1}\mathcal{R}^{\rm pr}(\mu)\rangle \qquad (5.194)\\ &= \sum_{i\in\mathcal{I}(\mu)} \alpha_i(\mu) \left\langle \sum_{q=1}^{Q_F} \varphi_F^q(\mu) F^q - \sum_{q=1}^{Q_\mathcal{A}}\sum_{n=1}^{N} \Theta^q(\mu)\, u_{N\,n}(\mu)\, \mathcal{A}^q\zeta_n,\right.\\ &\qquad \left. \mathcal{C}_i^{-1}\left(\sum_{q'=1}^{Q_F} \varphi_F^{q'}(\mu) F^{q'} - \sum_{q'=1}^{Q_\mathcal{A}}\sum_{n'=1}^{N} \Theta^{q'}(\mu)\, u_{N\,n'}(\mu)\, \mathcal{A}^{q'}\zeta_{n'}\right)\right\rangle\\ &= \sum_{i\in\mathcal{I}(\mu)} \alpha_i(\mu) \left[\sum_{q=1}^{Q_F}\sum_{q'=1}^{Q_F} \varphi_F^q(\mu)\varphi_F^{q'}(\mu)\langle F^q, \mathcal{C}_i^{-1}F^{q'}\rangle\right.\\ &\quad -\sum_{q=1}^{Q_\mathcal{A}}\sum_{q'=1}^{Q_F}\sum_{n=1}^{N} \Theta^q(\mu)\, \varphi_F^{q'}(\mu)\, u_{N\,n}(\mu) \left(\langle F^{q'}, \mathcal{C}_i^{-1}\mathcal{A}^q\zeta_n\rangle + \langle \mathcal{A}^q\zeta_n, \mathcal{C}_i^{-1}F^{q'}\rangle\right)\\ &\quad \left. +\sum_{q=1}^{Q_\mathcal{A}}\sum_{q'=1}^{Q_\mathcal{A}}\sum_{n=1}^{N}\sum_{n'=1}^{N} \Theta^q(\mu)\, \Theta^{q'}(\mu)\, u_{N\,n}(\mu)\, u_{N\,n'}(\mu)\, \langle \mathcal{A}^q\zeta_n, \mathcal{C}_i^{-1}\mathcal{A}^{q'}\zeta_{n'}\rangle\right];\end{aligned}$$

similarly, we have

$$\begin{aligned}\langle \mathcal{R}^{\rm du}(\mu), \mathcal{C}^{-1}(\mu)\mathcal{R}^{\rm du}(\mu)\rangle &= \sum_{i\in\mathcal{I}(\mu)} \alpha_i(\mu) \left[\sum_{q=1}^{Q_L}\sum_{q'=1}^{Q_L} \varphi_L^q(\mu)\varphi_L^{q'}(\mu)\langle L^q, \mathcal{C}_i^{-1}L^{q'}\rangle\right.\\ &\quad +\sum_{q=1}^{Q_\mathcal{A}}\sum_{q'=1}^{Q_L}\sum_{n=1}^{N} \Theta^q(\mu)\, \varphi_L^{q'}(\mu)\, \psi_{N\,n}(\mu) \left(\langle L^{q'}, \mathcal{C}_i^{-1}\mathcal{A}^q\zeta_n\rangle + \langle \mathcal{A}^q\zeta_n, \mathcal{C}_i^{-1}L^{q'}\rangle\right)\\ &\quad \left. +\sum_{q=1}^{Q_\mathcal{A}}\sum_{q'=1}^{Q_\mathcal{A}}\sum_{n=1}^{N}\sum_{n'=1}^{N} \Theta^q(\mu)\, \Theta^{q'}(\mu)\, \psi_{N\,n}(\mu)\, \psi_{N\,n'}(\mu)\, \langle \mathcal{A}^q\zeta_n, \mathcal{C}_i^{-1}\mathcal{A}^{q'}\zeta_{n'}\rangle\right].\end{aligned}$$

Thus, in the *off-line* stage, we compute the μ-independent inner products

$$c^{\rm pr}_{iqq'} = \langle F^q, \mathcal{C}_i^{-1} F^{q'} \rangle , \quad 1 \le i \le I,\ 1 \le q, q' \le Q_F \tag{5.195}$$

$$c^{\rm du}_{iqq'} = \langle L^q, \mathcal{C}_i^{-1} L^{q'} \rangle , \quad 1 \le i \le I,\ 1 \le q, q' \le Q_L \tag{5.196}$$

$$\Lambda^{\rm pr}_{iqq'n} = -\langle F^{q'}, \mathcal{C}_i^{-1} \mathcal{A}^q \zeta_n \rangle = -\langle \mathcal{A}^q \zeta_n, \mathcal{C}_i^{-1} F^{q'} \rangle , \quad 1 \le i \le I,\ 1 \le q \le Q_\mathcal{A}, \tag{5.197}$$

$$1 \le q' \le Q_F,\ 1 \le n \le N, \tag{5.198}$$

$$\Lambda^{\rm du}_{iqq'n} = \langle L^{q'}, \mathcal{C}_i^{-1} \mathcal{A}^q \zeta_n \rangle = \langle \mathcal{A}^q \zeta_n, \mathcal{C}_i^{-1} L^{q'} \rangle , \quad 1 \le i \le I,\ 1 \le q \le Q_\mathcal{A}, \tag{5.199}$$

$$1 \le q' \le Q_L,\ 1 \le n \le N, \tag{5.200}$$

and

$$\Gamma_{qq'nn'} = \langle \mathcal{A}^q \zeta_n, \mathcal{C}_i^{-1} \mathcal{A}^{q'} \zeta_{n'} \rangle , \quad 1 \le q, q' \le Q_\mathcal{A},\ 1 \le n, n' \le N, \tag{5.201}$$

where (5.197) and (5.199) follow from the symmetry of $\mathcal{C}_i$. Calculation of (5.195)-(5.201) then requires $Q_\mathcal{A} N$ $\mathcal{A}^q \zeta_n$ multiplications, $I(Q_\mathcal{A} N + Q_F + Q_L)$ $\mathcal{C}$-solves, and $I(Q_\mathcal{A}^2 N^2 + Q_\mathcal{A} N + Q_F Q_\mathcal{A} N + Q_L Q_\mathcal{A} N + Q_F^2 + Q_L^2)$ inner products.

In the *on-line* stage, given any new value of μ, we simply the compute

$$\Delta_N(\mu) = \left(\sum_{i \in \mathcal{I}(\mu)} \alpha_i(\mu) \left[\sum_{q=1}^{Q_F} \sum_{q'=1}^{Q_F} \varphi_F^q(\mu) \varphi_F^{q'}(\mu) c^{\rm pr}_{iqq'} + 2 \sum_{q=1}^{Q_\mathcal{A}} \sum_{q'=1}^{Q_F} \sum_{n=1}^{N} \Theta^q(\mu)\, \varphi_F^{q'}(\mu)\, u_{N\,n}(\mu)\, \Lambda^{\rm pr}_{iqq'n} \right.\right.$$
$$\left.\left. + \sum_{q=1}^{Q_\mathcal{A}} \sum_{q'=1}^{Q_\mathcal{A}} \sum_{n=1}^{N} \sum_{n'=1}^{N} \Theta^q(\mu)\, \Theta^{q'}(\mu)\, u_{N\,n}(\mu)\, u_{N\,n'}(\mu)\, \Gamma_{qq'nn'} \right] \right)^{1/2} \tag{5.202}$$

$$\left(\sum_{i \in \mathcal{I}(\mu)} \alpha_i(\mu) \left[\sum_{q=1}^{Q_L} \sum_{q'=1}^{Q_L} \varphi_L^q(\mu) \varphi_L^{q'}(\mu) c^{\rm du}_{iqq'} + 2 \sum_{q=1}^{Q_\mathcal{A}} \sum_{q'=1}^{Q_L} \sum_{n=1}^{N} \Theta^q(\mu)\, \varphi_L^{q'}(\mu)\, \psi_{N\,n}(\mu)\, \Lambda^{\rm du}_{iqq'n} \right.\right.$$
$$\left.\left. + \sum_{q=1}^{Q_\mathcal{A}} \sum_{q'=1}^{Q_\mathcal{A}} \sum_{n=1}^{N} \sum_{n'=1}^{N} \Theta^q(\mu)\, \Theta^{q'}(\mu)\, \psi_{N\,n}(\mu)\, \psi_{N\,n'}(\mu)\, \Gamma_{qq'nn'} \right] \right)^{1/2} . \tag{5.203}$$

The on-line complexity is thus, to leading order, $O(|\mathcal{I}(\mu)| Q_\mathcal{A}^2 N^2)$ — and hence, independent of $\mathcal{N}$.

We note from Section 5.7.1 that the guidelines presented in Chapter 4 for constructing and choosing the bound conditioner $\mathcal{C}(\mu)$ are the same for noncompliance as for the compliance case. We therefore need not repeat here the formulation of the different bound conditioner constructions. We present in Sections 5.7.5 to 5.7.9 numerical results for the examples of Section 5.4 *if* applicable, and remark on the limitations of the restrictions of each bound conditioner construction.

5.7.5 Minimum Coefficient Bound Conditioner

We present in Figure 5-10 the effectivities for Example 5 and $\bar{\theta} = \mu_0$ for different values of μ_0. Note that $\eta_N(\mu) - 1 > 0$ in all cases, and $\eta_N(\mu_0) = 1$ (as expected). The results also confirm our observations in Chapter 4: the sensitivity of the resulting effectivities to the choice of $\bar{\theta}$, and the increase in effectivities as $|\mu - \mu_0| \to \infty$. (Note, in this case the rate of increase is greater for $\mu > \mu_0$.) Nevertheless, the results show that even for $I = 1$ — *if* μ_0 is chosen "optimally" — $\eta_N(\mu) \le 3$ for all $\mu \in \mathcal{D}^\mu$ for $\mu_0 \sim 0.1$.

However, the disadvantages alluded to in Chapter 4 are even more restrictive in the case of linear elasticity. We recall that the minimum coefficient bound conditioner applies only to problems in which the A^q are positive semidefinite. Unlike in the case of the Laplace operator (heat conduction), the linear elasticity operator does not admit such a representation even for simple geometry variations. This particular technique — although it performs well for the property variation in Example 5 — is therefore only limitedly applicable to linear elasticity.

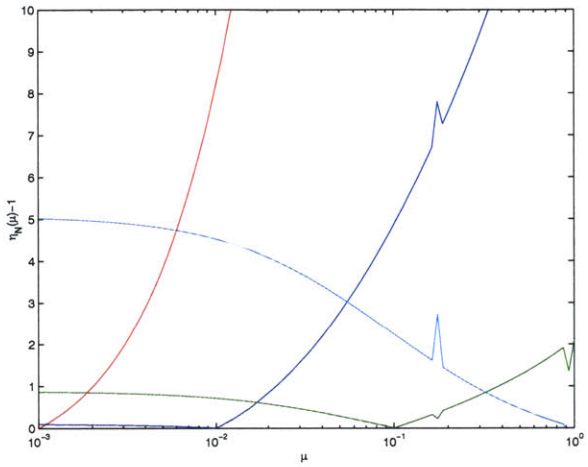


Figure 5-10: Effectivity as a function of μ for Example 5 calculated using the minimum coefficient bound conditioner with $\bar{\theta} = \Theta(\mu_0)$ for $\mu_0 = 0.001, 0.01, 0.1, 1.0$.

5.7.6 Eigenvalue Interpolation: Quasi-Concavity in μ

We again consider the possibility of constructing bound conditioners based on observations on the behavior of the relevant eigenvalues with respect to the parameter. We plot in Figures 5-11 (Example 6) and 5-12 (Example 7) the eigenvalues $\rho^i_{\min}$ as a functions of the parameters calculated using $\mathcal{C}_1 = \mathcal{A}(\mu_0)$ for $\mu_0 = 0.2$. We note that in both cases, the eigenvalues again appear to be quasi-concave with respect to the parameter. The slight "bump" in the curve for Example 7 (at $\mu \sim 0.65$) causes some concern, however, and potentially indicates that our hypothesis of quasi-concavity may in fact be false.

Furthermore, we observe that although the operators associated with Examples 6 and 7 are very similar, the eigenvalues for Example 7 are much worse in the sense that they approach zero much more rapidly. The difference lies in the boundary conditions: the Dirichlet boundary conditions applied in Example 6 eliminate the near-pure rotation modes which cause $\rho^i_{\min}$ to become very small.

We present in Figure 5-11 the effectivities for Examples 6 and 7 ; we present similar results in Table 5.3 for Example 8. We note that we obtain very good effectivities ((and therefore sharp bounds) except in Example 7 for small values of μ. The latter is due to the fact that the eigenvalues become very small for small thicknesses (due to near-pure rotation modes) in Example 7.

5.7.7 Eigenvalue Interpolation: Concavity θ

We plot in Figure 5-14 the resulting effectivities for Example 5. We note that with the proper choice of $\mathcal{C}_1$ (e.g., $\mathcal{A}(\mu_0)$ for $\mu_0 = 1.0$), the effectivities are very good even for $T = 2$. A substantial

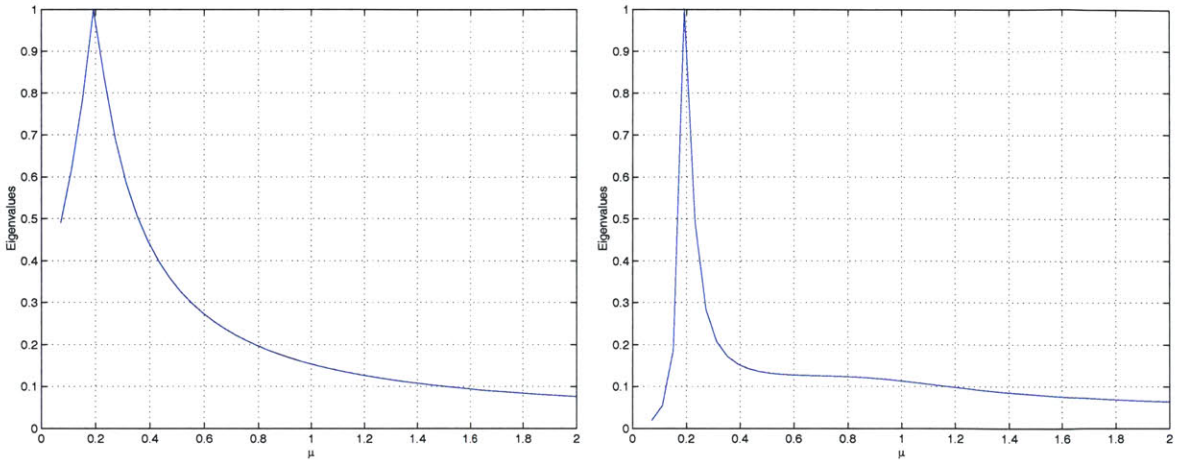


Figure 5-11: Plot of the eigenvalues $\rho^i_{\min}$ as a function of $\bar{t}$ for Examples 6 and 7, with $\mathcal{C}_1 = \mathcal{A}(\mu_0)$, and $\mu_0 = \{\bar{t}_0\} = 0.2$.

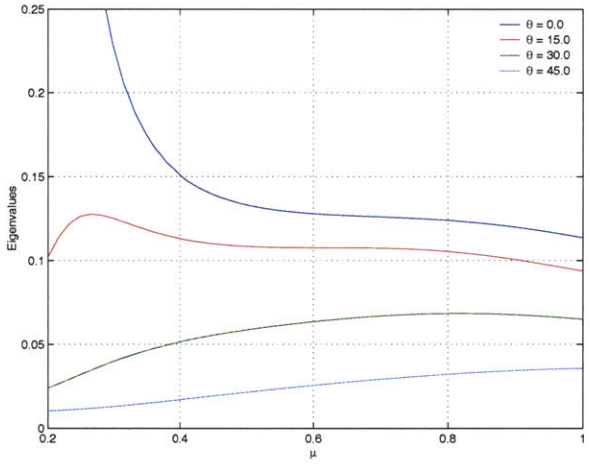


Figure 5-12: Contours of the eigenvalues $\rho^i_{\min}$ as a function of $\bar{t}$ for Example 8, with $\mathcal{C}_1 = \mathcal{A}(\mu_0)$, $\mu_0 = \{\bar{t}_0, \bar{\alpha}_0\} = \{0.2, 0.0\}$. The contours are calculated at constant $\bar{\alpha}$, for $\bar{\alpha} = 0°$, $15°$, $30°$, and $45°$

	$\bar{t}_0$	$\bar{\alpha}_0$	$\eta_N(\mu) - 1$
Trial 1	0.3	5.0	29.99
Trial 2	0.3	40.0	195.58
Trial 3	0.9	5.0	15.78
Trial 4	0.9	40.0	40.90

Table 5.3: Minimum, maximum, and average effectivity over $\mu \in \mathcal{D}^\mu$ for Example 8, obtained using the quasi-concave eigenvalue bound conditioner with $\mathcal{C}_1 = \mathcal{A}(\mu_0)$, $\mu_0 = \{\bar{t}_0, \bar{\alpha}_0\} = \{0.2, 0.0\}$, and $T = 9$.

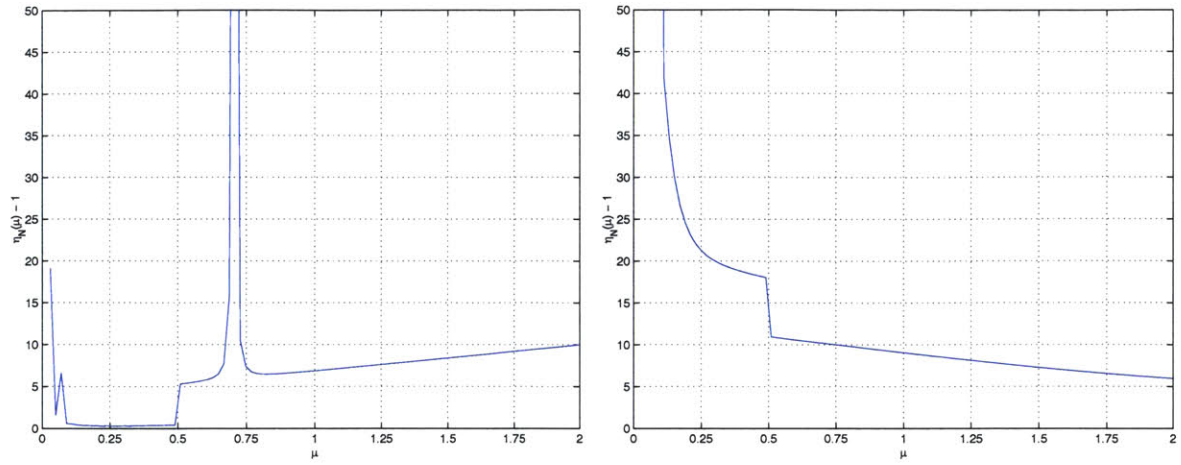


Figure 5-13: Effectivity as a function of μ for Examples 6 and 7, obtained using the quasi-concave eigenvalue bound conditioner with $\mathcal{C}_1 = \mathcal{A}(\mu_0)$, $\mu_0 = \bar{t}_0 = 0.2$.

decrease can also be observed for $T = 4$.

We also plot in Figure 5-15 the range $\mathcal{D}^\theta$, and in Figures 5-16 and 5-12 the eigenvalues over the three lines shown in Figure 5-15. These results agree with our theoretical results that the eigenvalues are concave with respect to θ. We also note that, particularly for Example 7, the eigenvalues become negative rather rapidly for the linear elasticity case. We then present in Figures 5-18 and 5-19 the resulting effectivities. We note that the effectivities for Example 6 are relatively high, and are even higher for Example 7 due to the speed with which the eigenvalues go to zero. However, it must be noted these effectivities are still quite good considering the fact that among all our bound conditioners, the concave eigenvalue bound conditioner is the only completely rigorous technique which is applicable even for general geometry variations.

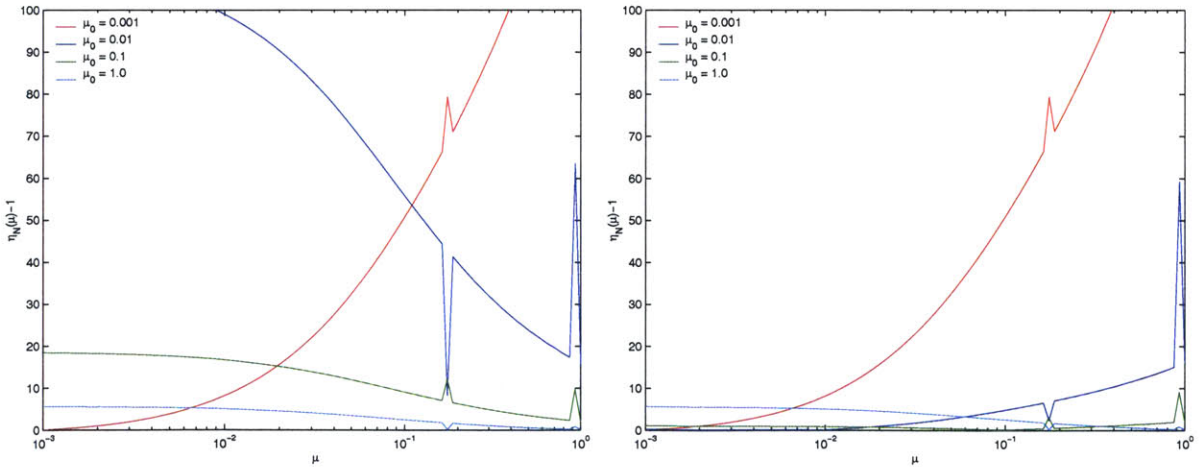


Figure 5-14: Effectivity as a function of μ for Example 5, calculated using the concave eigenvalue bound conditioner with $\mathcal{C}_1 = A(\theta_0)$ for different choices of θ_0 $(= \mu_0)$, and for $T = 2$ and $T = 4$.

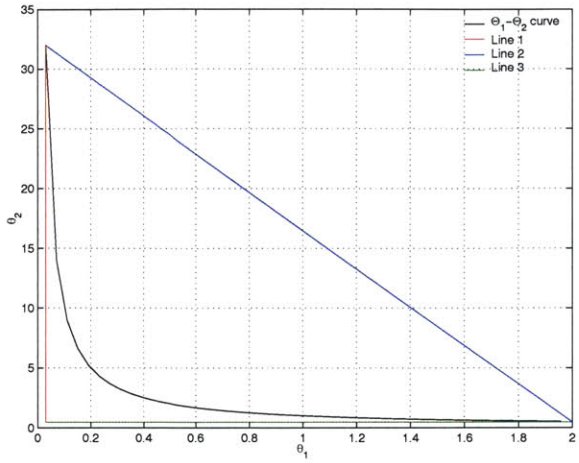


Figure 5-15: Θ_1-Θ_2 curve for Examples 6 and 7, with $\mu \in \mathcal{D}^{\mu} = [0.03125, 2.0]$.

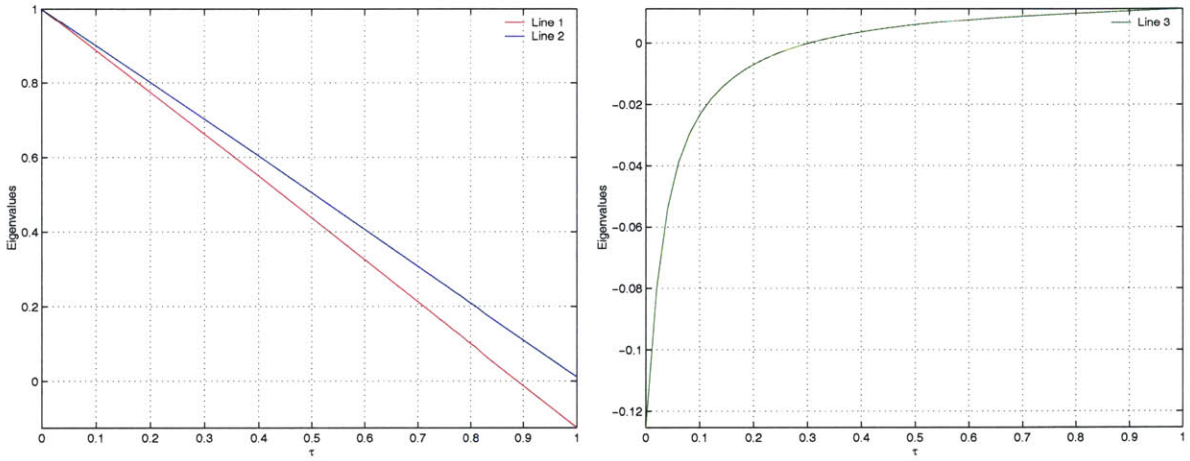


Figure 5-16: Concavity of the minimum eigenvalue with respect to θ for Example 6.

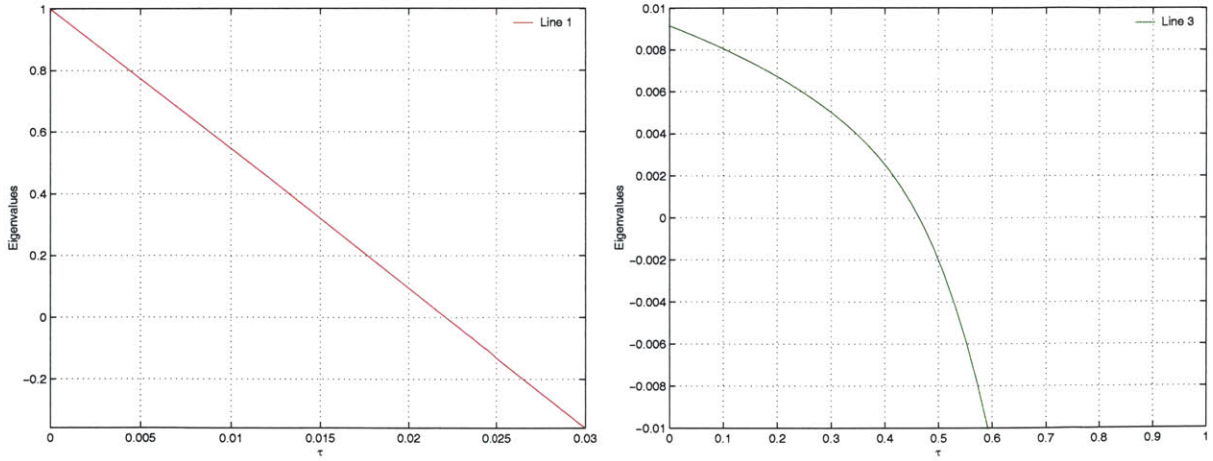


Figure 5-17: Concavity of the minimum eigenvalue with respect to θ for Example 6.

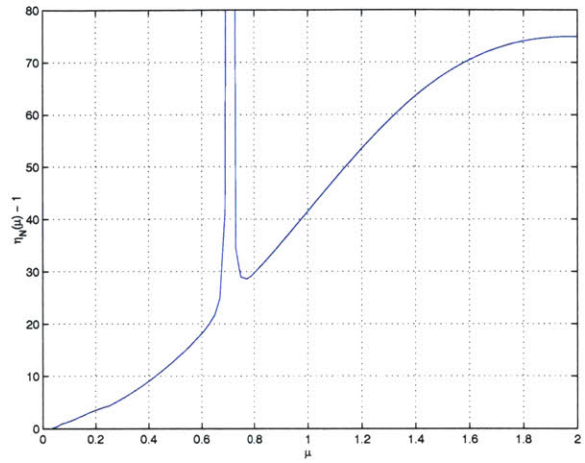


Figure 5-18: Effectivity as a function of μ for Example 6, calculated using the concave eigenvalue bound conditioner with $\mathcal{C}_1 = A(\theta_0)$.

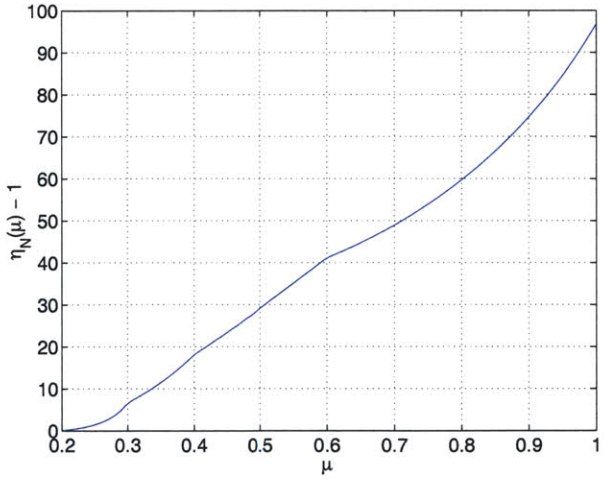


Figure 5-19: Effectivity as a function of μ for Example 7, calculated using the concave eigenvalue bound conditioner with $\mathcal{C}_1 = A(\theta_0)$.

5.7.8 Effective Property Bound Conditioner

We note that the effective property bound conditioner of Section 4.3.4 can be extended to problems involving property variations in the linear elasticity operator: instead of an effective diffusivity tensor, we now consider seek an effective elasticity tensor. We plot in Figure 5-20 the effectivity as a function of μ for Example 5. The results show that for $I = 3$, the effective property yields good effectivities — $\eta_N(\mu) \leq 10$ for all $\mu \in \mathcal{D}^\mu$.

However, we recall that the effective property bound conditioner requires that there exist an effective tensor (here, an effective elasticity tensor) which satisfies (4.103) for all $\mu \in \mathcal{D}^\mu$. In the case of geometry variations in the linear elasticity operator, however, the elasticity tensor is singular, with a nullspace that is parameter-dependent. It is therefore impossible for linear elasticity problems with geometry variations to find an effective elasticity tensor which satisfies our requirements.

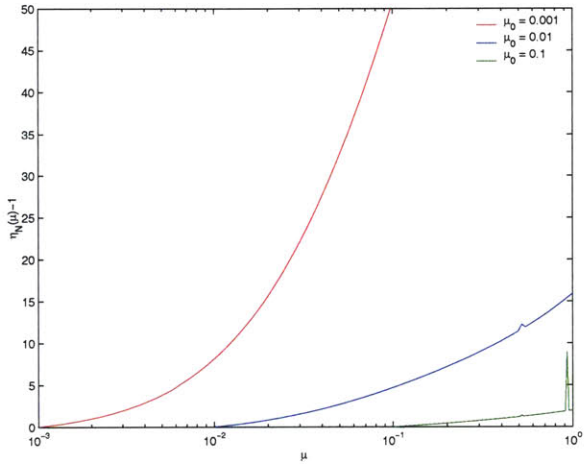


Figure 5-20: Effectivity as a function of μ for Example 5 obtained using the effective property bound conditioner.

5.7.9 Convex Inverse Bound Conditioners

We recall that the convex inverse Bound conditioners require that $\mathcal{A}(\mu)$ be of the form

$$\mathcal{A}(\mu) = \mathcal{A}^0 + \sum_{q=1}^{Q_\mathcal{A}} \Theta^q(\mu)\, \mathcal{A}^q \;, \quad \forall\, \mu \in \mathcal{D}^\mu \;, \tag{5.204}$$

where the $\Theta^q(\mu) : \mathcal{D}^\mu \rightarrow \mathbb{R}_{+,0}$, $q = 1, \ldots, Q_\mathcal{A}$; $\mathcal{A}^0 : Y \rightarrow Y'$ is symmetric, continuous, and coercive; and the $\mathcal{A}^q : Y \rightarrow Y'$, $q = 1, \ldots, Q_\mathcal{A}$ are symmetric, continuous, and positive-semidefinite.

We note that in all our model problems — Examples 5 to 8 — the operator cannot be expressed in the required form: either the $\mathcal{A}^q$ are all positive-semidefinite (as in Example 5), or the $\mathcal{A}^q$ are indefinite (as in Examples 6 to 8).

5.8 *A Posteriori* Error Estimation: Method II

We discuss here the extension of Method II of Chapter 4 to noncompliant outputs.

5.8.1 Error and Output Bounds

Nonsegregated Approach

Following [39], we set $M > N$, and introduce a parameter sample

$$S_M^\mu = \{\mu_1, \ldots, \mu_{M/2}\} \tag{5.205}$$

and associated reduced–basis approximation space

$$W_M = \text{span}\,\{u(\mu_m), \psi(\mu_m), m = 1, \ldots, M/2\} \tag{5.206}$$

where $u(\mu_m) \in Y$ and $\psi(\mu_m) \in Y$ are the solutions to (5.5) and (5.105), respectively, for $\mu = \mu_m$. For both theoretical and practical reasons we require $S_N^\mu \subset S_M^\mu$ and therefore $W_N \subset W_M$. The procedure is very simple: we first find $u_M(\mu) \in W_M$ such that

$$\langle \mathcal{A}(\mu) u_M(\mu), v\rangle = \langle F(\mu), v\rangle, \quad \forall\, v \in W_M\ ; \tag{5.207}$$

we then evaluate

$$s_M(\mu) = \langle L(\mu), u_M(\mu)\rangle\ . \tag{5.208}$$

Segregated Approach

We again set $M > N$, and introduce a parameter sample

$$S_M^\mu = \{\mu_1, \ldots, \mu_{M/2}\} \tag{5.209}$$

and associated segregated reduced–basis approximation spaces

$$W_M^{\text{pr}} = \text{span}\,\{\zeta_m^{\text{pr}} = u(\mu_m), m = 1, \ldots, M/2\} \tag{5.210}$$

$$W_M^{\text{du}} = \text{span}\left\{\zeta_m^{\text{du}} = \psi(\mu_m), m = 1, \ldots, M/2\right\}\ , \tag{5.211}$$

where $u(\mu_m) \in Y$ and $\psi(\mu_m) \in Y$ are the solutions to (5.5) and (5.105), respectively, for $\mu = \mu_m$. As before, we require $S_N^\mu \subset S_M^\mu$ and therefore $W_N^{\text{pr}} \subset W_M^{\text{pr}}$ and $W_N^{\text{du}} \subset W_M^{\text{du}}$. We first find $u_M(\mu) \in W_M^{\text{pr}}$ and $\psi_M(\mu) \in W_M^{\text{du}}$ such that

$$\langle \mathcal{A}(\mu) u_M(\mu), v\rangle = \quad \langle F(\mu), v\rangle, \quad \forall\, v \in W_M^{\text{pr}}\ , \tag{5.212}$$

$$\langle \mathcal{A}(\mu) \psi_M(\mu), v\rangle = -\langle L(\mu), v\rangle, \quad \forall\, v \in W_M^{\text{du}}\ ; \tag{5.213}$$

we then evaluate

$$s_M(\mu) = \langle L(\mu), u_M(\mu)\rangle - (\langle F(\mu), \psi_M(\mu)\rangle - \langle \mathcal{A}(\mu) u_M(\mu), \psi_M(\mu)\rangle)\ . \tag{5.214}$$

Output Bounds

Given $s_N(\mu)$ and $s_M(\mu)$ calculated using either the nonsegregated or segregated approaches, we evaluate the "improved" estimator $\bar{s}_N(\mu)$ given by

$$\bar{s}_N(\mu) = s_N(\mu) - \frac{1}{2\tau}\,(s_M(\mu) - s_N(\mu))\ ; \tag{5.215}$$

our lower and upper output bounds are then

$$s_N^-(\mu) = \bar{s}_N(\mu) - \frac{1}{2}\Delta_N(\mu) \tag{5.216}$$

$$s_N^+(\mu) = \bar{s}_N(\mu) + \frac{1}{2}\Delta_N(\mu) \tag{5.217}$$

where $\Delta_N(\mu)$ is given by

$$\Delta_{N,M}(\mu) = \frac{1}{\tau}\,|s_M(\mu) - s_N(\mu)| \tag{5.218}$$

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

$$\eta_{N,M}(\mu) = \frac{\Delta_{N,M}(\mu)}{|s(\mu) - s_N(\mu)|}\ . \tag{5.219}$$

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

5.8.2 Bounding Properties

As in Section 4.5, we would like to prove the effectivity inequality $1 \le \eta_{N,2N}(\mu) \le \rho$, for sufficiently large N; as in Section 4.5, we must again make the hypothesis (4.190). We first consider the lower effectivity inequality, $\eta_{N,M}(\mu) \ge 1$, and prove the equivalent statement

$$s_{N,2N}^-(\mu) \le s(\mu) \le s_{N,2N}^+(\mu), \quad N \to \infty. \tag{5.220}$$

Following [39], we write

$$s_{N,2N}^-(\mu) = s(\mu) - \frac{1}{1-\varepsilon_{N,2N}}|s_N(\mu) - s_{2N}(\mu)| \times \begin{cases} 1 & s_{2N}(\mu) \ge s_N(\mu) \\ \frac{1}{\tau}(1-\varepsilon_{N,2N}) - 1 & s_{2N}(\mu) < s_N(\mu)\ , \end{cases} \tag{5.221}$$

$$s_{N,2N}^+(\mu) = s(\mu) + \frac{1}{1-\varepsilon_{N,2N}}|s_N(\mu) - s_{2N}(\mu)| \times \begin{cases} \frac{1}{\tau}(1-\varepsilon_{N,2N}) - 1 & s_{2N}(\mu) \ge s_N(\mu) \\ 1 & s_{2N}(\mu) < s_N(\mu). \end{cases} \tag{5.222}$$

From our hypothesis on $\varepsilon_{N,2N}$, (4.190), and the range of τ, there exists an N^* such that for $N > N^*$, $\frac{1}{\tau}(1-\varepsilon_{N,2N}) > 1$; the result (5.220) directly follows.

We now consider the upper effectivity inequality (sharpness property). As in Section 4.5, we write

$$\eta_{N,2N}(\mu) = \frac{\frac{1}{\tau}|s_{2N} - s_N|}{|s - s_N|} = \frac{\frac{1}{\tau}|s_{2N} - s + s - s_N|}{|s - s_N|} \tag{5.223}$$

$$= \frac{1}{\tau}|1 - \varepsilon_{N,2N}|; \tag{5.224}$$

from our hypothesis (4.190) we may thus conclude that $\eta_{N,2N}(\mu) \to \frac{1}{\tau}$ as $N \to \infty$. Note that unlike in the compliant case, $\eta_N(\mu)$ no longer approaches $1/\tau$ strictly from below.

As before, 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. These claims are demonstrated in Section 5.8.4.

5.8.3 Off-line/On-line 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 5.6. As before, the calculation of the output approximation $s_N(\mu)$ and the output bounds are integrated: the matrices and vectors which yield $s_N(\mu)$ are sub-matrices and sub-vectors of those yielding $s_{2N}(\mu)$, $\Delta_{N,2N}(\mu)$, and $s^{\pm}_{N,2N}(\mu)$.

5.8.4 Numerical Results

As in Chapter 4, the essential approximation enabler is exponential convergence: we obtain bounds even for rather small N and relatively large τ. The rapid convergence in the reduced-basis approximation allows us to achieve both "near" certainty *and* good effectivities, as demonstrated in Tables 5.4, 5.5, and 5.6; the results tabulated correspond to the choice $\tau = 1/2$. In all cases, we clearly obtain bounds even for relatively small N; and we observe that $\eta_{N,2N}(\mu)$ does, indeed, rather quickly approach $1/\tau$. Furthermore, the results presented in Table 5.6 illustrate the fact that, in the noncompliance case, the effectivities are no longer bounded from above by $\tau = 1/2$.

N	$\min_{\mu} \eta_N(\mu)$	$\max_{\mu} \eta_{N,M}(\mu)$
1	1.98	2.00
2	1.96	2.00
3	2.00	2.00

N	$\min_{\mu} \eta_N(\mu)$	$\max_{\mu} \eta_{N,M}(\mu)$
1	0.32	2.00
2	1.55	2.00
3	1.44	2.00
4	1.18	2.00
5	1.93	2.00
6	1.97	2.00

Table 5.4: Minimum and maximum effectivities for the Method II error estimators for Examples 5 and 6.

N	$\min_{\mu} \eta_N(\mu)$	$\max_{\mu} \eta_{N,M}(\mu)$
1	0.18	1.99
2	0.92	1.99
3	1.55	2.00
4	1.56	2.00
5	1.96	2.00
6	1.92	2.00
7	1.78	2.00

N	$\min_{\mu} \eta_N(\mu)$	$\max_{\mu} \eta_{N,M}(\mu)$
1	0.01	2.00
2	1.25	1.97
3	1.47	1.99
4	0.37	2.00
5	1.06	2.00
6	1.57	2.00
7	1.54	2.00
8	1.71	2.00
9	1.60	2.00
10	1.70	2.00

Table 5.5: Minimum and maximum effectivities for the Method II error estimators for Examples 7 and 8.

N	$\min_{\mu} \eta_N(\mu)$	$\max_{\mu} \eta_{N,M}(\mu)$
4	0.30	3.60
8	0.16	2.00
12	1.63	2.01
16	1.43	2.00
20	1.81	4.15
24	1.54	2.04
28	1.93	2.01
32	1.96	2.06
36	1.96	2.44
40	1.97	2.00

Table 5.6: Minimum and maximum effectivities for the Method II error estimators for Example 9.

Chapter 6

Reduced-Basis Output Bounds for Eigenvalue Problems: An Elastic Stability Example

6.1 Introduction

There are many applications in which the posibility of unstable equilibria must be considered. Applied forces which exceed the critical loads may cause structures to buckle, resulting in large (and potentially dangerous) deformations; structural stability considerations must therefore be included in any design.

While analytical formulas exist for the critical loads of relatively simple geometries, they are of limited applicability particularly for more complex structures. However, exact (finite element) solution of the partial differential equations governing elastic stability is computationally too expensive, especially in the "many queries" context of design and optimization.

We consider here the reduced-basis approximation and (Method II) error estimation for the problem of elastic buckling. The output of interest is the critical buckling load, calculation of which requires solution of a nonlinear (partial differential) eigenvalue problem. Note the nonlinearity presents significant difficulties: the computational and storage requirements for the reduced-basis approximation are greater, and our rigorous error estimation procedures are currently inapplicable.

6.2 Abstraction

As in Chapter 5, we consider a suitably regular (smooth) parameter-independent domain $\Omega \subset \mathbb{R}^d, d = 1, 2$, or 3, the associated function space $Y \subset (H^1(\Omega))^d$, with inner product $(\cdot,\cdot)_Y$, norm $\|\cdot\|_Y = (\cdot,\cdot)_Y^{1/2}$, dual space Y', and duality pairing $\langle \cdot,\cdot \rangle = {}_{Y'}\langle \cdot,\cdot \rangle_Y$; as before, we define a parameter set $\mathcal{D}^\mu \in \mathbb{R}^P$, a particular point in which will be denoted μ.

We now introduce a symmetric, continuous, and coercive distributional operator $\mathcal{A}(\mu) : Y \to Y'$; a symmetric, continuous, and semi-definite distributional operator $\mathcal{B}(\mu; w) : Y \to Y'$; and bounded linear form $F(\mu) \in Y'$. We shall again make certain assumptions on the parametric dependence of $\mathcal{A}$, $\mathcal{B}$, and F. In particular, we shall suppose that, for some finite (preferably small) integers $Q_\mathcal{A}$,

Q_B, and Q_F, we may express $\mathcal{A}(\mu)$, $\mathcal{B}(\mu)$, and $F(\mu)$ for all $\mu \in \mathcal{D}^\mu$ as

$$\mathcal{A}(\mu) = \sum_{q=1}^{Q_\mathcal{A}} \Theta^q(\mu)\, \mathcal{A}^q \;, \tag{6.1}$$

$$\mathcal{B}(\mu; w) = \sum_{q=1}^{Q_\mathcal{A}} \Theta_\mathcal{B}^q(\mu)\, \mathcal{B}^q(w) \;, \tag{6.2}$$

$$F(\mu) = \sum_{q=1}^{Q_F} \varphi_F^q(\mu)\, F^q \;, \tag{6.3}$$

where $\Theta^q(\mu)\colon \mathcal{D}^\mu \to \mathbb{R}$, $\varphi_F^q(\mu)\colon \mathcal{D}^\mu \to \mathbb{R}$, and $\Theta_\mathcal{B}^q(\mu)\colon \mathcal{D}^\mu \to \mathbb{R}$. Here, $\mathcal{A}^q\colon Y \to Y'$, $F^q \in Y'$, and $\mathcal{B}^q(w)\colon Y \to Y'$. We further assume that the $\mathcal{B}^q$ are linear in w such that for any $\hat{a}_1, \hat{a}_2 \in \mathbb{R}$, and $\hat{w}_1, \hat{w}_2 \in Y$, $\mathcal{B}^q(\hat{a}_1\hat{w}_1 + \hat{a}_2\hat{w}_2) = \hat{a}_1\mathcal{B}^q(\hat{w}_1) + \hat{a}_2\mathcal{B}^q(\hat{w}_2)$. As before, the assumption of "separability" or affine parameter dependence is crucial to computational efficiency.

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

$$s(\mu) = \lambda^1(\mu), \tag{6.4}$$

where $\lambda^1(\mu)$ is the eigenvalue with the smallest magnitude satisfying

$$\langle \mathcal{A}(\mu)\xi(\mu), v\rangle = \lambda(\mu)\langle \mathcal{B}(\mu; u(\mu))\xi(\mu), v\rangle \;, \quad \forall\, v \in Y \;; \tag{6.5}$$

here, $\xi(\mu) \in Y$ is the eigenvector associated with $\lambda(\mu)$, and $u(\mu) \in Y$ is the solution of

$$\langle \mathcal{A}(\mu)u(\mu), v\rangle = \langle F(\mu), v\rangle \;, \quad \forall\, v \in Y \;. \tag{6.6}$$

Calculation of $s(\mu)$ thus requires solution of the partial differential equation (6.6) and the eigenvalue problem (6.5) — both computationally intensive.

6.3 Formulation of the Elastic Stability Problem

In this section we derive the eigenvalue problem for elastic bifurcation, using the weak form and the equivalent minimization principle as our point of departure. We then reformulate the problem in terms of a reference domain, thus recovering the abstract formulation of Section 6.2. We again use a bar to denote a general dependence on the parameter μ.

6.3.1 The Elastic Stability Eigenvalue Problem

Linear Elasticity

To begin, we consider a problem in linear elasticity described by

$$\langle \bar{\mathcal{A}}\bar{u}, \bar{v}\rangle = \langle \bar{F}, \bar{v}\rangle, \qquad \forall\, \bar{v} \in \bar{Y} \;, \tag{6.7}$$

where $\bar{Y} = \{\bar{v} \in H_1(\bar{\Omega}) \mid \bar{v}_i \bar{e}_i^n = 0 \text{ on } \bar{\Gamma}_D^n,\ \bar{v}_i \bar{e}_i^t = 0 \text{ on } \bar{\Gamma}_D^t\}$, $\bar{\mathcal{A}}$: $\bar{Y} \to \bar{Y}'$ is a symmetric, continuous, and coercive operator given by

$$\langle \bar{\mathcal{A}}\bar{w}, \bar{v}\rangle = \int_{\bar{\Omega}} \frac{\partial \bar{v}_i}{\partial \bar{x}_j} \bar{C}_{ijkl} \frac{\partial \bar{w}_k}{\partial \bar{x}_l}\, d\bar{\Omega}\ , \tag{6.8}$$

and the "loading" $\bar{F} \in \bar{Y}'$ is given by (5.25). We now introduce a load parameter λ and define

$$\langle \bar{F}'(\lambda), v\rangle = \lambda \langle \bar{F}, v\rangle, \qquad \forall\ \bar{v} \in \bar{Y}\ . \tag{6.9}$$

such that (6.7) may be written as

$$\langle \bar{\mathcal{A}}\bar{u}, \bar{v}\rangle = \langle \bar{F}'(1), \bar{v}\rangle, \qquad \forall\ \bar{v} \in \bar{Y}\ . \tag{6.10}$$

Furthermore, for any $\lambda \in \mathbb{R}$,

$$\langle \bar{\mathcal{A}}(\lambda\bar{u}), \bar{v}\rangle = \langle \bar{F}'(\lambda), \bar{v}\rangle, \qquad \forall\ \bar{v} \in \bar{Y}\ . \tag{6.11}$$

We note that (6.11) may also be written as a minimization principle: the solution $\lambda\bar{u}$ to (6.11) is the argument which minimizes the strain energy

$$\lambda\bar{u} = \arg\min_{\bar{w} \in \bar{Y}} \bar{J}(\bar{w}; \lambda) \tag{6.12}$$

where the strain energy $\bar{J}(\bar{w}; \lambda)$ is defined as

$$\bar{J}(\bar{w}; \lambda) = \frac{1}{2}\int_{\bar{\Omega}} \bar{\varepsilon}_{ij}(\bar{w}) \bar{C}_{ijkl} \bar{\varepsilon}_{kl}(\bar{w})\, d\bar{\Omega} - \lambda \left(\int_{\bar{\Omega}} \bar{b}_i \bar{w}_i\, d\bar{\Omega} + \int_{\bar{\Gamma}_N^n} \bar{v}_i \bar{f}_n \bar{e}_i^n\, d\bar{\Gamma} + \int_{\bar{\Gamma}_N^t} \bar{v}_i \bar{f}_t \bar{e}_i^t\, d\bar{\Gamma} \right) \tag{6.13}$$

$$= \frac{1}{2}\langle \bar{\mathcal{A}}\bar{w}, \bar{w}\rangle - \langle \bar{F}'(\lambda), \bar{w}\rangle\ . \tag{6.14}$$

To prove (6.12), we first note that any $\bar{w} \in \bar{Y}$ can be expressed as $\bar{w} = \lambda\bar{u} + \bar{v}$ where $\bar{v} \in \bar{Y}$. Substituting this expression for $\bar{w}$ into (6.13) yields

$$\bar{J}(\bar{w}; \lambda) = \frac{1}{2}\langle \bar{\mathcal{A}}(\lambda\bar{u} + \bar{v}), (\lambda\bar{u} + \bar{v})\rangle - \langle \bar{F}'(\lambda), (\lambda\bar{u} + \bar{v})\rangle \tag{6.15}$$

$$\begin{aligned} = {} & \frac{1}{2}\langle \bar{\mathcal{A}}(\lambda\bar{u}), (\lambda\bar{u})\rangle - \langle \bar{F}'(\lambda), (\lambda\bar{u})\rangle \\ & + \frac{1}{2}\langle \bar{\mathcal{A}}(\lambda\bar{u}), \bar{v}\rangle + \frac{1}{2}\langle \bar{\mathcal{A}}\bar{v}, (\lambda\bar{u})\rangle - \langle \bar{F}'(\lambda), \bar{v}\rangle \\ & + \frac{1}{2}\langle \bar{\mathcal{A}}\bar{v}, \bar{v}\rangle \end{aligned} \tag{6.16}$$

$$\begin{aligned} = {} & \bar{J}(\lambda\bar{u}; \lambda) \\ & + \langle \bar{\mathcal{A}}(\lambda\bar{u}), \bar{v}\rangle - \langle \bar{F}'(\lambda), \bar{v}\rangle \\ & + \frac{1}{2}\langle \bar{\mathcal{A}}\bar{v}, \bar{v}\rangle \end{aligned} \tag{6.17}$$

$$= \bar{J}(\lambda\bar{u}; \lambda) + \frac{1}{2}\langle \bar{\mathcal{A}}\bar{v}, \bar{v}\rangle\ . \tag{6.18}$$

Since $\langle A\bar{v}, \bar{v}\rangle > 0$ for all $\bar{v} \in \bar{Y}$, $\bar{v} \neq 0$, it follows that

$$\bar{J}(\bar{w};\lambda) > \bar{J}(\lambda\bar{u};\lambda) \ , \qquad \forall\ \bar{w} \in \bar{Y}\ ,\ \bar{w} \neq \bar{u}\ ; \tag{6.19}$$

this concludes the proof.

The equivalence of (6.11) and (6.12) is then clear: (6.11) is in fact the stationarity condition of (6.12). In other words, if we define $\delta_{\bar{v}}\bar{J}(\bar{w};\lambda)$ to be the first variation of $\bar{J}(\bar{w};\lambda)$ with respect to perturbations $\bar{v}$, i.e.,

$$\delta_{\bar{v}}\bar{J}(\bar{w};\lambda) = \langle \bar{\mathcal{A}}\bar{w}, \bar{v}\rangle - \langle \bar{F}'(\lambda), \bar{v}\rangle \ , \tag{6.20}$$

then the argument $\lambda\bar{u}$ which minimizes $\bar{J}(\bar{w},\lambda)$ over all $\bar{w} \in \bar{Y}$ renders the first variation zero for all perturbations $\bar{v}$:

$$\delta_{\bar{v}}\bar{J}(\bar{w};\lambda)\Big|_{\bar{w}=\bar{u}} = \langle \bar{\mathcal{A}}(\lambda\bar{u}), \bar{v}\rangle - \langle \bar{F}'(\lambda), \bar{v}\rangle = 0 \qquad \forall\ \bar{v} \in \bar{Y}\ . \tag{6.21}$$

We shall now generalize our statements to the case in which the displacements and rotations may be large; we shall use the minimization principle as the point of departure for our stability formulation.

Nonlinear Elasticity and Stability

We let $\tilde{\varepsilon}(\bar{w})$ denote the (general) nonlinear strains associated with the displacement field $\bar{w}$; $\tilde{\varepsilon}(\bar{w})$ is then given by

$$\tilde{\varepsilon}_{ij}(\bar{w}) = \frac{1}{2}\left(\frac{\partial \bar{w}_i}{\partial \bar{x}_j} + \frac{\partial \bar{w}_j}{\partial \bar{x}_i}\right) + \frac{1}{2}\frac{\partial \bar{w}_m}{\partial \bar{x}_i}\frac{\partial \bar{w}_m}{\partial \bar{x}_j}\ , \tag{6.22}$$

and the corresponding strain energy is

$$\tilde{J}(\bar{w};\lambda) = \frac{1}{2}\int_{\bar{\Omega}} \tilde{\varepsilon}_{ij}(\bar{w})\bar{C}_{ijkl}\tilde{\varepsilon}_{kl}(\bar{w})\, d\bar{\Omega} - \lambda\left(\int_{\bar{\Omega}} \bar{b}_i\bar{w}_i\, d\bar{\Omega} + \int_{\bar{\Gamma}^n_N} \bar{v}_i\bar{f}_n\bar{e}^n_i\, d\bar{\Gamma} + \int_{\bar{\Gamma}^t_N} \bar{v}_i\bar{f}_t\bar{e}^t_i\, d\bar{\Gamma}\right)\ . \tag{6.23}$$

In the case in which the strains and rotations induced by $\bar{w}$ are small, the nonlinear term in (6.22) may be neglected so that $\tilde{\varepsilon}(\bar{w}) \approx \bar{\varepsilon}(\bar{w})$ and $\tilde{J}(\bar{w};\lambda) \approx \bar{J}(\bar{w};\lambda)$.

We now consider *bifurcations* from the solution to the linearized equations. In particular, we seek values of the load parameter λ and corresponding perturbations $\bar{\xi}$ to the linearized displacements $\lambda\bar{u}$ such that $\tilde{J}(\lambda\bar{u} + \bar{\xi};\lambda)$ is stationary:

$$\delta_{\bar{v}}\tilde{J}(\bar{w};\lambda)\Big|_{\bar{w}=\lambda\bar{u}+\bar{\xi}} = 0 \qquad \forall\ \bar{v} \in \bar{Y}\ . \tag{6.24}$$

Taking $\bar{w} = \bar{u} + \bar{\xi} + \bar{v}$ (and after much manipulation) we obtain

$$\begin{aligned}
\tilde{J}(\bar{w};\lambda) &= \tilde{J}(\lambda\bar{u};\lambda) \\
&\quad +\lambda\left(\int_{\bar{\Omega}} \frac{\partial \bar{v}_i}{\partial \bar{x}_j}\bar{C}_{ijkl}\frac{\partial \bar{u}_k}{\partial \bar{x}_l}\,d\bar{\Omega} - \int_{\bar{\Omega}} \bar{b}_i\bar{v}_i\,d\bar{\Omega} + \int_{\bar{\Gamma}} \bar{f}_i\bar{v}_i\,d\bar{\Gamma}\right) \\
&\quad + \int_{\bar{\Omega}} \frac{\partial \bar{v}_i}{\partial \bar{x}_j}\bar{C}_{ijkl}\frac{\partial \bar{\xi}_k}{\partial \bar{x}_l}\,d\bar{\Omega} + \lambda\int_{\bar{\Omega}} \frac{\partial \bar{u}_i}{\partial \bar{x}_j}\bar{C}_{ijkl}\frac{\partial \bar{\xi}_m}{\partial \bar{x}_k}\frac{\partial \bar{v}_m}{\partial \bar{x}_l}\,d\bar{\Omega} \\
&\quad +\lambda\left(\int_{\bar{\Omega}} \frac{\partial \bar{\xi}_i}{\partial \bar{x}_j}\bar{C}_{ijkl}\frac{\partial \bar{u}_m}{\partial \bar{x}_k}\frac{\partial \bar{v}_m}{\partial \bar{x}_l}\,d\bar{\Omega} + \int_{\bar{\Omega}} \frac{\partial \bar{v}_i}{\partial \bar{x}_j}\bar{C}_{ijkl}\frac{\partial \bar{u}_m}{\partial \bar{x}_k}\frac{\partial \bar{\xi}_m}{\partial \bar{x}_l}\,d\bar{\Omega}\right) \\
&\quad +\text{h.o.t.} \qquad (6.25) \\
&= \tilde{J}(\lambda\bar{u};\lambda) \\
&\quad + \int_{\bar{\Omega}} \frac{\partial \bar{v}_i}{\partial \bar{x}_j}\bar{C}_{ijkl}\frac{\partial \bar{\xi}_k}{\partial \bar{x}_l}\,d\bar{\Omega} + \lambda\int_{\bar{\Omega}} \frac{\partial \bar{u}_i}{\partial \bar{x}_j}\bar{C}_{ijkl}\frac{\partial \bar{\xi}_m}{\partial \bar{x}_k}\frac{\partial \bar{v}_m}{\partial \bar{x}_l}\,d\bar{\Omega} \\
&\quad +\text{h.o.t.} \qquad (6.26)
\end{aligned}$$

Following classical elastic buckling theory [18], we neglect higher-order terms to obtain the stationarity condition

$$\int_{\bar{\Omega}} \frac{\partial \bar{v}_i}{\partial \bar{x}_j}\bar{C}_{ijkl}\frac{\partial \bar{\xi}_k}{\partial \bar{x}_l}\,d\bar{\Omega} + \lambda\int_{\bar{\Omega}} \frac{\partial \bar{u}_i}{\partial \bar{x}_j}\bar{C}_{ijkl}\frac{\partial \bar{\xi}_m}{\partial \bar{x}_k}\frac{\partial \bar{v}_m}{\partial \bar{x}_l}\,d\bar{\Omega} = 0, \quad \forall\, v \in \bar{Y} \qquad (6.27)$$

Numerical tests (see Section 6.4) confirm that the effect of higher order terms on the eigenvalues of interest is indeed minimal.

Our problem is then to find the "buckling load parameter" $\lambda^1(\mu)$ at which bifurcation first occurs; in particular, given the solution $\bar{u}$ to

$$\langle \bar{\mathcal{A}}\bar{u}, \bar{v}\rangle = \langle \bar{F}'(1), \bar{v}\rangle, \qquad \forall\, \bar{v} \in \bar{Y}\ , \qquad (6.28)$$

we wish to find $(\lambda^1, \bar{\xi}_1)$ where λ^1 is the eigenvalue with the smallest magnitude and $\bar{\xi}_1$ is the associated eigenvector satisfying

$$\langle \bar{\mathcal{A}}\bar{\xi}, \bar{v}\rangle = \lambda\langle \bar{\mathcal{B}}(\bar{u})\bar{\xi}, \bar{v}\rangle, \qquad \forall\, \bar{v} \in \bar{Y}\ ; \qquad (6.29)$$

here, $\bar{\mathcal{A}}$ is given by (6.8), $\bar{F}'(1)$ is given by (6.9) and

$$\langle \bar{\mathcal{B}}(\bar{z})\bar{w}, \bar{v}\rangle = -\int_{\bar{\Omega}} \frac{\partial \bar{z}_i}{\partial \bar{x}_j}\bar{C}_{ijkl}\frac{\partial \bar{w}_m}{\partial \bar{x}_k}\frac{\partial \bar{v}_m}{\partial \bar{x}_l}\,d\bar{\Omega}\ . \qquad (6.30)$$

Note that (6.29)-(6.30) is equivalent to the strong form given in Chapter 1.

6.3.2 Reduction to Abstract Form

In this section, we reformulate the eigenvalue problem defined by (6.28)-(6.30) so as to recover the abstract formulation of Section 6.2.

Affine Geometric Mapping

As before, we assume $\bar{\Omega}$ may be subdivided into R sub-domains $\bar{\Omega}^r$ such that there exists a reference domain $\Omega = \bigcup_{r=1}^{R} \overline{\Omega}^r$ where, for any $\underline{\bar{x}} \in \bar{\Omega}^r$, $r = 1, \ldots, R$, its image $\underline{x} \in \Omega^r$ is given by

$$\underline{x} = \mathcal{G}^r(\mu; \underline{\bar{x}}) = \underline{G}^r(\mu)\underline{\bar{x}} + \underline{g}^r(\mu); \tag{6.31}$$

we again write $\underline{x} = \mathcal{G}(\mu; \underline{\bar{x}}) = \underline{G}(\mu)\underline{\bar{x}} + \underline{g}(\mu)$, where $\underline{x} \in \Omega$, $\underline{\bar{x}} \in \bar{\Omega}$, $\underline{G}(\mu) \in \mathbb{R}^{d\times d}$ is a piecewise-constant matrix, $\underline{g}^r(\mu) \in \mathbb{R}^d$ is a piecewise-constant vector, and $\mathcal{G}(\mu)\colon \bar{\Omega} \to \Omega$ is a piecewise-affine geometric mapping.

Reference Domain Formulation

We also define the function space $Y = Y(\Omega) = \bar{Y}(\mathcal{G}^{-1}(\mu;\Omega)) = \bar{Y}(\bar{\Omega})$, and for any function $\bar{w} \in \bar{Y}$, we define $w \in Y$ such that $w(\underline{x}) = \bar{w}(\mathcal{G}^{-1}(\underline{x}))$. Furthermore, we recall that $d\bar{\Omega} = \det \underline{G}^{-1}(\mu)\, d\Omega$, and $d\bar{\Gamma} = \left|\underline{G}^{-1}(\mu)\, \underline{e}^t\right| d\Gamma$, where $\underline{e}^t$ is the unit vector tangent to the boundary Γ, and $\left|\underline{G}^{-1}(\mu)\, \underline{e}^t\right| = (\sum_{i=1}^{d}(G_{ij}e_j^t)^2)^{1/2}$. It then follows that $\langle \mathcal{A}(\mu)w, v\rangle = \langle \bar{A}\bar{w}, \bar{v}\rangle$ for $\bar{A}$ as in (6.8) and $\mathcal{A}(\mu)$ given by

$$\begin{aligned}
\langle \mathcal{A}(\mu)w, v\rangle &= \sum_{r=1}^{R} \int_{\Omega^r} \left(G^r_{jj'}(\mu)\frac{\partial w_i}{\partial x_j}\right) \bar{C}^r_{ij'kl'} \left(G^r_{ll'}(\mu)\frac{\partial v_k}{\partial x_l}\right) \det(\underline{G}^r(\mu))^{-1}\, d\Omega\ , \\
&= \sum_{r=1}^{R} \int_{\Omega^r} \frac{\partial w_i}{\partial x_j} \left(G^r_{jj'}(\mu)\bar{C}^r_{ij'kl'}G^r_{ll'}(\mu) \det(\underline{G}(\mu))^{-1}\right) \frac{\partial v_k}{\partial x_l}\, d\Omega, \quad \forall\ w, v \in Y\ ;
\end{aligned}$$

$\langle \mathcal{B}(\mu; z)w, v\rangle = \langle \bar{\mathcal{B}}(\bar{z})\bar{w}, \bar{v}\rangle$ for $\bar{\mathcal{B}}$ as in (6.30) and $\mathcal{B}(\mu)$ given by

$$\begin{aligned}
\langle \mathcal{B}(\mu; z)w, v\rangle &= \sum_{r=1}^{R} \int_{\Omega^r} \left(G^r_{jj'}(\mu)\frac{\partial z_i}{\partial x_j}\right) \bar{C}^r_{ij'k'l'} \left(G^r_{kk'}(\mu)\frac{\partial w_m}{\partial x_k}\right) \left(G^r_{ll'}(\mu)\frac{\partial v_m}{\partial x_l}\right) \det(\underline{G}^r(\mu))^{-1}\, d\Omega\ , \\
&= \sum_{r=1}^{R} \int_{\Omega^r} \frac{\partial z_i}{\partial x_j} \left(G^r_{jj'}(\mu)\bar{C}^r_{ij'k'l'}G^r_{kk'}(\mu)G^r_{ll'}(\mu) \det(\underline{G}^r(\mu))^{-1}\right) \frac{\partial w_m}{\partial x_k}\frac{\partial v_m}{\partial x_l}\, d\Omega\ , \\
&\qquad\qquad \forall\ w, v, z \in Y\ ;
\end{aligned}$$

and $\langle F'(\mu;\lambda), v\rangle = \langle \bar{F}'(\lambda), \bar{v}\rangle$ for $\bar{F}'(\lambda)$ as in (6.9), $F'(\mu;\lambda)$ given by

$$\langle F'(\mu;\lambda), v\rangle = \lambda\langle F(\mu), v\rangle\ , \tag{6.32}$$

and F given by (5.40). The abstract problem statement of Section 6.2 is then recovered for

$$\langle \mathcal{A}(\mu)w, v\rangle = \sum_{r=1}^{R} \int_{\Omega^r} \frac{\partial w_i}{\partial x_j} C^r_{ijkl}(\mu) \frac{\partial v_k}{\partial x_l}\, d\Omega \qquad \forall\ w, v \in Y\ , \tag{6.33}$$

$$\langle \mathcal{B}(\mu; z)w, v\rangle = \sum_{r=1}^{R} \int_{\Omega^r} \frac{\partial z_i}{\partial x_j} \tilde{E}^r_{ijkl}(\mu) \frac{\partial w_m}{\partial x_k}\frac{\partial v_m}{\partial x_l}\, d\Omega\ , \forall\ w, v, z \in Y\ , \tag{6.34}$$

where $C^r_{ijkl}(\mu)$ and $\tilde{C}^r_{ijkl}(\mu)$ are given by

$$C^r_{ijkl}(\mu) = G^r_{jj'}(\mu)\ \bar{C}^r_{ij'kl'}\ G^r_{ll'}(\mu)\ \det(\underline{G}^r(\mu))^{-1}\ , \tag{6.35}$$

$$\tilde{C}^r_{ijkl}(\mu) = G^r_{jj'}(\mu)\ \bar{C}^r_{ij'k'l'}\ G^r_{kk'}(\mu) G^r_{ll'}(\mu)\ \det(\underline{G}^r(\mu))^{-1}\ . \tag{6.36}$$

Furthermore,

$$\Theta^{q(i,j,k,l,r)}(\mu) = C^r_{ijkl}(\mu), \qquad \langle \mathcal{A}^{q(i,j,k,l,r)} w, v\rangle = \int_{\Omega^r} \frac{\partial v_i}{\partial x_j}\frac{\partial w_k}{\partial x_l}\, d\Omega\ , \tag{6.37}$$

for $1 \le i,j,k,l \le d$, $1 \le r \le R$, and $q : \{1,\ldots,d\}^4 \times \{1,\ldots,R\} \to \{1,\ldots,Q_{\mathcal{A}}\}$; and

$$\Theta_{\mathcal{B}}^{q'(i,j,k,l,r)}(\mu) = \tilde{C}^r_{ijkl}(\mu), \qquad \langle \mathcal{B}^{q'(i,j,k,l,r)}(z) w, v\rangle = \int_{\Omega^r} \frac{\partial z_i}{\partial x_j}\frac{\partial w_m}{\partial x_k}\frac{\partial v_m}{\partial x_l}\, d\bar{\Omega} \tag{6.38}$$

for $1 \le i,j,k,l \le d$, $1 \le r \le R$, and $q' : \{1,\ldots,d\}^4 \times \{1,\ldots,R\} \to \{1,\ldots,Q_{\mathcal{B}}\}$ (summation over $m = 1,\ldots,d$ is implied). Note the affine decomposition of $F'(\mu;\lambda)$ follows directly from the affine decomposition of $F(\mu)$ in Section 5.3.

6.4 Model Problem

In this section we consider a simple problem upon which we shall apply our reduced-basis approximation and error estimation methods.

6.4.1 Example 10

We again consider the homogeneous rod of Example 6, with Young's modulus $E = 1$ and Poisson's ratio $\nu = 0.25$ in the domain $\bar{\Omega} = (0,1)\times(0,\bar{t})$. The structure is clamped along $\bar{\Gamma}_{\rm D}$, and subjected to a compressive unit load uniformly distributed along $\bar{\Gamma}^n_{\rm N}$; the remainder of the boundary is stress-free. The equilibrium equations are given by

$$\frac{\partial \bar{\sigma}_{ij}}{\partial \bar{x}_j} = 0\ , \qquad \text{in } \bar{\Omega}\ , \tag{6.39}$$

with displacement boundary conditions

$$\bar{u}_i = 0, \quad \text{on } \bar{\Gamma}_{\rm D}\ , \tag{6.40}$$

and traction boundary conditions

$$\bar{\sigma}_{ij} e^n_j = \begin{cases} -\dfrac{1}{\bar{t}} & \text{on } \bar{\Gamma}^n_{\rm N}\ , \\ 0, & \text{on } \bar{\Gamma}\backslash(\bar{\Gamma}_{\rm D}\cup\bar{\Gamma}^n_{\rm N})\ , \end{cases} \qquad \bar{\sigma}_{ij} e^t_j = 0, \quad \text{on } \bar{\Gamma}\backslash(\bar{\Gamma}_{\rm D})\ . \tag{6.41}$$

Our output of interest is the critical load parameter

$$s(\mu) = \lambda^1(\mu) \quad \text{for } \mu = \{\bar{t}\} \in \mathcal{D}^\mu \equiv [0.03125, 0.2]\ . \tag{6.42}$$

Our problem can then be formulated as: given a $\mu \in \mathcal{D}^\mu \subset \mathbb{R}^{P=1}$, find $s(\mu) = \lambda^1(\mu)$ where

$\lambda^1(\mu)$ is the eigenvalue with the smallest magnitude satisfying

$$\langle \bar{\mathcal{A}}\bar{\xi}, \bar{v}\rangle = \lambda(\mu)\langle \bar{\mathcal{B}}(\bar{u})\bar{\xi}, \bar{v}\rangle \ , \qquad \forall\ \bar{v} \in \bar{Y} \ , \tag{6.43}$$

where $\bar{u} \in \bar{Y} = \{\bar{v} \in (H_1(\bar{\Omega}))^2 \mid \bar{v} = 0 \text{ on } \bar{\Gamma}_{\mathrm{D}}\}$ is the solution to

$$\langle \bar{\mathcal{A}}\bar{u}, \bar{v}\rangle = \langle \bar{F}'(1), \bar{v}\rangle, \quad \forall\ \bar{v} \in \bar{Y} \ ; \tag{6.44}$$

here, $\bar{\mathcal{A}}$ is given by (6.8), and $\bar{F}'$ is given by

$$\langle \bar{F}'(\lambda), \bar{v}\rangle = \lambda \left(-\frac{1}{\bar{t}} \int_{\bar{\Gamma}_{\mathrm{N}}^n} \bar{v}_1 \, d\bar{\Gamma} \right), \quad \forall\ \bar{v} \in \bar{Y} \ . \tag{6.45}$$

Finite Element Solution

We present in Figure 6-1 the buckling modes (eigenvectors) corresponding to the smallest three eigenvalues of (6.43) calculated using a finite element approximation with $\mathcal{N} = 600$. We also present in Table 6.1 the critical load parameter obtained by a finite element approximation of (6.43) for $\bar{t} = 0.1$. The results show good qualitative and quantitative agreement with the analytical results obtained using beam theory (for which $\lambda^1(\mu = 0.1) = \pi^2 EI/4L^2 = 2.056 \times 10^{-4}$) (see, for example, [44]).

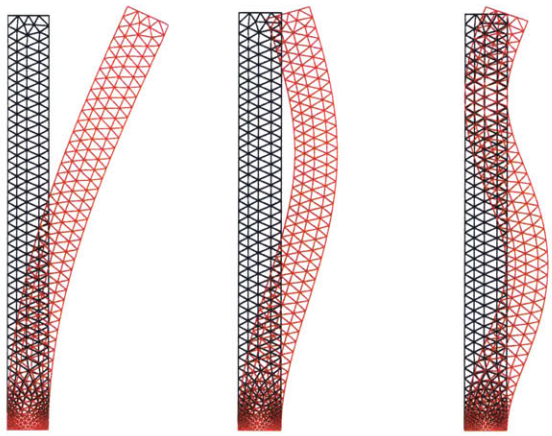


Figure 6-1: Buckling modes corresponding to the smallest three eigenvalues (Example 10).

Element size	λ^1
0.250	4.54×10^{-4}
0.150	2.47×10^{-4}
0.100	2.45×10^{-4}
0.075	2.20×10^{-4}

Table 6.1: Critical load parameter $\lambda^1(\mu = 0.1)$ obtained using a finite element approximation (Example 10).

Note the smallest eigenvalues are "uncontaminated" by the continuous spectrum induced by the partial differential nature of the problem. Furthermore, inclusion of the higher order terms in

(6.26) seem to affect only the continuous part of the spectrum, and the latter seems well separated from the smallest discrete eigenvalues. This is illustrated in Figure 6-2, in which the eigenvalues calculated using (6.26) (with higher order terms) are compared with those calculated using (6.27) (without the higher order terms).

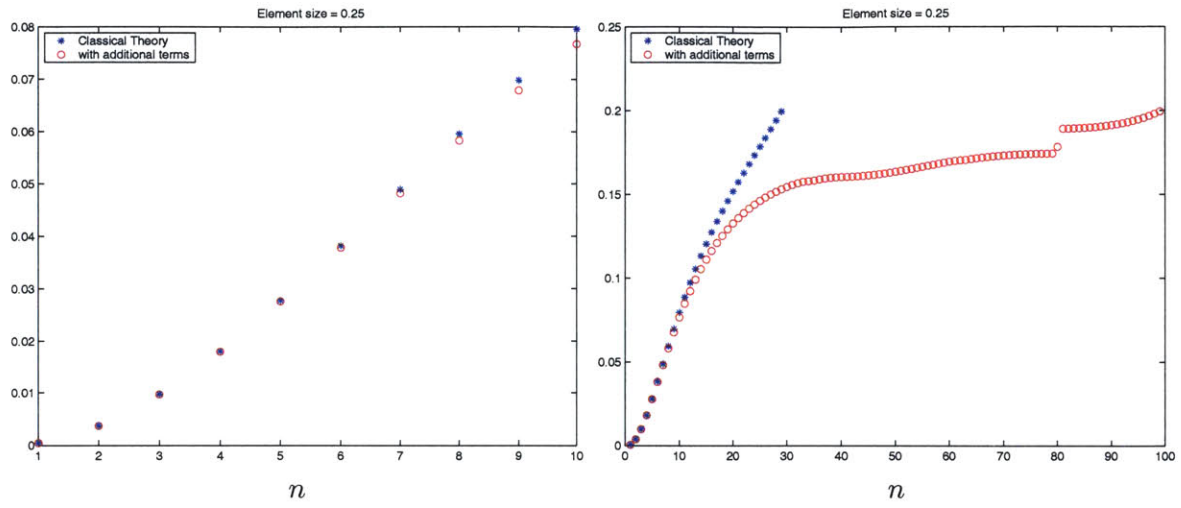


Figure 6-2: Plot of the eigenvalues λ^n, $n = 1, \ldots$, showing the effect of higher order terms on the spectrum.

Reduction to Abstract Form

Upon affine mapping onto a reference domain $\Omega = (0,1) \times (0, t^0)$, our problem may be reformulated as: find $s(\mu) = \lambda^1(\mu)$, the smallest eigenvalue satisfying

$$\langle \mathcal{A}(\mu)\xi(\mu), v\rangle = \lambda(\mu)\langle \mathcal{B}(u(\mu);\mu)\xi(\mu), v\rangle \ , \qquad \forall\ v \in Y \ , \tag{6.46}$$

where $u(\mu) \in Y = \{v \in (H_1(\Omega))^2 \mid v = 0 \text{ on } \Gamma_{\rm D}\}$ is the solution to

$$\langle \mathcal{A}(\mu)u(\mu), v\rangle = \langle F'(\mu;\lambda = 1), v\rangle, \quad \forall\ v \in Y \ ; \tag{6.47}$$

here, $\mathcal{A}(\mu)$ is given by (6.33), $\mathcal{B}(\mu) = 8$ is given by (6.34), and $\bar{F}'(\lambda)$ is given by

$$\langle F'(\mu;\lambda), v\rangle = \lambda\left(-\frac{1}{t^0}\int_{\Gamma_{\rm N}^n} v_1\, d\Gamma\right), \quad \forall\ v \in Y \ , \tag{6.48}$$

respectively.

The abstract problem statement of Section 6.2 is then recovered for $P = 1$, $Q_{\mathcal{A}} = 3$, $Q_F = 1$,

$$\Theta^1(\mu) = 1, \qquad \Theta^2(\mu) = \bar{t}, \qquad \Theta^3(\mu) = \frac{1}{\bar{t}}, \tag{6.49}$$

$$\langle \mathcal{A}^1 w, v\rangle = c_1 \int_\Omega \left(\frac{\partial v_1}{\partial x_1}\frac{\partial w_2}{\partial x_2} + \frac{\partial v_2}{\partial x_2}\frac{\partial w_1}{\partial x_1} \right) d\bar{\Omega} + c_2 \int_\Omega \left(\frac{\partial v_1}{\partial x_2}\frac{\partial w_2}{\partial x_1} + \frac{\partial v_2}{\partial x_1}\frac{\partial w_1}{\partial x_2} \right) d\bar{\Omega} \ , \quad (6.50)$$

$$\langle \mathcal{A}^2 w, v\rangle = (c_1 + 2c_2) \int_\Omega \frac{\partial v_1}{\partial x_1}\frac{\partial w_1}{\partial x_1}\, d\bar{\Omega} + c_2 \int_\Omega \frac{\partial v_2}{\partial x_1}\frac{\partial w_2}{\partial x_1}\, d\bar{\Omega} \ , \quad (6.51)$$

$$\langle \mathcal{A}^3 w, v\rangle = (c_1 + 2c_2) \int_\Omega \frac{\partial v_2}{\partial x_2}\frac{\partial w_2}{\partial x_2}\, d\bar{\Omega} + c_2 \int_\Omega \frac{\partial v_1}{\partial x_2}\frac{\partial w_1}{\partial x_2}\, d\bar{\Omega} \ . \quad (6.52)$$

In the case of the operator $\mathcal{B}$, the effective elasticity tensor $\tilde{C}_{ijkl}(\mu)$ is no longer symmetric. However, for this example (in which $\underline{G}(\mu)$ is diagonal), $\tilde{C}_{ijkl}(\mu)$ maintains the sparsity structure of $\bar{C}_{ijkl}(\mu)$; we therefore obtain $Q_{\mathcal{B}} = 10$ (the number of nonzero entries in $\tilde{C}_{ijkl}(\mu)$), and the $\mathcal{B}^q$ are the corresponding operators. (Note that for general $\underline{G}(\mu)$, $\tilde{C}_{ijkl}(\mu)$ is neither sparse nor symmetric, and $Q_{\mathcal{B}} = d^2 = 16$ for two-dimensional problems.)

6.5 Reduced-Basis Output Approximation

We consider here the extension of our reduced-basis approach to the eigenvalue problem described by Section 6.2. We recall that $\mathcal{A}(\mu)$ is continuous, coercive, symmetric, and affine in μ; and that $\mathcal{B}(\mu; w)$ is continous, symmetric, positive semi-definite, affine in μ, and linear in w.

6.5.1 Approximation Space

We again sample our design space $\mathcal{D}^\mu$ to create the parameter sample

$$S_N^\mu = \{\mu_1, \ldots, \mu_N\} \ ; \quad (6.53)$$

we then introduce the reduced-basis spaces

$$W_N^u = \text{span}\{\zeta_n = u(\mu_n), n = 1, \ldots, N\} \quad (6.54)$$

$$W_N^\xi = \text{span}\{\phi_n = \xi_1(\mu_n), n = 1, \ldots, N\} \quad (6.55)$$

where we recall $\xi_1(\mu)$ is the eigenfunction associated with $\lambda^1(\mu)$.

Our reduced-order approximation is then: for any $\mu \in \mathcal{D}^\mu$, find $u_N(\mu) \in W_N^u$ such that

$$\langle \mathcal{A} u_N(\mu), v\rangle = \langle F(\mu), v\rangle \ , \qquad \forall\ v \in W_N^u \ , \quad (6.56)$$

and $(\lambda_N(\mu), \xi_N(\mu))$ such that

$$\langle \mathcal{A}(\mu)\xi_N(\mu), v\rangle = \lambda_N(\mu)\langle \mathcal{B}(\mu; u_N(\mu))\xi_N(\mu), v\rangle \ , \qquad \forall\ v \in W_N^\xi \ ; \quad (6.57)$$

the output approximation is then

$$s_N(\mu) = \lambda_N^1(\mu) \ . \quad (6.58)$$

The dimensions of the reduced-basis spaces for $u(\mu)$ and $\xi(\mu)$ need not be the same. We note that the solution to (6.57) is the Galerkin projection onto W_N^ξ of the solution to

$$\langle \mathcal{A}(\mu)\xi_N'(\mu), v\rangle = \lambda_N'(\mu)\langle \mathcal{B}(\mu; u_N(\mu))\xi_N'(\mu), v\rangle \ , \qquad \forall\ v \in Y^\xi \ ; \quad (6.59)$$

we have effectively approximated $\mathcal{B}(\mu; u(\mu))$ in (6.5) with $\mathcal{B}(\mu; u_N(\mu))$ — $\lambda_N(\mu)$ is therefore an

approximation to $\lambda'_N(\mu)$, which in turn approaches $\lambda(\mu)$ only as $u_N(\mu) \to u(\mu)$ (and therefore $N \to \infty$).

We also note that $\lambda(\mu)$ may, in general, be negative since $\mathcal{B}$ is not strictly positive-definite. For instance, structures under pure shear may have "negative" critical load parameters — this implies that the buckling occurs when a load reaches the critical value and is applied in the direction opposite to that originally prescribed. In this case, it is not known *a priori* whether the critical load parameter is positive or negative; we must therefore approximate both λ^{1+} and λ^{1-}, the (respectively) positive and negative eigenvalues with the smallest magnitude. We then introduce (in place of W_N^ξ) the reduced-basis spaces

$$W_N^{\xi\pm} = \text{span}\{\phi_n^\pm = \xi^{1\pm}(\mu_n), n = 1, \ldots, N\} \ , \tag{6.60}$$

and calculate $(\lambda_N^{1\pm}(\mu), \xi_N^{1\pm}(\mu))$ such that

$$\langle \mathcal{A}(\mu)\xi_N^\pm(\mu), v\rangle = \lambda_N^\pm(\mu)\langle \mathcal{B}(\mu; u_N(\mu))\xi_N(\mu), v\rangle \ , \qquad \forall\, v \in W_N^\xi \ ; \tag{6.61}$$

the output approximation is then

$$s_N(\mu) = \lambda_N^{1\pm}(\mu) \ . \tag{6.62}$$

6.5.2 Offline/Online Computational Decomposition

We now develop off-line/on-line computational procedures that exploit this dimension reduction.

We first express our approximation $u_N(\mu) \in Y$ to be a linear combination of the appropriate basis functions,

$$u_N(\mu) = \sum_{j=1}^N u_{N\,j}(\mu)\, \zeta_j \tag{6.63}$$

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 (6.56) yields the desired algebraic equations for $\underline{u}_N(\mu) \in \mathbb{R}^N$,

$$\underline{A}_N^u(\mu)\, \underline{u}_N(\mu) \ = \ \underline{F}_N \ . \tag{6.64}$$

Here $\underline{A}_N^u(\mu) \in \mathbb{R}^{N\times N}$ is the symmetric positive-definite (SPD) matrix with entries

$$A_{N\,i,j}^u(\mu) \equiv \langle \mathcal{A}(\mu)\zeta_j, \zeta_i\rangle, \quad 1 \le i, j \le N \ , \tag{6.65}$$

and $\underline{F}_N \in \mathbb{R}^N$ is the "load" vector with entries

$$F_{N\,i} \equiv \langle F(\mu), \zeta_i\rangle, \quad 1 \le i \le N \ . \tag{6.66}$$

We also express our approximation $\xi_N(\mu) \in Y$ to be a linear combination of the corresponding basis functions,

$$\xi_N(\mu) = \sum_{j=1}^N \xi_{N\,j}(\mu)\, \phi_j \ , \tag{6.67}$$

where $\underline{\xi}_N(\mu) \in \mathbb{R}^N$. Then, in (6.57), we choose for test functions $v = \phi_i$, $i = 1, \ldots, N$, and approximate $u(\mu)$ by $u_N(\mu)$, yielding

$$\underline{A}_N^\xi(\mu)\, \underline{\xi}_N(\mu) \ = \ \underline{B}_N(\mu; \underline{u}_N(\mu)) \ . \tag{6.68}$$

Here $\underline{A}_N^{\xi}(\mu) \in \mathbb{R}^{N\times N}$ is the symmetric positive-definite (SPD) matrix with entries

$$A^{\xi}_{N\,i,j}(\mu) \equiv \langle \mathcal{A}(\mu)\phi_j, \phi_i\rangle, \quad 1 \le i,j \le N \ , \tag{6.69}$$

and $\underline{B}_N(\mu;\underline{u}_N(\mu)) \in \mathbb{R}^{N\times N}$ is the symmetric positive semi-definite (SPSD) matrix with entries

$$\begin{aligned} B_{N\,i,j}(\mu;\underline{u}_N(\mu)) &\equiv \langle \mathcal{B}(\mu;u_N(\mu))\phi_j, \phi_i\rangle, & (6.70)\\ &= \sum_{n=1}^{N} u_{N\,n}(\mu)\langle \mathcal{B}(\mu;\zeta_n)\phi_j, \phi_i\rangle, \quad 1 \le i,j \le N \ . & (6.71) \end{aligned}$$

We now invoke (6.1)-(6.3) to write

$$\begin{aligned} A^{u}_{N\,i,j}(\mu) &= \sum_{q=1}^{Q_{\mathcal{A}}} \Theta^q(\mu)\ \langle \mathcal{A}^q\zeta_j, \zeta_i\rangle \ , & (6.72)\\ A^{\xi}_{N\,i,j}(\mu) &= \sum_{q=1}^{Q_{\mathcal{A}}} \Theta^q(\mu)\ \langle \mathcal{A}^q\phi_j, \phi_i\rangle \ , & (6.73)\\ B_{N\,i,j}(\mu;\underline{u}_N(\mu)) &= \sum_{n=1}^{N}\sum_{q=1}^{Q_{\mathcal{B}}} u_{N\,n}(\mu)\ \Theta^q_{\mathcal{B}}(\mu)\ \langle \mathcal{B}^q(\mu;\zeta_n)\phi_j, \phi_i\rangle & (6.74)\\ F_{N\,i}(\mu) &= \sum_{q=1}^{Q_F} \varphi^q_F(\mu)\ \langle F^q, \zeta_i\rangle \ ; & (6.75) \end{aligned}$$

or, more concisely,

$$\begin{aligned} \underline{A}^{u}_N(\mu) &= \sum_{q=1}^{Q_{\mathcal{A}}} \Theta^q(\mu)\ \underline{A}^{u\,q}_N \ , & (6.76)\\ \underline{A}^{\xi}_N(\mu) &= \sum_{q=1}^{Q_{\mathcal{A}}} \Theta^q(\mu)\ \underline{A}^{\xi\,q}_N \ , & (6.77)\\ & & (6.78) \end{aligned}$$

$$\begin{aligned} \underline{B}_N(\mu;\underline{u}_N(\mu)) &= \sum_{n=1}^{N}\sum_{q=1}^{\mathcal{B}} \Theta^q_{\mathcal{B}}(\mu)\ u_{N\,n}(\mu)\underline{B}^{qn}_N \ , & (6.79)\\ \underline{F}_N(\mu) &= \sum_{q=1}^{Q_F} \alpha^q_F(\mu)\ \underline{F}^q_N \ , & (6.80) \end{aligned}$$

where the $\underline{A}_N^{u\,q} \in \mathbb{R}^{N\times N}$, $\underline{A}_N^{\xi\,q} \in \mathbb{R}^{N\times N}$, $\underline{B}_N^{qn} \in \mathbb{R}^{N\times N}$, and $\underline{F}_N^q \in \mathbb{R}^N$, are given by

$$A_{N\,i,j}^{u\,q} = \langle \mathcal{A}^q \zeta_j, \zeta_i \rangle , \quad 1 \le i,j \le N,\ 1 \le q \le Q_{\mathcal{A}} , \tag{6.81}$$

$$A_{N\,i,j}^{\xi\,q} = \langle \mathcal{A}^q \phi_j, \phi_i \rangle , \quad 1 \le i,j \le N,\ 1 \le q \le Q_{\mathcal{A}} , \tag{6.82}$$

$$B_{N\,i,j}^{qn} = \langle \mathcal{B}^q(\zeta_n)\phi_j, \phi_i \rangle , \quad 1 \le i,j,n \le N,\ 1 \le q \le Q_{\mathcal{B}} , \tag{6.83}$$

$$F_{N\,i}^q = \langle F^q, \zeta_i \rangle , \quad 1 \le i \le N,\ 1 \le q \le Q_F , \tag{6.84}$$

Hence, in the *off-line* stage, we compute the $u(\mu_n)$ and $\xi(\mu_n)$, and form the $\underline{A}_N^{u\,q}$, $\underline{A}_N^{\xi\,q}$, $\underline{B}_N^{qn}$, and $\underline{F}_N$: this requires N (expensive) finite element solutions of the equilibrium equation and the generalized eigenvalue problem, as well as $O(Q_{\mathcal{B}}N^3) + O(2Q_{\mathcal{A}}N^2) + O(Q_F N)$ finite-element-vector inner products. In the *on-line* stage, for any given new μ, we first form $\underline{A}_N^u(\mu)$ and $\underline{F}_N(\mu)$ from (6.76) and (6.80), respectively, then solve (6.64) for $\underline{u}_N(\mu)$; we subsequently form $\underline{A}_N^\xi(\mu)$ and $\underline{B}_N(\mu;\underline{u}_N(\mu))$ from (6.77)-(6.79), then solve (6.68) for $\underline{\xi}_N(\mu)$, and finally evaluate $s_N(\mu) = \lambda_N^1(\mu)$: this requires $O(Q_{\mathcal{B}}N^3) + O(Q_{\mathcal{A}}N^2) + O(Q_F N) + O(\frac{2}{3}N^3) + O(N^3)$ operations and $O(Q_{\mathcal{B}}N^3) + O(Q_{\mathcal{A}}N^2) + O(Q_F N)$ storage.

N	$\max \lvert s(\mu) - s_N(\mu)\rvert / s(\mu)$
3	0.129
4	0.055
5	0.053
6	0.054
7	0.042

Table 6.2: Maximum error in the reduced-basis approximation for the critical load parameter calculated over 20 samples (Example 10).

We present in Table 6.2 the maximum error in the reduced-basis approximation for the critical load parameter for Example 10. The convergence for this simple problem is rather slow, perhaps due to the additional error introduced by approximating $\mathcal{B}(\mu;u(\mu))$ with $\mathcal{B}(\mu;u_N(\mu))$. Note also that the error does not necessarily decrease monotonically with N.

6.6 A Posteriori Error Estimation (Method II)

Due to the strong nonlinearity of the eigenvalue problem (6.68), we consider only Method II error estimators. We thus set $M > N$, and introduce a parameter sample

$$S_M^\mu = \{\mu_1, \ldots, \mu_M\} \tag{6.85}$$

and associated reduced–basis approximation spaces

$$W_M^u = \text{span}\,\{\zeta_m \equiv u(\mu_m),\ m = 1,\ldots,M\} , \tag{6.86}$$

$$W_M^\xi = \text{span}\,\{\phi_m \equiv u(\mu_m),\ m = 1,\ldots,M\} ; \tag{6.87}$$

we again require $S_N^\mu \subset S_M^\mu$ and therefore $W_N^u \subset W_M^u$, $W_N^\xi \subset W_M^\xi$. We first find $u_M(\mu) \in W_M^u$ such that

$$\langle \mathcal{A}(\mu)u_M(\mu), v \rangle = \langle F, v \rangle, \quad \forall\ v \in W_M^u , \tag{6.88}$$

and $(\lambda_M(\mu), \xi_M(\mu)) \in \mathbb{R} \times W_M^\xi$, $i = 1, \ldots, M$, such that

$$\langle \mathcal{A}(\mu)\xi_M(\mu), v\rangle = \lambda_M(\mu)\langle \mathcal{B}(\mu; u_M(\mu))\xi_M(\mu), v\rangle, \quad \forall\, v \in W_M^\xi \ . \tag{6.89}$$

Letting

$$s_M(\mu) = \lambda_M^1(\mu) \tag{6.90}$$

we then evaluate the "improved" estimator $\bar{s}_N(\mu)$ given by

$$\bar{s}_N(\mu) = s_N(\mu) - \frac{1}{2\tau}\left(s_M(\mu) - s_N(\mu)\right) \ ; \tag{6.91}$$

our lower and upper output estimators are then

$$s_N^-(\mu) = \bar{s}_N(\mu) - \frac{1}{2}\Delta_N(\mu) \tag{6.92}$$

$$s_N^+(\mu) = \bar{s}_N(\mu) + \frac{1}{2}\Delta_N(\mu) \tag{6.93}$$

where $\Delta_N(\mu)$, the estimator for the error $|s(\mu) - \bar{s}_N(\mu)|$, is given by

$$\Delta_{N,M}(\mu) = \frac{1}{\tau}\left|s_M(\mu) - s_N(\mu)\right| \tag{6.94}$$

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

$$\eta_{N,M}(\mu) = \frac{\Delta_{N,M}(\mu)}{s(\mu) - s_N(\mu)} \ . \tag{6.95}$$

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

N	$\min \eta_N(\mu)$	$\max \eta_N(\mu)$
1	1.9773	2.0099
2	1.7067	2.3395
3	1.8811	2.8151
4	1.9772	2.2783

Table 6.3: Minimum and maximum effectivities for Method II error estimators (with $\tau = 0.5$) calculated over 20 samples (Example 10).

The analysis of the bounding properties of Method II estimators presented in Section 5.8.2 also applies here. However, the necessary hypothesis (4.190), while still plausible, is now *less* so than in Chapters 4 and 5. Nevertheless, the effectivity results presented in Table 6.3 show that $\eta_N(\mu)$ still tends to $1/\tau$.

Chapter 7

Reduced-Basis Methods for Analysis, Optimization, and Prognosis: A MicroTruss Example

7.1 Introduction

In Chapters 3 through 6 we present the reduced-basis method — a technique for the rapid and reliable prediction of linear-functional outputs of partial differential equations with affine parameter dependence. Reduced–basis output bound methods are intended to render partial–differential-equation solutions truly useful: essentially real–time as regards operation count; "blackbox" as regards reliability; and directly relevant as regards the (limited) input–output data required.

In this chapter we revisit the microtruss example presented in Chapter 1, focusing on the structural (solid mechanics) aspects of the problem. We apply reduced-basis methods to the prediction of the relevant outputs of interest; we then illustrate how these methods directly enable rapid and reliable solution of "real" optimization problems — even in the presence of significant uncertainty. In particular, we employ reduced-basis output bounds in the context of (*i*) design — optimizing a system at conception with respect to prescribed objectives, constraints, and worst-case environmental conditions — and (*ii*) prognosis — optimizing a system during operation subject to evolving system characteristics, dynamic system requirements, and changing environmental conditions.

7.2 Formulation

We begin with the formulation of the microtruss problem, focusing particularly on the structural aspects. In Sections 7.2.1 and 7.2.2, the circumflex ($\hat{}$) denotes dimensional quantities, and indices take the values $1, \ldots, d$.

7.2.1 Dimensional Formulation

We consider the periodic open cellular structure shown in Figure 7-1. As indicated in Chapter 1, the structure is simultaneously designed for heat-transfer as well as structural capability; we shall focus on the latter. The prismatic microtruss consists of a frame (upper, lower, and lateral faces) and a core of trusses. The structure transmits a force per unit depth $\hat{F}$ uniformly distributed over the tip $\hat{\Gamma}_{\rm N}^t$ through the truss system to the fixed left wall $\hat{\Gamma}_{\rm D}$. We assume that the structure

is sufficiently deep such that a physical model of plane strain (two-dimensional) linear elasticity suffices.

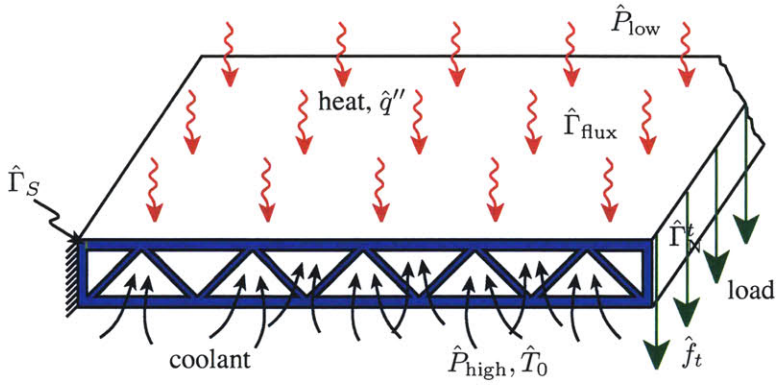


Figure 7-1: A microtruss structure.

The topology of the structure (see Figure 7-2) is characterized by several geometric parameters: the thickness $\hat{t}_t$ of the top frame, the thickness $\hat{t}_b$ of the bottom and side frames, the thickness $\hat{t}$ of the core trusses, the separation $\hat{H}$ between the top and bottom frames, the angle $\hat{\alpha}$ between the trusses and the frame, and the depth $\hat{L}$ of the structure. The structure is further characterized by the Young's modulus $\hat{C}$ and Poisson's ratio ν.

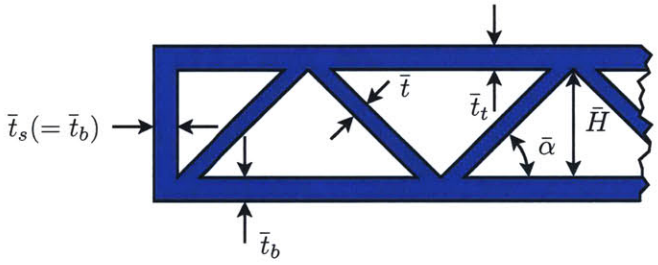


Figure 7-2: Geometry

Assuming that $\hat{H}/\hat{L} \ll 1$, the displacement $\hat{u}$ satisfies the partial differential equation

$$\frac{\partial}{\partial \hat{x}_j}\left(\hat{C}_{ijkl}\frac{\partial \hat{u}_k}{\partial \hat{x}_l}\right) = 0, \qquad \text{in } \hat{\Omega}, \tag{7.1}$$

where $\hat{\Omega}$ denotes the truss domain (see Figure 7-2), and $\hat{C}_{ijkl}$ is the elasticity tensor. The displacement and traction boundary conditions are

$$\hat{u}_i = 0\ , \qquad \text{on } \hat{\Gamma}_{\text{D}}\ , \tag{7.2}$$

and

$$\hat{\sigma}_{ij}\hat{\text{e}}^t_j = \begin{cases} -\hat{f}\,\hat{\text{e}}^t_i & \text{on } \hat{\Gamma}^t_{\text{N}} \\ 0 & \text{on } \hat{\Gamma}\backslash(\hat{\Gamma}_{\text{D}} \cup \hat{\Gamma}^t_{\text{N}})\ , \end{cases} \qquad \hat{\sigma}_{ij}\hat{\text{e}}^n_j = 0 \qquad \text{on } \hat{\Gamma}\backslash\hat{\Gamma}_{\text{D}} \tag{7.3}$$

respectively, where $\hat{\text{e}}^n_i$ and $\hat{\text{e}}^t_i$ are the unit normal and tangent vectors on the boundary $\hat{\Gamma}$, and

$\hat{\underline{e}}^t = (0,1)$ on $\hat{\Gamma}^t_{\rm N}$.

7.2.2 Nondimensional Formulation

We now introduce the following nondimensional quantities:

$$\bar{C}_{ijkl} = \frac{1}{\hat{E}}\hat{C}_{ijkl} \tag{7.4}$$

$$\bar{x} = \frac{1}{\hat{H}}\hat{x} \tag{7.5}$$

$$\bar{f}_t = \frac{\hat{H}}{\hat{F}}\hat{f}_t \tag{7.6}$$

$$\bar{u} = \frac{\hat{E}}{\hat{F}}\hat{u}\ ; \tag{7.7}$$

in addition, we let $\bar{t}_t$, $\bar{t}_b$, and $\bar{t}$, denote the thicknesses of the top frame, bottom frame (and sides), and core trusses, respectively, nondimensionalized relative to the separation $\hat{H}$. We also let $\bar{H}$ ($\equiv 1$) denote the separation between the top and bottom frames; note that $\bar{\alpha} = \hat{\alpha}$.

The nondimensional displacement $\bar{u}$ then satisfies the partial differential equation

$$\frac{\partial}{\partial \bar{x}_j}\left(\bar{C}_{ijkl}\frac{\partial \bar{u}_k}{\partial \bar{x}_l}\right) = 0, \qquad \text{in } \bar{\Omega}, \tag{7.8}$$

where $\bar{\Omega}$ denotes the truss domain. The displacement and traction boundary conditions are

$$\bar{u}_i = 0\ , \qquad \text{on } \bar{\Gamma}_{\rm D}\ , \tag{7.9}$$

and

$$\bar{\sigma}_{ij}\bar{\mathrm{e}}^t_j = \begin{cases} -\bar{f}_t\bar{\mathrm{e}}^t_i & \text{on } \bar{\Gamma}^t_{\rm N} \\ 0 & \text{on } \bar{\Gamma}\backslash(\bar{\Gamma}_{\rm D}\cup\bar{\Gamma}^t_{\rm N})\ , \end{cases} \qquad \bar{\sigma}_{ij}\bar{\mathrm{e}}^n_j = 0 \qquad \text{on } \bar{\Gamma}\backslash\bar{\Gamma}_{\rm D}\ , \tag{7.10}$$

respectively, where $\bar{\mathrm{e}}^n_i$ and $\bar{\mathrm{e}}^t_i$ are the unit normal and tangent vectors on $\bar{\Gamma}$, and $\bar{\underline{e}}^t = (0,1)$ on $\bar{\Gamma}^t_{\rm N}$.

Our problem is then: given $\mu = \{\bar{t}_t, \bar{t}_b, \bar{t}, \bar{\alpha}\}$, find the average deflection along $\Gamma^t_{\rm N}$,

$$\delta_{\rm ave}(\mu) = -\frac{1}{|\bar{\Gamma}^t_{\rm N}|}\int_{\bar{\Gamma}^t_{\rm N}} \bar{u}_2\, d\bar{\Gamma}\ , \tag{7.11}$$

the average normal stress along Γ_S,

$$\sigma_{\rm ave}(\mu) = -\frac{1}{\bar{t}_{\rm top}}\int_{\bar{\Gamma}_S} \bar{\sigma}_{11}(\bar{u})\, d\bar{\Omega}, \tag{7.12}$$

and the critical load parameter $\lambda^1(\mu)$. Here, $\bar{u}$ is the solution to

$$\langle \bar{\mathcal{A}}\bar{u}, \bar{v}\rangle = \langle \bar{F}'(1), \bar{v}\rangle\ , \quad \forall \bar{v} \in \bar{Y}\ , \tag{7.13}$$

and $\lambda^1(\mu)$ is the smallest eigenvalue of

$$\langle \bar{\mathcal{A}}\bar{\xi}, \bar{v}\rangle = \lambda(\mu)\langle \bar{\mathcal{B}}(\bar{u})\bar{\xi}, \bar{v}\rangle\ , \qquad \forall\ \bar{v} \in \bar{Y}\ , \tag{7.14}$$

where

$$\langle \bar{\mathcal{A}}\bar{w}, \bar{v}\rangle = \int_{\bar{\Omega}} \frac{\partial \bar{v}_i}{\partial \bar{x}_j} \bar{C}_{ijkl} \frac{\partial \bar{w}_k}{\partial \bar{x}_l}\, d\bar{\Omega}\ , \tag{7.15}$$

$$\langle \bar{\mathcal{B}}(\bar{z})\bar{w}, \bar{v}\rangle = -\int_{\bar{\Omega}} \frac{\partial \bar{z}_i}{\partial \bar{x}_j} \bar{C}_{ijkl} \frac{\partial \bar{w}_m}{\partial \bar{x}_k} \frac{\partial \bar{v}_m}{\partial \bar{x}_l}\, d\bar{\Omega}\ , \tag{7.16}$$

$$\langle \bar{F}(\lambda), \bar{v}\rangle = \lambda \left(-\int_{\bar{\Gamma}_{\mathrm{N}}^{\mathrm{t}}} \bar{f}_t v_2\, d\bar{\Gamma} \right)\ . \tag{7.17}$$

7.2.3 Reduction to Abstract Form

In this section, we reformulate the problem defined by (7.11)-(7.17) so as to obtain the affine parameter-dependence required by the reduced-basis method.

Affine Geometric Mapping

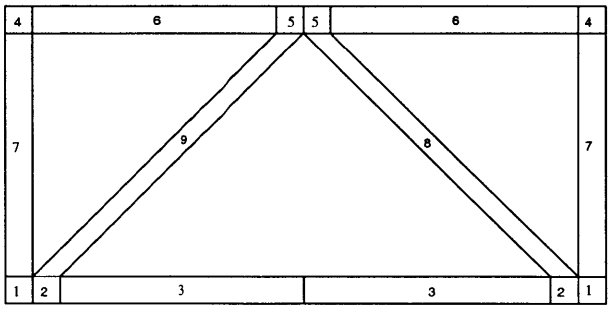


Figure 7-3: Subdomains and Reference Domain

We now partition the domain $\bar{\Omega} \equiv \bar{\Omega}(\mu)$ into $R = 9$ subdomains $\bar{\Omega}^r$ as shown in Figure 7-3. We take for the reference domain $\Omega = \bar{\Omega}(\mu_0)$, where $\mu_0 = \{t^0_{\mathrm{top}}, t^0_{\mathrm{bot}}, t^0, \alpha^0\}$. The affine mappings

$\mathcal{G}^r(\mu) : \bar{\Omega}^r \rightarrow \Omega^r$ are then given by $\mathcal{G}^r(\mu;\bar{x}) = \underline{G}^r(\mu)\bar{x} + \underline{b}^r(\mu)$, where

$$\underline{G}^1 = \begin{bmatrix} \frac{t_s^0}{\bar{t}_s} & 0 \\ 0 & \frac{t_{bot}^0}{\bar{t}_{bot}} \end{bmatrix} \qquad \underline{G}^2 = \begin{bmatrix} \frac{t_h^0}{\bar{t}_h} & 0 \\ 0 & \frac{t_{bot}^0}{\bar{t}_{bot}} \end{bmatrix} \qquad \underline{G}^3 = \begin{bmatrix} \frac{L^0}{\bar{L}} & 0 \\ 0 & \frac{t_{bot}^0}{\bar{t}_{bot}} \end{bmatrix} \tag{7.18}$$

$$\underline{G}^4 = \begin{bmatrix} \frac{t_s^0}{\bar{t}_s} & 0 \\ 0 & \frac{t_{top}^0}{\bar{t}_{top}} \end{bmatrix} \qquad \underline{G}^5 = \begin{bmatrix} \frac{t_h^0}{\bar{t}_h} & 0 \\ 0 & \frac{t_{top}^0}{\bar{t}_{top}} \end{bmatrix} \qquad \underline{G}^6 = \begin{bmatrix} \frac{L^0}{\bar{L}} & 0 \\ 0 & \frac{t_{top}^0}{\bar{t}_{top}} \end{bmatrix} \tag{7.19}$$

$$\underline{G}^7 = \begin{bmatrix} \frac{t_s^0}{\bar{t}_s} & 0 \\ 0 & \frac{H^0}{\bar{H}} \end{bmatrix} \tag{7.20}$$

$$\underline{G}^8 = \begin{bmatrix} 1 & -\tan\alpha^0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{t_h^0}{\bar{t}_h} & 0 \\ 0 & \frac{H^0}{\bar{H}} \end{bmatrix} \begin{bmatrix} 1 & \tan\bar{\alpha} \\ 0 & 1 \end{bmatrix} \tag{7.21}$$

$$\underline{G}^9 = \begin{bmatrix} 1 & \tan\alpha^0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{t_h^0}{\bar{t}_h} & 0 \\ 0 & \frac{H^0}{\bar{H}} \end{bmatrix} \begin{bmatrix} 1 & -\tan\bar{\alpha} \\ 0 & 1 \end{bmatrix}, \tag{7.22}$$

where the $\bar{t}_s = \bar{t}_{bot}$ is the thickness of the side trusses; $\bar{t}_h = \bar{t}/\sin\bar{\alpha}$ is the horizontal thickness of the diagonal trusses; and $\bar{L} = \bar{H}/\tan\bar{\alpha}$; similar formulas apply for t_s^0, t_h^0, and L^0.

We recall that the affine decomposition of the linear elasticity operator is related to the number of elements in the elasticity tensor. It would therefore seem that for our microtruss example $Q_{\mathcal{A}}$ might be very large — as large as 144, or the number of elements in $C_{ijkl}(\mu)$ (16) times the number of subdomains (9). However, as illustrated in Examples 1-9 of Chapters 3 and 5, the symmetry and sparsity of the elasticity tensor may be exploited to reduce $Q_{\mathcal{A}}$. In particular, we consider only the nonzero, unique elements of the $\bar{C}_{ijkl}^r(\mu)$, thus reducing $Q_{\mathcal{A}}$ to $[(3\times 7)+(10\times 2)] = 41$; the first term corresponds to the 7 subdomains undergoing only "stretch" transformations ($r = 1,\ldots,7$) while the second term coresponds to the 2 subdomains undergoing both "stretch" and "shear." Furthermore, we can further reduce $Q_{\mathcal{A}}$ by consolidating those terms among all the (unique,nonzero) elements of the $C_{ijkl}^r(\mu)$ which are the same. For example, we consider subdomains 1 through 7 for which the elements of the elasticity tensor are given in Table 7.1, where we $\alpha_x = G_{11}$ and $\alpha_y = G_{22}$ are the diagonal (nonzero) elements of the mapping matrix $\underline{G}$. We then note that $C_{1122}^r(\mu) = c_1$ and $C_{1221}^r(\mu) = c_2$ for $r = 1,\ldots,7$. Taking this into consideration, $Q_{\mathcal{A}}$ is further reduced to $[1 + (2\times 7) + (10\times 2)] = 35$. Further reductions can be obtained by making similar observations for $r = 8, 9$.

Reference Domain Formulation

Using the results obtained in Chapters 5 and 6, our problem may be reformulated as: for any given $\mu = \{\bar{t}_t, \bar{t}_b, \bar{t}, \bar{\alpha}\} \in \mathcal{D}^\mu = [0.22, 2.2]^3 \times [0°, 45°]$, find

$$\delta_{ave}(\mu) = -\frac{1}{\bar{t}_t + \bar{t}_b + \bar{H}} \int_{\Gamma_N^t} \bar{u}_2 |G^{-1}(\mu)\underline{e}^t| \, d\bar{\Gamma}\,, \tag{7.23}$$

$ij\backslash kl$	11	12	21	22
11	$\frac{\alpha_x}{\alpha_y}(c_1+2c_2)$	0	0	$1(c_1)$
12	·	$\frac{\alpha_y}{\alpha_x}(c_2)$	$1(c_2)$	0
21	·	·	$\frac{\alpha_x}{\alpha_y}(c_2)$	0
22	·	·	·	$\frac{\alpha_y}{\alpha_x}(c_1+2c_2)$

Table 7.1: Elements of the effective elasticity tensor for subdomains undergoing a "stretch" transformation.

the average normal stress along Γ_S (near the support),

$$\sigma_{\text{ave}}(\mu) = -\frac{1}{t_{\text{top}}}\int_{\Gamma_S} \sigma_{11}(\bar{u})\, d\bar{\Gamma}, \tag{7.24}$$

and the critical (positive and negative) load parameters, $\lambda^{1\pm}(\mu)$. Here, u is the solution to

$$\langle \mathcal{A}(\mu)u(\mu), v\rangle = \langle F'(\mu;1), v\rangle \ , \quad \forall\ v \in Y \ , \tag{7.25}$$

and $\lambda^{1\pm}(\mu)$ are the smallest positive and negative eigenvalues satisfying

$$\langle \mathcal{A}(\mu)\xi(\mu), v\rangle = \lambda(\mu)\langle \mathcal{B}(u(\mu);\mu)\xi(\mu), v\rangle \ , \qquad \forall\ v \in Y \ , \tag{7.26}$$

where

$$\langle \mathcal{A}(\mu)w, v\rangle = \sum_{r=1}^{R}\int_{\Omega^r} \frac{\partial w_i}{\partial x_j} C^r_{ijkl}(\mu)\frac{\partial v_k}{\partial x_l} \qquad \forall\ w, v \in Y \ , \tag{7.27}$$

$$\langle \mathcal{B}(z)w, v\rangle = -\sum_{r=1}^{R}\int_{\Omega} \frac{\partial z_i}{\partial x_j} C_{ijkl}\frac{\partial w_m}{\partial x_k}\frac{\partial v_m}{\partial x_l}\, d\bar{\Omega} \ , \tag{7.28}$$

$$\langle F'(\mu;\lambda), v\rangle = \lambda\sum_{r=1}^{R}\left(\int_{\Gamma^{t\,r}_{\text{N}}} f^{t\,r}(\mu) v_2\, d\Gamma\right) \ . \tag{7.29}$$

Here $C^r_{ijkl}(\mu) = C^r_{ijkl}(\mu) = G^r_{jj'}(\mu)\ \bar{C}^r_{ij'kl'}\ G^r_{ll'}(\mu)\ \det(\underline{G}^r(\mu))^{-1}$, $f^{t\,r}_i(\mu)$ is given by

$$f^{t\,r}_i(\mu) = \bar{f}_{t\,r}\bar{\mathrm{e}}^t_i \left|(\underline{G}^r(\mu))^{-1}\,\underline{\mathrm{e}}^t\right| \ , \tag{7.30}$$

and $Y(\Omega) = \{v \in \left(H^1(\Omega)\right)^2 \mid v_i \mathrm{e}^n_i = 0 \text{ on } \Gamma^n_{\text{D}},\ v_i\mathrm{e}^t_i = 0 \text{ on } \Gamma^t_{\text{D}}\}$.

7.3 Analysis

In this section we (i) examine the convergence of the reduced-basis approximation and the effectivity of the (Method II) error estimator for the microtruss example; and (ii) investigate the behavior of the microtruss structure using a reduced-basis model, focusing particularly on the dependence of the various outputs of interest on the parameters. Due to the relative complexity of the problem,

application of Method I error estimators is currently not feasible.

7.3.1 Average Deflection

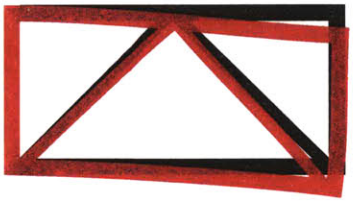


Figure 7-4: Example of a displacement field calculated using a finite element approximation ($\mathcal{N} = 13,000$)

We present in Figure 7-4 the resulting displacement field for several values of the parameter μ calculated using a finite element approximation (with $\mathcal{N} = 13,000$). Quite often in engineering problems, the entire displacement field is not of primary importance, but rather certain characteristics of the solution such as the average deflection. Furthermore, exact (i.e., finite element) solution of the partial differential equations is arguably very expensive, particularly when repeated solution (for many different values of the parameters) is required. The reduced-basis method is thus very attractive.

We present in Tables 7.2 and 7.3 the error in the reduced-basis approximation to the average deflection with respect to N for two representative values of μ. We also present in Table 7.4 the maximum error calculated over 20 randomly selected values of μ. We note that the error is indeed very small even for relatively small N, and observe the rapid convergence of the reduced-basis approximation. However, in practice an approximation is generally insufficient by itself — the approximation must also be certified, and its accuracy assessed. We therefore present in in Table 7.5 the minimum and maximum effectivities (over 20 randomly selected samples) of the Method II error estimator. We note that for $N \geq 10$ the effectivity is greater than unity (signifying valid bounds). Furthermore, since the output is compliant the effectivity is indeed bounded from above by $1/\tau$ (signifying sharp bounds) for all N; we also note that the effectivity approaches $1/\tau$ as N increases.

N	$\frac{\lvert s(\mu)-s_N(\mu)\rvert}{s(\mu)}$	$\frac{\Delta_N(\mu)}{s(\mu)}$	$\eta_N(\mu)$
10	5.49e-02	8.10e-02	1.48
20	1.44e-02	2.48e-02	1.72
30	4.84e-03	9.00e-03	1.86
40	2.00e-03	3.90e-03	1.94
50	9.57e-04	1.84e-03	1.92
60	3.51e-04	6.58e-04	1.87
70	1.19e-04	2.08e-04	1.75
80	5.39e-05	9.20e-05	1.71
90	4.15e-05	7.56e-05	1.82
100	3.38e-05	6.26e-05	1.85

Table 7.2: Error and effectivities for the deflection for $\mu = \mu^0$ (the reference parameter).

N	$\lvert s(\mu) - s_N(\mu)\rvert / s(\mu)$	$\Delta_N(\mu)/s(\mu)$	$\eta_N(\mu)$
10	3.67e-01	7.02e-01	1.91
20	1.64e-02	2.38e-02	1.45
30	7.89e-03	1.30e-02	1.64
40	4.46e-03	8.22e-03	1.84
50	2.16e-03	4.10e-03	1.90
60	1.41e-03	2.70e-03	1.92
70	5.85e-04	1.10e-03	1.89
80	3.50e-04	6.66e-04	1.91
90	1.21e-04	2.22e-04	1.83
100	1.10e-04	2.06e-04	1.88

Table 7.3: Error and effectivities for the deflection for a randomly selected value of μ.

N	$\max(s(\mu) - s_N(\mu))/s(\mu)$
10	4.80e-01
20	2.80e-01
40	1.40e-02
60	6.80e-03
80	1.50e-03
100	1.10e-03

Table 7.4: Maximum error in the deflection calculated over 20 randomly selected values of μ.

N	$\min \eta_N(\mu)$	$\max \eta_N(\mu)$
10	0.66	1.87
20	1.36	1.98
40	1.52	1.97
60	1.26	1.99
80	1.55	2.00
100	1.64	2.00

Table 7.5: Minimum and maximum effectivities for the deflection calculated over 20 randomly selected values of μ.

With a reduced-basis model at our disposal, we can now easily explore the parameter space — i.e., investigate the dependence of the average deflection on the different parameters. In Figure 7-5 we plot the average deflection with respect to $\bar{t}_{\text{top}}$ and $\bar{t}_{\text{bot}}$. While the fact that the deflection decreases with $\bar{t}_{\text{top}}$ and $\bar{t}_{\text{bot}}$ is rather obvious, we also observe the less intuitive result that the deflection is more sensitive to changes in $\bar{t}_{\text{top}}$ than to $\bar{t}_{\text{bot}}$. Furthermore, since the volume (and weight) increase with $\bar{t}_{\text{top}}$ and $\bar{t}_{\text{bot}}$, the results also indicate a trade-off between weight and deflection: the values of $\bar{t}_{\text{top}}$ and $\bar{t}_{\text{bot}}$ which are of greatest interest in design are those occuring at the "knee" of the trade-off curves (around $\bar{t}_{\text{top}} = 0.06$ and $\bar{t}_{\text{bot}} = 0.05$).

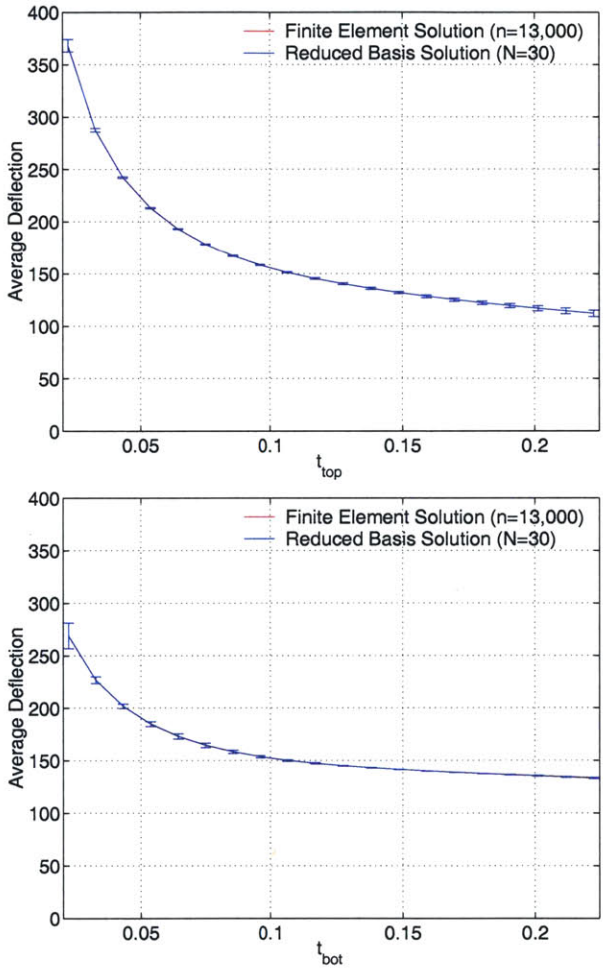


Figure 7-5: Plots of the average deflection as a function of $\bar{t}_{\text{top}}$ and $\bar{t}_{\text{bot}}$.

7.3.2 Average Stress

We now apply the reduced-basis method for noncompliant outputs (see Chapter 5) to the calculation of the stress averaged along Γ_S.

To begin, we apply a scheme similar to that used in Example 9 of Chapter 5 to express (7.24) as a bounded linear functional. We thus choose a function $\chi \in H^1(\Omega')$ such that $\chi|_{\Gamma_S} = 1$, and $\chi|_{\Gamma'_{\text{int}}} = 0$; here, Ω_S and Γ_S and Γ'_{int} are defined as in Figure 7-6. In general, χ must be chosen to

be any member of $H^1(\Omega')$ such that $\chi|_{\Gamma_S} = 1$, $\chi|_{\Gamma'_D} = 0$, and $\chi|_{\Gamma'_{int}} = 0$, where Γ' is the boundary of Ω', $\Gamma_S \subset \Gamma'$ is the surface over which the average stress is to be evaluated, $\Gamma'_D \equiv (\Gamma_D \cap \Gamma')\backslash\Gamma_S$, and $\Gamma'_{int} \equiv (\Gamma_{int} \cap \Gamma')\backslash\Gamma_S$. For simplicity, we shall also define $\chi \equiv 0$ on $\Omega\backslash\Omega'$.

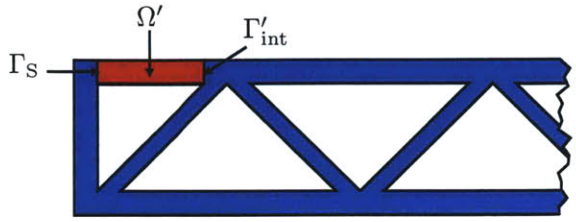


Figure 7-6: Geometry

We then let

$$\langle L(\mu), v\rangle \equiv \frac{1}{t_{top}} \left(\langle \mathcal{A}(\mu)\chi, v\rangle - \langle F(\mu), v\rangle\right) , \qquad \forall\, v \in Y \ . \tag{7.31}$$

Integrating by parts, and using the definition of χ and the boundary conditions, it can be shown that $\sigma_{ave}(\mu) = \langle L(\mu), u(\mu)\rangle$. We note that as in Example 9, $\langle L(\mu), v\rangle$ defined as in (7.31) is a bounded linear functional, while the more obvious alternative in (7.24) is not.

We now take for S_N and W_N a log-random sample and an integrated primal-dual approximation space. We present in Tables 7.6 and 7.7 the error and effectivities at two representative values of μ; we also present in Table 7.8 the maximum error, and in Table 7.9 the minimum and maximum effectivities for the Method II error estimator, calculated over 20 randomly selected values of μ. We note that even in this noncompliant case, the error is again very small even for relatively small N, and the reduced-basis approximation converges rapidly to the exact solution. For the particular values of μ taken in Tables 7.6 and 7.7, we obtain valid output bounds for $N \geq 10$, and the effectivities approach $1/\tau$ as desired. Since the output is noncompliant, the effectivity is no longer bounded from above by $1/\tau$ $(= 2)$; however it still approaches $1/\tau$ as N increases. However, the "worst-case" effectivities in Table 7.9 show that valid bounds are in some cases still not obtained even for $N = 120$. While this can be remedied by simply taking a smaller value for τ, these results reflect the need for developing rigorous (rather than asymptotic) error bounds.

N	$\lvert s(\mu) - s_N(\mu)\rvert / s(\mu)$	$\Delta_N(\mu)/s(\mu)$	$\eta_N(\mu)$
10	5.44e-02	1.09e-01	2.01
20	5.93e-05	2.21e-04	3.72
30	2.09e-04	4.36e-04	2.08
40	1.99e-04	3.69e-04	1.85
50	1.07e-04	1.89e-04	1.78
60	1.43e-05	2.27e-05	1.60
70	2.41e-05	4.57e-05	1.90
80	1.47e-05	2.99e-05	2.03
90	2.01e-05	4.06e-05	2.02
100	1.19e-05	2.43e-05	2.04

Table 7.6: Error and effectivities for the average stress for $\mu = \mu^0$ (the reference parameter).

In Figure 7-7 we plot the average normal stress with respect to $\bar{t}_{top}$. The results also indicate a trade-off between weight and the stress. In this case, the values of $\bar{t}_{top}$ which are of greatest

N	$\lvert s(\mu) - s_N(\mu)\rvert / s(\mu)$	$\Delta_N(\mu)/s(\mu)$	$\eta_N(\mu)$
10	9.84e-02	1.78e-01	1.81
20	9.43e-03	1.87e-02	1.99
30	1.14e-03	2.52e-03	2.20
40	6.85e-05	1.75e-04	2.55
50	1.18e-04	2.63e-04	2.24
60	1.20e-04	2.36e-04	1.96
70	4.61e-05	9.34e-05	2.03
80	1.58e-05	3.46e-05	2.19
90	1.73e-05	3.57e-05	2.07
100	1.40e-05	2.80e-05	1.99

Table 7.7: Error and effectivities for the average stress for a randomly selected value of μ.

N	$\lvert s(\mu) - s_N(\mu)\rvert / s(\mu)$
10	8.40e-01
20	2.50e-02
40	2.10e-02
60	2.30e-03
80	4.50e-04
100	1.90e-04
120	1.10e-04
140	3.97e-05
160	3.21e-05
180	3.22e-05
200	2.66e-05

Table 7.8: Maximum error in the average stress calculated over 20 randomly selected values of μ.

N	$\min \eta_N(\mu)$	$\max \eta_N(\mu)$
10	0.55	2.14
20	0.03	4.94
40	0.58	12.57
60	0.31	4.86
80	1.57	3.34
100	0.09	24.70
120	0.62	5.18
140	0.30	3.49
160	1.35	7.64
180	1.01	2.82
200	1.47	5.69

Table 7.9: Minimum and maximum effectivities for the average stress calculated over 20 randomly selected values of μ.

interest in design are those occuring at the "knee" of the trade-off curve (around $\bar{t}_{\text{top}} = 0.06$); taking $\bar{t}_{\text{top}} < 0.06$ decreases the weight but causes a rapid increase in the stress, and taking $\bar{t}_{\text{top}} > 0.06$ increases the weight but with very little effect on the stress.

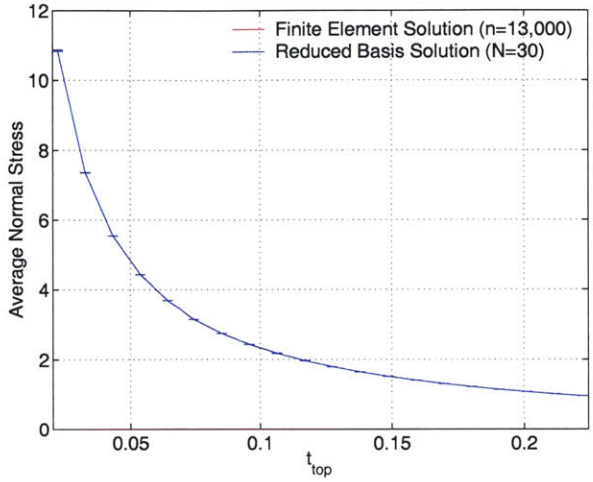


Figure 7-7: Plot of average stress as a function of $\bar{t}_{\text{top}}$.

Note that calculating the average stress at certain critical points in the structure allows us to ensure that the material does not yield. Predictions of the average stress may also be used in a stability analysis along with analytical formulas for the critical loads of beam-columns. However, such an analysis takes into account only *local* buckling modes within one member or truss, and requires the (sometimes inaccurate) approximation of the boundary conditions for the particular member (e.g., clamped-clamped, clamped-free, etc.). It is therefore desirable to be able to calculate the critical loads directly, without resorting to potentially oversimplified models.

7.3.3 Buckling Load

In this section we apply the methods presented in Chapter 6 to the calculation of the critical load parameter. We first note that for this particular problem, the operator $\mathcal{B}$ is indefinite, and the corresponding eigenvalues may take both positive and negative values. We illustrate this in Figure 7-8 which shows the eigenmodes associated with the smallest positive and the smallest negative eigenvalues (we recall that a negative eigenvalue means that the loading is applied in the direction opposite (in this case, upwards) to that prescribed (in this case, downwards)).

We therefore create reduced-basis spaces W_N^u, $W_N^{\xi+}$, and $W_N^{\xi-}$ using a log-random sample S_N. We present the error and effectivities in Tables 7.10 and 7.11 for $\lambda_N^{1+}(\mu)$, and in the Tables 7.12 and 7.13 for $\lambda_N^{1-}(\mu)$, at two representative values of μ. We note that the error is again small for relatively small N — in all cases the error is approximately 1% for $N = 50$. Although the errors are considerably larger than those for the average deflection and stress approximations, the slower convergence rate is probably due to additional error introduced by approximating $\mathcal{B}(\mu; u(\mu))$ with $\mathcal{B}(\mu; u_N(\mu))$. We also note that in most cases (Tables 7.11-7.13 in particular) the effectivities are greater than unity for relatively small N and approach $1/\tau$ rather quickly. In Table 7.10, however, $\eta_N(\mu) < 1$ even for $N = 50$. The inferiority of the effectivities are certainly related to the relatively slower convergence of the reduced-basis approximation for the buckling load: we recall

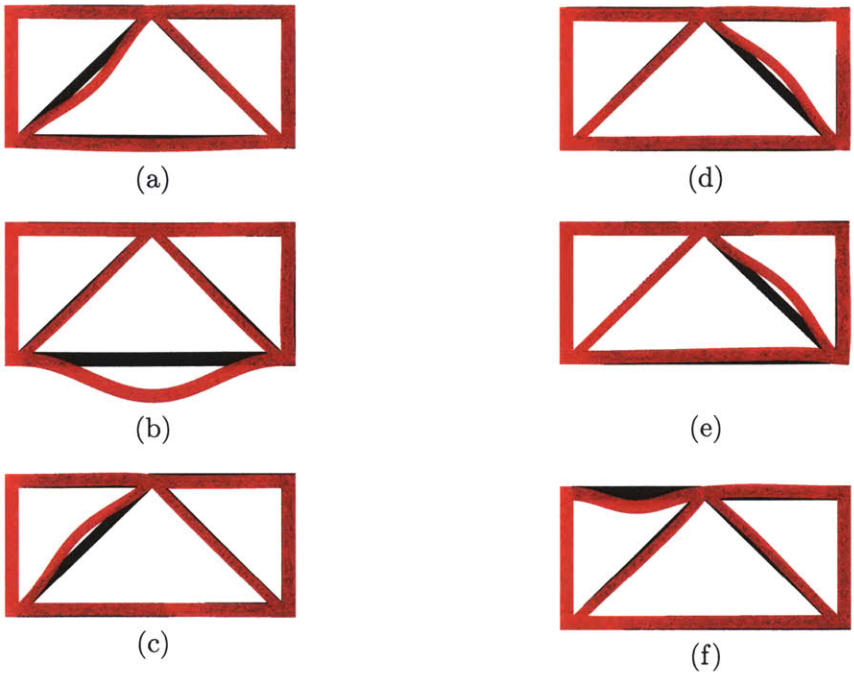


Figure 7-8: Buckling modes associated with the (a)-(c) smallest positive and (d)-(f) smallest negative eigenvalues for different values of μ.

from Chapter 4 that for Method II output bounds, exponential convergence of the reduced-basis approximation is essential for good bounds.

We now investigate the dependence of the minimum (positive and negative) eigenvalues on certain parameters. We plot in Figure 7-9 the smallest positive eigenvalues versus $\bar{t}_{\text{top}}$ and $\bar{t}_{\text{bot}}$. The plots reflect the intuitive fact that the critical positive load parameter is more sensitive to changes in $\bar{t}_{\text{bot}}$. Furthermore, for small values of $\bar{t}_{\text{bot}}$ (relative to the other dimensions) the eigenvalue is very sensitive to $\bar{t}_{\text{bot}}$, while for large $\bar{t}_{\text{bot}}$ the eigenvalue increases very slowly with $\bar{t}_{\text{bot}}$. Similar results are presented in Figure 7-10 for the smallest negative eigenvalue.

Knowledge of such trends and features are very important in the analysis and design of structures whether against yielding, elastic buckling, or other types of failure. Plots such as those presented in this and the previous sections also indicate to which parameters the behavior of the structure is most sensitive. However, in the design of structures with respect to many parameters and constraints, such plots are clearly insufficient — exploration of the entire design space and simultaneous consideration of all constraints is required. We therefore consider more general methods for design and optimization.

7.4 Design and Optimization

A particular microtruss design (that is, a particular value of μ) has associated with it operational and material costs, as well as performance merits reflecting its ability to support the applied structural loads. Furthermore, a design must meet certain constraints reflecting, for example, safety and manufacturability considerations. The goal of the design process is to minimize costs and optimize performance while ensuring that all design constraints are satisfied.

N	$\lvert s(\mu) - s_N(\mu)\rvert / s(\mu)$	$\Delta_N(\mu)/s(\mu)$	$\eta_N(\mu)$
5	9.35e-02	1.16e-01	1.25
10	3.53e-02	2.07e-01	5.87
15	1.79e-01	2.80e-01	1.56
20	1.39e-01	2.59e-01	1.87
25	1.02e-01	1.99e-01	1.95
30	3.92e-02	7.60e-02	1.94
35	2.70e-02	5.57e-02	2.06
40	9.36e-03	1.63e-02	1.75
45	9.70e-04	5.53e-03	5.70
50	2.33e-03	1.46e-03	0.62

Table 7.10: Error and effectivities for the smallest positive eigenvalue for $\mu = \mu^0$ (the reference parameter).

N	$\lvert s(\mu) - s_N(\mu)\rvert / s(\mu)$	$\Delta_N(\mu)/s(\mu)$	$\eta_N(\mu)$
5	3.52e 01	7.05e 01	2.00
10	6.71e-02	9.43e-02	1.41
15	2.92e-01	6.10e-01	2.09
20	2.88e-02	8.88e-03	0.31
25	4.11e-02	5.78e-02	1.41
30	2.08e-02	3.40e-02	1.64
35	2.99e-02	5.25e-02	1.75
40	2.44e-02	4.85e-02	1.99
45	1.26e-02	2.47e-02	1.96
50	1.22e-02	2.42e-02	1.99

Table 7.11: Error and effectivities for the smallest positive eigenvalue for a randomly selected value of μ.

N	$\lvert s(\mu) - s_N(\mu)\rvert / s(\mu)$	$\Delta_N(\mu)/s(\mu)$	$\eta_N(\mu)$
5	2.21e 00	3.19e 00	1.44
10	6.14e-01	1.15e 00	1.87
15	7.57e-02	1.11e-01	1.47
20	3.96e-02	3.40e-02	0.86
25	3.00e-02	5.00e-02	1.67
30	2.00e-02	3.81e-02	1.90
35	1.85e-02	3.41e-02	1.85
40	2.26e-02	4.86e-02	2.15
45	1.10e-02	2.47e-02	2.24
50	5.00e-03	1.12e-02	2.25

Table 7.12: Error and effectivities for the smallest negative eigenvalue for $\mu = \mu^0$ (the reference parameter).

N	$\lvert s(\mu) - s_N(\mu)\rvert / s(\mu)$	$\Delta_N(\mu)/s(\mu)$	$\eta_N(\mu)$
5	6.46e 00	2.46e 00	0.38
10	5.23e 00	1.03e 01	1.97
15	3.84e-01	7.22e-01	1.88
20	8.33e-02	1.08e-01	1.30
25	2.97e-02	5.07e-02	1.71
30	2.25e-02	4.86e-02	2.16
35	3.09e-02	6.83e-02	2.21
40	2.92e-02	6.19e-02	2.12
45	6.81e-03	1.37e-02	2.02
50	4.38e-03	9.09e-03	2.08

Table 7.13: Error and effectivities for the smallest negative eigenvalue for a randomly selected value of μ.

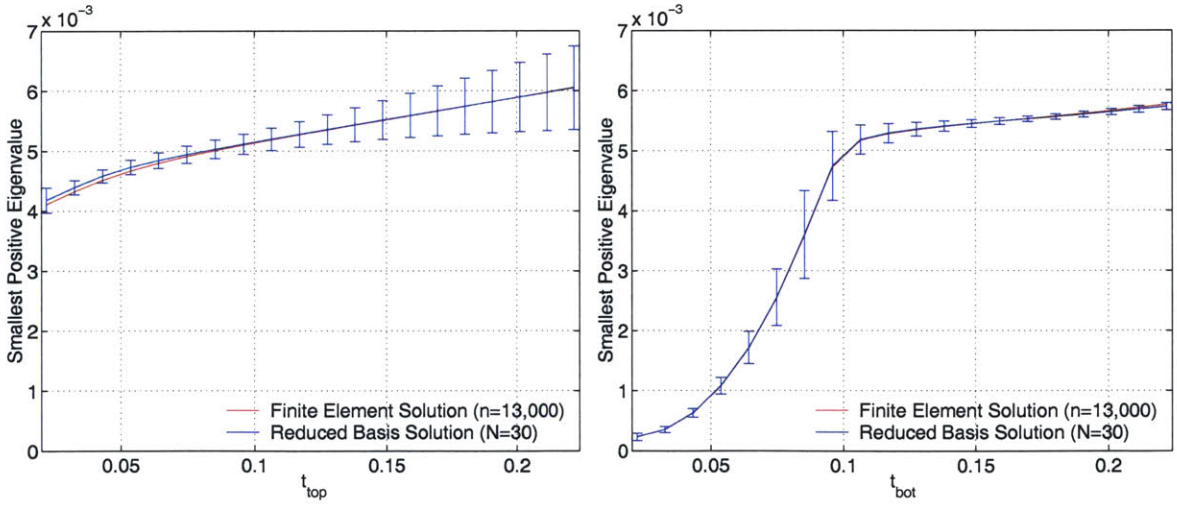


Figure 7-9: Plots of the smallest positive eigenvalue as function of $\bar{t}_{\text{top}}$ and $\bar{t}_{\text{bot}}$.

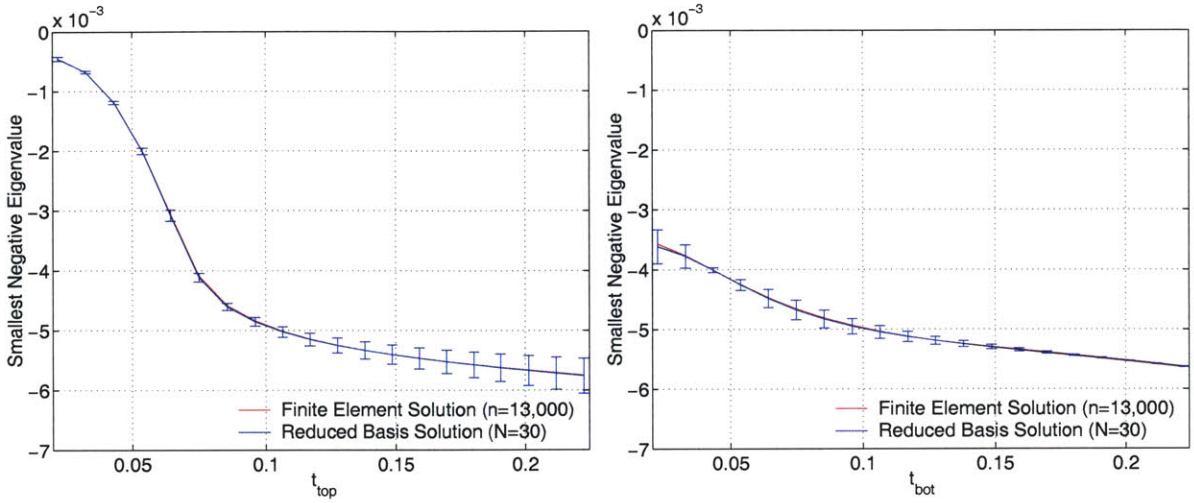


Figure 7-10: Plots of the smallest negative eigenvalue as function of $\bar{t}_{\text{top}}$ and $\bar{t}_{\text{bot}}$.

In this section we formulate the design-optimize problem and apply reduced-basis output bound methods to the solution of the optimization problem.

7.4.1 Design-Optimize Problem Formulation

We consider the following optimization problem:

$$\begin{aligned} \text{find} \qquad & \mu^* = \arg\min_{\mu} \mathcal{J}(\mu) \\ \text{subject to} \qquad & \begin{cases} f^i(\mu) \geq 0, & i = 1, \ldots, K_f, \\ g^j(\mu) \geq 0, & j = 1, \ldots, K_g, \\ h^k(\mu) \geq 0, & k = 1, \ldots, K_h, \end{cases} \end{aligned} \tag{7.32}$$

where $\mu \in \mathbb{R}^P$, the inequality constraints $f^i(\mu) \geq 0$ and (for simplicity) the equality constraints $h^i(\mu) = 0$ involve "simple" functions of μ — for instance,

$$f^1(\mu) = \mu_{\max} - \mu \geq 0 \tag{7.33}$$
$$f^2(\mu) = \mu - \mu_{\min} \geq 0 \ , \tag{7.34}$$

and the inequalities $g^i(\mu) \geq 0$ are constraints on outputs $s(\mu) = \langle L(\mu), u(\mu) \rangle$ — for instance,

$$g^1(\mu) = s_{\max} - s(\mu) \geq 0 \tag{7.35}$$
$$g^2(\mu) = s(\mu) - s_{\min} \geq 0, \tag{7.36}$$

where $u(\mu)$ is the solution to a partial differential equation. Note the *cost function* $\mathcal{J}(\mu)$ may in general be of the form

$$\mathcal{J}(\mu) = \sum_{i=1}^{n} a_i \mathcal{J}_i(\mu) \ , \tag{7.37}$$

where the a_i are weights and the $\mathcal{J}_i$ are different cost metrics.

Microtruss Design-Optimize Problem

We now consider a particular instantiation of our abstract optimization formulation. We consider here a simpler version of the optimization problem posed in Chapter 1 involving only outputs related to the structural aspects of the problem; the full multifunctional (thermo-structural) problem is addressed in [33].

As indicated in Chapter 1, one of the main advantages of the truss design is that it offers relatively high structural load capacities at low weights, making it useful for weight-sensitive applications such as next-generation rocket engines and unmanned autonomous vehicles (UAVs). Furthermore, the weight of the structure is directly related to the amount of material utilized, which in turn is related to material costs; in the case of rocket engines and UAVs, the weight is also directly related to fuel consumption (operational costs). We may thus define our *cost function* to be the area of the structure, $\mathcal{J}(\mu) = \mathcal{V}(\mu)$, where $\mu = \{\bar{t}_{\text{top}}, \bar{t}_{\text{bot}}, \bar{t}, \bar{\alpha}\}$, and

$$\mathcal{V}(\mu) = 2\left[(\bar{t}_t + \bar{t}_b)\left(\frac{\bar{H}}{\tan\alpha} + \frac{\bar{t}}{\sin\bar{\alpha}} + \bar{t}_b\right) + \bar{H}\left(\frac{\bar{t}}{\sin\alpha} + \bar{t}_b\right)\right] . \tag{7.38}$$

Furthermore, we require that (i) the parameter μ be in the design set, i.e., $\mu \in \mathcal{D}^\mu \subset \mathbb{R}^P$, where $\mathcal{D}^\mu = [0.022, 0.22]^3 \times [0°, 45°]$, (ii) the length of the structure be equal to ℓ_0,

$$\ell(\mu) = 2\left(\frac{\bar{H}}{\tan\alpha} + \frac{\bar{t}}{\sin\bar{\alpha}} + \bar{t}_b\right) = \ell_0 \ ; \tag{7.39}$$

(iii) the deflection be less than a prescribed limit,

$$\delta_{\text{ave}}(\mu) \leq \alpha_1 \delta_{\max}, \tag{7.40}$$

and (iv) the average stress near the support be less than the yield stress,

$$\sigma_{\text{ave}}(\mu) \leq \alpha_2 \sigma_Y \ ; \tag{7.41}$$

for this simpler formulation we shall not consider stability (buckling).

Our optimization problem can then be stated as: find $\mu^* = \{\bar{t}_t^*, \bar{t}_b^*, \bar{t}^*, \bar{\alpha}^*\}$, which satisfies

$$\text{find} \qquad \mu^* = \arg\min_{\mu} \mathcal{J}(\mu) \tag{7.42}$$

$$\text{subject to} \quad \begin{cases} f^0(\mu) = \ell(\mu) - \ell_0 = 0, \\ f^1(\mu) = \bar{t}_{\text{top}} - 0.022 \geq 0, \\ f^2(\mu) = 0.22 - \bar{t}_{\text{top}} \geq 0, \\ f^3(\mu) = \bar{t}_{\text{bot}} - 0.022 \geq 0, \\ f^4(\mu) = 0.22 - \bar{t}_{\text{bot}} \geq 0, \\ f^5(\mu) = \bar{t} - 0.022 \geq 0, \\ f^6(\mu) = 0.22 - \bar{t} \geq 0, \\ f^7(\mu) = \bar{\alpha} \geq 0, \\ f^8(\mu) = 45 - \bar{\alpha} \geq 0, \\ g^1(\mu) = \alpha_1 \delta_{\max} - \delta_{\text{ave}}(\mu) \geq 0, \\ g^2(\mu) = \alpha_2 \sigma_Y - \sigma_{\text{ave}}(\mu) \geq 0, \\ h^1(\mu) = \ell(\mu) - \ell_0 = 0, \end{cases}$$

7.4.2 Solution Methods

We now consider methods for solving general optimization problems of the form (7.32). In particular, we focus on *interior point methods*, computational methods for the solution of constrained optimization problems which essentially generate iteratetes which are strictly feasible (i.e., in the *interior* of the feasible region) and converge to the true solution. The constrained problem is replaced by a sequence of unconstrained problems which involve a *barrier function* which enforces strict feasibility and effectively prevents the approach to the boundary of the feasible region [9]. The solutions to these unconstrained problems then approximately follow a "central path" to the solution of the original constrained problem; this is depicted in Figure 7-11. We present here a particular variant of IPMs known as *primal-dual* algorithms.

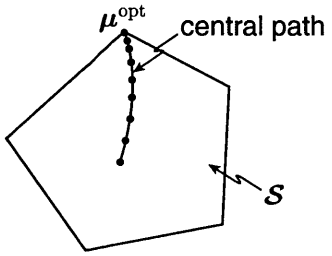


Figure 7-11: Central path.

To begin, we introduce the modified optimization problem

$$\text{find} \quad \mu_v^* = \arg\min_{\mu} \mathcal{J}_v(\mu) \ , \tag{7.43}$$

where the modified cost functional is given by

$$\mathcal{J}_\upsilon(\mu) \equiv \mathcal{J}(\mu) - \upsilon_b \left(\sum_{i=1}^{K_f} \ln f_i(\mu) + \sum_{j=1}^{K_g} \ln g_j(\mu) \right) + \frac{1}{2\upsilon_p} \sum_{k=1}^{K_h} h_k(\mu)^2; \tag{7.44}$$

it can then be shown [15] that as $\upsilon = (\upsilon_b, \upsilon_p) \to 0$, the solution to (7.43) approaches the solution to the original constrained optimization problem (7.32) — that is,

$$\mu^* = \lim_{\upsilon \to 0} \mu_\upsilon^* \ . \tag{7.45}$$

We therefore wish to find μ_υ^* such that

$$\nabla \mathcal{J}_\upsilon(\mu_\upsilon^*) = \nabla_\mu J(\mu_\upsilon^*) - \upsilon_b \left(\sum_{i=1}^{K_f} \frac{\nabla_\mu f_i(\mu_\upsilon^*)}{f_i(\mu_\upsilon^*)} + \sum_{j=1}^{K_g} \frac{\nabla_\mu g_j(\mu_\upsilon^*)}{g_j(\mu_\upsilon^*)} \right) + \frac{1}{\upsilon_p} \sum_{k=1}^{K_h} h_k(\mu) \nabla_\mu h_k(\mu) = 0 \tag{7.46}$$

in the limit as $\upsilon \to 0$.

We now introduce the "slack" variables $F \in \mathbb{R}^{K_h}$ and $G \in \mathbb{R}^{K_g}$ given by

$$F^i = \frac{\upsilon}{f_i(\mu)}, \qquad i = 1, \ldots, K_f, \tag{7.47}$$

$$G^i = \frac{\upsilon}{g_i(\mu)}, \qquad i = 1, \ldots, K_g, \tag{7.48}$$

and define $z = (\mu, F, G)$ and

$$\mathcal{F}(z; \upsilon) = \begin{bmatrix} \nabla_\mu \mathcal{J}(\mu) + \sum_{i=1}^{K_f} F^i \nabla_\mu f_i(\mu) + \sum_{i=1}^{K_g} G^i \nabla_\mu g_i(\mu) + \frac{1}{\upsilon_p} \sum_{i=1}^{K_h} h_i(\mu) \nabla_\mu h_i(\mu) \\ F_1 f_1(\mu) - \upsilon \\ \vdots \\ F_{K_f} f_{K_f}(\mu) - \upsilon \\ G_1 g_1(\mu) - \upsilon \\ \vdots \\ G_{K_g} g_{K_g}(\mu) - \upsilon \end{bmatrix}. \tag{7.49}$$

The optimizer $z^* = (\mu^*, F^*, G^*)$ then satisfies

$$\mathcal{F}(z^*; \upsilon) = 0, \qquad \text{as } \upsilon \to 0. \tag{7.50}$$

Applying Newton's method to (7.50) and letting $\upsilon \to 0$, the primal-dual interior point algorithm may thus be summarized as follows:

Set $\upsilon > 0$, $0 < \epsilon < 1$
for $i = 1,\ldots,$max_iter **do**
  while $||F(z;\upsilon)|| >$ tol **do**
    $\Delta z = [\nabla_z F(z;\upsilon)]^{-1} F(z)$;
    Set $\gamma = \min\left\{1, 0.95 \min_i \frac{z_i}{-\Delta z_i}\right\}$;
    $z = z + \gamma \Delta z$;
  end while
  $\upsilon = \epsilon\upsilon$
end for

Note the factor γ is chosen such that the iterate $z + \gamma\Delta z$ is strictly in the interior of the feasible set.

Clearly, solution of the full optimization problem is computationally very expensive since it requires repeated evaluation for many different values of the parameter μ of the outputs of interest. However, this is precisely the type of situation in which reduced-basis methods are most useful; we therefore pursue a reduced-basis approach.

7.4.3 Reduced-Basis Approach

To begin, we again consider the optimization problem (7.32) but replace $g^i(\mu)$ with the "appropriate" reduced-basis approximation such that feasibility is guaranteed,

$$\begin{aligned} &\text{find} && \mu_N^* = \arg\min_{\mu} \mathcal{J}(\mu) \\ &\text{subject to} && \begin{cases} f^i(\mu) \geq 0, & i = 1,\ldots,K_f, \\ g_N^i(\mu) \geq 0, & i = 1,\ldots,K_g. \end{cases} \end{aligned} \tag{7.51}$$

For example, in place of the $g^i(\mu)$ defined in (7.35)-(7.36) we would take

$$g_N^1(\mu) = s_{\max} - s_N^+(\mu), \tag{7.52}$$

$$g_N^2(\mu) = s_N^-(\mu) - s_{\min}\ . \tag{7.53}$$

Note if $s_N^-(\mu) \leq s(\mu) \leq s_N^+(\mu)$, and if the $g_N(\mu)$ are appropriately defined (as in (7.52)-(7.53)), the solution μ_N^* to (7.51) is guaranteed to be feasible in the sense of (7.32)

The advantage of using reduced-basis methods in solving our optimization problem is twofold: (i) the low computational cost of evaluating (online) the outputs of interest allow for fast and efficient solution of the optimization problem; and (ii) rigorous error estimators and output bounds guarantee feasibility of the solution obtained from (7.51).

7.4.4 Results

We present in Table 7.14 the results of our optimization procedure for several scenarios. For this example we take the yield stress to be 30MPa, and the maximum deflection to be 0.03mm. Note that in this case the angle effectively serves only to ensure that the equality constraint $\ell(\mu) = \ell_0(\mu)$ is satisfied. In Scenario 1, we obtain an initial guess which satisfies all the constraints but is not necessarily optimal. In Scenario 2, we minimize the area of the structure while allowing $\hat{t}_{\text{top}}$ to vary. We then find that the optimal value is $\hat{t}_{\text{top}} = 0.507$, resulting in a 30% reduction in the cost

function (area) compared to the results of Scenario 1. We also note that in this case, the yielding constraint on the stress is active. In Scenario 3, we allow both $\hat{t}_{\text{top}}$ and $\hat{t}_{\text{bot}}$ to vary. We then find that the cost can be reduced further by 30% compared to the results of Scenario 2. Finally, we allow $\hat{t}_{\text{top}}$, $\hat{t}_{\text{bot}}$, and t to vary, and find that the cost can still be reduced by another 10%; note that in this case, both the deflection and stress constraints are active.

Scenario	$\hat{t}_{\text{top}}$(mm)	$\hat{t}_{\text{bot}}$(mm)	$\hat{t}$(mm)	$\alpha(^\circ)$	$\mathcal{V}$(mm^2)	δ_N^+(mm)	σ_N^+(MPa)	time (s)
1	1.500	0.500	0.500	54.638	50.04	0.0146	09.227	0.680
2	0.507	0.500	0.500	54.638	35.14	0.0200	30.000*	1.020
3	0.523	0.200*	0.500	53.427	25.65	0.0277	30.000*	1.050
4	0.521	0.224	0.345	52.755	23.02	0.0300*	30.000*	1.330

Table 7.14: Optimization of the microtruss structure (for $\hat{H} = 9$mm) using reduced-basis output bounds. (These results were obtained in collaboration with Dr. Ivan Oliveira of MIT, and are used here with permission.)

The solution of the optimization problem for each scenario requires O(10) deflection and stress calculations. As shown in Table 7.14, our reduced-basis solution method therefore effectively solves — *on-line* — O(10) partial differential equations within a single second. In contrast, matrix assembly and solution (using non-commercial code) of the finite element equations for a single value of μ takes approximately 9 seconds. The online computational savings effected by the reduced-basis method is clearly no small economy.

7.5 Prognosis: An Assess-(Predict)-Optimize Approach

The design of an engineering system, as illustrated in Section 1.1.4, involves the determination of the system configuration based on system requirements and environment considerations. During operation, however, the state of the system may be unknown or evolving, and the system may be subjected to dynamic system requirements, as well as changing environmental conditions. The system must therefore be adaptively designed and optimized, taking into consideration the uncertainty and variability of system state, requirements, and environmental conditions.

For example, we assume that extended deployment of our microtruss structure (for instance, as a component in an airplane wing) has led to the developement of defects (e.g., cracks) shown in Figure 7-12. The characteristics of the defects (e.g., crack lengths) are unknown, but we assume that we are privy to a set of experimental measurements which serve to assess the state of the structure. Clearly, the defects may cause the deflection to reach unacceptably high values; a shim is therefore introduced so as to stiffen the structure and maintain the deflection at the desired levels. However, this intervention leads to an increase in both material and operational costs. Our goal is to find, given the uncertainties in the crack lengths, the shim dimensions which minimize the weight while honoring our deflection constraint.

7.5.1 Assess-Optimize Problem Formulation

More precisely, we characterize our system with a multiparameter $\mu = (\mu_{\text{crack}}, \mu_{\text{shim}})$ where $\mu_{\text{crack}} = (\bar{L}_1, \bar{L}_2)$ and $\mu_{\text{shim}} = (\bar{L}_{\text{shim}}, \bar{t}_{\text{shim}})$. As shown in Figure 1-5, $\bar{L}_1$ and $\bar{L}_2$ are our "current guesses" for the relative lengths of the cracks on the upper frame and truss, respectively, while $\bar{t}_{\text{shim}}$ and $\bar{L}_{\text{shim}}$ denote the thickness and length of the shim, respectively; we also denote by $\mu^*_{\text{crack}} = (\bar{L}^*_1, \bar{L}^*_2)$

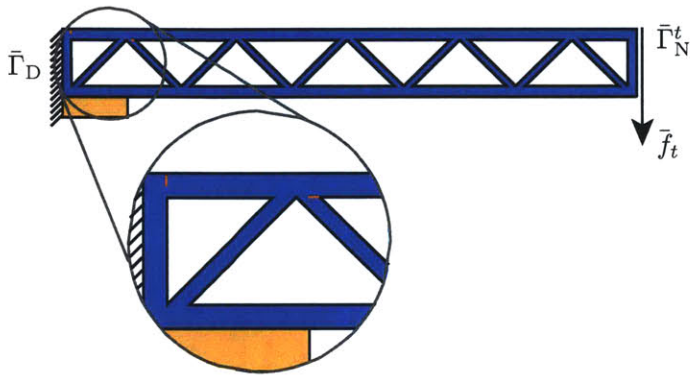


Figure 7-12: A "defective" microtruss structure. The insert highlights the defects (two cracks) and intervention (shim).

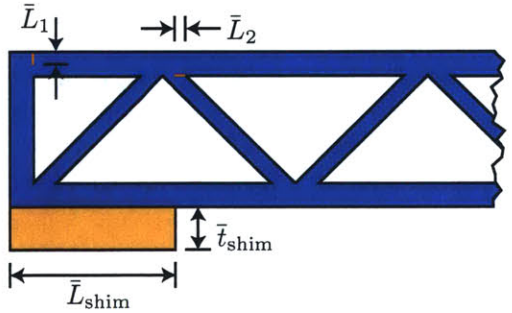


Figure 7-13: Parameters describing the defects, ($\bar{L}_1$ and $\bar{L}_2$), and the intervention, ($\bar{L}_{\mathrm{shim}}$ and $\bar{t}_{\mathrm{shim}}$).

the (real) *unknown* crack lengths. We further assume we are given M_{exp} intervals $[\delta_{LB}^m, \delta_{UB}^m]$, $m = 1, \ldots, M_{\text{exp}}$ representing experimental measurements of the deflection such that

$$\delta_{\text{ave}}(\mu_{\text{shim}}^m, \mu_{\text{crack}}^*) \in [\delta_{LB}^m, \delta_{UB}^m], \quad m = 1, \ldots, M_{\text{exp}} \ . \tag{7.54}$$

From these measurements, we then infer the existence of a set $\mathcal{L}^*$ such that $(\bar{L}_1^*, \bar{L}_2^*) \in \mathcal{L}^*$. We then wish to find the μ_{shim}^* which satisfies

$$\mu_{\text{shim}}^* = \arg \min_{\mu_{\text{shim}}} \mathcal{V}_{\text{shim}}(\mu_{\text{shim}}), \tag{7.55}$$

where $\mathcal{V}_{\text{shim}}(\mu_{\text{shim}}) = \bar{t}_{\text{shim}} \bar{L}_{\text{shim}}$ is simply the area of the shim, such that

$$\mu_{\text{shim}} \in \mathcal{D}^{\text{shim}} \tag{7.56}$$

$$\max_{\mu_{\text{crack}} \in \mathcal{L}^*} \delta_{\text{ave}}(\mu_{\text{shim}}, \mu_{\text{crack}}) \leq \delta_{\max} \quad . \tag{7.57}$$

In words, we wish to find the shim dimensions which minimizes the area of the shim such that the maximum deflection (over all crack lengths consistent with the experiments) is less than the prescribed deflection limit $\delta_{\max}$.

7.5.2 Solution Methods

Unfortunately, the set $\mathcal{L}^*$ is very costly to construct. However, we can readily calculate a conservative approximation to $\mathcal{L}^*$, $\hat{\mathcal{L}}^*$, by (*i*) applying our sharp yet computationally inexpensive output bounds, and (*ii*) exploiting the readily proven monotonicity (increasing) property of $\delta(\mu)$ with respect to $\bar{L}_1$ and $\bar{L}_2$.

By construction, the surrogate $\hat{\mathcal{L}}^*$ contains $\mathcal{L}^*$; we may thus replace (7.55) by:

$$\begin{aligned} \hat{\mu}_{\text{shim}}^* \quad = \quad & \arg \min_{\mu_{\text{shim}} \in \mathcal{D}^{\text{shim}}} \mathcal{V}_{\text{shim}}(\mu_{\text{shim}}) \\ & \text{s.t.} \quad \max_{\mu_{\text{crack}} \in \hat{\mathcal{L}}^*} \delta(\mu_{\text{shim}}, \mu_{\text{crack}}) \leq \delta_{\max}, \end{aligned} \tag{7.58}$$

since if $\hat{\mu}_{\text{shim}}^*$ satisfies $\max_{\mu_{\text{crack}} \in \hat{\mathcal{L}}^*} \delta(\mu_{\text{crack}}, \hat{\mu}_{\text{shim}}^*) \leq \delta_{\max}$ then it follows that $\max_{\mu_{\text{crack}} \in \mathcal{L}^*} \delta(\mu_{\text{crack}}, \hat{\mu}_{\text{shim}}^*) \leq \delta_{\max}$. By once again exploiting the monotonicity of $\delta(\mu)$ with respect to μ_{crack}, we can readily find the maximizer of δ over $\hat{\mathcal{L}}^*$; to wit, we define

$$\mu'_{\text{crack}} = \arg \max_{\mu_{\text{crack}} \in \hat{\mathcal{L}}^*} \delta(\mu_{\text{crack}}, \mu_{\text{shim}}),$$

where μ'_{crack} (which is independent of μ_{shim}) is calculated by a simple binary chop or bisection algorithm (see [2]). We may then write (7.58) as

$$\begin{aligned} \hat{\mu}_{\text{shim}}^* \quad = \quad & \arg \min_{\mu_{\text{shim}} \in \mathcal{D}_{\text{shim}}} \mathcal{V}_{\text{shim}}(\mu_{\text{shim}}) \\ & \text{s.t.} \ \delta(\mu'_{\text{crack}}, \mu_{\text{shim}}) \leq \delta_{\max}. \end{aligned} \tag{7.59}$$

Finally, we again apply our sharp yet inexpensive *a posteriori* output bounds to arrive at

$$\begin{aligned} \hat{\mu}_{\text{shim}}^* \quad = \quad & \arg \min_{\mu_{\text{shim}} \in \mathcal{D}_{\text{shim}}} \mathcal{V}_{\text{shim}}(\mu_{\text{shim}}) \quad (= \bar{L}_{\text{shim}} \bar{t}_{\text{shim}}). \\ & \text{s.t.} \ \delta_N^+(\mu'_{\text{crack}}, \mu_{\text{shim}}) \leq \delta_{\max}. \end{aligned} \tag{7.60}$$

The formulation (7.60) has several advantages. First, satisfaction of the constraints is guaranteed: if $\delta_N^+(\mu) \geq \delta_N(\mu)$, and μ^*_{shim} satisfies $\delta_N^+(\mu'_{\text{crack}}, \mu^*_{\text{shim}}) \leq \delta_{\max}$, then it follows that $\delta(\mu'_{\text{crack}}, \mu^*_{\text{shim}}) \leq \delta_{\max}$. Second, $\delta_N^+(\mu)$ can be calculated efficiently: the online complexity to calculate $\delta_N^+(\mu)$ is only $O(N^3)$ — independent of the very high dimension of the truth finite element space, $\mathcal{N}$. Finally, the constrained minimization problem (7.60) is readily treated by the interior point methods presented in Section 7.4.2.

We present in Table 7.15 the solution to the optimization problem (7.60) for several scenarios. The inputs to the algorithm are: the crack lengths $(\bar{L}_1^*, \bar{L}_2^*) \in [0.1, 0.75]^2$, which are "unknown" (to the algorithm); the error in the measurements, $\epsilon \in [0.1, 5]\%$; and the upper bound on the deflection, $\delta_{\max}$, "typically" chosen to be between (say) 6400 and 6600. The outputs are: the upper bound $\delta_N^+(\mu)$ and the shim volume $\mathcal{V}_{\text{shim}}(\mu_{\text{shim}})$ at optimality. We also present the computation time required to solve the complete optimization problem.

As a first test case, we assume that after several missions our truss structure has developed cracks of lengths $(\bar{L}_1^*, \bar{L}_2^*) = (0.5, 0.5)$. The plane is immediately needed for a new mission which requires $\delta(\mu) \leq \delta_{\max} = 6600.0$; a few crude deflection measurements ($\epsilon = 5.0\%$) are performed, from which we conclude that a shim of (optimal) dimensions $(\bar{L}^*_{\text{shim}}, \bar{t}^*_{\text{shim}}) = (15.0, 0.84)$ (and volume 12.6) is required to ensure feasibility.

Next, we assume that for the next mission, the constraint on the deflection is now tighter, $\delta_{\max} = 6400.0$; ($\bar{L}_1^*$, $\bar{L}_2^*$, and ϵ are unchanged.) Measurement and optimization confirm that, as anticipated, a larger shim with a volume of 20.9 is needed, with $(\bar{L}^*_{\text{shim}}, \bar{t}^*_{\text{shim}}) = (15.0, , 1.4)$. When the plane returns, more accurate measurements ($\epsilon = 0.1\%$) can now be afforded. (We assume that $\bar{L}_1^*$, $\bar{L}_2^*$, and $\delta_{\max}$ are unchanged.) On the basis of these new measurements, it is found that a shim of volume 8.5 and dimensions $(15.0, 0.57)$ actually suffices.

After a subsequent mission, we assume that the crack length $\bar{L}_2^*$ has increased to 0.7. Upon taking new measurements ($\epsilon = 0.1\%$), it is found that for the same deflection limit, $\delta_{\max} = 6400.0$, an (optimal) shim of volume 8.62 and dimensions $(\bar{L}^*_{\text{shim}}, \bar{t}^*_{\text{shim}}) = (15.0, 0.57)$ is now required.

Next, we assume that the crack length $\bar{L}_1^*$ has increased to 0.7. New measurements and optimization reveal that a much bigger shim of volume 16.1 and dimensions $(\bar{L}^*_{\text{shim}}, \bar{t}^*_{\text{shim}}) = (15.0, 1.1)$ is now required. The analysis automatically identifies these defects which most strongly affect performance.

Lastly, we assume that, for a final mission, the constraint on the deflection can be relaxed, $\delta_{\max} = 6500.0$. Measurement ($\epsilon = 0.1\%$) and optimization confirm that, as anticipated, a smaller shim is sufficient, with volume 12.4 and $(\bar{L}^*_{\text{shim}}, \bar{t}^*_{\text{shim}}) = (15.0, 0.82)$.

We note that solution of a single optimization problem (for a given set of constraints) requires solution of roughly 300 partial differential equations and associated sensitivity calculations (at different parameter values).

L_1^*	L_2^*	$\epsilon(\%)$	$\delta_{\max}$	$\delta_N^+(\mu_{\text{shim}}^*)$	$\mathcal{V}_{\text{shim}}(\mu_{\text{shim}}^*)$	Time(s)
0.5	0.5	5.0	6600	6600	12.6	9.4
0.5	0.5	5.0	6400	6400	20.9	7.1
0.5	0.5	0.1	6400	6400	8.5	7.6
0.5	0.7	0.1	6400	6400	8.6	7.5
0.7	0.7	0.1	6400	6400	16.1	9.5
0.7	0.7	0.1	6500	6500	12.4	10.1

Table 7.15: Solution to the assess-optimize problem with evolving (unknown) system characteristics and varying constraints.

Chapter 8

Summary and Future Work

8.1 Summary

The main goal of this thesis is to develop reduced-basis methods for problems in elasticity. The essential components are (*i*) rapidly convergent global reduced-basis approximations — projection onto a space W_N spanned by solutions of the governing partial differential equation at N selected points in parameter space; (*ii*) *a posteriori* error estimation — relaxations of the error-residual equation that provide inexpensive bounds for the error in the outputs of interest; and (*iii*) off-line/on-line computational procedures — methods which decouple the generation and projection stages of the approximation process. The operation count for the on-line stage — in which, given a new parameter value, we calculate the output of interest and associated error bound — depends only on N (typically very small) and the parametric complexity of the problem; the method is thus ideally suited for the repeated and rapid evaluations required in the context of parameter estimation, design, optimization, and real-time control.

Perhaps the biggest challenge in applying reduced-basis output bound methods to problems in linear elasticity lies in developing *rigorous* (Method I) and *inexpensive a posteriori* error estimators. The need for rigorous output bounds is clear: in many real-world applications, certifiability of approximations is required so as to rigorously satisfy the prescribed application constraints.

In this thesis, we find that our Method I *a posteriori* error estimation procedures rely critically on the existence of a "bound conditioner" — in essence, an operator preconditioner that (*i*) satisfies an additional spectral "bound" requirement, and (*ii*) admits the reduced-basis off-line/on-line computational stratagem. This improved understanding of reduced-basis *a posteriori* error estimation allows us to develop new techniques — in fact, recipes — for constructing the required bound conditioners.

To investigate the effectivity of these bound conditioner constructions, we first apply them in Chapter 4 to the problem of heat conduction, and in Chapter 5 to the problem of linear elasticity. Problems in heat conduction serve as a good test bed for our methods prior to application to linear elasticity: the Laplace operator is a "simpler version" of the linear elasticity operator. These theoretical and numerical tests show that our bound conditioners perform remarkably well — we achieve good effectivities, and therefore sharp output bounds; and enable a better understanding of the strengths and limitations of each of technique.

Although some of our Method I techniques have been found to be generally applicable, in some cases — particularly for linear elasticity — the computational cost is still relatively high, and the implementation difficult. We therefore also develop Method II error estimators — bounds which are simple and inexpensive, albeit at the loss of complete certainty. Nevertheless, numerical tests

show that these simple bounds perform quite well.

There are also many applications in which the posibility of unstable equilibria must be considered. While analytical formulas exist for the critical loads of relatively simple geometries, they are of limited applicability particularly for more complex structures. However, exact (finite element) solution of the partial differential equations governing elastic stability is computationally too expensive, especially in the "many queries" context of design and optimization. We therefore consider in Chapter 6 reduced-basis approximation and (Method II) error estimation methods for the problem of elastic buckling.

Finally, in Chapter 7 we apply our reduced-basis approximation and (Method II) error estimation methods to the structural aspects of the microtruss example put forth in Chapter 1. The advantages of the reduced-basis method in the "many queries" context of optimization is clear: we achieve substantial computational economy as compared to conventional finite element solution.

It must be noted that there are still many aspects of reduced-basis methods which must still be investigated and improved, some of which are briefly discussed in Chapter 2. In this chapter we present two areas which are related to the methods and problems presented here. First, we consider the problem of thermoelasticity, important for creating a full (linear) thermo-structural model of the microtruss example of Chapters 1 and 7. This problem is difficult due to the "propagation" of the reduced basis error: the temperature, which must also be approximated using reduced-basis methods, enters the linear elasticity equations as "data." Second, we consider the Helmholtz or wave equation, important in nondestructive evaluation, structural health monitoring, and prognosis — contexts for which our Assess-Predict-Optimize methodology of Chapter 7 is relevant. This problem is difficult since the operator is no longer coercive. We present initial ideas for these problems in the following sections.

8.2 Approximately Parametrized Data: Thermoelasticity

In this section we explore the situation in which the parametrized mathematical model is not exact. In particular, we permit error or imprecision in the data that define the linear functionals in our problem. In the case of thermoelasticity, for instance, this error is introduced by reduced-basis approximation of the temperature which couples into the linear elasticity equations through the linear functional. The ideas presented here extend the methods put forth in [38] for approximately parametrized operators to the simpler case in which the data (or loading) is approximately-parametrized; we develop these ideas for the particular case of thermoelasticity.

8.2.1 Formulation of the Thermoelasticity Problem

There are many important applications in which the changes in temperature may significantly affect the deformation of a structure — the multifunctional microtruss structure of Chapters 1 and 7 is one example, as are sensors and actuators.

When an isotropic body is subjected to a temperature change $\bar{T}$ from the uniform reference state temperature, it may develop thermal stresses such that the (linearized) stress-strain relations [29] become

$$\bar{\sigma}_{ij} = -\frac{\bar{E}\bar{\alpha}_\vartheta}{1-2\bar{\nu}}\delta_{ij}\bar{T} + \bar{C}_{ijkl}\bar{\varepsilon}_{kl} \ , \tag{8.1}$$

where $\bar{\alpha}_\vartheta$ is the linear coefficient of thermal expansion, the strains $\bar{\varepsilon}_{ij}$ are related to the deformation u by (5.8), $\bar{C}_{ijkl}$ is the elasticity tensor, and the temperature $\bar{T}$ satisfies the partial differential

equation

$$\langle \bar{\mathcal{A}}_\vartheta \bar{T}, \bar{v} \rangle = \langle \bar{F}_\vartheta, \bar{v} \rangle \ , \qquad \forall\ \bar{v} \in \bar{Y}_\vartheta \ . \tag{8.2}$$

Here, $\bar{\mathcal{A}}_\vartheta \colon \bar{Y}_\vartheta \to \bar{Y}_\vartheta'$ and $\bar{F}_\vartheta \in \bar{Y}_\vartheta'$ are given by (3.20) and (3.21), respectively. Note we assume that $\bar{\alpha}_\vartheta$ is constant and the $\bar{C}_{ijkl}$ independent of temperature change. Substituting (8.1) into (5.22), we find that the deformation satisfies the partial differential equation

$$\langle \bar{\mathcal{A}} \bar{u}, \bar{v} \rangle = \langle \bar{F}, \bar{v} \rangle + \langle \bar{\mathcal{H}} \bar{T}, \bar{v} \rangle \ , \qquad \forall\ \bar{v} \in \bar{Y} \ , \tag{8.3}$$

where $\bar{\mathcal{A}} \colon \bar{Y} \to \bar{Y}'$ and $\bar{F} \in \bar{Y}'$ are given by (5.24) and (5.25), respectively, and $\bar{\mathcal{H}}$ is given by

$$\langle \bar{\mathcal{H}}, \bar{v} \rangle = \int_{\bar{\Omega}} \frac{\bar{E}\bar{\alpha}_\vartheta}{1-2\bar{\nu}} \frac{\partial \bar{v}_i}{\partial \bar{x}_i} \bar{T} \, d\bar{\Omega} \ , \quad \forall\ \bar{v} \in \bar{Y} \ . \tag{8.4}$$

Upon application of a continuous piecewise-affine transformation from $\bar{\Omega}$ to a fixed (μ-independent) reference domain Ω we obtain from (8.2)

$$\langle \mathcal{A}_\vartheta(\mu) \vartheta(\mu), v \rangle = \langle F_\vartheta(\mu), v \rangle \ , \quad \forall\ v \in Y_\vartheta \ , \tag{8.5}$$

where $\mathcal{A}_\vartheta \colon Y_\vartheta \to Y_\vartheta'$ and $F_\vartheta \in Y_\vartheta'$ are given by (3.34) and (3.36), respectively; and (8.3)

$$\langle \mathcal{A}(\mu) u(\mu), v \rangle = \langle F(\mu), v \rangle + \langle \mathcal{H}(\mu) \vartheta(\mu), v \rangle \ , \quad \forall\ v \in Y \ , \tag{8.6}$$

where $\mathcal{A} \colon Y \to Y'$ and $F \in Y'$ are given by (5.38) and (5.40), respectively. Here,

$$\langle \mathcal{H}(\mu) \vartheta(\mu), v \rangle = \int_\Omega \beta_{ij}(\mu) \frac{\partial v_i}{\partial x_j} \vartheta(\mu) \, d\Omega \ , \tag{8.7}$$

where

$$\beta_{ij}(\mu) = \frac{\bar{E}\bar{\alpha}_\vartheta}{1-2\bar{\nu}} G_{ji} \left(\det \underline{G}(\mu) \right)^{-1} \ . \tag{8.8}$$

We note that $\mathcal{H}(\mu)$ is affinely dependent on the parameter since

$$\langle \mathcal{H}(\mu) w, v \rangle = \sum_{q=1}^{Q_\mathcal{H}} \alpha_\mathcal{H}^q(\mu) \langle \mathcal{H}^q w, v \rangle \ , \tag{8.9}$$

where $Q_\mathcal{H} = d^2$, and for $q(i,j) \colon \{1, \ldots, d\}^2 \to \{1, \ldots, Q_\mathcal{H}\}$,

$$\alpha_\mathcal{H}^{q(i,j)}(\mu) = \beta_{ij}(\mu) \ , \qquad \langle \mathcal{H}^q w, v \rangle = \int_\Omega \frac{\partial v_i}{\partial x_j} w \, d\Omega \ ; \tag{8.10}$$

we recall from Chapters 3 and 5 that $\mathcal{A}_\vartheta(\mu)$, $F_\vartheta(\mu)$, $\mathcal{A}(\mu)$, and $F(\mu)$ are also affine in the parameter.

8.2.2 Reduced-Basis Approximation

Temperature Approximation

As before, we sample our design space $\mathcal{D}^\mu$ to create the parameter sample

$$S_M^\mu = \{\mu_1, \ldots, \mu_M\} \ ; \tag{8.11}$$

we then introduce the reduced-basis approximation space for the temperature

$$W_M^T = \text{span}\{\phi_m = T(\mu_m), m = 1, \ldots, M\} \ . \tag{8.12}$$

Our reduced-order temperature approximation $T_M(\mu) \in W_M^T$ then satisfies

$$\langle \mathcal{A}_\vartheta(\mu) T_M(\mu), v\rangle = \langle F_\vartheta(\mu), v\rangle \ , \qquad \forall\ v \in W_M^T \ . \tag{8.13}$$

Deformation Approximation

We now introduce a "model-truncated" problem: given $\mu \in \mathcal{D}^\mu$, we find

$$\hat{s}(\mu) = \langle L(\mu), \hat{u}(\mu)\rangle \ , \tag{8.14}$$

where $\hat{u}(\mu) \in Y$ satisfies

$$\langle \mathcal{A}(\mu)\hat{u}(\mu), v\rangle = \langle F(\mu), v\rangle + \langle \mathcal{H}(\mu)\vartheta_M(\mu), v\rangle \ , \qquad \forall\ v \in Y \ . \tag{8.15}$$

We again sample the parameter space to create

$$S_N^\mu = \{\mu_1, \ldots, \mu_N\} \ ; \tag{8.16}$$

and introduce the reduced-basis approximation space for the deformation

$$W_N^u = \text{span}\{\zeta_n = \hat{u}(\mu_n), n = 1, \ldots, N\} \ . \tag{8.17}$$

Our output approximation is then

$$\hat{s}_N(\mu) = \langle L(\mu), \hat{u}_N(\mu)\rangle \ , \tag{8.18}$$

where $\hat{u}_N(\mu) \in W_N^u$ satisfies

$$\langle \mathcal{A}\hat{u}_N(\mu), v\rangle = \langle F(\mu), v\rangle + \langle \mathcal{H}(\mu)\vartheta_M(\mu), v\rangle \ , \qquad \forall\ v \in W_N^u \ . \tag{8.19}$$

The offline/online computational procedure to find $\vartheta_M(\mu)$, $\hat{u}_N(\mu)$, and $\hat{s}_N(\mu)$ is similar to that described in Chapters 3 and 5.

A Priori Theory

In this section, we find an *a priori* bound for $|s(\mu) - \hat{s}(\mu)|$. By way of preliminaries, we shall suppose that any member w of the temperature space Y_ϑ may be expressed as

$$w = \sum_{q=1}^{\tilde{Q}} w_{\tilde{Q}}^q \varphi_q \tag{8.20}$$

where the φ_q are basis functions of Y_ϑ, and $\tilde{Q}$ may be quite large. We may then write

$$\vartheta(\mu) - \vartheta_M(\mu) = \sum_{q=1}^{\tilde{Q}} \beta_q(\mu)\varphi_q \ . \tag{8.21}$$

Note that replacing (8.5) and (8.6) with a "truth approximation" defined over a finite element space $Y_{\mathcal{N}}$ of dimension $\mathcal{N}$ (as we do in actual practice) ensures that $\tilde{Q}$ may be presumed finite without loss of generality; we may, for instance, take $\tilde{Q} = \mathcal{N}$ and the φ_q to be the finite element nodal basis functions. (Recall that we assume $Y_{\mathcal{N}}$ is sufficiently rich such that $u_{\mathcal{N}}$, $s_{\mathcal{N}}$, $\hat{u}_{\mathcal{N}}$, and $\hat{s}_{\mathcal{N}}$ are indistinguishable from u, s, $\hat{u}$, and $\hat{s}$.)

Following [38], we also introduce $\tilde{Q}$ suitably regular open subdomains, $D_q \subset \Omega$, $1 \le q \le \tilde{Q}$, and assume that a given subdomain D_q intersects only finitely many other subdomains $D_{q'}$ (as $\tilde{Q} \to \infty$). We next define parameter-independent inner products and norms over D_q, $((\cdot,\cdot))_q$ and $|||\cdot|||_q \equiv ((\cdot,\cdot))_q^{1/2}$, respectively. We assume that $|||\cdot|||_q$ is uniformly equivalent to $\|\cdot\|_{H^1(D_q)}$ for all functions in $H^1(D_q)$ in the sense that the relevant constants may be bounded independent of μ and $\tilde{Q}$. It then follows from our assumptions that there exists a positive finite constant $\tilde{\rho}_\Sigma$ — *independent* of μ and $\tilde{Q}$ — such that

$$\sum_{q=1}^{\tilde{Q}} |||v|_{D_q}|||_q^2 < (\tilde{\rho}_\Sigma)^2 \|v\|_Y^2 \ , \quad \forall\, v \in Y \ . \tag{8.22}$$

We now define

$$\mathcal{E}_{L^2}(\mu) = \left(\sum_{q=1}^{\tilde{Q}} \beta_q^2(\mu) \int_\Omega \varphi_q^2 \right)^{1/2} \ , \tag{8.23}$$

and require that

$$\mathcal{E}_{L^2}(\mu) \le \tilde{\mathcal{E}} \tag{8.24}$$

for $\tilde{\mathcal{E}} \in \mathbb{R}_+$ independent of μ and $\tilde{Q}$ (and preferably small). We next define $\gamma_q \in \mathbb{R}$, $1 \le q \le \tilde{Q}$, as

$$\gamma_q(\mu) \equiv \frac{\sup_{v \in Y} \dfrac{\langle \mathcal{H}(\mu)\varphi_q, v\rangle}{|||v|_{D_q}|||_q}}{\left(\displaystyle\int_\Omega \varphi_q^2 \right)^{1/2}} \ , \tag{8.25}$$

and require that

$$\gamma_{\tilde{Q}}^{\max}(\mu) \equiv \max_{q \in \{1,\ldots,\tilde{Q}\}} \gamma_q(\mu) \le \tilde{\gamma} \ , \tag{8.26}$$

for $\tilde{\gamma}$ independent of μ.

We can now show that

$$|s(\mu) - \hat{s}(\mu)| \le \|L\|_{Y'} \left(C_1 \inf_{v_N \in W_N^u} \|\hat{u}(\mu) - v_N\|_Y + C_2\, \tilde{\rho}_\Sigma\, \tilde{\mathcal{E}}\, \tilde{\gamma} \right) \ , \quad \forall\, \mu \in \mathcal{D}^\mu \tag{8.27}$$

where C_1 and C_2 depend only on coercivity and continuity constants.

To prove (8.27), we first introduce $e(\mu) = u(\mu) - \hat{u}_N(\mu)$, and further define $\hat{e}(\mu) = \hat{u}(\mu) - \hat{u}_N(\mu)$,

and $\tilde{e}(\mu) = u(\mu) - \hat{u}(\mu)$, such that $e(\mu) = \hat{e}(\mu) + \tilde{e}(\mu)$. We can thus write

$$
\begin{aligned}
|s(\mu) - \hat{s}_N(\mu)| &= |\langle L(\mu), e(\mu)\rangle| && (8.28)\\
&= \left|\frac{\langle L, e(\mu)\rangle}{||e(\mu)||_Y}||e(\mu)||_Y\right| && (8.29)\\
&\leq \sup_{v\in Y}\frac{\langle L, v\rangle}{||v||_Y}||\hat{e}(\mu) + \tilde{e}(\mu)||_Y && (8.30)\\
&\leq ||L||_{Y'}\left(||\hat{e}(\mu)||_Y + ||\tilde{e}(\mu)||_Y\right) && (8.31)
\end{aligned}
$$

by the triangle inequality.

Furthermore, it immediately follows from standard Galerkin theory that

$$||\hat{e}(\mu)||_Y \leq C_1 \inf_{v\in W_N^u} ||\hat{u}(\mu) - v_N||_Y \ , \tag{8.32}$$

where C_1 depends only on the coercivity and continuity constants associated with $\mathcal{A}$.

To bound $||\tilde{e}(\mu)||_Y$, we note that

$$\langle \mathcal{A}(\mu)\tilde{e}(\mu), v\rangle = \langle \mathcal{H}(\mu)\left(\vartheta(\mu) - \vartheta_M(\mu)\right), v\rangle, \quad \forall\ v \in Y \ . \tag{8.33}$$

Expanding $\vartheta(\mu) - \vartheta_M(\mu)$, and choosing $v = \tilde{e}(\mu)$, we obtain

$$
\begin{aligned}
\langle \mathcal{A}(\mu)\tilde{e}(\mu), \tilde{e}(\mu)\rangle &= \sum_{q=1}^{\tilde{Q}} \beta_q(\mu)\langle \mathcal{H}(\mu)\phi_q, \tilde{e}(\mu)\rangle && (8.34)\\
&\leq \sum_{q=1}^{\tilde{Q}} |\beta_q(\mu)|\,\frac{\langle \mathcal{H}(\mu)\varphi_q, \tilde{e}(\mu)\rangle}{|||\ \tilde{e}(\mu)|_{D_q}\ |||_q}|||\ \tilde{e}(\mu)|_{D_q}\ |||_q && (8.35)\\
&\leq \sum_{q=1}^{\tilde{Q}} |\beta_q(\mu)|\left(\int_\Omega \varphi_q^2\right)^{1/2}\left(\frac{\sup_{v\in Y}\frac{\langle \mathcal{H}(\mu)\varphi_q, \tilde{v}\rangle}{|||\ \tilde{v}|_{D_q}\ |||_q}}{\left(\int_\Omega \varphi_q^2\right)^{1/2}}\right)|||\ \tilde{e}(\mu)|_{D_q}\ |||_q && (8.36)\\
&\leq \tilde{\gamma}\left(\sum_{q=1}^{\tilde{Q}} \beta_q^2(\mu)\int_\Omega \varphi_q^2\right)^{1/2}\left(\sum_{q=1}^{\tilde{Q}} |||\ \tilde{e}(\mu)|_{D_q}\ |||_q^2\right)^{1/2} && (8.37)\\
&\leq \tilde{\gamma}\,\tilde{\mathcal{E}}\,\tilde{\rho}_\Sigma\,||\tilde{e}(\mu)||_Y \ . && (8.38)
\end{aligned}
$$

Finally, we have

$$
\begin{aligned}
C_{\mathcal{A}}^0 &\leq \inf v \in Y \frac{\langle \mathcal{A}(\mu)v, v\rangle}{||v||_Y^2} && (8.39)\\
&\leq \frac{\langle \mathcal{A}(\mu)\tilde{e}(\mu), \tilde{e}(\mu)\rangle}{||\tilde{e}(\mu)||_Y^2} \ , && (8.40)
\end{aligned}
$$

from which it follows that

$$\langle \mathcal{A}(\mu)\tilde{e}(\mu), \tilde{e}(\mu)\rangle \leq C_2\,\tilde{\rho}_\Sigma\,\tilde{\mathcal{E}}\,\tilde{\gamma} \ , \quad \forall\ \mu \in \mathcal{D}^\mu \ , \tag{8.41}$$

where $C_2 = 1/C_{\mathcal{A}}^0$.

The bound (8.27) then follows from (8.31), (8.32), and (8.41). Note that we can, in fact, improve our bound to be quadratic (and not linear) in $||\hat{e}(\mu)||_Y$ by introduction of the reduced-basis adjoint techniques described in Chapter 5.

8.2.3 *A Posteriori* Error Estimation: Method I

We develop here the necessary *a posteriori* error bounds — a necessity for both efficiency and reliability.

Preliminaries

As in Chapters 4 and 5, we introduce a symmetric, coercive, continuous bound conditioner $\mathcal{C}(\mu): Y \rightarrow Y'$ that satisfies the spectral condition

$$1 \leq \frac{\langle \mathcal{A}(\mu)v, v\rangle}{\langle \mathcal{C}(\mu)v, v\rangle} \leq \rho, \quad \forall\, v \in Y, \forall\, \mu \in \mathcal{D}^\mu, \tag{8.42}$$

for some preferably small constant $\rho \in \mathbb{R}$, as well as the computationally invertibility hypothesis $\mathcal{C}^{-1}(\mu) = \sum_{i\in\mathcal{I}(\mu)} \alpha_i(\mu)\mathcal{C}_i^{-1}$. It then follows from the coercivity of $\mathcal{C}(\mu)$ and our assumptions in Section 8.2.2 that there exists a bound $\tilde{\rho}'_\Sigma(\mu)$ which is independent of $\tilde{Q}$, such that

$$\sum_{q=1}^{\tilde{Q}} |||v|_{D_q}|||_q^2 < (\tilde{\rho}'_\Sigma(\mu))^2 \langle \mathcal{C}(\mu)v, v\rangle \ , \quad \forall\, v \in Y \ . \tag{8.43}$$

We also introduce a symmetric, coercive, continuous bound conditioner for the thermal problem $\mathcal{C}_\vartheta(\mu): Y_\vartheta \rightarrow Y'_\vartheta$ that satisfies the (analogous) spectral and computational invertibility conditions. We then define

$$\Delta_M^\vartheta(\mu) = \langle R_\vartheta(\mu), \mathcal{C}_\vartheta^{-1}(\mu) R_\vartheta(\mu)\rangle \tag{8.44}$$

where $\langle R_\vartheta(\mu), v\rangle = \langle \mathcal{A}_\vartheta(\mu)\left(\vartheta(\mu) - \vartheta_M(\mu), v\right)\rangle$ for all $v \in Y_\vartheta$. From the results of Chapters 4 and 5, it follows that the error in the reduced-basis approximation for the temperature in the energy norm is bounded by $\Delta_M^\vartheta(\mu)$:

$$\langle \mathcal{A}_\vartheta(\mu)\left(\vartheta(\mu) - \vartheta_M(\mu)\right), \left(\vartheta(\mu) - \vartheta_M(\mu)\right)\rangle \ . \tag{8.45}$$

We next define the operator $\mathcal{D}$ as

$$\langle \mathcal{D}w, v\rangle = \sum_{q=1}^{\tilde{Q}} w_{\tilde{Q}}^q v_{\tilde{Q}}^q \int_\Omega \varphi_q^2 \, d\Omega \ , \quad \forall\, w, v \in Y_\vartheta \ , \tag{8.46}$$

where the φ_q are the nodal basis functions and $w_{\tilde{Q}}^q, v_{\tilde{Q}}^q$ the nodal values of w, v, respectively; and the eigenvalue $\lambda_{\min}^{L^2\text{-}H^1}(\mu)$ as

$$\lambda_{\min}^{L^2\text{-}H^1}(\mu) = \min_{v\in Y_\vartheta} \frac{\langle \mathcal{C}_\vartheta v, v\rangle}{\langle \mathcal{D}v, v\rangle}; \tag{8.47}$$

Furthermore, we have

$$\sup_{v\in Y}\frac{\langle\mathcal{H}(\mu)\varphi_q,v\rangle}{|||v|_{D_q}|||_q} = \sum_{k=1}^{Q_\mathcal{H}}\alpha_\mathcal{H}^k\sup_{v\in Y}\frac{\langle\mathcal{H}^k\varphi_q,v\rangle}{|||v|_{D_q}|||_q}\ . \tag{8.48}$$

Defining $z_{qk}\in Y_q$ as

$$z_{qk} = \arg\sup_{v\in Y}\frac{\langle\mathcal{H}^k\varphi_q,v\rangle}{|||v|_{D_q}|||_q}\ , \tag{8.49}$$

it follows that

$$((z_{qk},v|_{D_q})) = \langle\mathcal{H}^k\varphi_q,v\rangle,\quad\forall\ v\in Y_q\ . \tag{8.50}$$

We then have

$$\sup_{v\in Y}\left(\frac{\langle\mathcal{H}(\mu)\varphi_q,v\rangle}{|||v|_{D_q}|||_q}\right)^2 = \left(\sum_{k=1}^{Q_\mathcal{H}}\alpha_\mathcal{H}^k(\mu)\frac{\langle\mathcal{H}^k\varphi_q,z_{qk}\rangle}{|||z_{qk}|||_q}\right)^2 \tag{8.51}$$

$$= \left(\sum_{k=1}^{Q_\mathcal{H}}\alpha_\mathcal{H}^k((z_{qk},z_{qk}))^{1/2}\right)^2 \tag{8.52}$$

$$= \sum_{k=1}^{Q_\mathcal{H}}\sum_{k'=1}^{Q_\mathcal{H}}\alpha_\mathcal{H}^k\alpha_\mathcal{H}^{k'}((z_{qk},z_{qk}))^{1/2}((z_{qk'},z_{qk'}))^{1/2} \tag{8.53}$$

$$= \underline{a}^\vartheta(\mu)\underline{Z}^q\underline{a}(\mu)\ , \tag{8.54}$$

where $\underline{a}(\mu)\in\mathbb{R}^{Q_\mathcal{H}}$ and $\underline{Z}_q\in\mathbb{R}^{Q_\mathcal{H}\times Q_\mathcal{H}}$ are given by

$$\underline{a}(\mu) = \left[\alpha_\mathcal{H}^1(\mu)\ldots\alpha_\mathcal{H}^{Q_\mathcal{H}}(\mu)\right]^T \tag{8.55}$$

and

$$Z_{kk'}^q = ((z_{qk},z_{qk}))^{1/2}((z_{qk'},z_{qk'}))^{1/2}\ . \tag{8.56}$$

Finally, we let $\lambda_{\max}^{\underline{Z}^q}$ be the maximum eigenvalue of $\underline{Z}_q$,

$$\lambda_{\max}^{\underline{Z}^q} = \max_{\underline{v}\in\mathbb{R}^{Q_\mathcal{H}}}\frac{\underline{v}^T\underline{Z}_q\underline{v}}{\underline{v}^T\underline{v}} \tag{8.57}$$

and define

$$\Lambda_Z = \max_{q\in\{1,\ldots,\tilde{Q}\}}\left(\frac{\lambda_{\max}^{\underline{Z}^q}}{\int_\Omega\varphi_q^2}\right) \tag{8.58}$$

Error Bound

Our error estimator for $|s(\mu)-\hat{s}_N(\mu)|$ is then given by

$$\Delta_N(\mu)\equiv|||L|||_{Y'}\left(\hat{\delta}(\mu)+\tilde{\delta}(\mu)\right)\ , \tag{8.59}$$

where

$$|||L|||_{Y'} = \sup_{v \in Y} \frac{\langle L(\mu), v \rangle}{\langle \mathcal{C}(\mu)v, v \rangle^{1/2}} \tag{8.60}$$

$$\hat{\delta}(\mu) = \langle \mathcal{C}(\mu)\hat{E}(\mu), \hat{E}(\mu) \rangle^{1/2} \tag{8.61}$$

$$\tilde{\delta}(\mu) = \frac{\tilde{\rho}'_{\Sigma} \Delta_M^{\vartheta}(\mu) \Lambda_Z \left(\underline{a}^T(\mu) \underline{a}(\mu) \right)^{1/2}}{\lambda_{\min}^{L^2\text{-}H^1}(\mu)} \; . \tag{8.62}$$

Here $\hat{E}(\mu) \in Y$ satisfies

$$\langle \mathcal{C}(\mu)\hat{E}(\mu), v \rangle = \langle \mathcal{H}(\mu) \left(\vartheta(\mu) - \vartheta_M(\mu) \right), v \rangle, \quad \forall \; v \in Y \; . \tag{8.63}$$

Clearly, $\hat{\delta}(\mu)$ measures the error due to the reduced-basis approximation of $\hat{u}(\mu)$, and $\tilde{\delta}(\mu)$ measures the error due to model truncation — the error due to the reduced-basis approximation of $\vartheta(\mu)$ in the data.

We now show that our estimator $\Delta_N(\mu)$ is, in fact, a rigorous upper bound for $|s(\mu) - \hat{s}_N(\mu)|$. To begin, we note that

$$|s(\mu) - \hat{s}_N(\mu)| = |\langle L(\mu), e(\mu) \rangle| \tag{8.64}$$

$$\leq \left(\sup_{v \in X} \frac{\langle L(\mu), v \rangle}{\langle \mathcal{A}(\mu)v, v \rangle^{1/2}} \right) \langle \mathcal{A}(\mu)e(\mu), e(\mu) \rangle^{1/2} \tag{8.65}$$

$$\leq \left(\sup_{v \in X} \frac{\langle L(\mu), v \rangle}{\langle \mathcal{A}(\mu)v, v \rangle^{1/2}} \right) \left(\langle \mathcal{A}(\mu)\hat{e}(\mu), \hat{e}(\mu) \rangle^{1/2} + \langle \mathcal{A}(\mu)\tilde{e}(\mu), \tilde{e}(\mu) \rangle^{1/2} \right) \tag{8.66}$$

We next note from the results of Chapters 4 and 5 that

$$\langle \mathcal{A}(\mu)\hat{e}(\mu), \hat{e}(\mu) \rangle \leq \langle \mathcal{C}(\mu)\hat{E}(\mu), \hat{E}(\mu) \rangle \tag{8.67}$$

It follows that

$$|s(\mu) - \hat{s}_N(\mu)| \leq |||L|||_{Y'} \left(\hat{\delta}(\mu) + \langle \mathcal{A}(\mu)\tilde{e}(\mu), \tilde{e}(\mu) \rangle^{1/2} \right) \; ; \tag{8.68}$$

it thus remains only to show that $\langle \mathcal{A}(\mu)\tilde{e}(\mu), \tilde{e}(\mu) \rangle^{1/2} \leq \tilde{\delta}(\mu)$.

We first define $\tilde{E}(\mu) \in Y$ as the solution to

$$\langle \mathcal{C}(\mu)\tilde{E}(\mu), v \rangle = \langle \mathcal{H}(\mu) \left(\vartheta(\mu) - \vartheta_M(\mu) \right), v \rangle, \quad \forall \; v \in Y \; ; \tag{8.69}$$

it thus follows that

$$\langle \mathcal{A}(\mu)\tilde{e}(\mu), \tilde{e}(\mu) \rangle^{1/2} \leq \langle \mathcal{C}(\mu)\tilde{E}(\mu), \tilde{E}(\mu) \rangle^{1/2} \; . \tag{8.70}$$

We now note that

$$\begin{aligned}
\langle C(\mu)\tilde{E}(\mu),\tilde{E}(\mu)\rangle &= \langle \mathcal{H}(\mu)\left(\vartheta(\mu)-\vartheta_M(\mu)\right),\tilde{E}(\mu)\rangle && (8.71)\\
&\leq \sum_{q=1}^{\tilde{Q}} |\beta_q(\mu)| \frac{\langle \mathcal{H}(\mu)\varphi_q,\tilde{E}(\mu)\rangle}{|||\tilde{E}(\mu)\big|_{D_q}|||_q} |||\tilde{E}(\mu)|_{D_q}|||_q && (8.72)\\
&\leq \sum_{q=1}^{\tilde{Q}} |\beta_q(\mu)| \left(\int_\Omega \varphi_q^2\right)^{1/2} \left(\frac{\sup_{v\in Y}\frac{\langle \mathcal{H}(\mu)\varphi_q,\tilde{v}\rangle}{|||v|_{D_q}|||_q}}{\left(\int_\Omega \varphi_q^2\right)^{1/2}}\right) |||\tilde{E}(\mu)\big|_{D_q}|||_q && (8.73)\\
&= \sum_{q=1}^{\tilde{Q}} |\beta_q(\mu)| \left(\int_\Omega \varphi_q^2\right)^{1/2} \gamma^q(\mu) |||\tilde{E}(\mu)\big|_{D_q}|||_q && (8.74)\\
&\leq \gamma_{\tilde{Q}}^{\max}(\mu)\left(\sum_{q=1}^{\tilde{Q}}\beta_q^2(\mu)\int_\Omega\varphi_q^2\right)^{1/2}\left(\sum_{q=1}^{\tilde{Q}}|||\tilde{E}(\mu)\big|_{D_q}|||_q^2\right)^{1/2} && (8.75)\\
&\leq \gamma_{\tilde{Q}}^{\max}(\mu)\tilde{\mathcal{E}}_{L^2}(\mu)\tilde{\rho}'_\Sigma(\mu)\,\langle \mathcal{C}(\mu)\tilde{E}(\mu),\tilde{E}(\mu)\rangle^{1/2}\,, && (8.76)
\end{aligned}$$

and therefore

$$\langle \mathcal{C}(\mu)\tilde{E}(\mu),\tilde{E}(\mu)\rangle^{1/2} \leq \gamma_{\tilde{Q}}^{\max}(\mu)\tilde{\mathcal{E}}_{L^2}(\mu)\tilde{\rho}'_\Sigma(\mu)\,. \qquad (8.77)$$

From (8.21) and (8.46) we have

$$\begin{aligned}
\mathcal{E}_{L^2}(\mu) &= \sum_{q=1}^{\tilde{Q}}\beta_q^2(\mu)\int_\Omega\varphi_q^2 && (8.78)\\
&= \langle \mathcal{D}\left(\vartheta(\mu)-\vartheta_M(\mu)\right),\left(\vartheta(\mu)-\vartheta_M(\mu)\right)\rangle^{1/2} && (8.79)\\
&\leq \left(\sup_{v\in Y_\vartheta}\frac{\langle \mathcal{D}v,v\rangle}{\langle \mathcal{A}_\vartheta(\mu)v,v\rangle^{1/2}}\right)\langle \mathcal{A}_\vartheta(\mu)\left(\vartheta(\mu)-\vartheta_M(\mu)\right),\left(\vartheta(\mu)-\vartheta_M(\mu)\right)\rangle^{1/2} && (8.80)\\
&\leq \left(\sup_{v\in Y_\vartheta}\frac{\langle \mathcal{D}v,v\rangle}{\langle \mathcal{C}_\vartheta(\mu)v,v\rangle^{1/2}}\right)\Delta_M^\vartheta(\mu) && (8.81)\\
&= \frac{\Delta_M^\vartheta(\mu)}{\lambda_{\min}^{L^2\text{-}H^1}}\,. && (8.82)
\end{aligned}$$

We next note that

$$\left(\gamma_{\tilde{Q}}^{\max}(\mu)\right)^2 = \max_{q\in\{1,\ldots,\tilde{Q}\}} (\gamma^q(\mu))^2 \tag{8.83}$$

$$= \max_{q\in\{1,\ldots,\tilde{Q}\}} \frac{\underline{a}^T(\mu)\underline{Z}^q\underline{a}(\mu)}{\int_\Omega \varphi_q^2} \tag{8.84}$$

$$= \max_{q\in\{1,\ldots,\tilde{Q}\}} \frac{1}{\int_\Omega \varphi_q^2}\left(\frac{\underline{a}^T(\mu)\underline{Z}^q\underline{a}(\mu)}{\underline{a}^T(\mu)\underline{a}(\mu)}\right)\underline{a}^T(\mu)\underline{a}(\mu) \tag{8.85}$$

$$\leq \underline{a}^T(\mu)\underline{a}(\mu) \max_{q\in\{1,\ldots,\tilde{Q}\}} \frac{1}{\int_\Omega \varphi_q^2} \max_{\underline{v}\in\mathbb{R}^{Q_\mathcal{H}}} \frac{\underline{v}^T\underline{Z}^q\underline{v}}{\underline{v}^T\underline{v}} \tag{8.86}$$

$$\leq \underline{a}^T(\mu)\underline{a}(\mu) \max_{q\in\{1,\ldots,\tilde{Q}\}} \frac{1}{\int_\Omega \varphi_q^2}\lambda_{\max}^{\underline{Z}^q} \tag{8.87}$$

$$= \underline{a}^T(\mu)\underline{a}(\mu)\Lambda_Z^2 \tag{8.88}$$

Finally, we have

$$\langle \mathcal{A}(\mu)\tilde{e}(\mu),\tilde{e}(\mu)\rangle^{1/2} \leq \langle \mathcal{C}(\mu)\tilde{E}(\mu),\tilde{E}(\mu)\rangle^{1/2} \tag{8.89}$$

$$\leq \frac{\rho'_\Sigma(\mu)\Delta_M^\vartheta(\mu)\Lambda_Z\left(\underline{a}^T(\mu)\underline{a}(\mu)\right)^{1/2}}{\lambda_{\min}^{L^2\text{-}H^1}} \tag{8.90}$$

$$= \tilde{\delta}(\mu)\ . \tag{8.91}$$

We note that $\lambda_{\min}^{L^2\text{-}H^1}$ and Λ_Z can be calculated offline, while $\Delta_M^\vartheta(\mu)$ and $\underline{a}^T(\mu)$ may be calculated online.

8.3 Noncoercive Problems: The Reduced-Wave (Helmholtz) Equation

There are many important problems for which the coercivity of $\mathcal{A}$ is lost — a representative example is the Helmholtz, or reduced–wave, equation. For noncoercive problems, well–posedness is now ensured only by the inf–sup condition. This "weaker" stability condition presents difficulties in both approximation and error estimation; we shall address the latter in this section. In particular, we present some preliminary ideas for constructing approximations to the inf-sup parameter $\beta_\mathcal{A}(\mu)$ for (i) noncoercive problems in general, through application of certain bound conditioner constructions presented in Chapter 3; and (ii) the Helmholtz problem in particular, through ...

8.3.1 Abstract Formulation

We again consider a suitably regular (smooth) domain $\Omega \subset \mathbb{R}^d$, $d = 1, 2$, or 3, and associated function space $Y \subset (H^1(\Omega))^p$ with associated inner product and norm $(\cdot,\cdot)_Y$ and $||\cdot||_Y = (\cdot,\cdot)_Y^{1/2}$, respectively. We recall that the dual space of Y, Y', is defined as the set of all linear functionals F

such that the dual norm of F, defined as $||F||_{Y'}$,

$$||F||_{Y'} = \sup_{v \in Y} \frac{\langle F, v \rangle}{||v||_Y} \ , \tag{8.92}$$

is bounded.

We then consider the problem: for any $\mu \in \mathcal{D}^\mu$, find $s(\mu) = \langle L(\mu), u(\mu) \rangle$ where $u(\mu)$ satisfies

$$\langle \mathcal{A}(\mu) u(\mu), v \rangle = \langle F(\mu), v \rangle, \quad \forall \ v \in Y \ , \tag{8.93}$$

where the operator $\mathcal{A}(\mu)\colon Y \to Y'$ is symmetric, $\langle \mathcal{A}(\mu) w, v \rangle = \langle \mathcal{A}(\mu) v, w \rangle$, continuous,

$$\langle \mathcal{A}(\mu) w, v \rangle \leq \gamma_{\mathcal{A}}(\mu) ||w||_Y ||v||_Y \leq \gamma^0_{\mathcal{A}} ||w||_Y ||v||_Y, \quad \forall \ w, v \in Y, \ \forall \ \mu \in \mathcal{D}^\mu \ , \tag{8.94}$$

and depends affinely on the parameter

$$\mathcal{A}(\mu) = \sum_{q=1}^{Q_{\mathcal{A}}} \Theta^q(\mu) \mathcal{A}^q; \tag{8.95}$$

we further assume that $F(\mu) \in Y'$, $L(\mu) \in Y'$, and that F, L also depend affinely on the μ. In addition to the primal problem (8.93), we shall again require the dual problem

$$\langle \mathcal{A}(\mu) v, \psi(\mu) \rangle = \langle L(\mu), v \rangle, \quad \forall \ v \in Y \ , \tag{8.96}$$

where $\psi(\mu) \in Y$.

In this section, we no longer assume that $\mathcal{A}(\mu)$ is coercive, but instead ensure well-posedness by the inf-sup stability condition

$$0 < \beta^0_{\mathcal{A}} \leq \beta_{\mathcal{A}}(\mu) = \inf_{w \in Y} \sup_{v \in Y} \frac{\langle \mathcal{A}(\mu) w, v \rangle}{||w||_Y ||v||_Y}, \quad \forall \ \mu \in \mathcal{D}^\mu \ ; \tag{8.97}$$

Several 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. 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. Furthermore, loss of stability can lead to poor approximations — the inf-sup parameter enters in the denominator of the *a priori* convergence result (3.92). Nevertheless, it is posssible to resolve these issues by considering projections other than standard Galerkin, and "enriched" approximation spaces [26, 43]. The second numerical difficulty — which we shall address here — is estimation of the inf-sup parameter $\beta_{\mathcal{A}}(\mu)$, important for certain classes of Method I *a posteriori* techniques. Since $\beta_{\mathcal{A}}(\mu)$ can not typically be deduced analytically, it must therefore be approximated. To motivate our methods, we first consider an example.

8.3.2 Formulation of the Helmholtz Problem

There are many important applications in which the Helmholtz equation takes a central role — nondestructive evaluation, structural dynamics, and electromagnetics, for example.

We consider an isotropic body subjected to a harmonic loading such that the harmonic response

u is governed by the partial differential equation

$$\nabla^2 u + \omega^2 u = b, \qquad \text{in } \Omega \ , \tag{8.98}$$

with boundary conditions

$$u = 0 \qquad \text{on } \Gamma_{\rm D} \tag{8.99}$$

$$\frac{\partial u}{\partial x_j} \mathrm{e}_j^n = f \qquad \text{on } \Gamma_{\rm N} \ . \tag{8.100}$$

Furthermore, we assume that the output of interest is a linear functional of the response,

$$s(\mu) = \langle L(\mu), u(\mu) \rangle, \qquad \text{for } \mu = \{\omega\} \in \mathcal{D}^\mu \ . \tag{8.101}$$

The problem may then be (weakly) formulated as: find $s(\mu) = \langle L(\mu), u(\mu) \rangle$, where $u(\mu) \in\in Y = \{v \in H_1(\Omega) \mid v = 0 \text{ on } \Gamma_{\rm D}\}$ is the solution to

$$\langle \mathcal{A}(\mu) u(\mu), v \rangle = \langle F(\mu), v \rangle, \forall\ v \in Y \ ; \tag{8.102}$$

for this example,

$$\langle \mathcal{A}(\mu) w, v \rangle = \int_\Omega \frac{\partial v}{\partial x_i} \frac{\partial w}{\partial x_i} \, d\Omega + \omega^2 \int_\Omega w \, v \, d\Omega \ , \tag{8.103}$$

and

$$\langle F(\mu), v \rangle = \int_\Omega b \, v \, d\Omega + \int_\Gamma f v \, d\Gamma \ . \tag{8.104}$$

Affine parameter dependence of $\mathcal{A}(\mu)$ is then obtained for $Q_\mathcal{A} = 2$, and

$$\Theta^1(\mu) = 1, \qquad \langle \mathcal{A}^1 w, v \rangle = \int_\Omega \frac{\partial v}{\partial x_i} \frac{\partial w}{\partial x_i} \, d\Omega \ , \tag{8.105}$$

$$\Theta^2(\mu) = \omega^2, \qquad \langle \mathcal{A}^2 w, v \rangle = \int_\Omega w \, v \, d\Omega \ . \tag{8.106}$$

8.3.3 Reduced-Basis Approximation and Error Estimation

Following [43], we assume that we have the reduced-basis approximations $u_N(\mu) \in W_N$ and $\psi_N(\mu) \in W_N$ given by

$$u_N(\mu) = \arg \inf_{w \in W_N} \sup_{v \in V_N} \frac{\langle A(\mu)(u(\mu) - w), v \rangle}{||v||_Y} \ , \tag{8.107}$$

$$\psi_N(\mu) = \arg \inf_{w \in W_N} \sup_{v \in V_N} \frac{\langle A(\mu) v, (\psi(\mu) - w) \rangle}{||v||_Y} \ , \tag{8.108}$$

where W_N and V_N are the appropriate infimizing and supremizing spaces, respectively. Our output approximation is then

$$s_N(\mu) = \langle L(\mu), u_N(\mu) \rangle \ , \tag{8.109}$$

and the associated upper and lower bounds are $s_N^\pm(\mu) = s_N(\mu) \pm \Delta_N(\mu)$ where

$$\Delta_N(\mu) = \frac{1}{\beta_\mathcal{A}^{\rm LB}(\mu)} ||R^{\rm pr}(\mu)||_{Y'} ||R^{\rm du}(\mu)||_{Y'} \ , \tag{8.110}$$

and

$$0 < \beta_{\mathcal{A}}^{\mathrm{LB}}(\mu) \leq \beta_{\mathcal{A}}(\mu) \ ; \tag{8.111}$$

we assume for the moment that a lower bound $\beta_{\mathcal{A}}^{\mathrm{LB}}(\mu)$ can be readily found.

We now show that $\Delta_N(\mu)$ is indeed an error *bound*, that is,

$$|s(\mu) - s_N(\mu)| \leq \Delta_N(\mu) \ . \tag{8.112}$$

To begin, we define $T_\mu w$ as

$$T_\mu w = \arg\sup_{v \in Y} \frac{\langle \mathcal{A}(\mu) w, v \rangle}{||v||_Y} \ ; \tag{8.113}$$

it then follows that $T_\mu w$ satisfies

$$(T_\mu w, v)_Y = \langle \mathcal{A}(\mu) w, v \rangle, \quad \forall \ v \in Y \ , \tag{8.114}$$

and therefore

$$\begin{aligned}
||R^{\mathrm{pr}}(\mu)||_{Y'} &= \sup_{v \in Y} \frac{\langle R^{\mathrm{pr}}(\mu), v \rangle}{||v||_Y} && (8.115)\\
&= \sup_{v \in Y} \frac{\langle \mathcal{A}(\mu)(u(\mu) - u_N(\mu)), v \rangle}{||v||_Y} && (8.116)\\
&= \frac{\langle \mathcal{A}(\mu)(u(\mu) - u_N(\mu)), T_\mu(u(\mu) - u_N(\mu)) \rangle}{||T_\mu(u(\mu) - u_N(\mu))||_Y} && (8.117)\\
&= \frac{(T_\mu(u(\mu) - u_N(\mu)), T_\mu(u(\mu) - u_N(\mu)))_Y}{||T_\mu(u(\mu) - u_N(\mu))||_Y} && (8.118)\\
&= ||T_\mu(u(\mu) - u_N(\mu))||_Y \ . && (8.119)
\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
&\beta_{\mathcal{A}}(\mu)||\psi(\mu) - \psi_N(\mu)||_Y ||T_\mu(\psi(\mu) - \psi_N(\mu))||_Y \\
&\qquad \leq \langle \mathcal{A}(\mu) T_\mu(\psi(\mu) - \psi_N(\mu)), (\psi(\mu) - \psi_N(\mu)) \rangle && (8.120)\\
&\qquad = \langle R^{\mathrm{du}}(\mu), T_\mu(\psi(\mu) - \psi_N(\mu)) \rangle && (8.121)\\
&\qquad = \frac{\left| \langle R^{\mathrm{du}}(\mu), T_\mu(\psi(\mu) - \psi_N(\mu)) \rangle \right|}{||T_\mu(\psi(\mu) - \psi_N(\mu))||_{Y'}} ||T_\mu(\psi(\mu) - \psi_N(\mu))||_{Y'} && (8.122)\\
&\qquad \leq \sup_{v \in Y} \frac{\langle R^{\mathrm{du}}(\mu), v \rangle}{||v||_{Y'}} ||T_\mu(\psi(\mu) - \psi_N(\mu))||_{Y'} && (8.123)\\
&\qquad = ||R^{\mathrm{du}}(\mu)||_{Y'} ||T_\mu(\psi(\mu) - \psi_N(\mu))||_{Y'} \ , && (8.124)
\end{aligned}$$

and therefore

$$||\psi(\mu) - \psi_N(\mu)||_Y \leq \frac{1}{\beta_{\mathcal{A}}(\mu)} ||R^{\mathrm{du}}(\mu)||_{Y'} \ . \tag{8.125}$$

Finally, we have

$$
\begin{aligned}
|s(\mu) - s_N(\mu)| &= |\langle \mathcal{A}(\mu)(u(\mu) - u_N(\mu)), (\psi(\mu) - \psi_N(\mu))\rangle| && (8.126)\\
&= |\langle R^{\rm pr}(\mu), (\psi(\mu) - \psi_N(\mu))\rangle| && (8.127)\\
&= \frac{|\langle R^{\rm pr}(\mu), (\psi(\mu) - \psi_N(\mu))\rangle|}{||T_\mu(\psi(\mu) - \psi_N(\mu))||_Y} ||T_\mu(\psi(\mu) - \psi_N(\mu))||_Y && (8.128)\\
&\leq \sup_{v\in Y} \frac{\langle R^{\rm pr}(\mu), v\rangle}{||v||_{Y'}} ||\psi(\mu) - \psi_N(\mu)||_Y && (8.129)\\
&= ||R^{\rm pr}(\mu)||_{Y'} ||\psi(\mu) - \psi_N(\mu)||_Y && (8.130)\\
&\leq \frac{1}{\beta_{\mathcal{A}}(\mu)} ||R^{\rm pr}(\mu)||_{Y'} ||R^{\rm du}(\mu)||_{Y'} && (8.131)\\
&\leq \frac{1}{\beta^{\rm LB}_{\mathcal{A}}(\mu)} ||R^{\rm pr}(\mu)||_{Y'} ||R^{\rm du}(\mu)||_{Y'}; && (8.132)
\end{aligned}
$$

this completes the proof.

Approximation of the Stability Constant: Eigenvalue Interpolation

Since the stability constant $\beta_{\mathcal{A}}(\mu)$ can not typically be deduced analytically, an approximation — more precisely, a *lower bound* — must be calculated. We present here a method for calculating the requisite lower bound based on the the concave-eigenvalue interpolation methods of Chapter 4.

To begin, we assume that the inner product associated with the function space Y is given by

$$
(w, v)_Y \equiv \langle \hat{\mathcal{C}} w, v\rangle \ , \tag{8.133}
$$

where $\hat{\mathcal{C}}$: $Y \rightarrow Y'$ is a μ-independent, symmetric, continuous, and coercive operator; the associated norm is then $||v||_Y = \langle \hat{\mathcal{C}} v, v\rangle^{1/2}$. It then follows from (8.114) that

$$
\langle \hat{\mathcal{C}}(T_\mu w), v\rangle = \langle \mathcal{A}(\mu) w, v\rangle \ , \qquad \forall\ v \in Y \ , \tag{8.134}
$$

and therefore $T_\mu w = \hat{\mathcal{C}}^{-1}\mathcal{A}(\mu) w$. We then note that

$$
\begin{aligned}
(\beta_{\mathcal{A}}(\mu))^2 &= \inf_{w\in Y} \frac{\langle \mathcal{A}(\mu) w, T_\mu w\rangle^2}{||w||_Y^2 ||T_\mu w||_Y^2} && (8.135)\\
&= \inf_{w\in Y} \frac{\langle \mathcal{A}(\mu) w, T_\mu w\rangle\ (T_\mu w, T_\mu w)_Y}{(w, w)_Y (T_\mu w, T_\mu w)_Y} && (8.136)\\
&= \inf_{w\in Y} \frac{\langle \mathcal{A}(\mu) w, \hat{\mathcal{C}}^{-1}\mathcal{A}(\mu) w\rangle}{\langle \hat{\mathcal{C}} v, v\rangle} && (8.137)\\
&= \inf_{w\in Y} \frac{\langle \mathcal{L}(\mu) w, w\rangle}{\langle \hat{\mathcal{C}} v, v\rangle} \ , && (8.138)
\end{aligned}
$$

where

$$
\mathcal{L}(\mu) \equiv (\hat{\mathcal{C}}^{-1}\mathcal{A}(\mu))^T \mathcal{A}(\mu) = \mathcal{A}(\mu)\hat{\mathcal{C}}^{-1}\mathcal{A}(\mu) \ . \tag{8.139}
$$

We also note from our assumption of affine parameter dependence (8.95) that $\mathcal{L}(\mu)$ is also affine

in μ since

$$\mathcal{L}(\mu) = \left(\sum_{q=1}^{Q_\mathcal{A}} \Theta^q(\mu)\mathcal{A}^q\right) \hat{\mathcal{C}}^{-1} \left(\sum_{q'=1}^{Q_\mathcal{A}} \Theta^{q'}(\mu)\mathcal{A}^{q'}\right) \tag{8.140}$$

$$= \underbrace{\left(\Theta^1(\mu)\right)^2}_{\Phi^1(\mu)} \underbrace{\left[\mathcal{A}^1\hat{\mathcal{C}}^{-1}\mathcal{A}^1\right]}_{\mathcal{L}^1} + \underbrace{\left(2\Theta^1(\mu)\Theta^2(\mu)\right)}_{\Phi^2(\mu)} \underbrace{\left[\mathcal{A}^1\hat{\mathcal{C}}^{-1}\mathcal{A}^2\right]}_{\mathcal{L}^2} + \cdots \tag{8.141}$$

$$+ \underbrace{\left(2\Theta^{Q_\mathcal{A}-1}\Theta^{Q_\mathcal{A}}\right)}_{\Phi^{Q_\mathcal{L}-1}} \underbrace{\left[\mathcal{A}^{Q_\mathcal{A}-1}\hat{\mathcal{C}}^{-1}\mathcal{A}^{Q_\mathcal{A}}\right]}_{\mathcal{L}^{Q_\mathcal{L}-1}} + \underbrace{\left(\Theta^{Q_\mathcal{A}}\right)^2}_{\Phi^{Q_\mathcal{L}}} \underbrace{\left[\mathcal{A}^{Q_\mathcal{A}}\hat{\mathcal{C}}^{-1}\mathcal{A}^{Q_\mathcal{A}}\right]}_{\mathcal{L}^{Q_\mathcal{L}}} \tag{8.142}$$

$$\equiv \sum_{q=1}^{Q_\mathcal{L}} \Phi^q(\mu)\mathcal{L}^q \ , \tag{8.143}$$

where $Q_\mathcal{L} = Q_\mathcal{A}(Q_\mathcal{A}+1)/2$. We now define

$$L(\phi) \equiv \sum_{q=1}^{Q_\mathcal{L}} \phi^q \, L^q \tag{8.144}$$

where $\phi \in \mathbb{R}^{Q_\mathcal{L}}$, $L(\phi)\colon Y \to Y'$, and the $L^q \equiv \mathcal{L}^q$ are symmetric and continuous

$$\langle L^q w, v\rangle \le \gamma_{L^q}(\mu)||w||_Y||v||_Y \le \gamma^0_{L^q}||w||_Y||v||_Y, \quad \forall\, w, v \in Y \ . \tag{8.145}$$

We may then write

$$L(\Phi(\mu)) = \mathcal{L}(\mu) \ . \tag{8.146}$$

It then follows that the stability constant $\beta_\mathcal{A}(\mu)$ is related to the eigenvalue problem

$$\langle L(\phi)\xi(\phi), v\rangle = \lambda(\phi)\langle\hat{\mathcal{C}}\xi(\phi), v\rangle \tag{8.147}$$

In particular,

$$\beta_\mathcal{A}(\mu) = (\lambda_1(\Phi(\mu)))^{1/2} \ , \tag{8.148}$$

where λ_1 is the smallest eigenvalue satisfying (8.120). Furthermore, we know from Chapter 4 that λ_1 is concave in ϕ. We can thus construct a lower bound $\lambda_1^{\mathrm{LB}}(\Phi(\mu))$ to $\lambda_1(\Phi(\mu))$ as the convex combination of $\lambda_1(\phi^k)$:

$$\lambda_1^{\mathrm{LB}}(\Phi(\mu)) = \sum_{k\in\mathcal{T}(\Phi(\mu))} \alpha^k \lambda(\phi^k) \tag{8.149}$$

such that the $\alpha^k > 0$, and

$$\sum_{k\in\mathcal{T}(\Phi(\mu))} \alpha^k = 1 \ , \tag{8.150}$$

$$\sum_{k\in\mathcal{T}(\Phi(\mu))} \alpha^k \phi^k = \Phi(\mu) \ ; \tag{8.151}$$

The problem of finding the best (largest) convex combination $\lambda^{\mathrm{LB}}_{1\,\max}(\Phi(\mu))$ is again a linear programming problem (which may or may not be feasible). Clearly, the choice $\beta^{\mathrm{LB}}_\mathcal{A}(\mu) = (\lambda^{\mathrm{LB}}_{1\,\max}(\Phi(\mu)))^{1/2}$

then satisfies (8.111).

The potential difficulty with (8.149) is that the off-line stage may be expensive and also complicated, and relatedly, that there is little a priori guidance (for the ϕ^k or $\mathcal{T}(\Phi(\mu))$). Note in this noncoercive case the relevant operator yields $Q_{\mathcal{L}} = O(Q_{\mathcal{A}}^2)$ coefficients, thus aggravating the problems already identified in Chapter 4 — negative lower bounds, fine lookup tables, large Q.

Appendix A

Elementary Affine Geometric Transformations

We consider the case $\bar{\Omega}(\mu), \Omega \subset \mathbb{R}^d, d = 2$; the extension to $d = 1$ and $d = 3$ is straightforward.

Stretch

We consider the mapping from $\bar{\Omega}(\mu) \rightarrow \Omega$ shown in Figure A-1(a) and A-1(f). The forward and inverse affine mappings are then given by (here, α is the angle of $\bar{\Omega}(\mu)$, and the angle of Ω is 0.)

$$\underline{G}(\mu) = \begin{bmatrix} \frac{t_1}{\bar{t}_1(\mu)} & \cdot \\ \cdot & \frac{t_2}{\bar{t}_2(\mu)} \end{bmatrix}, \qquad \underline{g}(\mu) = \underline{0}\,, \tag{A.1}$$

$$\underline{G}^{-1}(\mu) = \begin{bmatrix} \frac{\bar{t}_1(\mu)}{t_1} & \cdot \\ \cdot & \frac{\bar{t}_2(\mu)}{t_2} \end{bmatrix}, \qquad \underline{g}'(\mu) = \underline{0}\,. \tag{A.2}$$

Horizontal Shear

We now consider the mapping from $\bar{\Omega}(\mu) \rightarrow \Omega$ shown in Figure A-1(b) and A-1(f). The forward and inverse affine mappings are then given by

$$\underline{G}(\mu) = \begin{bmatrix} 1 & -\tan\alpha \\ \cdot & 1 \end{bmatrix}, \qquad \underline{g}(\mu) = \underline{0}\,, \tag{A.3}$$

$$\underline{G}^{-1}(\mu) = \begin{bmatrix} 1 & \tan\alpha \\ \cdot & 1 \end{bmatrix}, \qquad \underline{g}'(\mu) = \underline{0}\,. \tag{A.4}$$

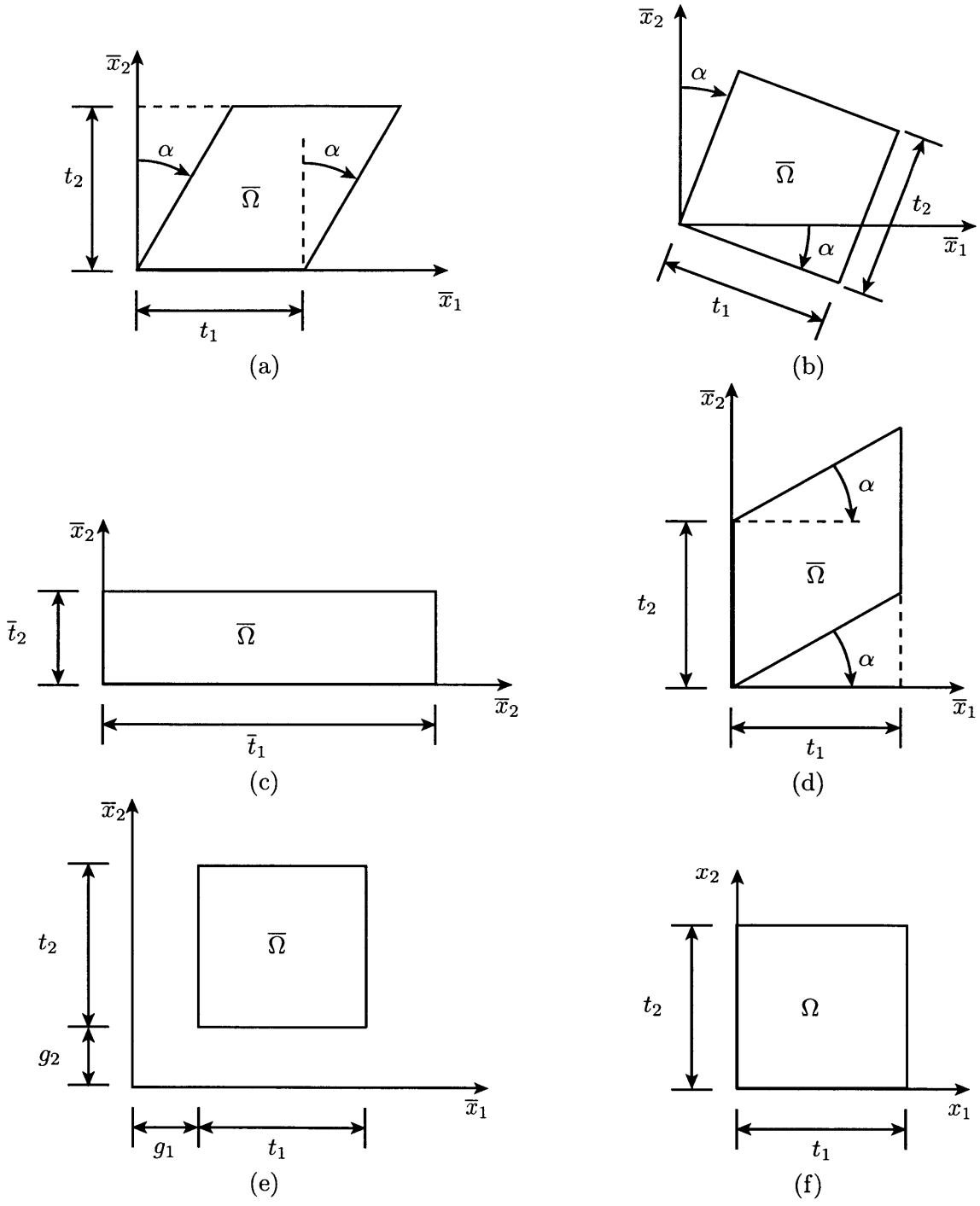


Figure A-1: Elementary two-dimensional affine transformations — (a) stretch, (b) horizontal (x_1-direction) shear , (c) vertical (x_2-direction) shear, (d) rotation, and (e) translation — between $\bar{\Omega}$, shown in (a)-(e), and Ω, shown in (f).

Vertical Shear

We consider the mapping from $\bar{\Omega}(\mu) \rightarrow \Omega$ shown in Figure A-1(c) and A-1(f). The forward and inverse affine mappings are then given by

$$\underline{G}(\mu) = \begin{bmatrix} 1 & \cdot \\ -\tan\alpha & 1 \end{bmatrix}, \quad \underline{g}(\mu) = \underline{0}\,, \tag{A.5}$$

$$\underline{G}^{-1}(\mu) = \begin{bmatrix} 1 & \cdot \\ \tan\alpha & 1 \end{bmatrix}, \quad \underline{g}'(\mu) = \underline{0}\,. \tag{A.6}$$

Rotation

We now consider the mapping from $\bar{\Omega}(\mu) \rightarrow \Omega$ shown in Figure A-1(d) and A-1(f). The forward and inverse affine mappings are then given by

$$\underline{G}(\mu) = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}, \quad \underline{g}(\mu) = \underline{0}\,, \tag{A.7}$$

$$\underline{G}^{-1}(\mu) = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}, \quad \underline{g}'(\mu) = \underline{0}\,. \tag{A.8}$$

Translation

We now consider the mapping from $\bar{\Omega}(\mu) \rightarrow \Omega$ shown in Figure A-1(e) and A-1(f). The forward and inverse affine mappings are then given by

$$\underline{G}(\mu) = \underline{0}\,, \quad \underline{g}\,(\mu) = \begin{bmatrix} -c_1 \\ -c_2 \end{bmatrix}, \tag{A.9}$$

$$\underline{G}^{-1}(\mu) = \underline{0}\,, \quad \underline{g}'(\mu) = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}. \tag{A.10}$$

Bibliography

[1] M. A. Akgun, J. H. Garcelon, and R. T. Haftka. Fast exact linear and non-linear structural reanalysis and the Sherman-Morrison-Woodbury formulas. *International Journal for Numerical Methods in Engineering*, 50(7):1587–1606, March 2001.

[2] S. Ali. *Real-time Optimal Parametric Design using the Assess-Predict-Optimize Strategy.* PhD thesis, Singapore-MIT Alliance, Nanyang Technological University, Singapore, 2003. In progress.

[3] E. Allgower and K. Georg. Simplicial and continuation methods for approximating fixed-points and solutions to systems of equations. *SIAM Review*, 22(1):28–85, 1980.

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

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

[6] M. Avriel. *Nonlinear Programming: Analysis and Methods.* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1976.

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

[8] A. Barrett and G. Reddien. On the reduced basis method. *Z. Angew. Math. Mech.*, 75(7):543–549, 1995.

[9] D. Bertsekas. *Nonlinear Programming.* Athena Scientific, Belmont, MA, 2nd edition, 1999.

[10] T. F. Chan and W. L. Wan. Analysis of projection methods for solving linear systems with multiple right-hand sides. *SIAM Journal on Scientific Computing*, 18(6):1698, 1721 1997.

[11] A. G. Evans, J. W. Hutchinson, N.A. Fleck, M. F. Ashby, and H. N. G. Wadley. The topological design of multifunctional cellular metals. *Progress in Materials Science*, 46(3-4):309–327, 2001.

[12] A.G. Evans. Private communication, 2001-2002.

[13] C. Farhat and F. X. Roux. Implicit parallel processing in structural mechanics. Technical Report CU-CSSC-93-26, Center for Aerospace Structures, University of Colorado, Boulder, CO, 1993.

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

[15] A. Forsgren, P.E. Gill, and M.H. Wright. Interior methods for nonlinear optimization. *SIAM Review*, 44(4):535–597, 2002.

[16] L. J. Gibson and M. F. Ashby. *Cellular Solids: Structure and Properties.* Cambridge University Press, 2nd edition, 1997.

[17] M. Grepl. PhD thesis, Massachusetts Institute of Technology, 2005. In progress.

[18] J. W. Hutchinson. Plastic buckling. In C.-S. Yih, editor, *Advances in Applied Mechanics*, volume 14, pages 67–144. Academic Press, Inc., 1974.

[19] J.W. Hutchinson. Private communication, 2001-2002.

[20] T. Leurent. Reduced basis output bounds for linear elasticity: Application to microtruss structures. Master's thesis, Massachusetts Institute of Technology, 2001.

[21] L. Machiels, Y. Maday, I. B. Oliveira, A.T. Patera, and D.V. Rovas. Output bounds for reduced-basis approximations of symmetric positive definite eigenvalue problems. *C. R. Acad. Sci. Paris, Série I*, 331(2):153–158, July 2000.

[22] L. Machiels, Y. Maday, and A. T. Patera. A "flux-free" nodal Neumann subproblem approach to output bounds for partial differential equations. *C. R. Acad. Sci. Paris, Série I*, 330(3):249–254, February 2000.

[23] L. Machiels, A.T. Patera, and D.V. Rovas. Reduced basis output bound methods for parabolic problems. *Computer Methods in Applied Mechanics and Engineering*, 2001. Submitted.

[24] L. Machiels, J. Peraire, and A. T. Patera. A posteriori finite element output bounds for the incompressible Navier-Stokes equations; Application to a natural convection problem. *Journal of Computational Physics*, 172:401–425, 2001.

[25] Y. Maday, A. T. Patera, and J. Peraire. A general formulation for a posteriori bounds for output functionals of partial differential equations; Application to the eigenvalue problem. *C. R. Acad. Sci. Paris, Série I*, 328:823–828, 1999.

[26] Y. Maday, A.T. Patera, and D.V. Rovas. A blackbox reduced-basis output bound method for noncoercive linear problems. In D. Cioranescu and J.-L. Lions, editors, *Nonlinear Partial Differential Equations and Their Applications, Collége de France Seminar Volume XIV*, pages 533–569. Elsevier Science B.V., 2002.

[27] Y. Maday, A.T. Patera, and G. Turinici. 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*, 335:1–6, 2002.

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

[29] L.E. Malvern. *Introduction to the Mechanics of a Continuous Medium.* Prentice-Hall, Inc., 1969.

[30] M.U. Nabi, V.R. Sule, and S.V. Kulkarni. An efficient computational modeling and solution scheme for repeated finite element analysis of composite domains with varying material property. *International Journal of Computational Engineering Science*, 3(3):257–275, 2002.

[31] A.W. Naylor and G.R. Sell. *Linear Operator Theory in Engineering and Science*, volume 40 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1982.

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

[33] I. Oliveira, C. Prud'homme, K. Veroy, and A.T. Patera. Thermo-structural analysis of multifunctinoal prismatic cellular (triangular) metals. Web Publication, `http://augustine.mit.edu`, 2002.

[34] I. B. Oliveira and A. T. Patera. Reliable real-time optimization of nonconvex systems described by parametrized partial differential equations. Technical report, Proceedings Singapore-MIT Alliance Symposium, January 2003.

[35] A. T. Patera and E. M. Rønquist. A general output bound result: Application to discretization and iteration error estimation and control. *Math. Models Methods Appl. Sci.*, 11(4):685–712, 2001.

[36] J. S. Peterson. The reduced basis method for incompressible viscous flow calculations. *SIAM J. Sci. Stat. Comput.*, 10(4):777–786, July 1989.

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

[38] C. Prud'homme and A.T. Patera. Reduced-basis output bounds for approximately parametrized elliptic coercive partial differential equations. *Computing and Visualization in Science*, 2002. Submitted.

[39] C. Prud'homme, D. Rovas, K. Veroy, Y. Maday, A.T. Patera, and G. Turinici. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. *Journal of Fluids Engineering*, 124(1):70–80, March 2002.

[40] C. Prud'homme, D. Rovas, K. Veroy, and A.T. Patera. Mathematical and computational framework for reliable real–time solution of parametrized partial differential equations. *M2AN*, 36(5):747–771, 2002.

[41] W.C. Rheinboldt. Numerical analysis of continuation methods for nonlinear structural problems. *Computers and Structures*, 13(1-3):103–113, 1981.

[42] W.C. Rheinboldt. On the theory and error estimation of the reduced basis method for multiparameter problems. *Nonlinear Analysis, Theory, Methods and Applications*, 21(11):849–858, 1993.

[43] D.V. Rovas. *Reduced-Basis Output Bound Methods for Parametrized Partial Differential Equations*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, October 2002.

[44] I.H. Shames. *Introduction to Solid Mechanics*. Prentice Hall, New Jersey, 2nd edition, 1989.

[45] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3):519–524, March 1987.

[46] Y. Solodukhov. *Reduced-Basis Methods Applied to Locally Non-Affine Problems.* PhD thesis, Massachusetts Institute of Technology, 2004. In progress.

[47] K. Veroy, D. Rovas, and A.T. Patera. A posteriori error estimation for reduced-basis approximation of parametrized elliptic coercive partial differential equations: "convex inverse" bound conditioners. *Control, Optimisation and Calculus of Variations*, 8:1007–1028, June 2002. Special Volume: A tribute to J.-L. Lions.

[48] E. L. Yip. A note on the stability of solving a rank-p modification of a linear system by the Sherman-Morrison-Woodbury formula. *SIAM Journal on Scientific and Statistical Computating*, 7(2):507–513, April 1986.