Reduced-Basis Approximation and *A Posteriori* Error Estimation for Parabolic Partial Differential Equations

by

Martin A. Grepl

S.M. Mechanical Engineering (2001)
Massachusetts Institute of Technology

Dipl.-Ing. Luft- und Raumfahrttechnik (2000)
Universität Stuttgart, Germany

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2005

Author ....
....
Department of Mechanical Engineering
April 15, 2005

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

Accepted by.................................
......
Lallit Anand
Professor of Mechanical Engineering
Chairman, Committee on Graduate Studies

Reduced-Basis Approximation and *A Posteriori* Error Estimation for Parabolic Partial Differential Equations

by
Martin A. Grepl

Submitted to the Department of Mechanical Engineering
on April 15, 2005, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

Modern engineering problems often require accurate, reliable, and efficient evaluation of quantities of interest, evaluation of which demands the solution of a partial differential equation. We present in this thesis a technique for the predicition of outputs of interest of parabolic partial differential equations. The essential ingredients are: (*i*) rapidly convergent reduced-basis approximations — Galerkin projection onto a space W_N spanned by solutions of the governing partial differential equation at N selected points in parameter-time space; (*ii*) *a posteriori* error estimation — relaxations of the error-residual equation that provide rigorous and sharp bounds for the error in specific outputs of interest: the error estimates serve *a priori* to construct our samples and *a posteriori* to confirm fidelity; and (*iii*) offline-online computional procedures — in the offline stage the reduced-basis approximation is generated; in the online stage, given a new parameter value, we calculate the reduced-basis output and associated error bound. The operation count for the online stage depends only on N (typically small) and the parametric complexity of the problem; the method is thus ideally suited for repeated, rapid, reliable evaluation of input-output relationships in the many-query or real-time contexts.

We first consider parabolic problems with affine parameter dependence and subsequently extend these results to nonaffine and certain classes of nonlinear parabolic problems. To this end, we introduce a collateral reduced-basis expansion for the nonaffine and nonlinear terms and employ an inexpensive interpolation procedure to calculate the coefficients for the function approximation — the approach permits an efficient offline-online computational decomposition even in the presence of nonaffine and highly nonlinear terms. Under certain restrictions on the function approximation, we also introduce rigorous *a posteriori* error estimators for nonaffine and nonlinear problems.

Finally, we apply our methods to the solution of inverse and optimal control problems. While the efficient evaluation of the input-output relationship is essential for the real-time solution of these problems, the *a posteriori* error bounds let us pursue a robust parameter estimation procedure which takes into account the uncertainty due to measurement and reduced-basis modeling errors explicitly (and rigorously). We consider several examples: the nondestructive evaluation of delamination in fiber-reinforced concrete, the dispersion of pollutants in a rectangular domain, the self-ignition of a coal stockpile, and the control of welding quality. Numerical results illustrate the applicability of our methods in the many-query contexts of optimization, characterization, and control.

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

Acknowledgments

I would first like to thank my thesis advisor, Professor Anthony T. Patera, for his support and guidance during my thesis work. I am truly grateful for his insights, flexibility, humor, and trust.

I am also very thankful to my thesis committee members, Professor Dimitris Bertsimas and Professor David Hardt, for their invaluable comments, suggestions, and encouragement throughout my studies. Furthermore, I would also like to thank Professor Yvon Maday for his help, support, and suggestions.

I am very grateful to my current and former colleagues: George Pau, Christophe Prud'homme, Dimitrios Rovas, Ivan Oliveira, Ngoc Cuong Nguyen, Simone Deparis, Sugata Sen, Gianluigi Rozza, and Yuri Solodukhov. I greatly appreciate the many helpful and interesting discussions on the topic and off the topic, and the time spent together in and out of the office. I am most grateful to Debra Blanchard for her invaluable help and support and her encouraging attitude throughout my thesis research. I must also acknowledge the friends I found during my time at MIT: Wei Wang, Tu Duc Nguyen, Thomas Grätsch, Jean-Louis Locsin, Winfried Lohmiller, and Daniel Dreyer.

Finally, above all I want to express my love and deepest gratitude to my family: my parents, Dorit and Hansjörg Grepl, my brother, Michael, my sister, Barbara, and my wife, Karen. Without their support, love and encouragement, I would not have been able to pursue my dreams. To them I dedicate this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The role of numerical simulation in engineering and science has become increasingly important. System or component behavior is often modeled using a set of partial differential equations and associated boundary conditions, the analytical solution to which is generally unavailable. In practice, a discretization procedure such as the finite element method (FEM) is used.

However, as the physical problems become more complex and the mathematical models more involved, current computational methods prove increasingly inadequate, especially in contexts requiring numerous solutions of parametrized partial differential equations for many different values of the parameter. Even for modest-complexity models, the computational cost to solve these problems is prohibitive.

For example, the design, optimization, control, and characterization of engineering components or systems often require repeated, reliable, and real-time prediction of performance metrics, or outputs, "s^{e}," such as heat fluxes or flowrates. These outputs are typically functionals of field variables, "y^{e}," — such as temperatures or velocities — associated with a parametrized partial differential equation; the parameters, or inputs, "μ," serve to identify a particular configuration of the component — such as boundary conditions, material properties, and geometry. The relevant system behaviour is thus described by an implicit input-output relationship, "$s^{\mathrm{e}}(\mu)$," evaluation of which demands solution of the underlying partial differential equation (PDE).

Classical approaches such as the finite element method can not typically satisfy the requirements of *real-time certified* prediction of the outputs of interest. In the finite element method, the infinite dimensional solution space is replaced by a finite dimensional "truth" approximation space of size $\mathcal{N}$. We shall assume — hence the appellation "truth" — that the approximation space is sufficiently rich such that the FEM approximation $y(\mu)$ (respectively, $s(\mu)$) is indistinguishable from the analytic, or exact, solution $y^{\mathrm{e}}(\mu)$ (respectively, $s^{\mathrm{e}}(\mu)$). Unfortunately, for any reasonable error tolerance, the dimension $\mathcal{N}$ needed to satisfy this condition — even with the application of appropriate (parameter-dependent) adaptive mesh refinement strategies — is typically extremely large, and in particular much too large to provide real-time response.

Our goal is the development of numerical methods that permit the *efficient* and *reliable* evaluation of this PDE-induced input-output relationship *in real-time* or *in the limit of many queries* — that is, in the design, optimization, control, and characterization contexts. To further motivate our methods and illustrate contexts in which we develop them, we consider several examples.

1.1.1 Application Problems (AP)

AP I: Nondestructive Evaluation of Delamination

Our first example is the nondestructive evaluation (NDE) of civil engineering structures. Fiber-reinforced polymer (FRP) composites are used widely in civil engineering for bridge column seismic retrofits to strengthen concrete and masonry structures, as well as in the rehabilitation of existing infrastructure [44]. FRP composites, which are also widely used in aerospace applications, are two-phase materials consisting of long unidirectional fibers within a polymer matrix. To enhance the structural capacity of the design, layers of FRP composites are bonded to concrete structures using adhesives such as epoxy resins. The success and performance of this reinforcement depends strongly on the quality of the bond between the FRP composite and the substrate. However, the mechanical properties of the composite are affected by environmental aspects such as temperature, moisture, and contaminants. Furthermore, the quality of the bond is influenced by the manufacturing process and the location in which the composite is fabricated; the latter is often "in the field" rather than in a controlled environment. Since debonds or delaminations at the composite-concrete interface often occur, effective quality control — i.e., providing reliable information about the thickness and fiber content of the composite, and the amount, location, and size of defects — is vital to safety. There is thus a need not only for the *detection* of flaws, but also for accurate *characterization* of the detected defects.

There exist two major classes of NDE techniques: electromagnetic methods and mechanical vibration methods. Infrared (IR) thermography, which belongs to the former, has been successfully applied to detect flaws in FRP composites bonded to concrete [114] due to the fact that it is sensitive to the presence of defects near the surface and allows for the efficient investigation of large surface areas. In active IR thermography, the structure is actively heated and the surface temperature is monitored using an IR imaging system. The heat transfer in the structure is affected by flaws in the structure, which gives rise to localized hot or cold spots on the surface. The goal is then to infer the location and size of defects from the surface temperature readings. The methods used thus have to be *reliable* (to guarantee safety) and *efficient* (to allow for real-time characterization in the field).

A model for a FRP composite bonded to concrete similar to the one discussed in [114] is shown in Figure 1-1. The concrete slab considered has length $l = 60$ and thickness $t_{\mathrm{C}} = 10$, and the FRP layer has thickness $t_{\mathrm{FRP}} = 1$. We assume that there exists a delamination of unknown width, w_D, centered at $x_1 = 0$ along the composite-concrete interface. The thermal conductivity of concrete is denoted k_{C}, and the thermal conductivity of FRP is denoted k_{FRP}; while the former is assumed known (the uncertainty is very small in practice), the latter depends on the direction of the fibers and the fiber content. We also define $\varkappa \equiv k_{\mathrm{FRP}}/k_{\mathrm{C}}$ to be the ratio of the two thermal conductivities. The (non-dimensionalized) heat flux, $q(t) = \overline{q}(t)t_{\mathrm{FRP}}/(k_{\mathrm{C}}(\overline{T}_{\mathrm{FRP,max}} - \overline{T}_0))$, is applied to the surface of the structure, Γ_{top}, for $t \in]0, 0.5]$; here, $\overline{q}(t)$ is the heat flux, $\overline{T}_0$ is the ambient temperature, and $\overline{T}_{\mathrm{FRP,max}}$ is the maximum allowable temperature of the FRP. The surface temperature is measured for $t \in]0, 10]$ at two locations on the surface: Measurement 1, denoted by s_1, is taken over a small region around $x_1 = 0$; and Measurement 2, denoted by s_2, is taken over a small region around $x_1 = 14.5$.

The temperature distribution, $T(x, t)$, in FRP and concrete is then governed by the (appropri-

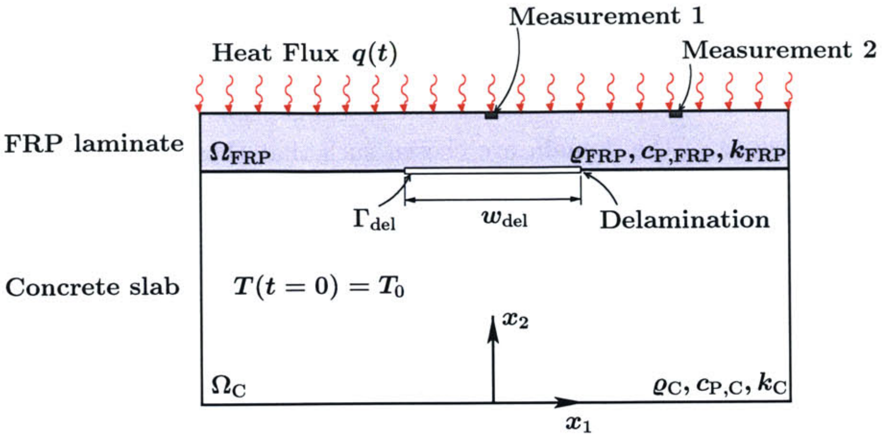


Figure 1-1: AP I: Delamination of FRP bonded to concrete.

ately) non-dimensionalized unsteady heat or diffusion equations

$$\frac{\partial T_{\text{FRP}}(x,t)}{\partial t} = \varkappa \nabla^2 T_{\text{FRP}}(x,t), \quad \text{in } \Omega_{\text{FRP}}, \tag{1.1}$$

$$\frac{\partial T_{\text{C}}(x,t)}{\partial t} = \nabla^2 T_{\text{C}}(x,t), \quad \text{in } \Omega_{\text{C}}, \tag{1.2}$$

with initial temperatures $T_{\text{FRP}}(x,t=0)=0$ and $T_{\text{C}}(x,t=0)=0$; here, $x=(x_1,x_2)$ is the spatial coordinate, and $T(x,t)=(\overline{T}-\overline{T}_0)/(\overline{T}_{\text{FRP,max}}-\overline{T}_0)$ is the non-dimensional temperature, $\overline{T}_0$ is the ambient temperature, and $\overline{T}_{\text{FRP,max}}$ is the maximum allowable temperature of the FRP. The temperature and heat flux at the composite-concrete interface (excluding the delaminated area) are continuous, i.e.,

$$T_{\text{FRP}} = T_{\text{C}}, \quad \text{on } \partial\Omega_{\text{FRP}} \cap \partial\Omega_{\text{C}}, \tag{1.3}$$

$$-\varkappa \nabla T_{\text{FRP}} \cdot \mathbf{n}\big|_{\partial\Omega_{\text{FRP}}} = -\nabla T_{\text{C}} \cdot \mathbf{n}\big|_{\partial\Omega_{\text{C}}}, \quad \text{on } \partial\Omega_{\text{FRP}} \cap \partial\Omega_{\text{C}}. \tag{1.4}$$

Since the conductivity of air is much larger than that of FRP and concrete, we may impose a homogeneous Neumann boundary condition at the delamination boundary, Γ_{del},

$$\nabla T_{\text{FRP}} \cdot \mathbf{n} = 0, \quad \text{on } \Gamma_{\text{del}}, \tag{1.5}$$

$$\nabla T_{\text{C}} \cdot \mathbf{n} = 0, \quad \text{on } \Gamma_{\text{del}}. \tag{1.6}$$

The heat source $q(t)$ is reflected in the boundary condition on the surface, given by

$$-\varkappa \nabla T_{\text{FRP}} \cdot \mathbf{n} = q(t), \quad \text{on } x_2 = 11, \; x_1 \in [-30, 30]. \tag{1.7}$$

Finally, we assume that the heat flux on the left and right boundaries is zero ,

$$\nabla T_{\text{C}} \cdot \mathbf{n} = 0, \quad \text{on } x_1 = \pm 30, \; x_2 \in [1, 10], \tag{1.8}$$

$$\nabla T_{\text{FRP}} \cdot \mathbf{n} = 0, \quad \text{on } x_1 = \pm 30, \; x_2 \in [10, 11], \tag{1.9}$$

and that the temperature at the bottom boundary is at ambient level,

$$T_{\mathrm{C}} \;=\; 0, \quad \text{on } x_1 \in [-30, 30],\; x_2 = 0. \tag{1.10}$$

Note that the dimensions of the domain are chosen such that those boundary conditions (1.8)-(1.10) which result from consideration of only a small section of the slab do not substantially affect the temperature measurements.

From the problem description and the governing equations (1.1)-(1.10) it is evident that the temperature distribution, $T_{\mathrm{C}}(x,t)$ and $T_{\mathrm{FRP}}(x,t)$, and hence the measured surface temperatures, s_1 and s_2, depend on the (known) heat flux, $q(t)$, the (unknown) delamination width w_{del}, and the (unknown) ratio of the conductivities, $\varkappa$. We can thus identify the input parameter set $\mu = (\mu_1, \mu_2) \equiv (w_{\mathrm{del}}, \varkappa) \in \mathcal{D} \subset \mathbb{R}^{P=2}$, where $\mathcal{D}$ is the (input) parameter domain. We assume that the delamination width satisfies $2 \leq w_{\mathrm{del}} \leq 20$ and that the ratio of the conductivities satisfies $0.4 \leq \varkappa \leq 1.8$; we thus have $\mathcal{D} \equiv [2, 20] \times [0.4, 1.8]$.

We note that given the input parameter μ, we directly determine the temperature outputs $s_1(\mu, t)$ and $s_2(\mu, t)$ — this is referred to as the "forward" problem. However, in the characterization context we have to infer the parameter μ from given measurements of $s_1(\mu, t)$ and $s_2(\mu, t)$ — this is referred to as the "inverse" problem. The solution of the inverse problem is usually obtained in an iterative procedure which requires repeated evaluation of the input-output relationship $s_1(\mu, t)$ and $s_2(\mu, t)$. The computational cost can be prohibitively large if classical discretization and solution approaches, such as finite element methods, are used to solve (1.1)-(1.10). We will return to this problem in Sections 4.7 and 7.5 where we will consider the reduced-basis approximation and solution of the inverse problem, respectively.

AP II: Dispersion of Pollutants

In this example we consider the dispersion of a pollutant or contaminant released at a specific location Ω_{P} in a two-dimensional domain Ω. We assume that the underlying flow or velocity field, $\mathbf{U} = (U_1, U_2)$, is fixed, i.e., the pollutant concentration has no effect on the nature of the flow. A sketch of the flow field is shown in Figure 1-2. The pollutant is released at the (possibly unknown) location Ω_{P} and the (nondimensional) concentration of the pollutant, c, is measured at eight sensors evenly distributed throughout the domain of interest.

The governing equation for the concentration is the unsteady convection-diffusion equation,

$$\frac{\partial c(x,t)}{\partial t} + \mathbf{U} \cdot \nabla c(x,t) = \kappa \nabla^2 c(x,t) + g^{\mathrm{PS}}(x)\, u(t), \tag{1.11}$$

with initial condition

$$c(x, t = 0) = c_0(x) = 0. \tag{1.12}$$

Here, $x = (x_1, x_2)$ is the spatial coordinate, κ is the (mass) diffusivity, $\mathbf{U}$ is the known velocity field which is incompressible $(\nabla \cdot \mathbf{U} = 0)^1$, and $u(t)$ is the control input which represents the strength of the pollution source term. Note that we nondimensionalize the concentration c by $c = (\bar{c} - c_0)/Q$ and the time via $L_c/\mathbf{U}_0$; here, $Q = \int\int g^{\mathrm{PS}}(x)\, u(t)\, d\Omega\, dt$, $L_c = 1$ is the characteristic length and $\mathbf{U}_0 = 1$ the average velocity. We further note that $\kappa = 1/\mathrm{Pe}$, where Pe is the Peclet number.

[1] The velocity field (provided by K. Veroy [121]) represents a natural convection (steady Navier-Stokes) flow with $\mathrm{Gr} = 10^5$ and $\mathrm{Pr} = 0$. Note that the solution is unique for these parameter values.

The pollution source, $g^{\mathrm{PS}}(x)$, is modeled as a Gaussian distribution with standard deviation σ^{PS} centered at $x^{\mathrm{PS}} = (x_1^{\mathrm{PS}}, x_2^{\mathrm{PS}})$, given by

$$g^{\mathrm{PS}}(x) = \frac{1}{2\pi(\sigma^{\mathrm{PS}})^2}\, e^{-((x_1 - x_1^{\mathrm{PS}})^2 + (x_2 - x_2^{\mathrm{PS}})^2)/(2(\sigma^{\mathrm{PS}})^2)}. \tag{1.13}$$

The outputs of interest are the concentrations, $s_i,\ 1 \le i \le 8$, measured at the eight sensor locations.

From (1.11)–(1.13) it follows that the concentration, $c(x,t)$, depends on the diffusivity κ, the control input $u(t)$, and the location $(x_1^{\mathrm{PS}}, x_2^{\mathrm{PS}})$ and standard deviation σ^{PS} of the source term. We can thus identify the input parameter $\mu = (\mu_1, \mu_2, \mu_3, \mu_4) \equiv (\kappa, x_1^{\mathrm{PS}}, x_2^{\mathrm{PS}}, \sigma^{\mathrm{PS}}) \in \mathcal{D} \subset \mathbb{R}^{P=4}$. Given the parameter μ, we can directly solve the forward problem to determine the output concentrations $s_i(\mu),\ 1 \le i \le 8$. However, inferring the source location from given measurements $s_i(\mu)$ requires the solution of an inverse problem and is therefore much more complex.

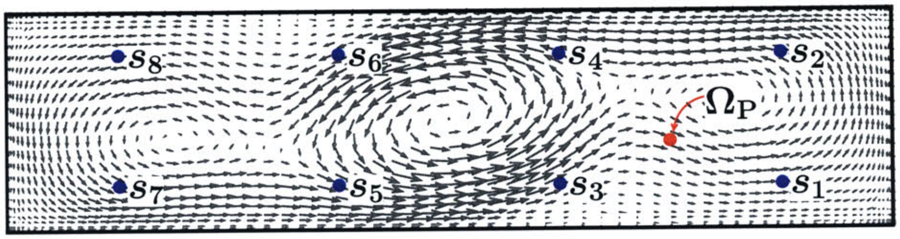


Figure 1-2: AP II: Dispersion of pollutants

We note that the Reynolds number for the specific application considered here is relatively low. However, problems of a similar kind — usually, considering turbulent flow with higher Reynolds numbers — have recently received a lot of attention in support of Homeland Security [17, 38, 119]. In this context a possible chemical indoor attack (for instance, in airports) is considered. The goal is the development of algorithms and capabilities for the real-time characterization of the unknown source location given the sensor measurement data [17]. First, the airflow is determined from the known location of supply and return vents in the airport. Second, the dispersion of the chemical agent is considered and the location of the source is estimated. Finally, given the known source location, containment and control strategies, or remediation and evacuation efforts have to be executed. Thus, real-time response in estimating the source location is critical for successful countermeasures.

We will revisit this problem several times in this thesis: in Section 4.8.5 we first consider the case where the source location is known and only the diffusivity varies; in Section 5.6 we additionally let the location of the source term vary in a certain region; and in Section 7.6 we discuss the inverse problem: locating the source from the concentration measurements.

AP III: Self-Ignition of a Coal Stockpile

For our next example we consider a one-dimensional non-isothermal reaction-diffusion model for the self-ignition of a coal stockpile with Arrhenius type nonlinearity [23, 103, 105]. In practice this problem arises if large piles of coal are stored, e.g., in harbors, over extended periods of time. As the oxygen in the air reacts with the coal, the pile starts to heat up and can eventually self-ignite if certain conditions — e.g., on porosity, oxygen concentration, and coal size — are met. We also note that similar models are used in combustion theory, biology, and in the description of porous catalysts. This problem is just one of many possible examples in the large class of reaction-diffusion systems [26]. Reaction-diffusion systems are an interesting area of applications because they have inherently many parameters and appear in a large number of real-world applications.

The field variables are the temperature of the reactive medium (here, the coal) normalized by the ambient temperature, $T(x,t) = (\overline{T}(x,t) - \overline{T}_\infty)/\overline{T}_\infty$, and the concentration of the reactant (here, the oxygen in the air) normalized by the concentration of oxygen in the ambient air, $c(x,t) = (\overline{c}(x,t) - \overline{c}_\infty)/\overline{c}_\infty$. The coupled set of equations governing the first order exothermic reaction in the one-dimensional layer of reactive medium through which a gaseous reactant diffuses are given by

$$\begin{aligned} \frac{\partial T(x,t)}{\partial t} &= \nabla^2 T(x,t) + \beta\,\Phi^2\,(c(x,t)+1)\,e^{-\gamma/(T(x,t)+1)}, && (1.14)\\ \frac{\partial c(x,t)}{\partial t} &= \mathrm{Le}\,\nabla^2 c(x,t) - \Phi^2\,(c(x,t)+1)\,e^{-\gamma/(T(x,t)+1)}, && (1.15) \end{aligned}$$

with initial conditions

$$\begin{aligned} T(x,t=0) &= T_0 = 0, && (1.16)\\ c(x,t=0) &= c_0 = \frac{1}{(3x+1)^2} - 1. && (1.17) \end{aligned}$$

Here, $x \in \Omega \subset \mathbb{R}^1$ is the spatial coordinate and $\Omega \equiv [0,1]$ is the spatial domain. A snapshot of the distribution of temperature and concentration for $t > 0$ is shown in Figure 1-3. The boundary conditions are

$$\begin{aligned} T(x,t)|_{x=0} &= 0, & T(x,t)|_{x=1} &= 0,\\ c(x,t)|_{x=0} &= 0, & \left.\frac{\partial c(x,t)}{\partial x}\right|_{x=1} &= 0. \end{aligned} \qquad (1.18)$$

Note that $x = 0$ corresponds to the top of the pile at which $T(x,t)$ and $c(x,t)$ are equal to the the ambient temperature and concentration, respectively; and $x = 1$ corresponds to the bottom of the pile at which $T(x,t)$ is equal to the ground (ambient) temperature, and the concentration gradient is zero. The output of interest is the temperature and concentration at $x = 0.2$ denoted by s_1 and s_2, respectively.

There are several parameters governing the dynamic behavior of the system: the Arrhenius number or activation energy, γ; the Prater temperature or nondimensional heat of reaction, β; the Lewis number, Le, which is the ratio of mass and heat diffusivities; and the Thiele modulus, Φ. Note that the Thiele modulus is related to the maximum possible temperature of the system: the temperature satisfies $1 \le T \le 1+\beta$ for Dirichlet boundary conditions and Le $= 1$. We can thus identify the input parameter $\mu = (\mu_1, \mu_2, \mu_3, \mu_4) \equiv (\gamma, \beta, \mathrm{Le}, \Phi^2) \in \mathcal{D} \subset \mathbb{R}^{P=4}$. (As we will see in Section 6.6.1, the model exhibits a very rich dynamic behavior for certain parameter ranges.)

We analyze and discuss the dynamic behavior of this model in Section 6.6.1.

AP IV: Control of Welding Quality

For our last example we look at the control of a gas metal arc welding (GMAW) process [113]. More specifically, we consider bead-on-plate welding of two metal plates being joined together

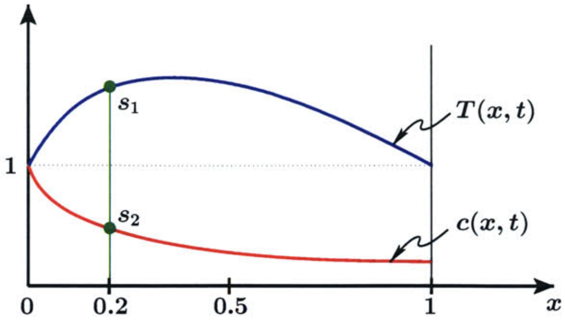


Figure 1-3: AP III: Model for the self-ignition of a coal stockpile. Sketch of temperature and concentration for $t > 0$.

with a partial penetration weld. A sketch of the joint-section and the welding torch, moving with (nondimensional) velocity Pe over the workpiece, is shown in Figure 1-4. One of the key geometric features for the quality of the weld is the welding depth, d_{W}, since it indicates the strength of the joint. The welding depth is defined as the depth of the joint penetration (through liquefaction and subsequent solidification). However, the welding depth cannot be directly measured for real-time control because (*i*) it is not directly visible during the welding process, and (*ii*) the weld pool boundary is a very hot solid-liquid interface. It is thus necessary to estimate the pool depth from quantities which are available for measurements. One such approach is proposed in [111, 112], where surface temperature measurements taken from the back of the workpiece are used in a real-time depth estimation algorithm — acceptable accuracy and speed of the depth estimate for in-process control are achieved.

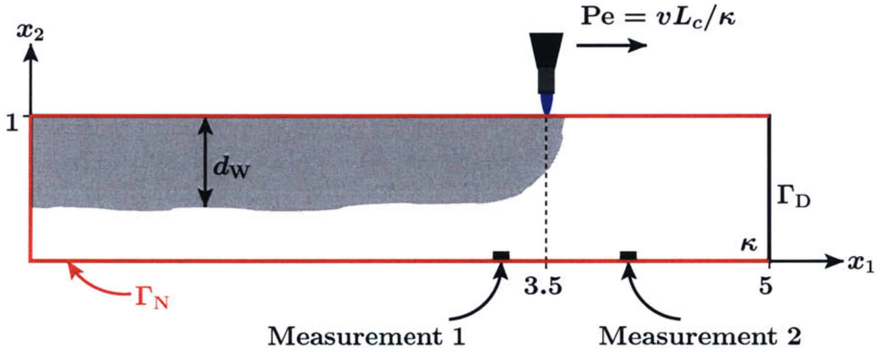


Figure 1-4: AP IV: Control of welding quality.

The depth estimation algorithm employs an inverse three-dimensional analytical heat conduction problem. To this end, the heat flux from the welding arc is modeled as a moving heat source with a dual Gaussian distribution [111]: a top Gaussian heat source — which is largely responsible for the width portion of the pool, and a lateral Gaussian heat source — which causes the "finger penetration" and accounts for the depth portion of the weld [112]. This model results in a good description of the entire weld pool shape and the temperature distribution in the workpiece. For the estimation of only the weld pool depth of moderately thick materials, however, it suffices to consider only the lateral Gaussian heat source.

To model the welding process we restrict our attention to the two-dimensional temperature distribution, $T(x,t)$, in the joint-section shown in Figure 1-4; here $T(x,t) \equiv (\overline{T}(x,t) - \overline{T}_\infty)/(\overline{T}_m - \overline{T}_\infty)$ is the non-dimensionalized temperature, where $\overline{T}_\infty$ and $\overline{T}_m$ are the ambient and melting temperatures, respectively, and $x = (x_1, x_2) \in \Omega \equiv [0,5] \times [0,1]$ is the nondimensional spatial coordinate. The melting point of the material is thus obtained for $T(x,t) = 1$. We note that the latent heat has only a minor effect and can therefore we neglected [62]. We consider a coordinate system moving with the same velocity as the torch; in this coordinate system, the torch is stationary and the velocity enters as a convective term in the governing equation. The (appropriately) non-dimensionalized governing equation is thus the unsteady convection-diffusion equation

$$\frac{\partial T(x,t)}{\partial t} + \mathrm{Pe}\,\frac{\partial T(x,t)}{\partial x_1} = \kappa\,\nabla^2 T(x,t) + q_{\mathrm{w}}(x)\,u(t), \tag{1.19}$$

with initial condition

$$T(x, t=0) = \overline{T}_0(x) = 0. \tag{1.20}$$

Here, $\mathrm{Pe} = vL_c/\kappa$ is the nondimensional velocity or Peclet number, v is the velocity of the torch, L_c is the characteristic length, κ is the thermal diffusivity, and $u(t)$ is the (nondimensional) heat input. Note that we consider the start-up process and the temperature is thus zero initially. The spatial (Gaussian) distribution of the heat input, centered at the torch position $x^{\mathrm{T}} \equiv (3.5, 1)$, is given by

$$q_{\mathrm{w}}(x) = \frac{\eta_{\mathrm{w}}}{2\pi\sigma_{\mathrm{w}}^2}\, e^{-((x_1 - x_1^{\mathrm{T}})^2 + (x_2 - x_2^{\mathrm{T}})^2)/(2\sigma_{\mathrm{w}}^2)}, \tag{1.21}$$

where η_{w} is the efficiency and σ_{w} is the distribution parameter. The outputs of interest, $s_i(\mu, t)$, $i = 1, 2$, are the temperatures at the two measurement locations 1 and 2, respectively. We assume homogeneous Neumann boundary conditions on Γ_{N}, and homogeneous Dirichlet boundary conditions on Γ_{D}, i.e., the temperature of the workpiece is equal to the ambient temperature sufficiently far upwind from the torch position.

The velocity, Pe, and total heat input, $u(t)$, can be controlled during the process while the efficiency, η_{w}, and distribution, σ_{w}, are the system parameters that have to be estimated. Given estimates for η_{w} and σ_{w} (and the known inputs Pe and $u(t)$) we can solve (1.19)–(1.20), search for the isotherm $T(x,t) = 1$ [2] (corresponding to the melting temperature), and determine the weld pool depth d_{w}. We thus identify the parameter $\mu \equiv (\mu_1, \mu_2) \equiv (\eta_{\mathrm{w}}, \sigma_{\mathrm{w}}) \in \mathcal{D} \subset \mathbb{R}^{P=2}$.

The in-process control of the weld pool depth d_{w} hence requires the real-time and reliable solution of the input-output relationship $s(\mu, t)$ for (i) the online parameter estimation algorithm and (ii) the subsequent control action to obtain the desired weld pool depth $d_{\mathrm{w,d}}$. We will discuss the parameter estimation as well as control problem in detail in Chapter 8.

1.1.2 Computational Challenge/Thesis Objectives

The applications described above have several common features: first, the governing equations are time-dependent (parabolic) partial differential equations whose dynamic behavior strongly depends on the parameters characterizing the problem; second, the *transient behavior* of the system (and not only the steady-state case) is crucial for the solution of these problems; and third, the efficient

[2] Note that in the case where an analytic expression for the temperature field $T(x,t)$ is known, we can (analytically or numerically) solve $T(x,t) = 1$ for x at each time t.

solution of the inverse and/or control problem requires the fast (possibly real-time) and reliable evaluation of the input-output map $\mu \to s(\mu, t^k)$, $k \in \mathbb{K}$.

In actual practice, of course, we do not have access to the analytic solution of these problems and a discretization procedure such as the finite element method is used. The algebraic equations obtained using these procedures are, in general, very high-dimensional and their solution is expensive — applying these methods in the many-query context is thus prohibitive, and real-time solution infeasible.

Our goals are twofold. Our first goal is the development of computational methods that permit *accurate*, *reliable*, and *rapid* evaluation of input-output relationships induced by parabolic partial differential equations in *real-time* and *in the limit of many queries*. In particular, we seek to develop techniques 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 online 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 techniques to solve representative problems involving the control and characterization of engineered systems. To achieve these goals we pursue the reduced-basis method.

1.2 Earlier Work

1.2.1 Reduced-Basis Method

The reduced-basis method was first introduced in the late 1970s for nonlinear structural analysis [4, 77], and subsequently abstracted and analyzed [9, 16, 37, 85, 98] and extended [43, 48, 83] to a much larger class of parametrized partial differential equations. The reduced-basis method recognizes that the field variable is not, in fact, some arbitrary member of the infinite-dimensional solution space associated with the partial differential equation; rather, it resides, or "evolves," on a much lower-dimensional manifold induced by the parametric dependence.

The reduced-basis approach as earlier articulated is local in parameter space in both practice and theory. To wit, Lagrangian or Taylor approximation spaces for the low-dimensional manifold are typically defined relative to a particular parameter point; and the associated *a priori* convergence theory relies on asymptotic arguments in sufficiently small neighborhoods [37]. As a result, the computational improvements — relative to conventional (say) finite element approximation — are often quite modest [85]. Later work [40, 60, 61, 76, 91, 121, 123, 124] differs from these earlier efforts in several important ways: first, *global* approximation spaces are developed; second, rigorous *a posteriori error estimators* are introduced; and third, *off-line/on-line* computational decompositions are exploited (see [9] for an earlier application of this strategy within the reduced-basis context). These three ingredients allow us — for the restricted but important class of "parameter-affine" problems — to reliably decouple the generation and projection stages of reduced-basis approximation, thereby effecting computational economies of several orders of magnitude.

Much progress has been made in *a posteriori* error estimation for reduced-basis approximations. In particular, *a posteriori* error bounds have been successfully developed for (*i*) linear [40, 60, 61, 91, 124] and (*ii*) at most quadratically nonlinear [76, 121, 123] elliptic partial differential equations that are affine in the parameter. These two assumptions enable the development of very efficient offline-online computational strategies relevant in the many-query and real-time contexts. The operation count for the online stage — in which, given a new parameter value, we calculate the reduced-basis output and associated error bound — depends only on the dimension of the reduced-basis space (typically small) and on the parametric complexity of the problem, but is *independent* of the dimension of the underlying "truth" finite element approximation space (typically very large).

The case of non-affine parameter dependence has also been recently addressed. In particular, problems which are *locally* non-affine — i.e., non-affine only in a small part of the domain — are treated in [110]. More general non-affine problems are addressed in Barrault *et al.* [15] in which a technique is introduced that recovers the efficient offline-online decomposition even in the presence of general non-affine parameter dependence. In this approach, a (necessarily affine) "collateral" reduced-basis approximation is developed for the non-affine terms. The essential ingredients to this approach are (*i*) a "good" collateral reduced-basis approximation space, (*ii*) a stable and inexpensive interpolation procedure by which to determine the approximation, and (*iii*) an effective *a posteriori* estimator with which to quantify the newly introduced error terms.

Finally, reduced-basis approximations and error estimators have also been developed in [101, 102] for parabolic partial differential equations in which (*i*) the temporal forcing or controls are known, and (*ii*) the outputs of interest are independent of time; see also [56, 86] for an application of the reduced-basis method to initial value problems. However, because of these limiting assumptions this earlier work is, in general, not applicable to the control, optimization, or characterization context. One of the contributions of this thesis is to address and lift these restrictions — thus allowing for a much wider field of applications — and to considerably simplify the methodology developed in [101, 102]. Furthermore, we also extend the theory to consider nonaffine and certain classes of nonlinear problems.

1.2.2 Model Order Reduction

Many model-order reduction techniques for linear time-dependent systems are proposed in the literature: the most well-known are proper orthogonal decomposition (POD or Karhunen-Loève decomposition) [109], balanced truncation [69], and various related hybrid [55, 127] techniques. In POD — probably the most popular model-order reduction technique — time is considered the varying parameter, and "snapshots" of the field variable at different times are obtained from either a numerical or experimental procedure. The optimal approximation space is constructed by applying the singular value decomposition to these vectors, and keeping only the N vectors corresponding to the largest singular values. Since the singular values are related to the "energy" of the system, only the modes preserving the most energy are preserved. The reduced-order model is then obtained by a Galerkin projection onto the space spanned by these vectors. POD has been successfully applied in many fields: turbulent flows [59], fluid structure-interaction [35], non-linear structural mechanics [54], turbo-machinery flows [127]. On the other hand, balanced truncation is a very popular method on control theory. In this approach, the Hankel Singular Values (HSV) of the controllability and observability gramians of the system are computed. The state-space dimensions with low HSVs are truncated, leading to a reduced-order model. For high-dimensional systems, computation of the required gramians is very expensive; combining POD and balanced truncation can overcome this limitation.

A large number of model-order reduction techniques has also been developed in particular to treat nonlinear time-dependent problems [8, 29, 28, 68, 84, 97, 108, 125]. Linearization approaches [125], for example, usually suffer from a lack of efficient representation of the nonlinear terms, whereas polynomial approximation approaches [84, 29] usually exhibit a fast exponential growth of computational complexity with the degree of the nonlinear approximation order. These two methods are thus quite expensive and do not address strong nonlinearities efficiently; other approaches for highly nonlinear systems (such as piecewise-linearization) have also been proposed [104, 97] but also at the expense of high computational cost and little control over model accuracy.

Furthermore, although *a priori* error bounds to quantify the error due to model reduction have been derived in the linear case, *a posteriori* error bounds have not yet been adequately considered even for the linear case, let alone the nonlinear case, for most model-order reduction approaches.

Finally, it is important to note that most model-order reduction techniques focus mainly on reduced-order modeling of dynamical systems in which time is considered the *only* "variable;" the development of reduced-order models for parametric applications is much less common [30, 25].

Our focus is (*i*) the simultaneous dependence of the field variable (and output) on both time and parameters, and (*ii*) the introduction of rigorous *a posteriori* error estimators.

1.2.3 Inverse Problems

Inverse problems are pervasive in engineering and science: ranging from geophysics [115, 128], to ecology [14], image processing [27], heat transfer [18, 19, 2, 80],physiology [11], continuum mechanics [12], medicine (e.g., hyperthermia treatment) [89, 31], and nondestructive evaluation [44]. The objective of the inverse problem is to determine unknown system parameters from observations (or measurements) of the state variables or outputs of the system.

Because of its practical importance, many methods have been developed to solve inverse problems; see [116] for a very recent review. Unfortunately, inverse problems are generally ill-posed and their solution thus difficult. One solution approach employs statistics: the to-be-estimated parameter is considered a random variable with unknown statistics which are estimated using, e.g., Monte Carlo Methods [52, 90] or simulated annealing [99, 100]. Probably the most common approach is to consider inverse problems as an optimization problems: a cost functional is defined to measure the difference between the measurements and the computed outputs from the system model. The parameter estimate is then found by minimizing the cost functional subject to the governing equations being satisfied. However, if the computational cost to solve the governing equations is high — as is the case for partial differential equations — the solution of the optimization problem may become unattainable [34]. Furthermore, regularization techniques are often employed to obtain a well-posed problem. Since the regularization changes the nature of the problem, the regularized solution differs from the original solution and valuable information can be lost.

In [74] a solution method for inverse problems governed by elliptic partial differential equations is proposed which explicitly quantifies the uncertainty in the problem formulation. In Chapter 7, we extend and generalize these ideas to the parabolic case: the outputs are then functions of time and measurements are taken at several discrete points in time. We consider several application and present numerical results that show the validity and indeed very good performance of this approach.

1.3 Scope

1.3.1 Thesis Contributions

In this thesis we focus on the development of reduced-basis output bound methods and associated *a posteriori* error estimation for parametrized parabolic partial differential equations. We improve and extend on earlier work [101, 102] in this field in several directions.

First, we consider a new class of problems and output families — we rigorously treat (a) temporal forcing/control inputs that are *not known a priori* (often a problem within the model reduction context) and (b) outputs, or functionals of the time-dependent field variable, that are also (scalar) *functions of (discrete) time.* We thus need to develop a new *a posteriori* error estimation procedure that provides rigorous bounds for the error in the energy norm and in the output at all (discrete) points in time. This generalization allows us to treat a much wider class of applications.

Second, based on our new *a posteriori* error bounds, we propose a "greedy" adaptive procedure to optimally construct the parameter-time sample set. This sampling procedure can help avoid ill-conditioning of the reduced-order model — a problem that easily occurs for a random sample set without *a priori* knowledge of the temporal forcing (e.g., given a periodic forcing in time only samples within one period should be chosen). Furthermore, under the assumption of linear time-invariance (LTI), we follow an impulse approach to construct our basis. The resulting reduced-basis approximation is then valid *for all* control input histories and the method applicable to (say) optimal control.

Third, we extend the methodology to treat nonaffine and (certain classes of) nonlinear problems. To this end, we will employ an empirical interpolation method introduced earlier [15] to approximate the nonaffine and nonlinear terms. We will introduce *a posteriori* error bounds which are rigorous under certain conditions on the function approximation, and offline-online computational procedures which are valid even in the presence of nonaffine and highly nonlinear terms.

Finally, we apply our methods to inverse and optimal control 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, and how they allow the development of new methods (e.g. quantifying the uncertainty in the solution of inverse problems) which would have been intractable with conventional methods.

1.3.2 Thesis Outline

In Chapter 2 we introduce the necessary mathematical background and give a short overview of the finite element method. We also present a short review of the empirical interpolation method for nonaffine coefficient functions. In Chapter 3 we summarize the reduced-basis method formulation and associated *a posteriori* error estimation for linear coercive elliptic problems.

Linear parabolic problems with affine parameter dependence are discussed in Chapter 4. We develop reduced-basis approximations and associated *a posteriori* error estimation and adjoint procedures for symmetric as well as nonsymmetric problems. We also propose a new greedy adaptive procedure to "optimally" construct the parameter-time sample set. In Chapter 5 we relax the condition on affine parameter dependence and extend the results from Chapter 4 to problems with nonaffine parameter dependence. We particularly focus on *a posteriori* error estimation and on efficient offline-online computational procedures. We extend our results to nonlinear parabolic problems in Chapter 6. Since nonlinear problems do not allow the same generality as linear problems, we focus on a certain class of problems with monotonic nonlinearity. At the end of this chapter we apply the proposed method to a particular nonlinear reaction-diffusion system.

In Chapter 7 we then apply the method developed in this thesis to several parameter identification problems. We first briefly discuss solution techniques for inverse problems and summarize a new method to characterize the uncertainty in the parameter estimation. The application of the proposed method to an optimal control problem is considered in Chapter 8. Our specific example is the control of welding quality, which combines parameter estimation and control techniques.

Finally, in Chapter 9 we summarize our work and conclude with some suggestions for future work.

Chapter 2

Preliminaries

2.1 Introduction

This chapter serves to provide some necessary background information for the work to follow. In Section 2.2 we review the basis function spaces used throughout this thesis; in Section 2.3 we introduce our "truth" approximation which is the point of departure for the reduced-basis method; and in Section 2.4 we review the empirical interpolation method introduced in [15] which is an essential building block for our discussion of nonaffine and nonlinear problems in Chapters 5 and 6.

2.2 Function Spaces

In this section, we introduce some notation and review some basic definitions that will be used in the following. The summary provided here is largely based on [71, 102]. To begin, let $\Omega \subset \mathbb{R}^d$, $d = 1, \ldots, 3$ be an open bounded domain with Lipschitz-continuous boundary $\partial\Omega$; we denote the closed domain by $\overline{\Omega}$.

We can then define the following function spaces:

2.2.1 Spaces of Continuous Functions

Definition 1. *Let k a non-negative integer. The space $C^k(\overline{\Omega})$ is then defined as*

$$C^k(\overline{\Omega}) \equiv \{v \mid D^\alpha v \textit{ is bounded and uniformly continuous on } \Omega, \ \forall\, \alpha \textit{ s.t. } 0 \leq |\alpha| \leq k\}\,;$$

where, for given multi-index $\alpha \equiv (\alpha_1, \ldots, \alpha_d)$, $\alpha_i \geq 0$, $1 \leq i \leq d$,

$$D^\alpha \equiv \frac{\partial^{|\alpha|}}{\partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d}}, \qquad |\alpha| = \sum_{i=1}^{d} \alpha_i.$$

Then $C^k(\overline{\Omega})$ is a Banach space (i.e., a complete normed linear space) with a norm

$$\|v\|_{C^k(\overline{\Omega})} = \max_{0 \leq |\alpha| \leq k} \sup_{x \in \Omega} |D^\alpha v(x)|.$$

We recall that $C_0^\infty(\Omega)$ is the space of continuous, infinitely differentiable functions with compact support, i.e., vanishing outside a bounded open set $\Omega' \subset \Omega$.

2.2.2 Lebesgue Spaces

Definition 2. *Let $p \geq 1$. The Lebesgue space $L^p(\Omega)$ is then defined as*

$$L^p(\Omega) \equiv \{v \mid \|v\|_{L^p(\Omega)} < \infty\}$$

where

$$\begin{aligned} \|v\|_{L^p(\Omega)} &\equiv \left(\int_\Omega |v|^p \, dx\right)^{\frac{1}{p}}, \quad 1 \leq p < \infty, \\ \|v\|_{L^\infty(\Omega)} &\equiv \operatorname*{ess\,sup}_{x \in \Omega} |v(x)|, \qquad p = \infty. \end{aligned}$$

We note that Lebesgue spaces are also Banach spaces. Here (and in subsequent chapters), "$\int_\Omega$" denotes the Lebesgue integral, and that, in theory, v is not a function but rather an (equivalence) class of functions that differ over a set of measure zero. Finally, the essential supremum of a function v, $\operatorname{ess\,sup}_{x\in\Omega} v(x)$, is defined as the greatest lower bound $C_{\max}$ of the set of all constants C, such that $|v(x)| \leq C$ "almost everywhere" on Ω.

2.2.3 Hilbert Spaces

Definition 3. *Let k be a non-negative integer. The Hilbert space $H^k(\Omega)$ is then defined as*

$$H^k(\Omega) \equiv \{v \mid D^\alpha v \in L^2(\Omega), \ \forall\, \alpha \ s.t. \ |\alpha| \leq k\},$$

with associated inner product

$$(w, v)_{H^k(\Omega)} \equiv \sum_{|\alpha| \leq k} \int_\Omega D^\alpha w \, D^\alpha v \, dx,$$

and induced norm

$$\|v\|_{H^k(\Omega)} \equiv \left(\sum_{|\alpha| \leq k} \int_\Omega |D^\alpha v|^2 \, dx\right)^{\frac{1}{2}}.$$

Hilbert spaces, which are the natural generalization of Euclidean spaces in the functional setting, will be used extensively in the subsequent chapters. We note that $L^2(\Omega)$ ($\equiv H^0(\Omega)$) is the only Lebesgue space that is a Hilbert space. Finally, since the Hilbert norm is induced by an inner-product, the Cauchy-Schwarz inequality holds:

$$|(w, v)_{H^k(\Omega)}| \leq \|w\|_{H^k(\Omega)} \|v\|_{H^k(\Omega)}.$$

2.2.4 Sobolev Spaces

Definition 4. *Let k be a non-negative integer and $p \geq 1$. The Sobolev space $W^{k,p}(\Omega)$ is then defined as*

$$W^{k,p}(\Omega) = \{v \mid D^\alpha v \in L^p(\Omega), \ \forall\ \alpha \ s.t. \ |\alpha| \leq k\};$$

the Sobolev spaces are Banach spaces with norms

$$\begin{aligned} \|v\|_{W^{k,p}(\Omega)} &\equiv \left(\sum_{|\alpha| \leq k} \int_\Omega |D^\alpha v|^p \, dx \right)^{\frac{1}{p}}, && 1 \leq p < \infty, \\ \|v\|_{W^{k,\infty}(\Omega)} &\equiv \max_{|\alpha| \leq k} \operatorname{ess\,sup}_{x \in \Omega} |D^\alpha v(x)|, && p = \infty. \end{aligned}$$

The Sobolev spaces are the natural setting for the variational formulation of partial differential equations. The case $k = 0$ for which $W^{0,p}(\Omega) \equiv L^p(\Omega)$ (and hence the Lebesgue spaces are included in the Sobolev spaces), and the case $p = 2$ which corresponds to a family of Hilbert Spaces, are of particular interest. Finally, we note that the derivatives here should be interpreted in the proper distributional sense [46].

2.2.5 Dual Hilbert Spaces

Definition 5. *Given a functional f, Hilbert space Y, and associated inner product and norm, $(\cdot, \cdot)_Y$ and $\| \cdot \|_Y$, respectively, we define the corresponding dual space Y' as*

$$Y' \equiv \{ f \mid \|f\|_{Y'} < \infty \},$$

where the dual norm $\| \cdot \|_{Y'}$ is given by

$$\|f\|_{Y'} \equiv \sup_{v \in Y} \frac{f(v)}{\|v\|_Y}. \tag{2.1}$$

The space Y' is also a Hilbert space, and for $Y = H^k(\Omega)$, we denote $Y' = H^{-k}(\Omega)$; in general:

$$H^k(\Omega) \subset \cdots \subset H^1(\Omega) \subset H^0(\Omega) \subset H^{-1}(\Omega) \subset \cdots \subset H^{-k}(\Omega).$$

From the Riesz representation theorem we know that for every $f \in Y'$ there exists a unique $u_f^Y \in Y$ such that

$$(u_f^Y, v)_Y = f(v), \quad \forall\, v \in Y.$$

It follows from the Cauchy-Schwarz inequality applied to the Y-inner product that

$$\|f\|_{Y'} = \sup_{v \in Y} \frac{(u_f^Y, v)_Y}{\|v\|_Y} = \|u_f^Y\|_Y.$$

This result is widely used in subsequent chapters.

2.3 "Truth" Approximation

In general, solving a partial differential equation exactly is difficult: a closed form solution is often unavailable. Classical discretization methods (such as the finite element method or the finite difference method) are therefore employed to obtain numerical approximations to the exact solution.

The point of departure for the methods presented in this thesis is the "truth" approximation — a numerical approximation that is sufficiently accurate such that the resulting approximate solution is "indistinguishable" from the exact solution. We now describe how we obtain such an approximation and introduce associated notation.

2.3.1 Exact Problem: A Simple Example

To begin, we consider, for the sake of illustration, the (linear) partial differential equation

$$\frac{\partial \tilde{y}^{\mathrm{e}}(x,t)}{\partial t} = \frac{\partial^2 \tilde{y}^{\mathrm{e}}(x,t)}{\partial x^2}, \quad x \in \Omega \equiv]0,1[\,, \ t \in I \equiv]0,t_f], \tag{2.2}$$

with initial condition

$$\tilde{y}^{\mathrm{e}}(x,0) = \tilde{y}^{\mathrm{e}}_0(x), \quad x \in \Omega, \tag{2.3}$$

and boundary conditions

$$\tilde{y}^{\mathrm{e}}(0,t) \ = \ 0, \quad t \in I, \tag{2.4}$$

$$\frac{\partial \tilde{y}^{\mathrm{e}}(1,t)}{\partial x} \ = \ 1, \quad t \in I; \tag{2.5}$$

(2.2) represents (say) the one-dimensional heat equation. Equations (2.2)–(2.4) are called the strong form of the initial-boundary-value problem.

2.3.2 Temporal Discretization: Finite Difference Approximation

Throughout this thesis, we directly consider a time-discrete framework associated to the time interval $I \equiv]0,t_f]$ ($\bar{I} \equiv [0,t_f]$). We divide $\bar{I}$ into K subintervals of equal length $\Delta t = t_f/K$ and define $t^k \equiv k\Delta t$, $0 \le k \le K \equiv t_f/\Delta t$; for notational convenience, we also introduce $\mathbb{K} \equiv \{1,\ldots,K\}$, and $\mathbb{I} \equiv \{t^0,\ldots,t^k\}$. Clearly, our results must be stable as $\Delta t \to 0$, $K \to \infty$.

We now employ a finite difference scheme to our initial-boundary-value problem. In particular, we approximate the time-derivative of a function g at time t^k by a first-order difference:

$$\frac{\partial g(t^k)}{\partial t} \approx \frac{1}{\Delta t}\left(g(t^k) - g(t^{k-1})\right). \tag{2.6}$$

Our finite-difference approximation, $y^{\mathrm{e}}(x,t^k)$, $k \in \mathbb{K}$, then satisfies

$$\frac{y^{\mathrm{e}}(x,t^k) - y^{\mathrm{e}}(x,t^{k-1})}{\Delta t} = \frac{\partial^2 y^{\mathrm{e}}(x,t^k)}{\partial x^2}, \quad x \in \Omega, \ \forall\, k \in \mathbb{K}, \tag{2.7}$$

with initial condition

$$y^{\mathrm{e}}(x,0) = y^{\mathrm{e}}_0(x), \quad x \in \Omega, \tag{2.8}$$

and boundary conditions

$$y^{\mathrm{e}}(0,t^k) \ = \ 0, \quad \forall\, k \in \mathbb{K}, \tag{2.9}$$

$$\frac{\partial y^{\mathrm{e}}(1,t^k)}{\partial x} \ = \ 1, \quad \forall\, kin\mathbb{K}; \tag{2.10}$$

It can be shown that this Euler-Backward scheme is unconditionally stable [94].

Note that we can also employ higher-order schemes such as Crank-Nicolson; details for the Crank-Nicolson scheme are summarized in Appendix B. There are also other time-discretization methods (aside from finite difference methods) which may be employed for time-dependent problems [51, 94, 117]. For example, if the finite element method is used for the spatial discretization, the discontinuous Galerkin method often lends itself well to the time discretization because of its variational origin. It was first introduced in the context of time-dependent problems in [50] and further analyzed in [63, 106]. The discontinuous Galerkin method has also been successfully applied to reduced-basis approximations of parabolic PDEs [101, 102]. However, finite difference methods are also widely used [42] and are the method of choice employed in this thesis.

2.3.3 Spatial Discretization: Finite Element Approximation

In this section, we provide a brief overview of the finite element method as applied to our simple one-dimensional example (2.7)–(2.10). Detailed treatment of the finite element method for partial differential equations may be found in [94], for example.

Variational or Weak Form

We now derive the weak form of the problem (2.7)–(2.10). In this and the following sections, we shall omit the dependence on the spatial variable, x; we thus write $y^{\mathrm{e}}(t^k)$ for $y^{\mathrm{e}}(x, t^k)$.

To begin, we multiply (2.7) by an arbitrary function v ($\equiv v(x)$) and integrate over the domain Ω to obtain

$$\frac{1}{\Delta t} \int_\Omega v \left(y^{\mathrm{e}}(t^k) - y^{\mathrm{e}}(x, t^{k-1}) \right) = \int_\Omega v \frac{\partial^2 y^{\mathrm{e}}(t^k)}{\partial x^2}. \tag{2.11}$$

Integrating the right-hand side by parts, we have

$$\frac{1}{\Delta t} \int_\Omega v \left(y^{\mathrm{e}}(t^k) - y^{\mathrm{e}}(t^{k-1}) \right) = \int_\Omega \frac{\partial}{\partial x} \left(v \frac{\partial y^{\mathrm{e}}(t^k)}{\partial x} \right) - \int_\Omega \frac{\partial v}{\partial x} \frac{\partial y^{\mathrm{e}}(t^k)}{\partial x}. \tag{2.12}$$

Rearranging terms and applying Green's Theorem, we obtain

$$\frac{1}{\Delta t} \int_\Omega v \left(y^{\mathrm{e}}(t^k) - y^{\mathrm{e}}(t^{k-1}) \right) + \int_\Omega \frac{\partial v}{\partial x} \frac{\partial y^{\mathrm{e}}(t^k)}{\partial x} = \left[v \frac{\partial y^{\mathrm{e}}(t^k)}{\partial x} \right]_{x=0}^{x=1}. \tag{2.13}$$

From the boundary conditions (2.4) and (2.5), we have

$$\frac{1}{\Delta t} \int_\Omega v \left(y^{\mathrm{e}}(t^k) - y^{\mathrm{e}}(t^{k-1}) \right) + \int_\Omega \frac{\partial v}{\partial x} \frac{\partial y^{\mathrm{e}}(t^k)}{\partial x} = v|_{x=1} - v|_{x=0} \left. \frac{\partial y^{\mathrm{e}}(0, t^k)}{\partial x} \right|_{x=0}. \tag{2.14}$$

It then follows that

$$m(y^{\mathrm{e}}(t^k), v) + \Delta t \, a(y^{\mathrm{e}}(t^k), v) = m(y^{\mathrm{e}}(t^{k-1}), v) + \Delta t \, f(v), \quad \forall \, v \in Y^{\mathrm{e}}, \tag{2.15}$$

where the function space Y^{e} is given by

$$Y^{\mathrm{e}} \equiv \{ v \in H^1(\Omega) \mid v(0) = 0 \}, \tag{2.16}$$

and the linear form, $f \in Y^{\mathrm{e}\prime}$, and bilinear forms, $a \colon Y^{\mathrm{e}} \times Y^{\mathrm{e}} \to \mathbb{R}$ and $m \colon Y^{\mathrm{e}} \times Y^{\mathrm{e}} \to \mathbb{R}$, are defined as

$$\begin{aligned} f(v) &= v|_{x=1}, & \forall\, v \in Y^{\mathrm{e}}, \\ a(w,v) &= \int_\Omega \frac{\partial v}{\partial x}\frac{\partial w}{\partial x}, & \forall\, w, v \in Y^{\mathrm{e}}, \\ m(w,v) &= \int_\Omega wv, & \forall\, w, v \in Y^{\mathrm{e}}. \end{aligned} \tag{2.17}$$

Equation (2.15) is the weak form of (2.7)–(2.10); the weak form is the point of departure for the finite element method, which will be discussed in the following section.

Triangulation

We now decompose the domain Ω into a collection of J *elements* (in this one-dimensional case, segments; in general, simplices), T_h^j, $1 \le j \le J$. Such a decomposition, which we denote $\mathcal{T}_h$ is known as a *triangulation* of Ω; an example is shown in Figure 2-1(a). Note that the elements are open (i.e., the T_h^j exclude the nodes) and satisfy $T_h^j \cap T_h^{j'} = \emptyset$ for $j \neq j'$; furthermore, the union of the (closure of the) elements reconstitutes the original domain, that is,

$$\overline{\Omega} = \bigcup_{T_h \in \mathcal{T}_h} \overline{T_h}, \tag{2.18}$$

where T_h refers to any particular member of the triangulation $\mathcal{T}_h$. In general, the sizes of the elements in a mesh are different; here, the subscript "h" denotes the maximum diameter over all elements.

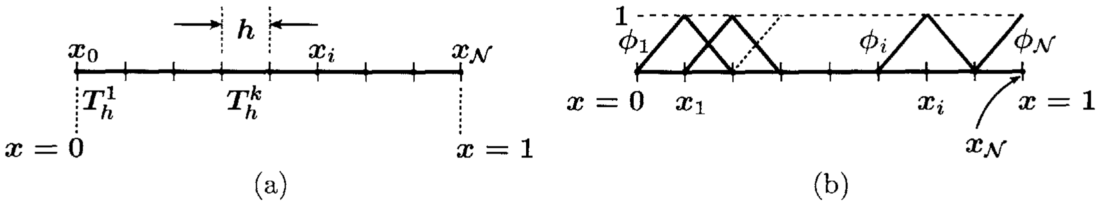


Figure 2-1: (a) Triangulation of the domain $\Omega \equiv]0,1[$; and (b) nodal basis functions $\phi_i(x)$, $1 \le i \le \mathcal{N}$, for Y.

Finite Element Approximation

We now define the "truth" finite element approximation space $Y \subset Y^{\mathrm{e}}$ as

$$Y = \{v \in X \mid v|_{T_h^j} \in \mathbb{P}_1(T_h^j),\ 1 \le j \le J\}; \tag{2.19}$$

in other words, Y is the space of functions which are linear over each element. We note that any element v of Y may be written as

$$v(x) = \sum_{n=1}^{\mathcal{N}} v_n \phi_n(x), \tag{2.20}$$

where $\mathcal{N}$ is the dimension of Y, the $\phi_n(x)$, $n = 1, \ldots, \mathcal{N}$ form a basis for Y, and the coefficients $\underline{v} \equiv [v_1, \ldots, v_{\mathcal{N}}]^T \in \mathbb{R}^N$ are unique. For our one-dimensional example, (2.7)–(2.10), $\mathcal{N} = J$, and we take the $\phi_n(x)$ to be the hat functions shown in Figure 2-1(b).

We may now define our "truth" finite element approximation to $y^{\mathrm{e}}(t^k)$ of (2.15): we calculate $y(t^k) \in Y$ which satisfies

$$m(y(t^k), v) + \Delta t\, a(y(t^k), v) = m(y(t^{k-1}), v) + \Delta t\, f(v), \quad \forall\, v \in Y,\ k \in \mathbb{K}. \tag{2.21}$$

It can be shown that as $h \to 0$ (and therefore $\mathcal{N} \to \infty$), $y(x, t^k) \to y^{\mathrm{e}}(t^k)$, $\forall\, k \in \mathbb{K}$; we assume that Y is sufficiently rich that $y(x, t^k)$ is sufficiently close to $y^{\mathrm{e}}(t^k)$, $\forall\ k \in \mathbb{K}$. The truth approximation of the form (2.21) is the point of departure for the reduced-basis method. The reduced-basis approximation shall be build upon the truth approximation, and the reduced-basis error will thus be evaluated with respect to $y(t^k) \in Y$.

Discrete Equations

We note that since $y(x, t^k) \in Y$, we can express $y(x, t^k)$ in terms of the basis functions $\phi_j(x)$:

$$y(x, t^k) = \sum_{j=1}^{\mathcal{N}} y_j(t^k)\phi_j(x), \tag{2.22}$$

where the coefficients $\underline{y}(t^k) \equiv [y_1(t^k), \ldots, y_{\mathcal{N}}(t^k)]^T \in \mathbb{R}^N$ associated with the basis functions $\phi_n(x)$, $n = 1, \ldots, N$, $\forall k \in \mathbb{K}$. Substituting (2.22) into (2.21), and choosing for the test functions $v = \phi_i(x)$, we obtain the algebraic system of equations

$$M\underline{y}(t^k) + \Delta t\, A\underline{y}(t^k) = M\underline{y}(t^{k-1}) + \Delta t\, F, \tag{2.23}$$

where

$$\begin{aligned} M_{ij} &= m(\phi_j, \phi_i), \quad 1 \le i, j \le \mathcal{N} \\ A_{ij} &= a(\phi_j, \phi_i), \quad 1 \le i, j \le \mathcal{N} \\ F_i &= f(\phi_i), \qquad 1 \le i \le \mathcal{N}. \end{aligned} \tag{2.24}$$

2.4 Empirical Interpolation Method

2.4.1 Coefficient-Function Approximation

We begin by summarizing the results in [15]. We consider the problem of approximating a given μ-dependent function of sufficient regularity, $g(\,\cdot\,; \mu) \in L^\infty(\Omega)$, $\forall\, \mu \in \mathcal{D}$, by a reduced-basis expansion $g_M(\,\cdot\,; \mu)$. To this end, we introduce the sample sets

$$S_M^g \equiv \{\mu_1^g \in \mathcal{D}, \ldots, \mu_M^g \in \mathcal{D}\}, \quad 1 \le M \le M_{\max}, \tag{2.25}$$

and associated reduced-basis spaces

$$W_M^g = \text{span}\ \{\xi_m \equiv g(x; \mu_m^g), 1 \le m \le M\}, \quad 1 \le M \le M_{\max}, \tag{2.26}$$

in which our approximation g_M shall reside. Note that the sample sets and therefore the reduced-basis spaces are, by definition, *nested*: $S_1^g \subset S_M^g \subset S_{M_{\max}}^g$, and $W_1^g \subset W_M^g \subset W_{M_{\max}}^g$. We also introduce the best approximation

$$g_M^*(\,\cdot\,;\mu) \equiv \arg\min_{z \in W_M^g} \|g(\,\cdot\,;\mu) - z\|_{L^\infty(\Omega)} \tag{2.27}$$

and the associated error

$$\varepsilon_M^*(\mu) \equiv \|g(\,\cdot\,;\mu) - g_M^*(\,\cdot\,;\mu)\|_{L^\infty(\Omega)}. \tag{2.28}$$

The construction of S_M^g and W_M^g is based on a greedy algorithm. To begin, we select our first sample point to be $\mu_1^g = \arg\max_{\mu\in\Xi^g} \|g(\,\cdot\,;\mu)\|_{L^\infty(\Omega)}$, and define $S_1^g = \{\mu_1^g\}$, $\xi_1 \equiv g(x;\mu_1^g)$, and $W_1^g = \text{span}\,\{\xi_1\}$, where Ξ^g is a suitably fine parameter sample over $\mathcal{D}$. Then, for $M \geq 2$, we set $\mu_M^g = \arg\max_{\mu\in\Xi^g} \varepsilon_{M-1}^*(\mu)$, and define $S_M^g = S_{M-1}^g \cup \mu_M^g$, $\xi_M = g(x;\mu_M^g)$, and $W_M^g = \text{span}\,\{\xi_m,\ 1 \leq m \leq M\}$. In essence, W_M^g comprises basis functions on the parametrically induced manifold $\mathcal{M}^g \equiv \{g(\,\cdot\,;\mu) \mid \mu \in \mathcal{D}\}$. Thanks to our truth approximation, solving for $g_{M-1}^*(\,\cdot\,;\mu)$ and hence $\varepsilon_{M-1}^*(\mu)$ is a *standard linear program.*

Before we proceed, we note that the evaluation of $\varepsilon_M^*(\mu)$, $1 \leq M \leq M_{\max}$, requires the solution of a linear program for *each* parameter sample in Ξ^g; the computational cost involved thus depends strongly on the size of Ξ^g as well as on $M_{\max}$. In the parabolic case this cost may become prohibitively large — at least in our current implementation — if the function g is itself time-varying either through (i) an explicit dependence on time, or (ii) an implicit dependence on time in nonlinear problems, where g is a function of the time-dependent field variable $u(\mu, t^k)$. In these cases the parameter sample Ξ^g is in effect replaced by the *parameter-time* sample $\tilde{\Xi}^g \equiv \Xi^g \times \mathbb{I}$, i.e., the number of samples in Ξ^g is multiplied by the number of timesteps K; even for modest K the computational cost can be very high. We thus propose an alternative way of constructing S_M^g: we simply replace the $L^\infty(\Omega)$-norm in our best approximation by the $L^2(\Omega)$-norm, where $L^2(\Omega)$ is the space of functions square integrable over Ω — our next sample point is thus based on $\mu_M^g = \arg\max_{\mu\in\Xi^g} \inf_{z\in W_{M-1}^g} \|g(\,\cdot\,;\mu) - z\|_{L^2(\Omega)}$ — which is relatively inexpensive to evaluate; the computational cost is $O(M\mathcal{N}) + O(M^3)$: we first solve for the coefficients $\lambda_{M-1,m}(\mu)$, $1 \leq m \leq M-1$, from $((Z^{M-1})^T Z^{M-1})\,\lambda_{M-1}(\mu) = (Z^{M-1})^T g(\cdot;\mu)$, where $Z^{M-1} = [\xi_1 \ldots \xi_{M-1}]$, and then evaluate the norm $\|g(\,\cdot\,;\mu) - Z^{M-1}\lambda_{M-1}(\mu)\|_{L^2(\Omega)}$. Although the following analysis is not rigorous for this alternative (or "surrogate") construction of S_M^g, we in fact obtain very similar convergence results in practice (see Section 2.4.3).

We begin the analysis of our greedy procedure with the following Lemma.

Lemma 1. *Suppose that $M_{\max}$ is chosen such that the dimension of $\mathcal{M}^g$ exceeds $M_{\max}$; then the space W_M^g is of dimension M.*

Proof. It directly follows from our hypothesis on $M_{\max}$ that $\varepsilon_0 \equiv \varepsilon_{M_{\max}}^*(\mu_{M_{\max}+1}^g) > 0$; our "arg max" construction then implies $\varepsilon_{M-1}^*(\mu_M^g) \geq \varepsilon_0$, $2 \leq M \leq M_{\max}$, since $\varepsilon_{M-1}^*(\mu_M^g) \geq \varepsilon_{M-1}^*(\mu_{M+1}^g) \geq \varepsilon_M^*(\mu_{M+1}^g)$. We now prove Lemma 1 by induction. Clearly, $\dim(W_1^g) = 1$. Assume $\dim(W_{M-1}^g) = M-1$; then if $\dim(W_M^g) \neq M$, we have $g(\,\cdot\,;\mu_M^g) \in W_{M-1}^g$ and thus $\varepsilon_{M-1}^*(\mu_M^g) = 0$; however, the latter contradicts $\varepsilon_{M-1}^*(\mu_M^g) \geq \varepsilon_0 > 0$. □

We now construct nested sets of interpolation points $T_M = \{t_1, \ldots, t_M\}$, $1 \leq M \leq M_{\max}$. We first set $t_1 = \arg\,\text{ess}\,\sup_{x\in\Omega} |\xi_1(x)|$, $q_1 = \xi_1(x)/\xi_1(t_1)$, $B_{11}^1 = 1$. Then for $M = 2, \ldots, M_{\max}$, we solve the linear system $\sum_{j=1}^{M-1} \sigma_j^{M-1}\, q_j(t_i) = \xi_M(t_i)$, $1 \leq i \leq M-1$, and set $r_M(x) = \xi_M(x) -$

$\sum_{j=1}^{M-1} \sigma_j^{M-1}\, q_j(x)$, $t_M = \arg \operatorname{ess}\sup_{x\in\Omega} |r_M(x)|$, $q_M(x) = r_M(x)/r_M(t_M)$, and $B_{i\,j}^M = q_j(t_i)$, $1 \le i,j \le M$. It remains to demonstrate

Lemma 2. *The construction of the interpolation points is well-defined, and the functions* $\{q_1,\dots,q_M\}$ *form a basis for* W_M^g.

Proof. We shall proceed by induction. Clearly, we have $W_1^g = \operatorname{span}\{q_1\}$. Next we assume $W_{M-1}^g = \operatorname{span}\{q_1,\dots,q_{M-1}\}$; if (i) $|r_M(t_M))| > 0$ and (ii) B^{M-1} is invertible, then our construction may proceed and we may form $W_M^g = \operatorname{span}\{q_1,\dots,q_M\}$. To prove (i), we observe that $|r_M(t_M)| \ge \varepsilon_{M-1}^*(\mu_M^g) \ge \varepsilon_0 > 0$ since $\varepsilon_{M-1}^*(\mu_M^g)$ is the error associated with the best approximation. To prove (ii), we just note by the construction procedure that $B_{i\,j}^{M-1} = r_j(t_i)/r_j(t_j) = 0$ for $i < j$; that $B_{i\,j}^{M-1} = r_j(t_i)/r_j(t_j) = 1$ for $i = j$; and that $\left|B_{i\,j}^{M-1}\right| = |r_j(t_i)/r_j(t_j)| \le 1$ for $i > j$ since $t_i = \arg \operatorname{ess}\sup_{x\in\Omega} |r_i(x)|, 1 \le i \le M$. Hence, B^{M-1} is lower triangular with unity diagonal. □

Lemma 3. *For any M-tuple $(\alpha_i)_{i=1,\dots,M}$ of real numbers, there exists a unique element $w \in W_M^g$ such that $w(t_i) = \alpha_i$, $1 \le i \le M$.*

Proof. Since the functions $\{q_1,\dots,q_M\}$ form a basis for W_M^g (Lemma 2), any member of W_M^g can be expressed as $w = \sum_{j=1}^M q_j(x)\,\kappa_j$. Recalling that B^M is invertible, we may now consider the particular function w corresponding to the choice of coefficients κ_j, $1 \le j \le M$, such that $\sum_{j=1}^M B_{ij}^M \kappa_j = \alpha_i$, $1 \le i \le M$; but since $B_{ij}^M = q_j(t_i)$, $w(t_i) = \sum_{j=1}^M q_j(t_i)\,\kappa_j = \sum_{j=1}^M B_{ij}^M\,\kappa_j = \alpha_i$, $1 \le i \le M$, which hence proves existence. To prove uniqueness, we need only consider two possible candidates and again invoke the invertibility of B^M. □

It remains to develop an *efficient* procedure for obtaining a *good* collateral reduced-basis expansion $g_M(\cdot;\mu)$. Based on the approximation space W_M^g and set of interpolation points T_M, we can readily construct an approximation to $g(x;\mu)$. Indeed, our coefficient function approximation is the interpolant of g over T_M as provided for from Lemma 3:

$$g_M(x;\mu) = \sum_{m=1}^M \varphi_{M\,m}(\mu)\, q_m(x), \tag{2.29}$$

where $\varphi_M(\mu) \in \mathbb{R}^M$ is given by

$$\sum_{j=1}^M B_{i\,j}^M\, \varphi_{M\,j}(\mu) = g(t_i;\mu),\ 1 \le i \le M; \tag{2.30}$$

note that $g_M(t_i;\mu) = g(t_i;\mu), 1 \le i \le M$. We define the associated error as

$$\varepsilon_M(\mu) \equiv \|g(\,\cdot\,;\mu) - g_M(\,\cdot\,;\mu)\|_{L^\infty(\Omega)}. \tag{2.31}$$

2.4.2 Error Analysis

A Priori Stability: Lebesgue Constant

To begin, we define a "Lebesgue constant" [93] $\Lambda_M = \sup_{x\in\Omega} \sum_{m=1}^M |V_m^M(x)|$. Here, the $V_m^M(x) \in W_M^g$ are characteristic functions satisfying $V_m^M(t_n) = \delta_{mn}$, the existence and uniqueness of which is guaranteed by Lemma 3. It can be shown that

Lemma 4. *The characteristic functions V^M_m are a basis for W^g_M. And the two bases q_m, $1 \leq m \leq M$, and V^M_m, $1 \leq m \leq M$, are related by*

$$q_i(x) = \sum_{j=1}^{M} B^M_{j\,i}\, V^M_j(x), \quad 1 \leq i \leq M \ . \tag{2.32}$$

Proof. We first consider $x = t_n$, $1 \leq n \leq M$, and note that $\sum_{j=1}^M B^M_{j\,i}\, V^M_j(t_n) = \sum_{j=1}^M B^M_{j\,i}\, \delta_{jn} = B^M_{n\,i} = q_i(t_n)$, $1 \leq i \leq M$; it thus follows from Lemma 3 that (2.32) holds. It further follows from Lemma 2 and from Lemma 3 that any $w \in W^g_M$ can be uniquely expressed as $w = \sum_{i=1}^M \kappa_i q_i(x) = \sum_{i=1}^M \kappa_i (\sum_{j=1}^M B^M_{j\,i} V^M_j(x)) = \sum_{j=1}^M (\sum_{i=1}^M \kappa_i B^M_{j\,i}) V^M_j(x) = \sum_{j=1}^M \alpha_j V^M_j(x)$, where $\alpha_j = w(t_j)$, $1 \leq j \leq M$; thus the V^M_j, $1 \leq j \leq M$, form a ("nodal") basis for W^g_M. □

We observe that Λ_M depends on W^g_M and T_M, but not on μ nor on our choice of basis for W^g_M. We can further prove

Lemma 5. *The interpolation error $\varepsilon_M(\mu)$ satisfies $\varepsilon_M(\mu) \leq \varepsilon^*_M(\mu)(1 + \Lambda_M)$, $\forall\, \mu \in \mathcal{D}$.*

Proof. We first introduce $e^*_M(x;\mu) = g(x;\mu) - g^*_M(x;\mu)$ and $g_M(x;\mu) - g^*_M(x;\mu) = \sum_{m=1}^M \kappa_m(\mu) q_m(x)$. It then follows that

$$\begin{aligned} e^*_M(t_i;\mu) &= (g(t_i;\mu) - g_M(t_i;\mu)) + (g_M(t_i;\mu) - g^*_M(t_i;\mu)) \\ &= \sum_{m=1}^{M} B^M_{i\,m} \kappa_m(\mu), \quad 1 \leq i \leq M \ . \end{aligned} \tag{2.33}$$

Furthermore, from the definition of $\varepsilon_M(\mu)$ and $\varepsilon^*_M(\mu)$, and the triangle inequality, we obtain

$$\begin{aligned} \varepsilon_M(\mu) &= \|g(\,\cdot\,;\mu) - g_M(\,\cdot\,;\mu)\|_{L^\infty(\Omega)} \\ &= \|g(\,\cdot\,;\mu) - g^*_M(\,\cdot\,;\mu) + g^*_M(\,\cdot\,;\mu) - g_M(\,\cdot\,;\mu)\|_{L^\infty(\Omega)} \\ &\leq \varepsilon^*_M(\mu) + \|g_M(\,\cdot\,;\mu) - g^*_M(\,\cdot\,;\mu)\|_{L^\infty(\Omega)}. \end{aligned}$$

The desired result

$$\begin{aligned} \varepsilon_M(\mu) - \varepsilon^*_M(\mu) &\leq \|g_M(\,\cdot\,;\mu) - g^*_M(\,\cdot\,;\mu)\|_{L^\infty(\Omega)} \\ &= \|\sum_{k=1}^{M} \sum_{m=1}^{M} B^M_{k\,m}\, \kappa_m(\mu)\, V^M_k(x)\|_{L^\infty(\Omega)} \\ &= \|\sum_{i=1}^{M} e^*_M(t_i;\mu)\, V^M_i(x)\|_{L^\infty(\Omega)} \\ &\leq \varepsilon^*_M(\mu)\, \Lambda_M \end{aligned}$$

then immediately follows from (2.32), (2.33), and $|e^*_M(t_i;\mu)| \leq \varepsilon^*_M(\mu)$, $1 \leq i \leq M$. □

We can further show

Proposition 1. *The Lebesgue constant Λ_M satisfies $\Lambda_M \leq 2^M - 1$.*

Proof. We first recall two crucial properties of the matrix B^M: (i) B^M is lower triangular with unity diagonal — $q_m(t_m) = 1$, $1 \leq m \leq M$, and (ii) all entries of B^M are of modulus no greater than unity — $\|q_m\|_{L^\infty(\Omega)} \leq 1$, $1 \leq m \leq M$. Hence, from (2.32) we can write

$$\begin{aligned} |V_m^M(x)| &= \left| q_m(x) - \sum_{i=m+1}^{M} B_{i\,m}^M V_i^M(x) \right| \\ &\leq 1 + \sum_{i=m+1}^{M} |V_i^M(x)|, \quad 1 \leq m \leq M-1. \end{aligned}$$

It follows that, starting from $|V_M^M(x)| = |q_M(x)| \leq 1$, we can deduce $|V_{M+1-m}^M(x)| \leq 1 + |V_M^M(x)| + \ldots + |V_{M+2-m}^M(x)| \leq 2^{m-1}$, $2 \leq m \leq M$, and thus obtain $\sum_{m=1}^{M} |V_m^M(x)| \leq 2^M - 1$. □

Proposition 1 is very pessimistic and of little practical value (though $\varepsilon_M^*(\mu)$ does often converge sufficiently rapidly that $\varepsilon_M^*(\mu)\, 2^M \to 0$ as $M \to \infty$); this is not surprising given analogous results in the theory of polynomial interpolation [93]. However, Proposition 1 does provide some notion of stability.

A Posteriori Error Estimation

Given an approximation $g_M(x;\mu)$ for $M \leq M_{\max} - 1$, we define $\mathcal{E}_M(x;\mu) \equiv \hat{\varepsilon}_M(\mu)\, q_{M+1}(x)$, where $\hat{\varepsilon}_M(\mu) \equiv |g(t_{M+1};\mu) - g_M(t_{M+1};\mu)|$. In general, $\varepsilon_M(\mu) \geq \hat{\varepsilon}_M(\mu)$, since $\varepsilon_M(\mu) = \|g(\cdot;\mu) - g_M(\cdot;\mu)\|_{L^\infty(\Omega)} \geq |g(x;\mu) - g_M(x;\mu)|$ for all $x \in \Omega$, and thus also for $x = t_{M+1}$. However, we can prove

Proposition 2. *If $g(\,\cdot\,;\mu) \in W_{M+1}^g$, then (i) $g(x;\mu) - g_M(x;\mu) = \pm\mathcal{E}_M(x;\mu)$ (either $\mathcal{E}_M(x;\mu)$ or $-\mathcal{E}_M(x;\mu)$), and (ii) $\|g(\,\cdot\,;\mu) - g_M(\,\cdot\,;\mu)\|_{L^\infty(\Omega)} = \hat{\varepsilon}_M(\mu)$.*

Proof. By our assumption $g(\cdot;\mu) \in W_{M+1}^g$, there exists $\kappa(\mu) \in \mathbb{R}^{M+1}$ such that $g(x;\mu) - g_M(x;\mu) = \sum_{m=1}^{M+1} \kappa_m(\mu)\, q_m(x)$. We now consider $x = t_i, 1 \leq i \leq M+1$, and arrive at

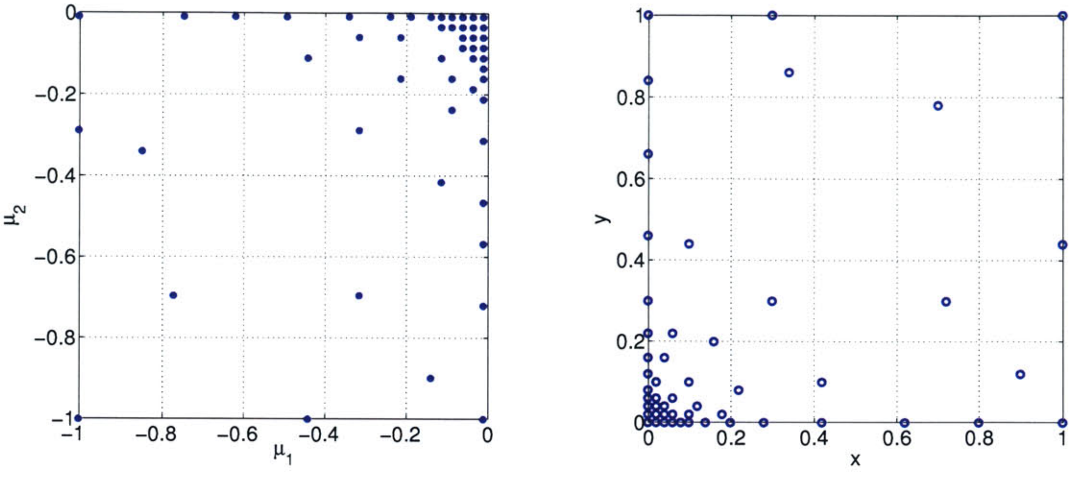
$$\sum_{m=1}^{M+1} \kappa_m(\mu)\, q_m(t_i) = g(t_i;\mu) - g_M(t_i;\mu), \quad 1 \leq i \leq M+1\ . \tag{2.34}$$

It thus follows that $\kappa_m(\mu) = 0$, $1 \leq m \leq M$, since $g(t_i;\mu) - g_M(t_i;\mu) = 0, 1 \leq i \leq M$ and the matrix $q_m(t_i) (= B_{im}^M)$ is lower triangular, and that $\kappa_{M+1}(\mu) = g(t_{M+1};\mu) - g_M(t_{M+1};\mu)$ since $q_{M+1}(t_{M+1}) = 1$; this concludes the proof of (i). The proof of (ii) then directly follows from $\|q_{M+1}\|_{L^\infty(\Omega)} = 1$. □

Of course, in general $g(\,\cdot\,;\mu) \notin W_{M+1}^g$, and hence our estimator $\hat{\varepsilon}_M(\mu)$ is indeed a lower bound; however, if $\varepsilon_M(\mu) \to 0$ very fast, we expect that the effectivity,

$$\eta_M(\mu) \equiv \frac{\hat{\varepsilon}_M(\mu)}{\varepsilon_M(\mu)}\,, \tag{2.35}$$

shall be close to unity. Furthermore, the estimator is very inexpensive – *one additional evaluation* of $g(\,\cdot\,;\mu)$ at a single point in Ω.



(a) (b)

Figure 2-2: NE 1: (a) Parameter sample set S_M^g, $M_{\max} = 51$, and (b) interpolation points t_m, $1 \leq m \leq M_{\max}$, for the nonaffine function (2.36).

2.4.3 Numerical Exercise 1: Approximation of a Nonaffine Function

Problem Formulation

We consider the nonaffine function

$$G(x;\mu) \equiv \frac{1}{\sqrt{(x_1 - \mu_{(1)})^2 + (x_2 - \mu_{(2)})^2}} \tag{2.36}$$

for $x = (x_{(1)}, x_{(2)}) \in \Omega \equiv]0,1[^{\,2}$ and $\mu \in \mathcal{D} \equiv [-1,-0.01]^2$. We choose for Ξ^g a deterministic grid of 40×40 parameter points over $\mathcal{D}$ and we take $\mu_1^g = (-0.01,-0.01)$. Next, we pursue the empirical interpolation procedure described in Section 2.4.1 to construct S_M^g, W_M^g, T_M, and B^M, $1 \leq M \leq M_{\max}$, for $M_{\max} = 51$. We note that the parameter points in S_M^g, shown in Figure 2-2(a), are mainly distributed around the corner $(-0.01,-0.01)$ of the parameter domain; and that the interpolation points in T_M, plotted in Figure 2-2(b), are largely allocated around the corner $(0,0)$ of the physical domain Ω.

Numerical Results

We now introduce a parameter test sample Ξ_{Test}^g of size $Q_{\text{Test}} = 225$, and define

$$\varepsilon_{M,\max}^* \equiv \max_{\mu \in \Xi_{\text{Test}}^g} \varepsilon_M^*(\mu), \tag{2.37}$$

$$\overline{\rho}_M \equiv Q_{\text{Test}}^{-1} \sum_{\mu \in \Xi_{\text{Test}}^g} \frac{\varepsilon_M(\mu)}{\varepsilon_M^*(\mu)(1+\Lambda_M)} \tag{2.38}$$

$$\overline{\eta}_M \equiv Q_{\text{Test}}^{-1} \sum_{\mu \in \Xi_{\text{Test}}^g} \eta_M(\mu); \tag{2.39}$$

here $\eta_M(\mu)$ is the effectivity defined in (2.35), and $\varkappa_M$ is the condition number of B^M. We present in Table 2.1 $\varepsilon^*_{M,\max}$, $\overline{\rho}_M$, Λ_M, $\overline{\eta}_M$, and $\varkappa_M$ as a function of M. We observe that $\varepsilon^*_{M,\max}$ converges rapidly with M; that the Lebesgue constant provides a reasonably sharp measure of the interpolation-induced error; that the Lebesgue constant grows very slowly — $\varepsilon_M(\mu)$ is *only slightly larger that the min-max result* $\varepsilon^*_M(\mu)$; that the error estimator effectivity is reasonably close to unity; and that B^M is quite well-conditioned for our choice of basis. (For the non-orthogonalized basis ξ_m, $1 \leq m \leq M$, the condition number of B^M will grow exponentially with M.) These results are expected since the given function $G(x;\mu)$ is quite regular and smooth in the parameter μ.

M	$\varepsilon^*_{M,\max}$	$\overline{\rho}_M$	Λ_M	$\overline{\eta}_M$	κ_M
8	8.30 E – 02	0.68	1.76	0.17	3.65
16	4.22 E – 03	0.67	2.63	0.10	6.08
24	2.68 E – 04	0.49	4.42	0.28	9.19
32	5.64 E – 05	0.48	5.15	0.20	12.86
40	3.66 E – 06	0.54	4.98	0.60	18.37
48	6.08 E – 07	0.37	7.43	0.29	20.41

Table 2.1: NE 1: $\varepsilon^*_{M,\max}$, $\overline{\rho}_M$, Λ_M, $\overline{\eta}_M$, and $\varkappa_M$ as a function of M.

Using the $L^2(\Omega)$-norm surrogate in our best approximation we can construct S^g_M much less expensively than using the $L^\infty(\Omega)$-norm. We present in Table 2.2 numerical results obtained from this alternative construction of S^g_M. The results are very similar to those in Table 2.1, which implies that the approximation quality of our empirical interpolation approach is relatively insensitive to the choice of norms exploited in constructing the sample.

M	$\varepsilon^*_{M,\max}$	$\overline{\rho}_M$	Λ_M	$\overline{\eta}_M$	$\varkappa_M$
8	1.18 E – 01	0.66	2.26	0.23	3.82
16	3.96 E – 03	0.45	4.86	0.81	7.58
24	3.83 E – 04	0.43	3.89	0.28	13.53
32	3.92 E – 05	0.45	7.07	0.47	16.60
40	4.10 E – 06	0.43	6.40	0.25	18.84
48	6.59 E – 07	0.30	8.86	0.18	21.88

Table 2.2: NE 1: $\varepsilon^*_{M,\max}$, $\overline{\rho}_M$, Λ_M, $\overline{\eta}_M$, and $\varkappa_M$ as a function of M; here S^g_M is constructed using the $L^2(\Omega)$-norm as a surrogate for the $L^\infty(\Omega)$-norm.

Chapter 3

Reduced-Basis Method for Elliptic Problems

3.1 Introduction

In this chapter, we present a more detailed discussion of the reduced-basis output approximation method for linear coercive elliptic problems. We focus particularly on the *global* approximation spaces, *a priori* convergence theory, and the assumption of affine parameter dependence.

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

3.2 Abstraction

3.2.1 "Exact" Problem Statement

We consider a suitably regular (smooth) domain $\Omega \subset \mathbb{R}^d, d = 1, 2$, or 3[1] with Lipschitz-continuous boundary $\partial\Omega$, and associated (infinite-dimensional) Hilbert space Y^{e} satisfying $H_0^1(\Omega) \subset Y^{\mathrm{e}} \subset (H^1(\Omega))^d$. The inner product and norm associated with Y^{e} are given by $(\cdot,\cdot)_{Y^{\mathrm{e}}}$ and $||\cdot||_{Y^{\mathrm{e}}} \equiv (\cdot,\cdot)_{Y^{\mathrm{e}}}^{1/2}$, respectively. The corresponding dual space of Y^{e} is denoted $Y^{\mathrm{e}\prime}$. We also define a parameter set $\mathcal{D} \subset \mathbb{R}^P$, a particular point in which will be denoted μ.

Our "exact" problem may then be stated as: Given a parameter $\mu \in \mathcal{D} \subset \mathbb{R}^P$, we evaluate the output of interest,

$$s^{\mathrm{e}}(\mu) = \ell(y^{\mathrm{e}}(\mu)); \tag{3.1}$$

here the field variable, $y^{\mathrm{e}}(\mu) \in Y^{\mathrm{e}}$, satisfies the weak form of the μ-parametrized linear elliptic partial differential equation

$$a(y^{\mathrm{e}}(\mu), v; \mu) = f(v), \quad \forall\, v \in Y^{\mathrm{e}}. \tag{3.2}$$

The form $a(\cdot,\cdot;\mu)\colon Y^{\mathrm{e}} \times Y^{\mathrm{e}} \to \mathbb{R}$ is bilinear — i.e., linear in the first and second argument; and f is a bounded linear functional.

[1] Note that Ω is a *reference domain* and hence does *not* depend on the parameter.

3.2.2 "Truth" Finite Element Approximation

In actual practice, we replace Y^{e} with a "truth" finite element approximation space $Y \subset Y^{\mathrm{e}}$ of finite (but large) dimension $\mathcal{N}$ — as discussed in the last chapter. Note that the inner product and norm associated with Y, $(\cdot,\cdot)_Y$ and $||\cdot||_Y \equiv (\cdot,\cdot)_Y^{1/2}$, respectively, are inherited from Y^{e}. Our "truth" finite element approximation $y(\mu) \in Y$ to $y^{\mathrm{e}}(\mu)$ is then defined as the Galerkin projection of $y^{\mathrm{e}}(\mu)$ onto Y:

$$a(y(\mu), v; \mu) = f(v), \quad \forall\, v \in Y; \tag{3.3}$$

our output approximation is then given by

$$s(\mu) = \ell(y(\mu)). \tag{3.4}$$

We shall assume — hence the appellation "truth" — that the finite element discretization is sufficiently rich such that $y(\mu)$ and $y^{\mathrm{e}}(\mu)$ (and hence $s(\mu)$ and $s^{\mathrm{e}}(\mu)$) are indistinguishable. The reduced-basis approximation shall be built upon our "truth" finite element approximation, and the reduced-basis error will thus be evaluated with respect to $y(\mu) \in Y$ and $s(\mu)$. Clearly, our methods must remain computationally efficient and stable as $\mathcal{N} \to \infty$.

We now make several assumptions on the well-posedness and the nature of the parametric dependence of our problem.

3.2.3 Well-posedness

We first assume that the bilinear form $a(\cdot,\cdot;\mu)$ is continuous,

$$\gamma_a(\mu) \equiv \sup_{v \in Y} \frac{a(v,v;\mu)}{||v||_Y^2} \le \gamma_a^0 < \infty, \quad \forall\, \mu \in \mathcal{D}; \tag{3.5}$$

coercive,

$$\alpha_a(\mu) \equiv \inf_{v \in Y} \frac{a(v,v;\mu)}{||v||_Y^2} \ge \alpha_a^0 > 0 \quad \forall\, \mu \in \mathcal{D}; \tag{3.6}$$

and symmetric, $a(w,v;\mu) = a(v,w;\mu)$, $\forall w, v \in Y$, $\forall\, \mu \in \mathcal{D}$; and that the linear functionals f and ℓ are bounded. Note that the coercivity constant $\alpha_a(\mu)$ and continuity constant $\gamma_a(\mu)$ are functions of μ.

3.2.4 Affine Parameter Dependence

We shall now make certain assumptions on the parametric dependence of our problem. In particular, we shall suppose that, for some finite (preferably small) integer Q_a, $a(\cdot,\cdot;\mu)$ may be expressed as

$$a(w,v;\mu) = \sum_{q=1}^{Q_a} \Theta_a^q(\mu)\; a^q(w,v), \quad \forall\, w, v \in Y,\ \forall\, \mu \in \mathcal{D}, \tag{3.7}$$

where the functions $\Theta_a^q(\mu)\colon \mathcal{D} \to \mathbb{R}$ depend on μ, but the bilinear forms $a^q(\cdot,\cdot)\colon Y \times Y \to \mathbb{R}$ are *independent* of μ. This assumption of affine parameter dependence is crucial for the computational efficiency of our method. Finally, for simplicity of exposition, we assume that the linear forms f and ℓ do not depend on the parameter; however, (affine) parameter dependence is readily admitted (see for example [120]).[2]

3.3 Reduced-Basis Approximation

3.3.1 Critical Observation: Dimension Reduction

The reduced-basis method recognizes that the field variable $y(\mu)$ is not an arbitrary member of the $\mathcal{N}$-dimensional solution space Y ($\mathcal{N} \gg 1$) associated with the partial differential equation; rather, it resides, or "evolves," on a much lower-dimensional and typically very smooth manifold $\mathcal{M} \equiv \{y(\mu) | \mu \in \mathcal{D}\}$ induced by the parametric dependence [91]. In the case of a single parameter ($P = 1$), for instance, $y(\mu)$ describes a one-dimensional filament that winds through Y; this is illustrated in Figure 3-1(a).

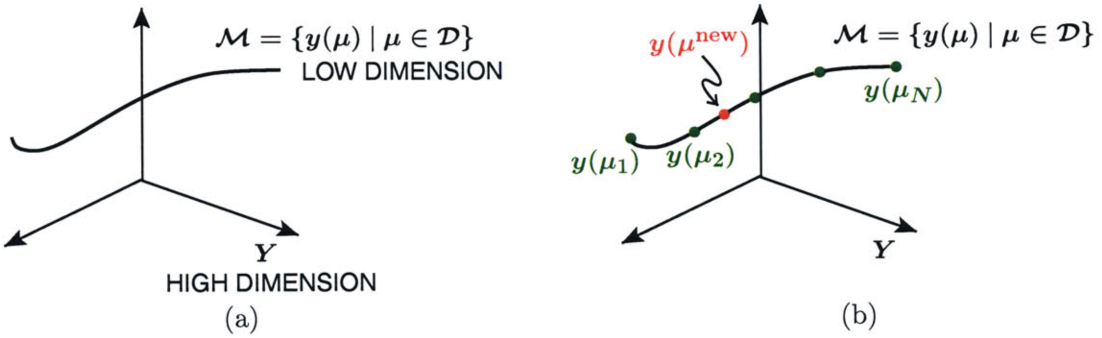

$\mathcal{M} = \{y(\mu) \mid \mu \in \mathcal{D}\}$
LOW DIMENSION
Y
HIGH DIMENSION
(a)
$y(\mu^{\text{new}})$
$\mathcal{M} = \{y(\mu) \mid \mu \in \mathcal{D}\}$
$y(\mu_N)$
$y(\mu_1)$
$y(\mu_2)$
Y
(b)

Figure 3-1: (a) Low-dimensional solution manifold $\mathcal{M}$ induced by the parametric dependence; and (b) dimension reduction obtained by restricting attention to $\mathcal{M}$.

The finite element approximation space Y is thus much too general — Y includes many functions that do not reside on the manifold of interest. Hence, to approximate $y(\mu)$, we need not represent every single function in Y, but rather only those which lie on $\mathcal{M}$. This observation presents a clear opportunity: we can effect significant (in many cases, Draconian) dimension reduction and therefore considerable computational economies if we restrict attention to the parameter-induced low-dimensional solution manifold. We may therefore pre-compute N "points" $y(\mu_n)$, $n = 1, \ldots, N$ along $\mathcal{M}$ as shown in Figure 3-1(b), and approximate $y(\mu^{\text{new}})$ by taking an appropriate linear combination of the sample points $y(\mu_n)$. We now make these ideas more precise.

3.3.2 Formulation

We first introduce a set of nested samples in parameter space,

$$S_N = \{\mu_1 \in \mathcal{D}, \ldots, \mu_N \in \mathcal{D}\}, \quad 1 \leq N \leq N_{\max}, \tag{3.8}$$

such that $S_1 \subset S_N \subset S_{N_{\max}}$. We then define the associated Lagrangian [85] reduced-basis approximation space as

$$W_N = \text{span}\{\zeta_n \equiv y(\mu_n), 1 \leq n \leq N\}, \quad 1 \leq N \leq N_{\max}, \tag{3.9}$$

[2] Note that the assumption of affine parameter dependence can be relaxed; see [15, 121] for extensions to problems exhibiting non-affine parameter dependence or nonlinearities.

where $y(\mu_n) \in Y$ is the solution to (3.3) for $\mu = \mu_n$. Note that, by construction, $W_1 \subset W_N \subset W_{N_{\max}}$.

Our reduced-basis approximation $y_N(\mu)$ is then obtained by a standard Galerkin projection: for any $\mu \in \mathcal{D}$, $y_N(\mu) \in W_N$ satisfies

$$a(y_N(\mu), v; \mu) = f(v), \quad \forall\, v \in W_N; \tag{3.10}$$

our output approximation is then

$$s_N(\mu) = \ell(y_N(\mu)). \tag{3.11}$$

Since $\mathcal{M}$ is low-dimensional and smooth, we thus anticipate that $y_N(\mu) \to y(\mu)$ very rapidly, and therefore we may choose $N \ll \mathcal{N}$. We now attempt to qualify our claim.

3.3.3 *A Priori* Convergence Theory

We consider here the rate at which $y_N(\mu)$ and $s_N(\mu)$ converge to $y(\mu)$ and $s(\mu)$, respectively.

Optimality

To begin, it is standard to demonstrate the optimality of $y_N(\mu)$ in the sense that

$$||y(\mu) - y_N(\mu)||_Y \leq \sqrt{\frac{\gamma_a(\mu)}{\alpha_a(\mu)}} \inf_{w_N \in W_N} ||y(\mu) - w_N||_Y \;. \tag{3.12}$$

To prove (3.12), we first note from (3.3) and (3.10) that

$$a(y(\mu) - y_N(\mu)), v; \mu) = 0, \quad \forall\, v \in W_N. \tag{3.13}$$

It then follows that for any $w_N = u_N + v_N \in W_N$ $(v_N \neq 0)$,

$$\begin{aligned} a(y(\mu) - w_N, y(\mu) - w_N; \mu) &= a(y(\mu) - y_N(\mu) - v_N, y(\mu) - y_N(\mu) - v_N; \mu) \\ &= a(y(\mu) - y_N(\mu), y(\mu) - y_N(\mu); \mu) - 2a(y(\mu) - y_N(\mu), v_N; \mu) + a(v_N, v_N; \mu) \\ &= a(y(\mu) - y_N(\mu), y(\mu) - y_N(\mu); \mu) + a(v_N, v_N; \mu) \\ &> a(y(\mu) - y_N(\mu), y(\mu) - y_N(\mu); \mu) \end{aligned} \tag{3.14}$$

from the symmetry of a, Galerkin orthogonality (3.13), and coercivity (3.6). Furthermore, from (3.6), (3.14), and (3.7), we have

$$\begin{aligned} \alpha_a(\mu)||y(\mu) - y_N(\mu)||_Y^2 &\leq a(y(\mu) - y_N(\mu), y(\mu) - y_N(\mu); \mu) \\ &= \inf_{w_N \in W_N} a(y(\mu) - w_N, y(\mu) - w_N; \mu) \\ &\leq \gamma_a(\mu) \inf_{w_N \in W_N} ||y(\mu) - w_N||_Y^2. \end{aligned} \tag{3.15}$$

This concludes the proof.

Furthermore, for the case of compliance, $\ell = f$, we have

$$\begin{aligned} |s(\mu) - s_N(\mu)| &= |\ell(y(\mu) - y_N(\mu))| \\ &= |a(y(\mu), y(\mu) - y_N(\mu); \mu)| \\ &= |a(y(\mu) - y_N(\mu), y(\mu) - y_N(\mu); \mu)| \\ &\leq \gamma_a(\mu) ||y(\mu) - y_N(\mu)||_Y^2 \\ &\leq \frac{\gamma_a^2(\mu)}{\alpha_a(\mu)} \inf_{w_N \in W_N} ||y(\mu) - w_N||_Y^2 \end{aligned} \tag{3.16}$$

from (3.4), (3.11), (3.3), the symmetry of a, Galerkin orthogonality (3.13), (3.7), and (3.15). The output approximation, $s_N(\mu)$, thus converges to $s(\mu)$ as the square of the error in $y_N(\mu)$.

We observed in Section 3.3.1 that the parametrically induced manifold $\mathcal{M}$ is typically low-dimensional, and note that $\mathcal{M}$ is smooth under our hypotheses on stability and continuity — we consider the detailed proof for the parabolic case in Section 4.3.2. We thus expect that the best approximation will converge to $y(\mu)$ very rapidly, and hence N may be chosen small.

3.3.4 Offline-Online Computational Procedure

The arguments of Sections 3.3.1 and 3.3.3 suggest that to obtain an accurate reduced-basis approximation $y_N(\mu)$, N need not be very large. We now develop off-line/on-line computational procedures that exploit this dimension reduction and enable us to evaluate our approximations in real-time.

We first note that since $y_N(\mu) \in W_N$, there exists a unique set of coefficients $y_{N\,j}(\mu)$, $1 \leq j \leq N$, such that

$$y_N(\mu) = \sum_{j=1}^{N} y_{N\,j}(\mu)\, \zeta_j. \tag{3.17}$$

We then choose as test functions in (3.10) $v = \zeta_i$, $i = 1, \ldots, N$; it then follows from (3.10) that $\underline{y}_N(\mu) \equiv [y_{N\,1}(\mu), \ldots, y_{N\,N}(\mu)]^T \in \mathbb{R}^N$ satisfies

$$A_N(\mu) \underline{y}_N(\mu) = F_N \tag{3.18}$$

where $A_N(\mu) \in \mathbb{R}^{N \times N}$ and $F_N \in \mathbb{R}^N$ are given by $A_{N\,i,j}(\mu) = a(\zeta_j, \zeta_i; \mu)$, $1 \leq i, j \leq N$, and $F_{N\,i} = f(\zeta_i)$, $1 \leq i \leq N$, respectively; note that $A_N(\mu)$ is symmetric positive-definite.

Invoking the affine decomposition (3.7) we obtain

$$A_{N\,i,j}(\mu) = \sum_{q=1}^{Q_a} \Theta_a^q(\mu)\; a^q(\zeta_i^{\rm pr}, \zeta_j^{\rm pr}); \tag{3.19}$$

we may therefore write

$$A_N(\mu) = \sum_{q=1}^{Q_a} \Theta_a^q(\mu)\; A_N^q, \tag{3.20}$$

where the *parameter independent* quantities matrices $A_N^q \in \mathbb{R}^{N \times N}$, are given by

$$A_{N\,i,j}^q = a^q(\zeta_i, \zeta_j), \quad 1 \leq i, j \leq N,\ 1 \leq q \leq Q_a. \tag{3.21}$$

Our output approximation can then be evaluated as

$$\begin{aligned} s_N(\mu) &= \sum_{j=1}^{N} y_{N\,j}(\mu)\, \ell(\zeta_j)\ , \\ &= \sum_{j=1}^{N} y_{N\,j}(\mu)\, L_{N\,j}\ , \end{aligned} \tag{3.22}$$

where

$$L_{N\,j} \equiv \ell(\zeta_j), \quad 1 \le j \le N. \tag{3.23}$$

We now observe that the A_N^q and L_N are *independent* of the parameter μ; we may thus pursue an offline-online computational strategy.

In the *offline* stage — performed *once* — we compute the ζ_i, $1 \le i \le N_{\max}$: this requires $N_{\max}$ expensive finite element solves; we then form *and store* $L_{N\,j}$, $1 \le j \le N_{\max}$ and $A^q_{N\,i,j}$, $1 \le i,j \le N_{\max}$, $1 \le q \le Q$: this requires $O(QN^2_{\max}\mathcal{N})$ operations and $O(QN^2_{\max})$ storage.

In the *on-line* stage — performed *many times*, for each new value of μ, we perform the summation (3.20) for $A_N(\mu)$: this requires $O(QN^2)$ operations; we then solve (3.18) for the reduced-basis coefficients $\underline{y}_{N\,j}(\mu)$, $j \in \mathbb{N}$: this requires $O(N^3)$ operations; and finally we evaluate the output approximation from (3.22): this requires $O(N)$ operations.

The essential point is that, as required in the many-query or real-time contexts, the online complexity depends only on Q and N and is *independent* of $\mathcal{N}$, the dimension of the underlying "truth" finite element approximation space. Since $N \ll \mathcal{N}$, we expect — and often realize — significant (orders-of-magnitude) computational economies relevant to classical discretization approaches.

Finally, we note that classical model-order reduction techniques, such as modal decomposition [36] and POD [7], require the evaluation of a new set of eigenmodes or basis functions — and thus a return to the (very fine) "truth" approximation — for each new parameter value encountered. In contrast, reduced-basis methods need not invoke the "truth" approximation in the online stage, and are therefore far more efficient in evaluating input-output relationships in the limit of many queries.

3.4 *A Posteriori* Error Estimation

A posteriori error estimation procedures are very well developed for classical approximations of, and solution procedures for, (say) partial differential equations [20, 81, 1] and algebraic systems [33]. However, until quite recently, there has been essentially no way to rigorously, quantitatively, sharply, and efficiently assess the accuracy of reduced-basis approximations.

As a result, for any given new μ, the reduced-basis solution $y_N(\mu)$ typically *raises* many more questions than it *answers*. Is there even a solution $y(\mu)$ near $y_N(\mu)$? Is $|s(\mu) - s_N(\mu)| \le \epsilon_{\text{tol}}$, where ϵ_{tol} is the maximum acceptable error? Is a crucial feasibility condition $s(\mu) \le S$ (say, in a constrained optimization exercise) satisfied — not just for the reduced-basis approximation, $s_N(\mu)$, but also for the "true" output, $s(\mu)$? If these questions can not be affirmatively answered, we may propose the wrong — and unsafe or infeasible — action in the deployed context. A fourth question is also important: Is N too large, $|s(\mu) - s_N(\mu)| \ll \epsilon_{\text{tol}}$, with an associated steep (N^3) penalty on computational efficiency? An overly conservative approximation may jeopardize the real-time response and associated action — with corresponding detriment to the deployed systems.

We may also consider the approximation properties and efficiency of the parameter samples and associated reduced-basis approximation spaces, S_N and W_N, $1 \leq N \leq N_{\max}$. Do we satisfy our global "acceptable error level" condition, $|s(\mu) - s_N(\mu)| \leq \epsilon_{\text{tol}}$, $\forall\, \mu \in \mathcal{D}$, for (close to) the smallest possible value of N? And a related question: For our given tolerance ϵ_{tol}, are the reduced-basis stiffness matrices (or, in the nonlinear case, Newton Jacobians) as well-conditioned as possible — given that *by construction* W_N will be increasingly colinear with increasing N? If the answers are not affirmative, then our reduced-basis approximations are more expensive (and unstable) than necessary — and perhaps too expensive to provide real-time response.

In short, the pre-asymptotic and essentially *ad hoc* or empirical nature of reduced-basis discretizations, the strongly superlinear scaling (with N) of the reduced-basis online complexity, and the particular needs of deployed real-time systems virtually demand rigorous *a posteriori* error estimators. Absent such certification, we must either err on the side of computational pessimism — and compromise real-time response — or err on the side of computational optimism — and risk sub-optimal, infeasible, or potentially unsafe decisions.

In [60, 91], and [121, 122, 123], a family of rigorous error estimators for reduced-basis approximation of a wide class of elliptic partial differential equations is introduced (see also [68] for an alternative approach). As in almost all error estimation contexts, the enabling (trivial) observation is that, whereas a 100% error in the *field variable* $y(\mu)$ or output $s(\mu)$ is clearly unacceptable, a 100% or even larger (conservative) error in the *error* is tolerable and not at all useless; we may thus pursue "relaxations" of the equation governing the error and residual that would be bootless for the original equation governing the field variable $y(\mu)$.

We now present further details for the particular case of elliptic linear problems with exact affine parameter dependence (3.7): the truth solution satisfies (3.3) and (3.4), and the corresponding reduced-basis approximation satisfies (3.10) and (3.11).

3.4.1 Formulation

Error Bounds

To begin, we assume we are given a positive lower bound $\tilde{\alpha}_a(\mu)$ for the coercivity constant $\alpha_a(\mu)$:

$$\alpha_a(\mu) \geq \hat{\alpha}_a(\mu) > 0, \quad \forall\, \mu \in \mathcal{D}. \tag{3.24}$$

We next introduce the dual norm of the residual

$$\varepsilon_N(\mu) \equiv \|R(\cdot;\mu)\|_{Y'} = \sup_{v \in Y} \frac{R(v;\mu)}{\|v\|_Y}, \tag{3.25}$$

where

$$R(v;\mu) \equiv f(v) - a(y_N(\mu), v; \mu), \quad \forall\, v \in Y, \tag{3.26}$$

is the residual associated to $y_N(\mu)$. We also specify the inner product $(w,v)_Y \equiv a(w,v;\mu_{\text{ref(s)}})$ for some reference value(s) $\mu_{\text{ref(s)}}$. We then define our "energy" error bound

$$\Delta_N(\mu) \equiv \frac{\varepsilon_N(\mu)}{\hat{\alpha}_a(\mu)}\ , \tag{3.27}$$

the effectivity of which is given by

$$\eta_N(\mu) \equiv \frac{\Delta_N(\mu)}{\|e(\mu)\|_Y}. \tag{3.28}$$

We may then state

Proposition 3. *For the error bound $\Delta_N(\mu)$ of (3.27), the effectivity satisfies [91, 123]*

$$1 \leq \eta_N(\mu) \leq \frac{\gamma_a(\mu)}{\hat{\alpha}_a(\mu)}, \quad \forall\, \mu \in \mathcal{D} \ . \tag{3.29}$$

for all $1 \leq N \leq N_{\max}$.

Proof. Given our reduced-basis primal solution $y_N(\mu)$, it is readily derived that the error $e(\mu) \equiv y(\mu) - y_N(\mu) \in Y$ satisfies

$$a(e(\mu), v; \mu) = R(v; \mu), \quad \forall\, v \in Y. \tag{3.30}$$

Furthermore, we note from standard duality arguments that

$$\varepsilon_N(\mu) = \|\hat{e}(\mu)\|_Y, \tag{3.31}$$

where $\hat{e}(\mu)$ is given by

$$(\hat{e}(\mu), v)_Y = R(v; \mu), \quad \forall\, v \in Y. \tag{3.32}$$

We then note that

$$\begin{aligned} \alpha_a(\mu)\|e(\mu)\|_Y^2 &\leq a(e(\mu), e(\mu); \mu) \\ &= (\hat{e}(\mu), e(\mu))_Y \\ &\leq \|\hat{e}(\mu)\|_Y \|e(\mu)\|_Y \end{aligned} \tag{3.33}$$

from (3.6), (3.32), and the Cauchy-Schwarz inequality; and

$$\begin{aligned} \|\hat{e}(\mu)\|_Y^2 &= a(e(\mu), \hat{e}(\mu); \mu) \\ &\leq \gamma_a(\mu)\|\hat{e}(\mu)\|_Y \|e(\mu)\|_Y \end{aligned} \tag{3.34}$$

from (3.30), (3.32) and (3.5). The desired result directly follows from (3.33), (3.34), (3.28), (3.31), and (3.24). □

From the left inequality of (3.29), we deduce that $\|e(\mu)\|_Y \leq \Delta_N(\mu)$, $\forall\, \mu \in \mathcal{D}$, and hence that $\Delta_N(\mu)$ is a rigorous upper bound for the true error[3] measured in the $\|\cdot\|_Y$ norm — this provides certification: feasibility and "safety" are guaranteed. From the right inequality, we deduce that $\Delta_N(\mu)$ overestimates the true error by at most $\gamma_a(\mu)/\hat{\alpha}_a(\mu)$,[4] *independent of* N — this relates to efficiency: an overly conservative error bound will be manifested in an unnecessarily large N and unduly expensive RB approximation, or (even worse) an overly conservative or expensive decision or action "in the field."

[3] Note however that these error bounds are relative to our underlying "truth" approximation, $y(\mu) \in Y$, not to the exact solution, $y^{\mathrm{e}}(\mu) \in Y^{\mathrm{e}}$.

[4] The upper bound on the effectivity can be large. In many cases, this effectivity bound is in fact quite pessimistic; in many other cases, the effectivity (bound) may be improved by judicious choice of (multi-point) inner product $(\cdot, \cdot)_Y$ — in effect, a "bound conditioner" [124].

We can now define a (simple) error bound for the output of interest in

Proposition 4. *For $\Delta_N(\mu)$ of (3.27),*

$$|s(\mu) - s_N(\mu)| \leq \Delta_N^s(\mu), \qquad \forall\, \mu \in \mathcal{D}\ , \tag{3.35}$$

where

$$\Delta_N^s(\mu) \equiv \|\ell\|_{Y'} \Delta_N(\mu)\ . \tag{3.36}$$

Proof. We note that

$$\begin{aligned} |s(\mu) - s_N(\mu)| &= |\ell(e(\mu))| \\ &= \frac{|\ell(e(\mu))|}{\|e(\mu)\|_Y}\|e(\mu)\|_Y \\ &\leq \sup_{v\in Y}\frac{\ell(v)}{\|v\|_Y}\|e(\mu)\|_Y \end{aligned} \tag{3.37}$$

The result directly follows from the (3.4), (3.11), (2.1) and Proposition 3. □

We note that this output bound — although very easy and efficient to evaluate — might not provide very sharp bounds and is thus not very useful in actual practice; this will also become evident from the numerical results. However, we can obtain a more rapid convergence of the reduced-basis output approximation as well as a sharper bound by introducing a dual (or adjoint) problem. We will discuss the adjoint formulation in detail in Chapter 4 in the parabolic case.

Offline-Online Procedure

The real challenge in *a posteriori* error estimation is not the presentation of these rather classical results, but rather the development of efficient computational approaches for the evaluation of the necessary constituents. In our particular deployed context, "efficient" translates to "online complexity *independent* of $\mathcal{N}$," and "necessary constituents" translates to "dual norm of the primal residual, $\varepsilon_N(\mu) \equiv \|R(\cdot;\mu)\|_{Y'}$, and lower bound for the coercivity constant, $\hat{\alpha}_a(\mu)$." We now turn to these issues.

The Dual Norm of the Residual

To begin, we note from duality (3.31), (3.32), our reduced-basis expansion (3.17), and our assumption of affine parameter dependence (3.7), that $\hat{e}(\mu)$ satisfies

$$(\hat{e}(\mu), v)_Y = f(v) - \sum_{q=1}^{Q}\sum_{n=1}^{N}\Theta^q(\mu)\, y_{N\,n}(\mu)\, a^q(\zeta_n, v), \quad \forall\, v \in Y. \tag{3.38}$$

It then follows from linear superposition that we may write $\hat{e}(\mu) \in Y$ as

$$\hat{e}(\mu) = \mathcal{F} + \sum_{q=1}^{Q}\sum_{n=1}^{N}\Theta^q(\mu)\, y_{N\,n}(\mu)\, \mathcal{A}_n^q\ ,$$

where $\mathcal{F} \in Y$ and $\mathcal{A}_n^q \in Y$ satisfy

$$\begin{aligned}(\mathcal{F}, v)_Y &= f(v), \quad \forall\, v \in Y, & (3.39)\\ (\mathcal{A}_n^q, v)_Y &= -a^q(\zeta_n, v), \quad \forall\, v \in Y,\ n \in \mathbb{N},\ 1 \le q \le Q, & (3.40)\end{aligned}$$

respectively; note that (3.39),(3.40) are simple *parameter-independent* (scalar or vector) Poisson, or Poisson-like, problems. It thus follows that

$$\begin{aligned}\|\hat{e}(\mu)\|_Y^2 = (\mathcal{F}, \mathcal{F})_Y + \sum_{q=1}^{Q} \sum_{n=1}^{N} \Theta^q(\mu)\, y_{N\,n}(\mu) \Big\{ 2(\mathcal{F}, \mathcal{A}_n^q)_Y \\ + \sum_{q'=1}^{Q} \sum_{n'=1}^{N} \Theta^{q'}(\mu)\, y_{N\,n'}(\mu)\, (\mathcal{A}_n^q, \mathcal{A}_{n'}^{q'})_Y \Big\}\,. \qquad (3.41)\end{aligned}$$

The critical observation [60, 91] is that the expression (3.41) — which we relate to the requisite dual norm of the residual through (3.31) — is the sum of products of parameter-dependent (simple, known) functions and parameter-independent inner products. The offline-online decomposition is now clear.

In the offline stage — performed once — we first solve (3.39), (3.40) for $\mathcal{F}$ and $\mathcal{A}_n^q$, $1 \le n \le N_{\max}$, $1 \le q \le Q$; we then evaluate and save the relevant parameter-independent inner products $(\mathcal{F}, \mathcal{F})_Y$, $(\mathcal{F}, \mathcal{A}_n^q)_Y$, $(\mathcal{A}_n^q, \mathcal{A}_{n'}^{q'})_Y$, $1 \le n, n' \le N_{\max}$, $1 \le q, q' \le Q$. Note that all quantities computed in the offline stage are independent of the parameter μ.

In the online stage — performed many times, for each new value of μ "in the field" — we simply evaluate the sum (3.41) in terms of the $\Theta^q(\mu)$, $y_{N\,n}(\mu)$ and the precalculated and stored (parameter-independent) $(\cdot,\cdot)_Y$ inner products. The operation count for the online stage is only $O(Q^2N^2)$ — again, the essential point is that the online complexity is *independent of* $\mathcal{N}$, the dimension of the underlying truth finite element approximation space. We further note that, unless Q is quite large, the online cost associated with the calculation of the dual norm of the residual is commensurate with the online cost associated with the calculation of $s_N(\mu)$.

Lower Bound for the Coercivity Parameter

Obviously, from the definition (3.6), we may readily obtain by a variety of techniques effective *upper bounds* for $\alpha_a(\mu)$; however, lower bounds are much more difficult to construct. We do note that in the case of *symmetric coercive* operators we *can* often determine $\hat{\alpha}_a(\mu)$ "by inspection." In particular, we define our "minimum coefficient" coercivity lower bound [60, 124] in

Lemma 6. *Assume that $a(w, v; \mu)$ is given by (3.7) where $\Theta^q(\mu) > 0$, $\forall\, \mu \in \mathcal{D}$, and $a^q(v, v) \ge 0$, $\forall\, v \in Y$, $1 \le q \le Q$. Then, given $\overline{\mu} \in \mathcal{D}$,*

$$\hat{\alpha}_a(\mu) \equiv \left(\min_{q \in \{1,\ldots,Q\}} \frac{\Theta^q(\mu)}{\Theta^q(\overline{\mu})} \right) \alpha_a(\overline{\mu}) \le \alpha_a(\mu), \quad \forall\, \mu \in \mathcal{D}. \qquad (3.42)$$

Proof. We note that for any $\mu \in \mathcal{D}$,

$$\begin{aligned}
\alpha_a(\mu) &\equiv \inf_{v\in Y} \frac{a(v,v;\mu)}{\|v\|_Y} \\
&= \inf_{v\in Y} \frac{\sum_{q=1}^{Q_a} \Theta^q(\mu) a^q(v,v)}{\|v\|_Y} \\
&= \inf_{v\in Y} \frac{\sum_{q=1}^{Q_a} \left(\frac{\Theta^q(\mu)}{\Theta^q(\overline{\mu})}\right) \Theta^q(\overline{\mu}) a^q(v,v)}{\|v\|_Y} \\
&\geq \left(\min_{q\in\{1,\ldots,Q\}} \frac{\Theta^q(\mu)}{\Theta^q(\overline{\mu})}\right) \inf_{v\in Y} \frac{\sum_{q=1}^{Q_a} \Theta^q(\overline{\mu}) a^q(v,v)}{\|v\|_Y} \\
&= \left(\min_{q\in\{1,\ldots,Q\}} \frac{\Theta^q(\mu)}{\Theta^q(\overline{\mu})}\right) \inf_{v\in Y} \frac{a(v,v;\overline{\mu})}{\|v\|_Y} \\
&= \left(\min_{q\in\{1,\ldots,Q\}} \frac{\Theta^q(\mu)}{\Theta^q(\overline{\mu})}\right) \alpha_a(\overline{\mu}) \\
&\equiv \hat{\alpha}_a(\mu) \ . \qquad (3.43)
\end{aligned}$$

□

Finally, we note that the choice of the Y-norm and associated bound conditioner does affect the quality of the error bound. The effectivity at a parameter value μ close to $\overline{\mu}$ in (3.42) will, in general, be smaller than for a value μ farther away from $\overline{\mu}$ — in fact, for $\mu = \overline{\mu}$ the effectivity is one. If the parameter domain $\mathcal{D}$ is large, multi-point bound conditioners can be introduced to exploit this fact [124].

3.5 Construction of Samples: A "Greedy" Algorithm

Our error estimation procedures also allow us to pursue more rational constructions of our parameter samples S_N (and hence spaces W_N) [123]. We denote the smallest error tolerance anticipated as $\epsilon_{\text{tol, min}}$ — this must be determined *a priori* offline; we then permit $\epsilon_{\text{tol}} \in [\epsilon_{\text{tol, min}}, \infty[$ to be specified online. We also introduce $\Xi_{\text{F}} \in \mathcal{D}^{n_{\text{F}}}$, a very fine random sample over the parameter domain $\mathcal{D}$ of size $n_{\text{F}} \gg 1$.

We first consider the offline stage. We assume that we are given a sample S_N, and hence space W_N and associated reduced-basis approximation (procedure to determine) $y_N(\mu)$, $\forall\, \mu \in \mathcal{D}$. We then calculate

$$\mu_N^* = \arg \max_{\mu \in \Xi_{\text{F}}} \Delta_N(\mu);$$

here $\Delta_N(\mu)$ is our "online" error bound (3.27) that, in the limit of $n_{\text{F}} \to \infty$ queries, may be evaluated (on average) in $O(N^2Q^2)$ operations; we next append μ_N^* to S_N to form S_{N+1}, and hence W_{N+1}. We now continue this process until $N = N_{\max}$ such that $\epsilon^*_{N_{\max}} = \epsilon_{\text{tol,min}}$, where $\epsilon_N^* \equiv \Delta_N(\mu_N^*)$, $1 \leq N \leq N_{\max}$.

In the online stage, given any desired $\epsilon_{\text{tol}} \in [\epsilon_{\text{tol, min}}, \infty[$ and any new value of $\mu \in \mathcal{D}$ "in the field," we first choose N from a pre-tabulated array such that $\epsilon_N^* \, (\equiv \Delta_N(\mu_N^*)) = \epsilon_{\text{tol}}$. We next calculate $y_N(\mu)$ and $\Delta_N(\mu)$, and then verify that — and if necessary, subsequently increase N *such* *that* — the condition $\Delta_N(\mu) \leq \epsilon_{\text{tol}}$ is indeed satisfied. (We should not and do not rely on the finite sample Ξ_{F} for either rigor or sharpness.)

The crucial point is that $\Delta_N(\mu)$ is an accurate and "online-inexpensive" — $O(1)$ effectivity and $\mathcal{N}$-independent asymptotic complexity — surrogate for the true (very-expensive-to-calculate) error $\|y(\mu) - y_N(\mu)\|_Y$. This surrogate permits us to (i) offline — here we exploit low average cost — perform a much more exhaustive ($n_{\text{F}} \gg 1$) and hence meaningful search for the best samples S_N and hence most rapidly *uniformly* convergent spaces W_N,[5] and (ii) online — here we exploit low marginal cost — determine the smallest N, and hence the most efficient approximation, for which we *rigorously* achieve the desired accuracy.

[5] We may in fact view our offline sampling process as a (greedy, parameter space, "$L^\infty(\mathcal{D})$") variant of the POD economization procedure [108] in which — thanks to $\Delta_N(\mu)$ — *we need never construct* the "rejected" snapshots.

Chapter 4

Linear Parabolic Equations

4.1 Introduction

In Chapter 3 we discussed the reduced-basis method and associated *a posteriori* error estimation for linear coercive elliptic problems with affine parameter dependence. In this chapter, we will extend these results to parabolic problems with affine parameter dependence [41]. The essential new ingredient is the presence of time in the formulation and solution of the problem — we shall "simply" treat time as an additional, albeit special, parameter. In the first part of this chapter we focus on symmetric problems: we introduce the reduced-basis method and associated *a posteriori* error estimation, and we propose adjoint procedures that provide rigorous and sharp bound for the error in specific outputs of interest. We then develop a new greedy adaptive procedure to "optimally" construct the parameter-time sample set. Finally, we extend our results to non-symmetric problems, such as the convection-diffusion equation. Based on the assumption of affine parameter dependence, we develop offline-online computation procedures by construction rather similar to the elliptic case. Problems with nonaffine parameter dependence are addressed in Chapter 5.

4.2 Abstract Formulation

We first recall the Hilbert spaces $Y^{\mathrm{e}} \equiv H_0^1(\Omega)$ — or, more generally, $H_0^1(\Omega) \subset Y^{\mathrm{e}} \subset H^1(\Omega)$ — and $X^{\mathrm{e}} \equiv L^2(\Omega)$, where $H^1(\Omega) \equiv \{v \mid v \in L^2(\Omega), \nabla v \in (L^2(\Omega))^d\}$, $H_0^1(\Omega) \equiv \{v \mid v \in H^1(\Omega), v|_{\partial\Omega} = 0\}$, and $L^2(\Omega)$ is the space of square integrable functions over Ω [94]; here Ω is a bounded domain in $\mathbb{R}^d$ with Lipschitz continuous boundary $\partial\Omega$. The inner product and norm associated with Y^{e} (X^{e}) are given by $(\cdot,\cdot)_{Y^{\mathrm{e}}}$ ($(\cdot,\cdot)_{X^{\mathrm{e}}}$) and $\|\cdot\|_{Y^{\mathrm{e}}} = (\cdot,\cdot)_{Y^{\mathrm{e}}}^{1/2}$ ($\|\cdot\|_{X^{\mathrm{e}}} = (\cdot,\cdot)_{X^{\mathrm{e}}}^{1/2}$) , respectively; for example, $(w,v)_{Y^{\mathrm{e}}} \equiv \int_\Omega \nabla w \cdot \nabla v + \int_\Omega w\, v,\ \forall\, w,v \in Y^{\mathrm{e}}$, and $(w,v)_{X^{\mathrm{e}}} \equiv \int_\Omega w\, v,\ \forall\, w,v \in X^{\mathrm{e}}$.

4.2.1 Primal Problem

We may now introduce the "exact" (superscript e) — more precisely, semi-discrete — problem: given a parameter $\mu \in \mathcal{D} \subset \mathbb{R}^P$, we evaluate the (here, single) output of interest

$$s^{\mathrm{e}}(\mu, t^k) = \ell(y^{\mathrm{e}}(\mu, t^k)), \quad \forall\, k \in \mathbb{K}, \tag{4.1}$$

where the field variable, $y^{\mathrm{e}}(\mu,t^k) \in Y^{\mathrm{e}}$, $\forall\, k \in \mathbb{K}$, satisfies the weak form of the μ-parametrized parabolic PDE [14]

$$m(y^{\mathrm{e}}(\mu,t^k),v;\mu) + \Delta t\ a(y^{\mathrm{e}}(\mu,t^k),v;\mu) = m(y^{\mathrm{e}}(\mu,t^{k-1}),v;\mu) + \Delta t\ b(v;\mu)\ u(t^k),$$
$$\forall\, v \in Y^{\mathrm{e}},\ \forall\, k \in \mathbb{K}, \quad (4.2)$$

with initial condition (say) $y^{\mathrm{e}}(\mu,t^0) = y_0(\mu) = 0$. Here μ and $\mathcal{D}$ are the input and input domain; $a(\cdot,\cdot;\mu)$ and $b(\cdot;\mu)$ are Y^{e}-continuous bilinear and linear forms, respectively; $m(\cdot,\cdot;\mu)$ and $\ell(\cdot)$ are X^{e}-continuous bilinear and linear forms, respectively; and $u(t^k)$ denotes the (here, single) control input at time $t = t^k$.

We next introduce a reference finite element approximation space $Y \subset Y^{\mathrm{e}}$ $(\subset X^{\mathrm{e}})$ of very large dimension $\mathcal{N}$; we further define $X \equiv X^{\mathrm{e}}$. Note that Y and X shall inherit the inner product and norm from Y^{e} and X^{e}, respectively. Our reference (or "truth") finite element approximation $y(\mu,t^k) \in Y$ to the semi-discrete problem (4.2) is then given by

$$m(y(\mu,t^k),v;\mu) + \Delta t\ a(y(\mu,t^k),v;\mu) = m(y(\mu,t^{k-1}),v;\mu) + \Delta t\ b(v;\mu)\ u(t^k),$$
$$\forall\, v \in Y,\ \forall\, k \in \mathbb{K}, \quad (4.3)$$

with initial condition $y(\mu,t^0) = 0$; we then evaluate the output $s(\mu,t^k) \in \boldsymbol{R}$ from

$$s(\mu,t^k) = \ell(y(\mu,t^k)), \quad \forall\, k \in \mathbb{K}. \quad (4.4)$$

We shall assume — hence the appellation "truth" — that the discretization is sufficiently rich such that $y(\mu,t^k)$ and $y^{\mathrm{e}}(\mu,t^k)$ and hence $s(\mu,t^k)$ and $s^{\mathrm{e}}(\mu,t^k)$ are indistinguishable. The reduced-basis approximation shall be built upon our reference finite element approximation, and the reduced-basis error will thus be evaluated with respect to $y(\mu,t^k) \in Y$. Clearly, our methods must remain computationally efficient and stable as $\mathcal{N} \to \infty$.

We shall make the following assumptions. First, we assume that the bilinear forms $a(\cdot,\cdot;\mu)$ and $m(\cdot,\cdot;\mu)$ are symmetric, $a(v,w;\mu) = a(w,v;\mu)$, $\forall\, w,v \in Y$, $\forall\, \mu \in \mathcal{D}$, and $m(v,w;\mu) = m(w,v;\mu)$, $\forall\, w,v \in X$, $\forall\, \mu \in \mathcal{D}$. For the sake of well-posedness, we assume that a and m are continuous,

$$\begin{aligned} a(w,v;\mu) &\le \gamma_a(\mu)\|w\|_Y\|v\|_Y \le \gamma_a^0\|w\|_Y\|v\|_Y, \quad \forall\, w,v \in Y,\ \forall\, \mu \in \mathcal{D}, && (4.5)\\ m(w,v;\mu) &\le \gamma_m(\mu)\|w\|_X\|v\|_X \le \gamma_m^0\|w\|_X\|v\|_X, \quad \forall\, w,v \in Y,\ \forall\, \mu \in \mathcal{D}; && (4.6) \end{aligned}$$

and coercive,

$$\begin{aligned} 0 &< \alpha_a^0 \le \alpha_a(\mu) \equiv \inf_{v \in Y} \frac{a(v,v;\mu)}{\|v\|_Y^2}, \quad \forall\, \mu \in \mathcal{D}, && (4.7)\\ 0 &< \alpha_m^0 \le \alpha_m(\mu) \equiv \inf_{v \in Y} \frac{m(v,v;\mu)}{\|v\|_X^2}, \quad \forall\, \mu \in \mathcal{D}. && (4.8) \end{aligned}$$

(We (plausibly) suppose that γ_a^0, γ_m^0, α_a^0, and α_m^0 may be chosen independent of $\mathcal{N}$.) We also require that the linear forms $b(\cdot;\mu) : Y \to \boldsymbol{R}$ and $\ell(\cdot) : Y \to \boldsymbol{R}$ be bounded with respect to $\|\cdot\|_Y$ and $\|\cdot\|_X$, respectively. It thus follows that a solution to (4.3) exists and is unique [94].

Second, we shall assume that a, m, and b depend affinely on the parameter μ and can be expressed as

$$
\begin{aligned}
a(w,v;\mu) &= \sum_{q=1}^{Q_a} \Theta_a^q(\mu)\; a^q(w,v), \quad \forall\, w,v \in Y,\ \forall\, \mu \in \mathcal{D}, && (4.9)\\
m(w,v;\mu) &= \sum_{q=1}^{Q_m} \Theta_m^q(\mu)\; m^q(w,v), \quad \forall\, w,v \in Y,\ \forall\, \mu \in \mathcal{D}, && (4.10)\\
b(v;\mu) &= \sum_{q=1}^{Q_b} \Theta_b^q(\mu)\; b^q(v), \quad \forall\, v \in Y,\ \forall\, \mu \in \mathcal{D}, && (4.11)
\end{aligned}
$$

for some (preferably) small integers $Q_{a,m,b}$. Here, the functions $\Theta^q_{a,m,b}(\mu) : \mathcal{D} \to \boldsymbol{R}$ depend on μ, but the continuous forms a^q, m^q, and b^q do *not* depend on μ. This affine parameter dependence is crucial for the computational efficiency of the proposed method; however, in Chapter 5 and Chapter 6 we extend the method to the nonaffine and nonlinear case, respectively (for a discussion of nonaffine and nonlinear elliptic problems, see [15, 121]). For simplicity of exposition, we assume that the linear form ℓ does not depend on the parameter; however, (affine) parameter dependence is readily admitted.

Third, and finally, we require that all linear and bilinear forms are independent of time — the system is thus linear time-invariant (LTI). This is true for many physical problems governed by parabolic PDEs, with the most notable exception of deforming domains. We point out that an important application which often satisfies all of our assumptions is the classical heat equation [94]; we shall provide a detailed example in Section 4.2.4

We note that the method presented here easily extends to nonzero initial conditions with affine parameter dependence and to multiple control inputs and outputs. We will comment on the extension to nonzero initial conditions in Section 4.5.2. The extension to nonsymmetric problems such as the convection-diffusion equation is discussed in Section 4.8.

4.2.2 Dual Problem

To ensure rapid convergence of the reduced-basis output approximation we introduce a dual (or adjoint) problem which shall evolve backward in time [20]. Invoking the LTI property we can express the adjoint for the output at time t^L, $1 \le L \le K$, as $\psi_L(\mu, t^k) = \Psi(\mu, t^{K-L+k})$, $1 \le k \le L$, where $\Psi(\mu, t^k) \in Y$ satisfies

$$
m(v, \Psi(\mu,t^k);\mu) + \Delta t\ a(v, \Psi(\mu,t^k);\mu) = m(v, \Psi(\mu, t^{k+1});\mu), \qquad \forall\, v \in Y,\ \forall\, k \in \mathbb{K}, \qquad (4.12)
$$

with final condition

$$
m(v, \Psi(\mu, t^{K+1});\mu) \equiv \ell(v), \qquad \forall\, v \in Y. \qquad (4.13)
$$

Thus, to obtain $\psi_L(\mu,t^k)$, $1 \le k \le L$, $\forall\, L \in \mathbb{K}$, we solve *once* for $\Psi(\mu,t^k)$, $\forall\, k \in \mathbb{K}$, and then appropriately shift the result — we do not need to solve K separate dual problems. A sketch of the shifting property is presented in Figure 4-1. The primal problem, $y(\mu,t^k)$, shown on top evolves forward in time from t^0 to t^K, whereas the dual problems, $\psi_L(\mu,t^k)$, $\forall\, L \in \mathbb{K}$, evolve backward in time from t^L to t^1. On the bottom of the sketch we show $\Psi(\mu,t^k)$ evolving backward from t^K to t^1. The blue and green brackets signify how $\Psi(\mu,t^k)$ is related to the $\psi_L(\mu,t^k)$, $\forall\, L \in \mathbb{K}$, through an appropriate shift in time.

We note that the issue of "rough" final conditions — output functionals — is implicitly addressed in our temporal discretization and truth approximation. However, we stress that the output functional, ℓ, has to be bounded in X, otherwise $\Psi(\mu, t^{K+1})$ in (4.13) is not bounded as $\mathcal{N} \to \infty$.

We also note that, given a *specific* input $u(t^k)$, $\forall\, k \in \mathbb{K}$, our results directly carry over to the linear time-varying (LTV) case; we can no longer, however, invoke the shift property of the dual problem — which renders the calculation of our output bound more cumbersome.

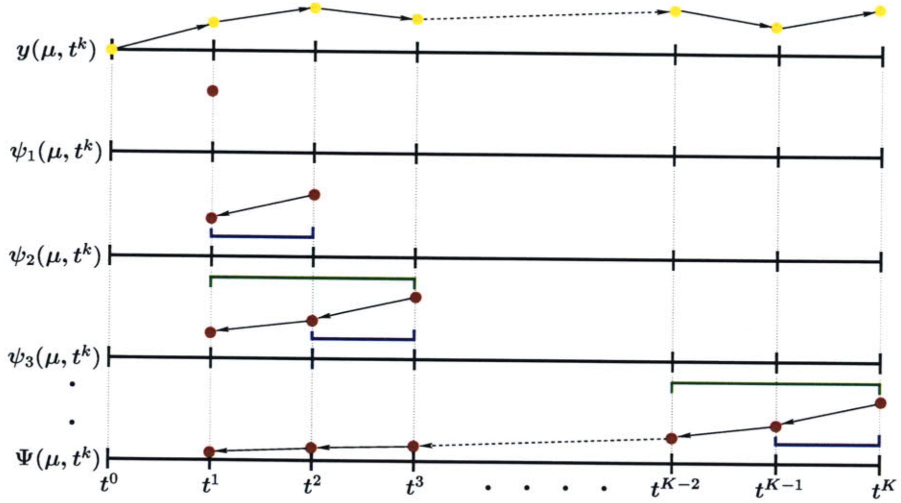


Figure 4-1: Shifting property of the dual problem.

4.2.3 Impulse Response

The reduced-basis subspace shall be developed as the span of solutions $y(\mu, t^k)$ of our "truth" approximation (4.3) at selected points in parameter-time space. In many cases, however, the input $u(t^k)$ will not be known in advance and thus we cannot solve for $y(\mu, t^k)$ — one such example is the optimal control problem described in the Introduction. In such situations, fortunately, we may appeal to the LTI hypothesis to justify an impulse approach, as we now describe.

We first note that the solution of any LTI system can be written as the convolution of the impulse response with the control input (Duhamel's Principle): for any control input $u(t^k)$, $\forall\, k \in \mathbb{K}$, we can obtain $y(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, from

$$y(\mu, t^k) = \sum_{j=1}^{k} g(\mu, t^{k-j+1})\, u(t^j), \quad \forall\, k \in \mathbb{K}, \tag{4.14}$$

where the impulse response, $g(\mu, t^k)$, is the solution of (4.3) for a unit impulse control input $u(t^k) = \delta_{1k}$, $\forall\, k \in \mathbb{K}$. Equation (4.14) simply states that $y(\mu, t^k)$ is a linear combination of the impulse response $g(\mu, t^j)$, $1 \leq j \leq k$; it is thus sufficient that the reduced-basis subspace approximates well the (parameter-dependent) impulse response. It still remains to select which basis functions to retain, i.e., to determine the "best" sampling points in parameter-time space for the basis; we will address this issue in Section 4.5

4.2.4 Numerical Exercise 2: Design of a Heat Shield

We now turn to a particular numerical example related to transient heat conduction. We consider the design of a heat shield, one segment of which is shown in Figure 4-2. The domain Ω, a typical point of which is (x_1, x_2), is thus given by $\Omega \equiv \{[0,10] \times [0,4]\} \backslash \{(]1,3[\times]1,3[) \cup (]4,6[\times]1,3[) \cup (]7,9[\times]1,3[)\}$. The left boundary, $\partial\Omega_{\text{out}}$ ($x_1 = 0$), is exposed to a hot temperature (here normalized to unity) for $t \in]0, t_f]$; the right boundary as well as the top and bottom boundaries are insulated. The internal boundaries $\partial\Omega_{\text{in}}$ — corresponding to the surfaces of the three square cooling channels $]1,3[\times]1,3[$, $]4,6[\times]1,3[$, and $]7,9[\times]1,3[$ — are exposed to a (normalized) zero-temperature air flow. The (non-dimensionalized) heat transfer coefficients for the non-insulated boundaries $\partial\Omega_{\text{out}}$ and $\partial\Omega_{\text{in}}$ are given by the Biot numbers Bi_{out} and Bi_{in}, respectively. Our input parameter is hence $\mu \equiv (\mu_{(1)}, \mu_{(2)}) \equiv (\text{Bi}_{\text{out}}, \text{Bi}_{\text{in}}) \in \mathcal{D} \equiv [0.01, 0.5] \times [0.001, 0.1] \subset \mathbb{R}^{P=2}$. Our output is the average temperature of the structure, which serves as a surrogate for the maximum possible temperature of the (to-be-protected) right boundary for $t \in [0, \infty[$.

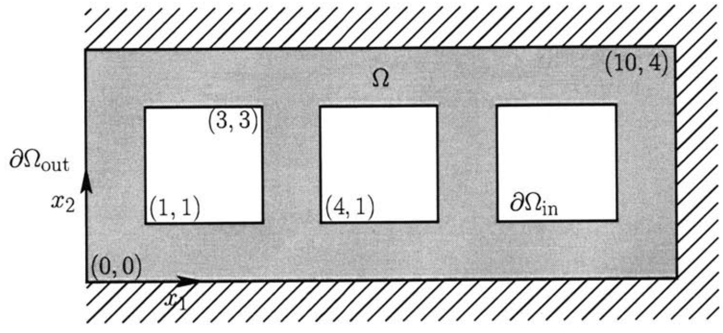


Figure 4-2: NE 2: One segment of the heat shield.

The underlying partial differential equation is the heat (diffusion) equation. The (appropriately non-dimensionalized) governing equation for the temperature $y(\mu, t^k) \in Y$ is thus (4.3), where $Y \subset Y^{\text{e}} \equiv H^1(\Omega)$ is a linear finite element truth approximation subspace of dimension (exploiting symmetry) $\mathcal{N} = 1396$ shown in Figure 4-3. The bilinear and linear forms are given by $m(w, v; \mu) \equiv \int_\Omega w\, v$, $a(w, v; \mu) \equiv \int_\Omega \nabla w\, \nabla v + \mu_{(1)} \int_{\partial\Omega_{\text{out}}} w\, v + \mu_{(2)} \int_{\partial\Omega_{\text{in}}} w\, v$, and $b(v; \mu) \equiv \mu_{(1)} \int_{\partial\Omega_{\text{out}}} v$; these forms admit obvious affine representations (4.9)-(4.11) with $Q_m = 1$, $Q_a = 3$, and $Q_b = 1$. The output can be written in the form (4.4), $s(\mu, t^k) = \ell(y(\mu, t^k))$, $\forall\, k \in \mathbb{K}$, where $\ell(v) = |\Omega|^{-1} \int_\Omega v$ is clearly a very smooth functional. We shall consider the time interval $\bar{I} = [0, 20]$ and a timestep $\Delta t = 0.2$; we thus have $K = 100$.

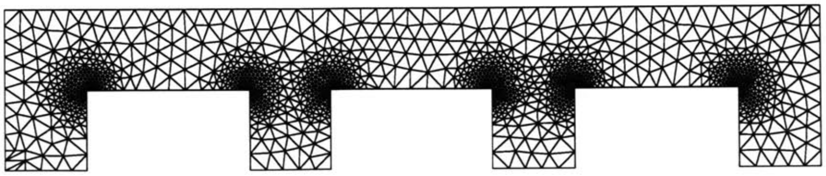
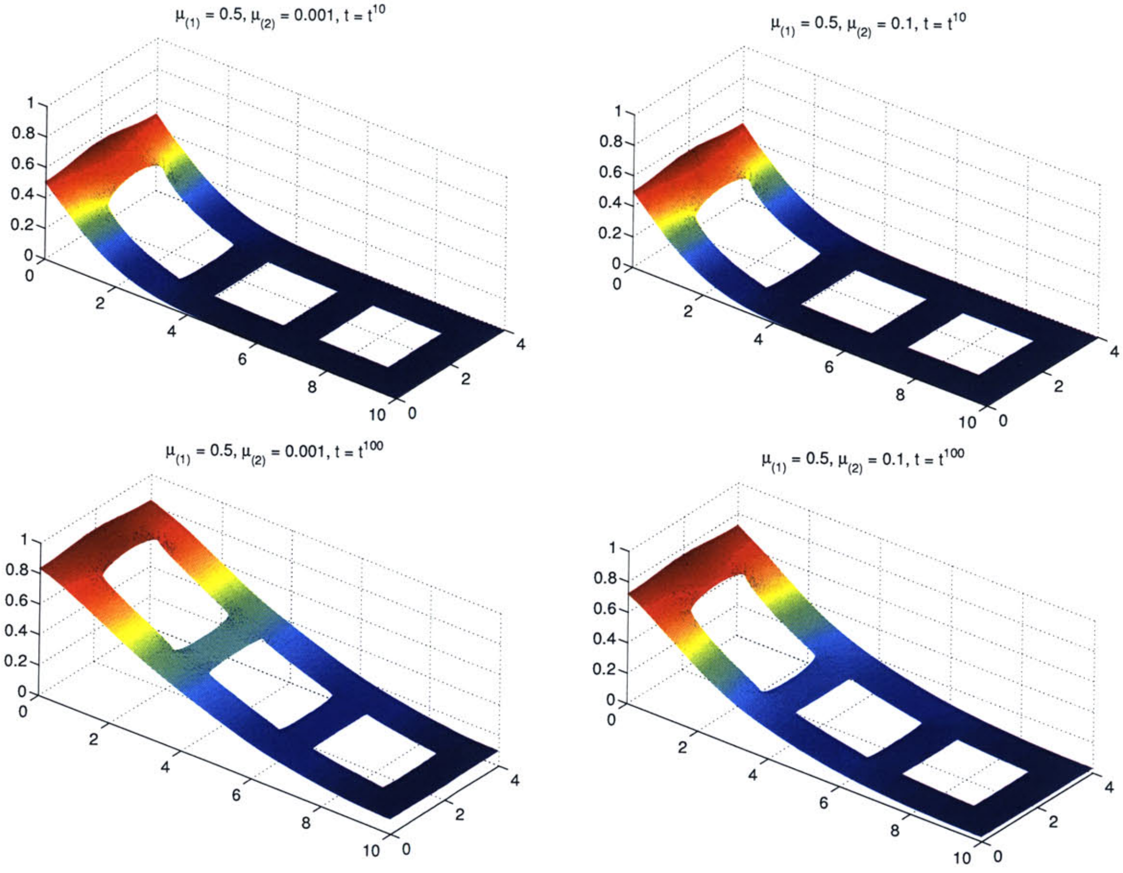


Figure 4-3: NE 2: Finite element truth approximation mesh.

In Figures 4-4 and 4-5 we show the temperature variation over the heat shield at different points in time and for different parameter combinations. We first note that for larger values of $\mu_{(1)}$ the temperature is, overall, much higher than for smaller values of $\mu_{(1)}$. Also, for larger values of $\mu_{(2)}$ more heat is removed through the first cooling channel; for smaller values of $\mu_{(2)}$, however, the heat penetrates deeper into the structure and the temperature tends to be higher and more uniform over the heat shield.



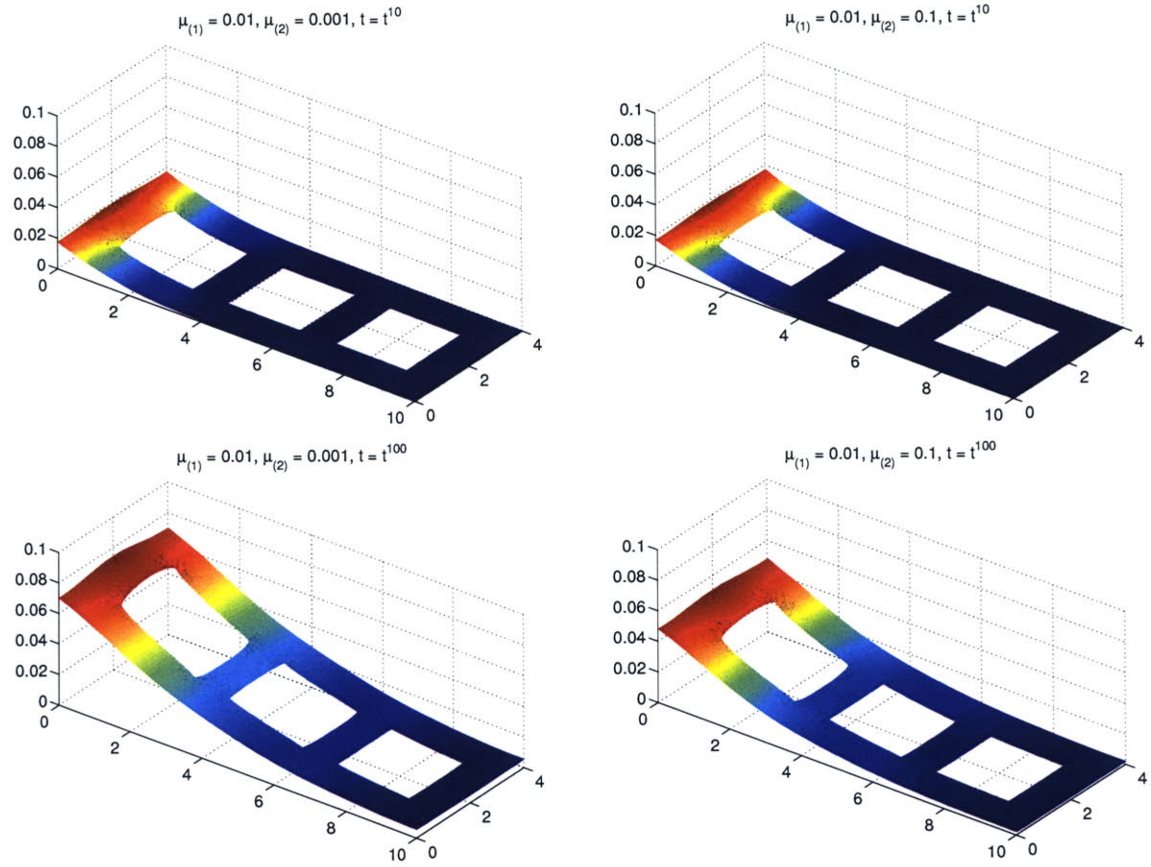
(a) (b)

Figure 4-4: NE 2: Temperature in the heat shield at $t = t^{10} = 2$ and $t = t^{100} = 20$ over the domain Ω for (a) $\mu = (0.5, 0.001)$ and (b) $\mu = (0.5, 0.1)$.

4.3 Reduced-Basis Approximation

4.3.1 Formulation

We first introduce the nested sample sets $S^{\rm pr}_{N_{\rm pr}} = \{\tilde{\mu}^{\rm pr}_1 \in \tilde{\mathcal{D}}, \ldots, \tilde{\mu}^{\rm pr}_{N_{\rm pr}} \in \tilde{\mathcal{D}}\}$, $1 \le N_{\rm pr} \le N_{\rm pr,max}$, and $S^{\rm du}_{N_{\rm du}} = \{\tilde{\mu}^{\rm du}_1 \in \tilde{\mathcal{D}}, \ldots, \tilde{\mu}^{\rm du}_{N_{\rm du}} \in \tilde{\mathcal{D}}\}$, $1 \le N_{\rm du} \le N_{\rm du,max}$, where $\tilde{\mu} \equiv (\mu, t^k)$ and $\tilde{\mathcal{D}} \equiv \mathcal{D} \times \mathbb{I}$; note that the samples must reside in the *parameter-time* space, $\tilde{\mathcal{D}}$. Here, $N_{\rm pr}$ and $N_{\rm du}$ are the dimensions



(a)

(b)

Figure 4-5: NE 2: Temperature in the heat shield at $t = t^{10} = 2$ and $t = t^{100} = 20$ over the domain Ω for (a) $\mu = (0.01, 0.001)$ and (b) $\mu = (0.01, 0.1)$.

of the reduced basis space for the primal and dual variables, respectively; in general, $S^{\rm pr}_{N_{\rm pr}} \neq S^{\rm du}_{N_{\rm du}}$ and in fact $N_{\rm pr} \neq N_{\rm du}$. We then define the associated nested Lagrangian [85] reduced-basis spaces

$$W^{\rm pr}_{N_{\rm pr}} = \operatorname{span}\{\zeta^{\rm pr}_n \equiv y(\tilde{\mu}^{\rm pr}_n),\ 1 \le n \le N_{\rm pr}\}, \quad 1 \le N_{\rm pr} \le N_{\rm pr,max}, \tag{4.15}$$

and

$$W^{\rm du}_{N_{\rm du}} = \operatorname{span}\{\zeta^{\rm du}_n \equiv \Psi(\tilde{\mu}^{\rm du}_n),\ 1 \le n \le N_{\rm du}\}, \quad 1 \le N_{\rm du} \le N_{\rm du,max}, \tag{4.16}$$

where $y(\tilde{\mu}^{\rm pr}_n)$ is the solution of (4.3) at time $t = t^{k^{\rm pr}_n}$ for $\mu = \mu^{\rm pr}_n$ and $\Psi(\tilde{\mu}^{\rm du}_n)$ is the solution of (4.12) at time $t = t^{k^{\rm du}_n}$ for $\mu = \mu^{\rm du}_n$.

Our reduced-basis approximation $y_N(\mu, t^k)$ to $y(\mu, t^k)$ is then obtained by a standard Galerkin projection: given $\mu \in \mathcal{D}$, $y_N(\mu, t^k) \in W^{\rm pr}_{N_{\rm pr}}$, $\forall\, k \in \mathbb{K}$, satisfies

$$m(y_N(\mu, t^k), v; \mu) + \Delta t\ a(y_N(\mu, t^k), v; \mu) = m(y_N(\mu, t^{k-1}), v; \mu) + \Delta t\ b(v; \mu)\ u(t^k), \quad \forall\, v \in W^{\rm pr}_{N_{\rm pr}}, \tag{4.17}$$

with initial condition $y_N(\mu, t^0) = 0$. Similarly, we obtain the dual reduced-basis approximation $\Psi_N(\mu, t^k)$ $\in W^{\rm du}_{N_{\rm du}}$ to $\Psi(\mu, t^k)$ as the solution of

$$m(v, \Psi_N(\mu, t^k); \mu) + \Delta t\ a(v, \Psi_N(\mu, t^k); \mu) = m(v, \Psi_N(\mu, t^{k+1}); \mu), \quad \forall\, v \in W^{\rm du}_{N_{\rm du}},\ \forall\, k \in \mathbb{K}, \tag{4.18}$$

with final condition

$$m(v, \Psi_N(\mu, t^{K+1}); \mu) \equiv \ell(v), \quad \forall\, v \in W^{\rm du}_{N_{\rm du}}. \tag{4.19}$$

Finally, we evaluate the output estimate, $s_N(\mu, t^k)$, from

$$s_N(\mu, t^k) \equiv \ell(y_N(\mu, t^k)) + \sum_{k'=1}^{k} R^{\rm pr}(\Psi_N(\mu, t^{K-k+k'}); \mu, t^{k'})\ \Delta t, \quad \forall\, k \in \mathbb{K}, \tag{4.20}$$

where

$$R^{\rm pr}(v; \mu, t^k) \equiv b(v; \mu)\ u(t^k) - a(y_N(\mu, t^k), v; \mu) - \frac{1}{\Delta t} m(y_N(\mu, t^k) - y_N(\mu, t^{k-1}), v; \mu), \quad \forall\, v \in Y,\ \forall\, k \in \mathbb{K}, \tag{4.21}$$

is the primal residual. Note that here $N \equiv (N_{\rm pr}, N_{\rm du})$.

The critical observation is that the field variable $y(\mu, t^k)$, $\forall k \in \mathbb{K}$, is not, in fact, some arbitrary member of the very high dimensional finite element space Y; rather, it resides, or "evolves," on a much lower dimensional manifold — in effect, a $P + 1$ dimensional manifold — induced by the parametric and temporal dependence. Thus, by restricting our attention to this manifold, we can adequately approximate the field variable by a space of dimension $N_{\rm pr}, N_{\rm du} \ll \mathcal{N}$. In the next section we will show that the field variable is indeed smooth in μ which may be deduced from the equation for the sensitivity derivatives — the stability and continuity properties of the partial differential operator are crucial. Note, however, that the proposed method *does not* require great regularity of the field variable in x; hence non-smooth domains (sharp corners) pose no impediment to rapid convergence. This observation is fundamental to our approach, and is the basis of our approximation; we confirm the rapid convergence in Section 4.6.

4.3.2 *A Priori* Convergence Theory

We consider here the rate at which $y_N(\mu, t^k)$ converges to $y(\mu, t^k)$. The results for the dual variable are very similar and therefore omitted.

Proposition 5. *Assume that the "truth" solution $y(\mu, t^k)$ and the corresponding reduced-basis solution $y_N(\mu, t^k)$ satisfy (4.3) and (4.17), respectively. The error, $e^{\mathrm{pr}}(\mu, t^k) = y(\mu, t^k) - y_N(\mu, t^k)$, is bounded by*

$$\alpha_m(\mu)\ \|e^{\mathrm{pr}}(\mu, t^k)\|_X^2 + \alpha_a(\mu)\ \Delta t \sum_{k'=1}^{k} \|e^{\mathrm{pr}}(\mu, t^{k'})\|_Y^2$$

$$\leq \cancel{\inf}_{w_N(t^k) \in W_N} \left\{ \gamma_m(\mu)\ \|y(\mu, t^k) - w_N(t^k)\|_X^2 + \gamma_a(\mu)\ \Delta t\ \|y(\mu, t^k) - w_N(t^k)\|_Y^2 \right\}$$

$$\left(w_N(t^{k'}) = \arg\inf_{\substack{v \in X \\ 1 \leq k' \leq k}} \|y(\mu, t^{k'}) - v\| \right. + \sum_{k'=1}^{k-1} \cancel{\inf}_{w_N(t^{k'}) \in W_N} \gamma_a(\mu)\ \Delta t\ \|y(\mu, t^{k'}) - w_N(t^{k'})\|_Y^2 \qquad (4.22)$$

Before presenting the proof, we note that in the case of a non-zero initial condition, $y(\mu, t^0) \neq 0$, the additional term $\gamma_m(\mu) \|y(\mu, t^0) - y_N(\mu, t^0)\|_X^2$ — representing the error due to the initial condition — will usually appear on the right hand side of (4.22). However, if $y(\mu, t^0) \in W_{N_{\mathrm{pr}}}^{\mathrm{pr}}$, the initial condition error, $y(\mu, t^0) - y_N(\mu, t^0)$, is identically zero for all $\mu \in \mathcal{D}$ even for $y(\mu, t^0) \neq 0$, and (4.22) remains unchanged.

Proof. From (4.3) and (4.17) it directly follows that the error, $e^{\mathrm{pr}}(\mu, t^k) = y(\mu, t^k) - y_N(\mu, t^k)$, satisfies

$$m(e^{\mathrm{pr}}(\mu, t^k), v; \mu) + \Delta t\ a(e^{\mathrm{pr}}(\mu, t^k), v; \mu) = m(e^{\mathrm{pr}}(\mu, t^{k-1}), v; \mu), \quad \forall\, v \in W_{N_{\mathrm{pr}}}^{\mathrm{pr}}, \qquad (4.23)$$

with initial condition $e^{\mathrm{pr}}(\mu, t^0) = y(\mu, t^0) - y_N(\mu, t^0) = 0$, since $y(\mu, t^0) = y_N(\mu, t^0) = 0$ by assumption. Let $w_N(t^k) \in W_{N_{\mathrm{pr}}}^{\mathrm{pr}}$ be the projection of $y(\mu, t^k)$ with respect to the "m" scalar product and choose $v \equiv w_N(t^k) - y_N(\mu, t^k) = e^{\mathrm{pr}}(\mu, t^k) - (y(\mu, t^k) - w_N(t^k))$ in (4.23). We then obtain

$$\begin{aligned} &m(e^{\mathrm{pr}}(\mu, t^k) - e^{\mathrm{pr}}(\mu, t^{k-1}), e^{\mathrm{pr}}(\mu, t^k); \mu) + \Delta t\ a(e^{\mathrm{pr}}(\mu, t^k), e^{\mathrm{pr}}(\mu, t^k); \mu) \\ &\quad = m(e^{\mathrm{pr}}(\mu, t^k) - e^{\mathrm{pr}}(\mu, t^{k-1}), y(\mu, t^k) - w_N(t^k); \mu) + \Delta t\ a(e^{\mathrm{pr}}(\mu, t^k), y(\mu, t^k) - w_N(t^k); \mu), \end{aligned}$$

or again

$$\begin{aligned} m(&e^{\mathrm{pr}}(\mu, t^k), e^{\mathrm{pr}}(\mu, t^k); \mu) - m(e^{\mathrm{pr}}(\mu, t^{k-1}), e^{\mathrm{pr}}(\mu, t^{k-1}); \mu) \\ &+ m(e^{\mathrm{pr}}(\mu, t^k) - e^{\mathrm{pr}}(\mu, t^{k-1}), e^{\mathrm{pr}}(\mu, t^k) - e^{\mathrm{pr}}(\mu, t^{k-1}); \mu) + 2\ \Delta t\ a(e^{\mathrm{pr}}(\mu, t^k), e^{\mathrm{pr}}(\mu, t^k); \mu) \\ &= \ 2\ m(y(\mu, t^k) - w_N(t^k) - (y(\mu, t^{k-1}) - w_N(t^{k-1})), y(\mu, t^k) - w_N(t^k); \mu) \\ &\quad + 2\ \Delta t\ a(e^{\mathrm{pr}}(\mu, t^k), y(\mu, t^k) - w_N(t^k); \mu), \end{aligned} \qquad (4.24)$$

since $m(z, y(\mu, t^k) - w_N(t^k)) = 0,\ \forall\, z \in W^{\rm pr}_{N_{\rm pr}}$. We next note that

$$\begin{aligned}
2\ m(&y(\mu,t^k) - w_N(t^k) - (y(\mu,t^{k-1}) - w_N(t^{k-1})), y(\mu,t^k) - w_N(t^k); \mu) \\
&= m(y(\mu,t^k) - w_N(t^k), y(\mu,t^k) - w_N(t^k); \mu) \\
&\quad -m(y(\mu,t^{k-1}) - w_N(t^{k-1}), y(\mu,t^{k-1}) - w_N(t^{k-1}); \mu) \\
&\quad +m(y(\mu,t^k) - w_N(t^k) - (y(\mu,t^{k-1}) - w_N(t^{k-1})), \\
&\qquad\qquad y(\mu,t^k) - w_N(t^k) - (y(\mu,t^{k-1}) - w_N(t^{k-1})); \mu) \\
&\le m(y(\mu,t^k) - w_N(t^k), y(\mu,t^k) - w_N(t^k); \mu) \\
&\quad -m(y(\mu,t^{k-1}) - w_N(t^{k-1}), y(\mu,t^{k-1}) - w_N(t^{k-1}); \mu) \\
&\quad +m(e^{\rm pr}(\mu,t^k) - e^{\rm pr}(\mu,t^{k-1}), e^{\rm pr}(\mu,t^k) - e^{\rm pr}(\mu,t^{k-1}); \mu), \qquad (4.25)
\end{aligned}$$

where the last inequality follows since the projection gives the minimum distance, and

$$\begin{aligned}
2\ \Delta t\ a(&e^{\rm pr}(\mu,t^k), y(\mu,t^k) - w_N(t^k); \mu) \\
&= \Delta t\ a(e^{\rm pr}(\mu,t^k), e^{\rm pr}(\mu,t^k); \mu) \\
&\quad +\Delta t\ a(y(\mu,t^k) - w_N(t^k), y(\mu,t^k) - w_N(t^k); \mu) - \Delta t\ a(v, v; \mu) \\
&\le \Delta t\ a(e^{\rm pr}(\mu,t^k), e^{\rm pr}(\mu,t^k); \mu) + \Delta t\ a(y(\mu,t^k) - w_N(t^k), y(\mu,t^k) - w_N(t^k); \mu), \quad (4.26)
\end{aligned}$$

where we used the symmetry and coercivity of a. From (4.24), (4.25), and (4.26) it thus follows that

$$\begin{aligned}
m(e^{\rm pr}(\mu,t^k), e^{\rm pr}(\mu,t^k); \mu) - m(e^{\rm pr}(\mu,t^{k-1}), e^{\rm pr}(\mu,t^{k-1}); \mu) + \Delta t\ a(e^{\rm pr}(\mu,t^k), e^{\rm pr}(\mu,t^k); \mu) \\
\le m(y(\mu,t^k) - w_N(t^k), y(\mu,t^k) - w_N(t^k); \mu) \\
-m(y(\mu,t^{k-1}) - w_N(t^{k-1}), y(\mu,t^{k-1}) - w_N(t^{k-1}); \mu) \\
+\Delta t\ a(y(\mu,t^k) - w_N(t^k), y(\mu,t^k) - w_N(t^k); \mu). \qquad (4.27)
\end{aligned}$$

The desired results directly follows by summing from 1 to k and invoking the coercivity and continuity of the bilinear forms a and m. □

Proposition 5 states that $y_N(\mu, t^k)$ is the best approximation among all members of $W^{\rm pr}_{N_{\rm pr}}$ in the sense of (4.22). However, only if the field variable $y(\mu, t^k)$ is indeed smooth in μ can we expect a rapid convergence of our reduced-basis approximation. We want to obtain an approximation of sufficient accuracy for only modest $N_{\rm pr}$ — this is crucial for our method since a small value of $N_{\rm pr}$ is vital for the computational efficiency.

We will deduce the smoothness of $y(\mu, t^k)$ from the equation for the sensitivity derivatives. As a preliminary result, however, we have to prove that $y(\mu, t^k)$ is bounded: choosing $v = y(\mu, t^k)$ in (4.3) we have

$$\begin{aligned}
m(y(\mu,t^k), y(\mu,t^k; \mu) + \Delta t\ a(y(\mu,t^k), y(\mu,t^k); \mu) \\
= m(y(\mu,t^{k-1}), y(\mu,t^k); \mu) + \Delta t\ b(y(\mu,t^k); \mu)\ u(t^k), \quad \forall\, k \in \mathbb{K}, \quad (4.28)
\end{aligned}$$

where $y(\mu, t^0) = 0$ by assumption. We now invoke the Cauchy-Schwarz inequality for the cross term $m(y(\mu,t^{k-1}),y(\mu,t^k);\mu)$ to obtain

$$\begin{aligned} m(y(\mu,t^k),y(\mu,t^k);\mu) &+ \Delta t\ a(y(\mu,t^k),y(\mu,t^k);\mu) \\ &\leq m^{\frac{1}{2}}(y(\mu,t^k),y(\mu,t^k);\mu)\ m^{\frac{1}{2}}(y(\mu,t^{k-1}),y(\mu,t^{k-1});\mu) \\ &\qquad + \Delta t\ \|b(\cdot;\mu)\|_{Y'}\ |u(t^k)|\ \|y(\mu,t^k)\|_Y, \quad \forall\, k \in \mathbb{K}. \end{aligned} \tag{4.29}$$

We now recall the identity (for $c \in \mathbb{R},\ d \in \mathbb{R},\ \rho \in \mathbb{R}_+$)

$$2\ |c|\ |d| \leq \frac{1}{\rho^2}c^2 + \rho^2\ d^2, \tag{4.30}$$

which we apply twice: first, choosing $c = m^{\frac{1}{2}}(y(\mu,t^k),y(\mu,t^k);\mu)$, $d = m^{\frac{1}{2}}(y(\mu,t^{k-1}),y(\mu,t^{k-1});\mu)$, and $\rho = 1$, we obtain

$$\begin{aligned} 2\ m^{\frac{1}{2}}&(y(\mu,t^k),y(\mu,t^k);\mu)\ m^{\frac{1}{2}}(y(\mu,t^{k-1}),y(\mu,t^{k-1});\mu) \\ &\leq m(y(\mu,t^{k-1}),y(\mu,t^{k-1});\mu) + m(y(\mu,t^k),y(\mu,t^k);\mu); \end{aligned} \tag{4.31}$$

and second, choosing $c = \|b(\cdot;\mu)\|_{Y'}\,|u(t^k)|$, $d = \|y(\mu,t^k)\|_Y$, and $\rho = \alpha_a(\mu)^{\frac{1}{2}}$ we have

$$2\ \|b(\cdot;\mu)\|_{Y'}\ |u(t^k)|\ \|y(\mu,t^k)\|_Y \leq \frac{1}{\alpha_a(\mu)}\ \|b(\cdot;\mu)\|^2_{Y'}\ |u(t^k)|^2 + \alpha_a(\mu)\ \|y(\mu,t^k)\|^2_Y. \tag{4.32}$$

Combining (4.29), (4.31), and (4.32), and invoking (4.7), we obtain

$$\begin{aligned} m(y(\mu,t^k),y(\mu,t^k);\mu) &- m(y(\mu,t^{k-1}),y(\mu,t^{k-1});\mu) \\ &+ \Delta t\ a(y(\mu,t^k),y(\mu,t^k);\mu) \leq \frac{\Delta t}{\alpha_a(\mu)}\,\|b(\cdot;\mu)\|^2_{Y'}\ |u(t^k)|^2 \quad \forall\, k \in \mathbb{K}, \end{aligned} \tag{4.33}$$

We now perform the sum from $k' = 1$ to k and recall that $y(\mu,t^0) = 0$, leading to

$$|||y(\mu,t^k)|||^{\mathrm{pr}^2} \leq \frac{\Delta t}{\alpha_a(\mu)} \sum_{k'=1}^{k} \|b(\cdot;\mu)\|^2_{Y'}\ |u(t^k)|^2, \quad \forall\, k \in \mathbb{K}, \tag{4.34}$$

where the spatio-temporal energy norm, $|||\cdot|||^{\mathrm{pr}}$, is defined as

$$|||v(\mu,t^k)|||^{\mathrm{pr}} \equiv \left(m(v(\mu,t^k),v(\mu,t^k);\mu) + \sum_{k'=1}^{k} a(v(\mu,t^{k'}),v(\mu,t^{k'});\mu)\ \Delta t \right)^{\frac{1}{2}}, \quad \forall\, v \in Y. \tag{4.35}$$

Thus, since $\|b(\cdot;\mu)\|_{Y'}$ is bounded by assumption, $y(\mu,t^k)$ is bounded in the energy norm as long as the control input $u(t^k)$ remains bounded.

We now turn to the sensitivity derivatives. To begin, we take the derivative of (4.3) with respect to the parameter μ twice. We first obtain

$$m(y_\mu(\mu,t^k),v;\mu) + \Delta t\ a(y_\mu(\mu,t^k),v;\mu) = m(y_\mu(\mu,t^{k-1}),v;\mu) + \Delta t\ b_\mu(v;\mu)\ u(t^k) \\ - [m_\mu(y(\mu,t^k) - y(\mu,t^{k-1});v;\mu) - \Delta t\ a_\mu(y(\mu,t^k),v;\mu)], \quad \forall\, v \in Y \quad (4.36)$$

with initial condition $y_\mu(\mu,t^0) = 0$, and taking the second derivative results in

$$m(y_{\mu\mu}(\mu,t^k),v;\mu) + \Delta t\ a(y_{\mu\mu}(\mu,t^k),v;\mu) = m(y_{\mu\mu}(\mu,t^{k-1}),v;\mu) + \Delta t\ b_{\mu\mu}(v;\mu)\ u(t^k) \\ - 2\left\{m_\mu(y_\mu(\mu,t^k) - y_\mu(\mu,t^{k-1});v;\mu) + \Delta t\ a_\mu(y_\mu(\mu,t^k),v;\mu)\right\} \\ - \left\{m_{\mu\mu}(y(\mu,t^k) - y(\mu,t^{k-1}),v;\mu) + a_{\mu\mu}(y(\mu,t^k),v;\mu)\right\}, \quad \forall\, v \in Y \quad (4.37)$$

with initial condition $y_{\mu\mu}(\mu,t^0) = 0$. From the assumption of affine parameter dependence (4.9) it follows that

$$a_\mu(w,v;\mu) = \sum_{q=1}^{Q_a} \frac{\partial \Theta_a^q(\mu)}{\partial \mu}\ a^q(w,v), \quad \forall\, w,v \in Y,\ \forall\, \mu \in \mathcal{D}, \quad (4.38)$$

$$a_{\mu\mu}(w,v;\mu) = \sum_{q=1}^{Q_a} \frac{\partial^2 \Theta_a^q(\mu)}{\partial \mu^2}\ a^q(w,v), \quad \forall\, w,v \in Y,\ \forall\, \mu \in \mathcal{D}. \quad (4.39)$$

A similar result follows for the bilinear form m and the linear form b from (4.10) and (4.11), respectively. We shall also make the assumption that the bilinear forms $a^q(\cdot,\cdot)$, $q = 1,\ldots,Q_a$ and $m^q(\cdot,\cdot)$, $q = 1,\ldots,Q_m$ are continuous, i.e.,

$$a^q(w,v) \le \gamma_a^q(\mu)\|w\|_Y\|v\|_Y \le \gamma_a^{0\,q}\|w\|_Y\|v\|_Y, \quad \forall\, w,v \in Y,\ \forall\, \mu \in \mathcal{D},\ q = 1,\ldots,Q_a \quad (4.40)$$

$$m^q(w,v) \le \gamma_m^q(\mu)\|w\|_X\|v\|_X \le \gamma_m^{0\,q}\|w\|_X\|v\|_X, \quad \forall\, w,v \in Y,\ \forall\, \mu \in \mathcal{D},\ q = 1,\ldots,Q_m. (4.41)$$

We first prove the boundedness of $y_\mu(\mu,t^k)$: choosing $v = y_\mu(\mu,t^k)$ in (4.36) we obtain

$$m(y_\mu(\mu,t^k),y_\mu(\mu,t^k);\mu) + \Delta t\ a(y_\mu(\mu,t^k),y_\mu(\mu,t^k);\mu) \\ = m(y_\mu(\mu,t^{k-1}),y_\mu(\mu,t^k);\mu) + \Delta t\ b_\mu(y_\mu(\mu,t^k);\mu)\ u(t) \\ - \left\{m_\mu(y(\mu,t^k) - y(\mu,t^{k-1}),y_\mu(\mu,t^k);\mu) + \Delta t\ a_\mu(y(\mu,t^k),y_\mu(\mu,t^k);\mu)\right\}. \quad (4.42)$$

We now invoke the Cauchy-Schwarz inequality for the cross-term $m(y_\mu(\mu,t^{k-1}),y_\mu(\mu,t^k);\mu)$ which results in

$$2\ m(y_\mu(\mu,t^{k-1}),y_\mu(\mu,t^k);\mu) \le 2\ m^{1/2}(y_\mu(\mu,t^{k-1}),y_\mu(\mu,t^{k-1});\mu)\ m^{1/2}(y_\mu(\mu,t^k),y_\mu(\mu,t^k);\mu) \\ \le m(y_\mu(\mu,t^{k-1}),y_\mu(\mu,t^{k-1});\mu) + m(y_\mu(\mu,t^k),y_\mu(\mu,t^k);\mu); \quad (4.43)$$

where the second step follows (4.30) with $\rho = 1$, $c = m^{1/2}(y_\mu(\mu,t^{k-1}),y_\mu(\mu,t^{k-1});\mu)$, and $d =$

$m^{1/2}(y_\mu(\mu,t^k),y_\mu(\mu,t^k);\mu)$. From (4.42) and (4.43) we then obtain

$$\begin{aligned} m(y_\mu(\mu,t^k),y_\mu(\mu,t^k);\mu) - m(y_\mu(\mu,t^{k-1}),y_\mu(\mu,t^{k-1});\mu) + 2\Delta t\ a(y_\mu(\mu,t^k),y_\mu(\mu,t^k);\mu) \\ \leq \left|2\ \Delta t\ b_\mu(y_\mu(\mu,t^k);\mu)\ u(t^k)\right| + \left|2m_\mu(y(\mu,t^k)-y(\mu,t^{k-1}),y_\mu(\mu,t^k);\mu)\right| \\ + \left|2\ \Delta t\ a_\mu(y(\mu,t^k),y_\mu(\mu,t^k);\mu)\right|. \quad (4.44) \end{aligned}$$

From the continuity (4.40) of a^q and (4.38) it follows that

$$\begin{aligned} \left|a_\mu(y(\mu,t^k),y_\mu(\mu,t^k);\mu)\right| &= \left|\sum_{q=1}^{Q_a}\frac{\partial\Theta_a^q(\mu)}{\partial\mu}\ a^q(y(\mu,t^k),y_\mu(\mu,t^k))\right| \\ &\leq \left|\sum_{q=1}^{Q_a}\frac{\partial\Theta_a^q(\mu)}{\partial\mu}\ \gamma_a^q(\mu)\right|\ \|y(\mu,t^k)\|_Y\ \|y_\mu(\mu,t^k)\|_Y; \qquad (4.45) \end{aligned}$$

similarly, from the continuity (4.41) of m^q and the affine decomposition (4.10) we obtain

$$\begin{aligned} \left|m_\mu(y(\mu,t^k)-y(\mu,t^{k-1}),y_\mu(\mu,t^k);\mu)\right| &= \left|\sum_{q=1}^{Q_m}\frac{\partial\Theta_m^q(\mu)}{\partial\mu}\ m^q(y(\mu,t^k)-y(\mu,t^{k-1}),y_\mu(\mu,t^k)\right| \\ &\leq \left|\sum_{q=1}^{Q_m}\frac{\partial\Theta_m^q(\mu)}{\partial\mu}\ \gamma_m^q(\mu)\right| \\ &\qquad \times\|y(\mu,t^k)-y(\mu,t^{k-1})\|_X\ \|y_\mu(\mu,t^k)\|_X; (4.46) \end{aligned}$$

We now apply (4.30) thrice: first, with $\rho=\epsilon_b$, $c=\|b_\mu(\cdot;\mu)\|_{Y'}|u(t^k)|$, and $d=\|y_\mu(\mu,t^k)\|_Y$; second, with $\rho=\epsilon_a$, $c=\left|\sum_{q=1}^{Q_a}\frac{\partial\Theta_a^q(\mu)}{\partial\mu}\gamma_a^q(\mu)\right|\ \|y(\mu,t^k)\|_Y$, and $d=\|y_\mu(\mu,t^k)\|_Y$; and third, with $\rho=\epsilon_m\sqrt{\Delta t}$, $c=\left|\sum_{q=1}^{Q_m}\frac{\partial\Theta_m^q(\mu)}{\partial\mu}\gamma_m^q(\mu)\right|\ \|y(\mu,t^k)-y(\mu,t^{k-1})\|_X$, and $d=\|y_\mu(\mu,t^k)\|_X$. We thus obtain from (4.44)-(4.46) after summing from $k' = 1$ that

$$\begin{aligned}
m(y_\mu(\mu,t^k), y_\mu(\mu,t^k);\mu) &+ 2\ \Delta t \sum_{k'=1}^{k} a(y_\mu(\mu,t^{k'}), y_\mu(\mu,t^{k'});\mu) \\
&\le \frac{\Delta t}{\epsilon_b^2} \sum_{k'=1}^{k} \|b_\mu(\cdot;\mu)\|_{Y'}^2\ |u(t^{k'})|^2 + \epsilon_b^2\ \Delta t \sum_{k'=1}^{k} \|y_\mu(\mu,t^{k'})\|_Y^2 \\
&+ \frac{1}{\epsilon_m^2\ \Delta t} \sum_{k'=1}^{k} \left(\sum_{q=1}^{Q_m} \frac{\partial \Theta_m^q(\mu)}{\partial \mu}\ \gamma_m^q(\mu) \right)^2 \|y(\mu,t^{k'}) - y(\mu,t^{k'-1})\|_X^2 \\
&+ \epsilon_m^2\ \Delta t \sum_{k'=1}^{k} \|y_\mu(\mu,t^{k'})\|_X^2 + \frac{\Delta t}{\epsilon_a^2} \sum_{k'=1}^{k} \left(\sum_{q=1}^{Q_a} \frac{\partial \Theta_a^q(\mu)}{\partial \mu}\ \gamma_a^q(\mu) \right)^2 \|y(\mu,t^{k'})\|_Y^2 \\
&+ \epsilon_a^2\ \Delta t \sum_{k'=1}^{k} \|y_\mu(\mu,t^{k'})\|_Y^2. \qquad (4.47)
\end{aligned}$$

From the Poincaré-Friedrich's inequality, $\|v\|_X \le C_{PF}\|v\|_Y$, choosing ϵ_b, ϵ_a, and ϵ_m such that $\epsilon_b^2 + C_{PF}\epsilon_a^2 + \epsilon_m^2 = \alpha_a(\mu)$, $\epsilon_{a,b,m} > 0$, and invoking (4.5), we finally obtain

$$\begin{aligned}
|||y_\mu(\mu,t^k)|||^{\mathrm{pr}^2} &= m(y_\mu(\mu,t^k), y_\mu(\mu,t^k);\mu) + \sum_{k'=1}^{k} a(y_\mu(\mu,t^{k'}), y_\mu(\mu,t^{k'});\mu) \\
&\le \frac{\Delta t}{\epsilon_b^2} \sum_{k'=1}^{k} \|b_\mu(\cdot;\mu)\|_{Y'}^2\ |u(t^{k'})|^2 + \frac{\Delta t}{\epsilon_a^2} \sum_{k'=1}^{k} \left(\sum_{q=1}^{Q_a} \frac{\partial \Theta_a^q(\mu)}{\partial \mu}\ \gamma_a^q(\mu) \right)^2 \|y(\mu,t)\|_Y^2 \\
&+ \frac{\Delta t}{\epsilon_m^2} \sum_{k'=1}^{k} \left(\sum_{q=1}^{Q_m} \frac{\partial \Theta_m^q(\mu)}{\partial \mu}\ \gamma_m^q(\mu) \right)^2 \|\frac{1}{\Delta t}(y(\mu,t^{k'}) - y(\mu,t^{k'-1}))\|_X^2. \qquad (4.48)
\end{aligned}$$

It thus follows that the derivative of $y(\mu,t^k)$ with respect to the parameter μ is bounded. A similar result can be derived for the second derivative, $y_{\mu\mu}(\mu,t^k)$, by starting from (4.37) with $v = y_{\mu\mu}(\mu,t^k)$ and following the same steps as above — the field variable $y(\mu,t^k)$ is indeed smooth in μ and we can expect a rapid convergence of our reduced-basis approximation.

4.3.3 Offline-Online Computational Procedure

In this section we develop offline-online computational procedures in order to fully exploit the dimension reduction of the problem [9, 48, 60, 91]. We first express $y_N(\mu,t^k)$ and $\Psi_N(\mu,t^k)$ as

$$y_N(\mu,t^k) = \sum_{n=1}^{N_{\mathrm{pr}}} y_{Nn}(\mu,t^k)\ \zeta_n^{\mathrm{pr}}, \qquad (4.49)$$

and

$$\Psi_N(\mu, t^k) = \sum_{n=1}^{N_{\rm du}} \Psi_{N\,n}(\mu, t^k)\ \zeta_n^{\rm du}, \tag{4.50}$$

respectively. We then choose as test functions $v = \zeta_n^{\rm pr}$, $1 \le n \le N_{\rm pr}$, for the primal problem (4.17) and $v = \zeta_n^{\rm du}$, $1 \le n \le N_{\rm du}$, for the dual problem (4.18). (We prefer Galerkin over Petrov-Galerkin for purposes of stability.)

It then follows from (4.17) that $\underline{y}_N(\mu, t^k) = [y_{N\,1}(\mu, t^k)\ \ y_{N\,2}(\mu, t^k)\ \ \ldots\ \ y_{N\,N_{\rm pr}}(\mu, t^k)]^T \in \mathbb{R}^{N_{\rm pr}}$ satisfies

$$(M_N^{\rm pr}(\mu) + \Delta t\ A_N^{\rm pr}(\mu))\ \underline{y}_N(\mu, t^k) = M_N^{\rm pr}(\mu)\ \underline{y}_N(\mu, t^{k-1}) + \Delta t\ B_N^{\rm pr}(\mu)\ u(t^k), \quad \forall\, k \in \mathbb{K}, \tag{4.51}$$

with initial condition $y_{N\,n}(\mu, t^0) = 0$, $1 \le n \le N_{\rm pr}$. Here, $M_N^{\rm pr}(\mu) \in \mathbb{R}^{N_{\rm pr}\times N_{\rm pr}}$ and $A_N^{\rm pr}(\mu) \in \mathbb{R}^{N_{\rm pr}\times N_{\rm pr}}$ are SPD matrices with entries $M_{N\,i,j}^{\rm pr}(\mu) = m(\zeta_i^{\rm pr}, \zeta_j^{\rm pr}; \mu)$, $1 \le i,j \le N_{\rm pr}$, and $A_{N\,i,j}^{\rm pr}(\mu) = a(\zeta_i^{\rm pr}, \zeta_j^{\rm pr}; \mu)$, $1 \le i,j \le N_{\rm pr}$, respectively; and $B_N^{\rm pr}(\mu) \in \mathbb{R}^{N_{\rm pr}}$ is the control vector with entries $B_{N\,i}^{\rm pr}(\mu) = b(\zeta_i^{\rm pr}; \mu)$, $1 \le i \le N_{\rm pr}$.

Invoking the affine decomposition (4.9)-(4.11) we obtain

$$\begin{aligned}
M_{N\,i,j}^{\rm pr}(\mu) &= m(\zeta_i^{\rm pr}, \zeta_j^{\rm pr}; \mu) = \sum_{q=1}^{Q_m} \Theta_m^q(\mu)\ m^q(\zeta_i^{\rm pr}, \zeta_j^{\rm pr}), && (4.52)\\
A_{N\,i,j}^{\rm pr}(\mu) &= a(\zeta_i^{\rm pr}, \zeta_j^{\rm pr}; \mu) = \sum_{q=1}^{Q_a} \Theta_a^q(\mu)\ a^q(\zeta_i^{\rm pr}, \zeta_j^{\rm pr}), && (4.53)\\
B_{N\,i}^{\rm pr}(\mu) &= b^q(\zeta_i^{\rm pr}; \mu) = \sum_{q=1}^{Q_b} \Theta_b^q(\mu)\ b^q(\zeta_i^{\rm pr}), && (4.54)
\end{aligned}$$

which can be written as

$$M_N^{\rm pr}(\mu) = \sum_{q=1}^{Q_m} \Theta_m^q(\mu)\ M_N^{{\rm pr}\,q}, \quad A_N^{\rm pr}(\mu) = \sum_{q=1}^{Q_a} \Theta_a^q(\mu)\ A_N^{{\rm pr}\,q}, \quad B_N^{\rm pr}(\mu) = \sum_{q=1}^{Q_b} \Theta_b^q(\mu)\ B_N^{{\rm pr}\,q}, \tag{4.55}$$

where the *parameter independent* quantities $M_N^{{\rm pr}\,q} \in \mathbb{R}^{N_{\rm pr}\times N_{\rm pr}}$, $A_N^{{\rm pr}\,q} \in \mathbb{R}^{N_{\rm pr}\times N_{\rm pr}}$, and $B_N^{{\rm pr}\,q} \in \mathbb{R}^{N_{\rm pr}}$ are given by

$$\begin{aligned}
M_{N\,i,j}^{{\rm pr}\,q} &= m^q(\zeta_i^{\rm pr}, \zeta_j^{\rm pr}), && 1 \le i,j \le N_{\rm pr,max},\ 1 \le q \le Q_m,\\
A_{N\,i,j}^{{\rm pr}\,q} &= a^q(\zeta_i^{\rm pr}, \zeta_j^{\rm pr}), && 1 \le i,j \le N_{\rm pr,max},\ 1 \le q \le Q_a,\\
B_{N\,i}^{{\rm pr}\,q} &= b^q(\zeta_i^{\rm pr}), && 1 \le i \le N_{\rm pr,max},\ 1 \le q \le Q_b,
\end{aligned} \tag{4.56}$$

respectively.

A similar computational procedure for the dual problem (4.18)-(4.19) and the residual correction term in (4.20) can also be developed. The details of this derivation and the definitions of the necessary quantities are summarized in Appendix A.

The offline-online decomposition is now clear. In the offline stage — performed only *once* — we first solve for the $\zeta_n^{\rm pr}$, $1 \le n \le N_{\rm pr,max}$ and $\zeta_n^{\rm du}$, $1 \le n \le N_{\rm du,max}$; we then compute and store the μ-independent quantities in (4.56) for the primal problem, (A.3) for the dual problem, and (A.6) for the output estimate. The computational cost is therefore $O(K(N_{\mathrm{pr,max}}+N_{\mathrm{du,max}}))$ solutions of the underlying $\mathcal{N}$-dimensional "truth" finite element approximation and $O((N^2_{\mathrm{pr,max}}+N^2_{\mathrm{du,max}}+N_{\mathrm{pr,max}}N_{\mathrm{du,max}})(Q_a+Q_m))$ $\mathcal{N}$-inner products; the storage requirements are also $O((N^2_{\mathrm{pr,max}}+N^2_{\mathrm{du,max}}+N_{\mathrm{pr,max}}N_{\mathrm{du,max}})(Q_a+Q_m))$.

In the online stage — performed many times, for each new parameter value μ — we first assemble the reduced-basis matrices (4.55), (A.2), and (A.5); this requires $O((N^2_{\mathrm{pr}}+N^2_{\mathrm{du}}+N_{\mathrm{pr}}N_{\mathrm{du}})(Q_a+Q_m))$ operations. We then solve the primal and dual problem for $\underline{y}_N(\mu,t^k)$ and $\underline{\Psi}_N(\mu,t^k)$, respectively; since the reduced-basis matrices are in general full, the operation count (based on LU factorization and our LTI assumption) is $O(N^3_{\mathrm{pr}}+N^3_{\mathrm{du}}+K(N^2_{\mathrm{pr}}+N^2_{\mathrm{du}}))$. Finally, given $\underline{y}_N(\mu,t^k)$ and $\underline{\Psi}_N(\mu,t^k)$ we evaluate the output estimate $s_N(\mu,t^k)$ from (A.4) at a cost of $O(2kN_{\mathrm{pr}}N_{\mathrm{du}})$; note that the calculation of all outputs $s_N(\mu,t^k)$, $\forall\, k\in\mathbb{K}$, requires $O(K(K+1)N_{\mathrm{pr}}N_{\mathrm{du}})$ operations.

Thus, as required in the many-query or real-time contexts, the online complexity is *independent* of $\mathcal{N}$, the dimension of the underlying "truth" finite element approximation space. Since $N_{\mathrm{pr}}, N_{\mathrm{du}} \ll \mathcal{N}$ we expect significant computational savings in the online stage relative to classical discretization and solution approaches.

Finally, we note that classical model-order reduction techniques, such as modal decomposition [36] and POD [7], require the evaluation of a new set of eigenmodes or basis functions — and thus a return to the (very fine) "truth" approximation — for each new parameter value encountered. In contrast, reduced-basis methods do not need to return to the "truth" approximation in the online stage, and are therefore far more efficient in evaluating input-output relationships for many different parameter values.

4.4 *A Posteriori* Error Estimation

From Section 4.3 we know that we can efficiently obtain the output estimate, $s_N(\mu,t^k)$, for the output of interest, $s(\mu,t^k)$: the online complexity depends only on N_{pr} and N_{du}, the dimensions of the reduced-basis spaces for the primal and dual variable, respectively. However, we do not yet know if $s_N(\mu,t^k)$ is indeed a good approximation to $s(\mu,t^k)$, i.e., is $|s(\mu,t^k)-s_N(\mu,t^k)|\le\epsilon^s_{\mathrm{tol}}$, where $\epsilon^s_{\mathrm{tol}}$ is a maximum acceptable error? Or conversely, is our approximation "too good," i.e., is $|s(\mu,t^k)-s_N(\mu,t^k)|\ll\epsilon^s_{\mathrm{tol}}$ — that is, is N_{pr} or N_{du} too large, with associated detriment to the online efficiency? It should also be evident that the approximation properties do not only depend on the size of N_{pr} and N_{du}, but also on the choice of the sampling sets $S^{\mathrm{pr}}_{N_{\mathrm{pr}}}$ and $S^{\mathrm{du}}_{N_{\mathrm{du}}}$ and associated reduced-basis spaces $W^{\mathrm{pr}}_{N_{\mathrm{pr}}}$ and $W^{\mathrm{du}}_{N_{\mathrm{du}}}$.

We thus need to develop rigorous *a posteriori* error estimators which will help us to (*i*) assess the error introduced by our reduced-basis approximation (relative to the "truth" finite element approximation); and (*ii*) devise an "optimal" and efficient procedure for selecting the sample sets $S^{\mathrm{pr}}_{N_{\mathrm{pr}}}$ and $S^{\mathrm{du}}_{N_{\mathrm{du}}}$. Surprisingly, *a posteriori* error estimation for reduced-basis approximations has received very little attention in the past. A family of rigorous error estimators for reduced-basis approximations of a wide class of elliptic PDEs is introduced in [60, 91, 121, 122, 123]; we will now extend these ideas to time-dependent (parabolic) partial differential equations. Our approach here is a simplification and generalization of earlier efforts in this direction [102].

We remark that the development of the error bounds presented below is not limited to the reduced-basis approximation described in this paper: with suitable hypotheses, we may consider "any" stable ODE or PDE system and associated reduced-order model.

4.4.1 Preliminaries

To begin, we assume that we are given positive lower bounds for the coercivity constants, $\alpha_a(\mu)$ and $\alpha_m(\mu)$: $\hat{\alpha}_a(\mu) : \mathcal{D} \rightarrow \boldsymbol{R}_+$ satisfies

$$\alpha_a(\mu) \geq \hat{\alpha}_a(\mu) \geq \hat{\alpha}_a^0 > 0, \quad \forall\, \mu \in \mathcal{D}, \tag{4.57}$$

and $\hat{\alpha}_m(\mu) : \mathcal{D} \rightarrow \boldsymbol{R}_+$ satisfies

$$\alpha_m(\mu) \geq \hat{\alpha}_m(\mu) \geq \hat{\alpha}_m^0 > 0, \quad \forall\, \mu \in \mathcal{D}; \tag{4.58}$$

various recipes for this construction can be found in [91, 124]. We next introduce the dual norm of the primal residual

$$\varepsilon^{\text{pr}}_{N_{\text{pr}}}(\mu, t^k) \equiv \sup_{v \in Y} \frac{R^{\text{pr}}(v; \mu, t^k)}{\|v\|_Y}, \quad \forall\, k \in \mathbb{K}, \tag{4.59}$$

and the dual norm of the dual residual

$$\varepsilon^{\text{du}}_{N_{\text{du}}}(\mu, t^k) \equiv \sup_{v \in Y} \frac{R^{\text{du}}(v; \mu, t^k)}{\|v\|_Y}, \quad \forall\, k \in \mathbb{K}, \tag{4.60}$$

where

$$R^{\text{du}}(v; \mu, t^k) \equiv -a(v, \Psi_N(\mu, t^k); \mu) - \frac{1}{\Delta t} m(v, \Psi_N(\mu, t^k) - \Psi_N(\mu, t^{k+1}); \mu), \quad \forall v \in Y,\ \forall k \in \mathbb{K}, \tag{4.61}$$

is the dual residual. We also specify the inner products

$$(v, w)_Y \equiv a(v, w; \mu_{\text{ref(s)}}), \quad \forall\, v, w \in Y, \tag{4.62}$$

and

$$(v, w)_X \equiv m(v, w; \mu_{\text{ref(s)}}), \quad \forall\, v, w \in Y, \tag{4.63}$$

for some constant reference value(s) $\mu_{\text{ref(s)}}$, and recall that $\|\cdot\|_Y = (\cdot,\cdot)_Y^{1/2}$, $\|\cdot\|_X = (\cdot,\cdot)_X^{1/2}$.

We now present and prove the bounding properties for the errors in the primal variable, the dual variable, and the output estimate. Throughout this section we assume that the "truth" solutions $y(\mu, t^k)$ and $\Psi(\mu, t^k)$ satisfy (4.3) and (4.12), respectively, and the corresponding reduced-basis approximations $y_N(\mu, t^k)$ and $\Psi_N(\mu, t^k)$ satisfy (4.17) and (4.18), respectively. We emphasize that our error bounds are very classical, based entirely on standard stability results invoked in *a priori* analyses [94]; the critical new ingredient — tailored to the reduced-basis context — is the offline-online computational procedure of Section 4.4.4

4.4.2 Error Bound Formulation

Primal Variable

We obtain the following result for the error in the primal variable.

Proposition 6. *Let $e^{\text{pr}}(\mu, t^k) \equiv y(\mu, t^k) - y_N(\mu, t^k)$ be the error in the primal variable and define* *the "spatio-temporal" energy norm*

$$|||v(\mu,t^k)|||^{\mathrm{pr}} \equiv \left(m(v(\mu,t^k),v(\mu,t^k);\mu) + \sum_{k'=1}^{k} a(v(\mu,t^{k'}),v(\mu,t^{k'});\mu)\ \Delta t \right)^{\frac{1}{2}}, \quad \forall\, v \in Y. \quad (4.64)$$

The error in the primal variable is then bounded by

$$|||e^{\mathrm{pr}}(\mu,t^k)|||^{\mathrm{pr}} \le \Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^k), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \quad (4.65)$$

where the error bound $\Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^k)$ is defined as

$$\Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^k) \equiv \left(\frac{\Delta t}{\hat{\alpha}_a(\mu)} \sum_{k'=1}^{k} \varepsilon^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^{k'})^2 \right)^{\frac{1}{2}}, \quad (4.66)$$

and $\varepsilon^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^k)$ is the dual norm of the primal residual defined in (4.59).

Proof. We immediately derive from (4.3) and (4.21) that $e^{\mathrm{pr}}(\mu,t^k) = y(\mu,t^k) - y_N(\mu,t^k)$ satisfies

$$m(e^{\mathrm{pr}}(\mu,t^k),v;\mu) + \Delta t\ a(e^{\mathrm{pr}}(\mu,t^k),v;\mu) = m(e^{\mathrm{pr}}(\mu,t^{k-1}),v;\mu) + \Delta t\ R^{\mathrm{pr}}(v;\mu,t^k), \qquad \forall\, v \in Y,\ \forall\, k \in \mathbb{K}, \quad (4.67)$$

where $e^{\mathrm{pr}}(\mu,t^0) = 0$ since $y(\mu,t^0) = y_N(\mu,t^0) = 0$ by assumption. We now choose $v = e^{\mathrm{pr}}(\mu,t^k)$, invoke the Cauchy-Schwarz inequality for the cross term $m(e^{\mathrm{pr}}(\mu,t^{k-1}),e^{\mathrm{pr}}(\mu,t^k);\mu)$, and apply (4.59) to obtain

$$\begin{aligned} m(e^{\mathrm{pr}}(\mu,t^k),e^{\mathrm{pr}}(\mu,t^k);\mu) &+ \Delta t\ a(e^{\mathrm{pr}}(\mu,t^k),e^{\mathrm{pr}}(\mu,t^k);\mu) \\ &\le m^{\frac{1}{2}}(e^{\mathrm{pr}}(\mu,t^k),e^{\mathrm{pr}}(\mu,t^k);\mu)\ m^{\frac{1}{2}}(e^{\mathrm{pr}}(\mu,t^{k-1}),e^{\mathrm{pr}}(\mu,t^{k-1});\mu) \\ &\quad + \Delta t\ \varepsilon^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^k)\ \|e^{\mathrm{pr}}(\mu,t^k)\|_Y, \quad \forall\, k \in \mathbb{K}. \end{aligned} \quad (4.68)$$

We now recall the identity (for $c \in \mathbb{R},\ d \in \mathbb{R},\ \rho \in \mathbb{R}_+$)

$$2\ |c|\ |d| \le \frac{1}{\rho^2}c^2 + \rho^2\ d^2, \quad (4.69)$$

which we apply twice: first, choosing

$$c = m^{\frac{1}{2}}(e^{\mathrm{pr}}(\mu,t^k),e^{\mathrm{pr}}(\mu,t^k);\mu),$$

$$d = m^{\frac{1}{2}}(e^{\mathrm{pr}}(\mu,t^{k-1}),e^{\mathrm{pr}}(\mu,t^{k-1});\mu),$$

and $\rho = 1$, we obtain

$$\begin{aligned} 2\ m^{\frac{1}{2}}(e^{\mathrm{pr}}(\mu,t^k),e^{\mathrm{pr}}(\mu,t^k);\mu)\ & m^{\frac{1}{2}}(e^{\mathrm{pr}}(\mu,t^{k-1}),e^{\mathrm{pr}}(\mu,t^{k-1});\mu) \\ &\le m(e^{\mathrm{pr}}(\mu,t^{k-1}),e^{\mathrm{pr}}(\mu,t^{k-1});\mu) + m(e^{\mathrm{pr}}(\mu,t^k),e^{\mathrm{pr}}(\mu,t^k);\mu); \end{aligned} \quad (4.70)$$

and second, choosing $c = \varepsilon^{\rm pr}_{N_{\rm pr}}(\mu, t^k)$, $d = \|e^{\rm pr}(\mu, t^k)\|_Y$, and $\rho = \hat{\alpha}_a(\mu)^{\frac{1}{2}}$ we have

$$2\ \varepsilon^{\rm pr}_{N_{\rm pr}}(\mu, t^k)\ \|e^{\rm pr}(\mu, t^k)\|_Y \leq \frac{1}{\hat{\alpha}_a(\mu)}\ \varepsilon^{\rm pr}_{N_{\rm pr}}(\mu, t^k)^2 + \hat{\alpha}_a(\mu)\ \|e^{\rm pr}(\mu, t^k)\|^2_Y. \tag{4.71}$$

Combining (4.68), (4.70), and (4.71), and invoking (4.7) and (4.57), we obtain

$$\begin{aligned} m(e^{\rm pr}(\mu, t^k), e^{\rm pr}(\mu, t^k); \mu) - m(e^{\rm pr}(\mu, t^{k-1}), e^{\rm pr}(\mu, t^{k-1}); \mu) \\ + \Delta t\ a(e^{\rm pr}(\mu, t^k), e^{\rm pr}(\mu, t^k); \mu) \leq \frac{\Delta t}{\hat{\alpha}_a(\mu)}\ \varepsilon^{\rm pr}_{N_{\rm pr}}(\mu, t^k)^2, \quad \forall\, k \in \mathbb{K}, \end{aligned} \tag{4.72}$$

We now perform the sum from $k' = 1$ to k and recall that $e^{\rm pr}(\mu, t^0) = 0$, leading to

$$\begin{aligned} m(e^{\rm pr}(\mu, t^k), e^{\rm pr}(\mu, t^k); \mu) + \sum_{k'=1}^{k} \Delta t\ a(e^{\rm pr}(\mu, t^{k'}), e^{\rm pr}(\mu, t^{k'}); \mu) \\ \leq \frac{\Delta t}{\hat{\alpha}_a(\mu)} \sum_{k'=1}^{k} \varepsilon^{\rm pr}_{N_{\rm pr}}(\mu, t^{k'})^2, \quad \forall\, k \in \mathbb{K}, \end{aligned} \tag{4.73}$$

which is the result stated in Proposition 6. □

Dual Variable

Before proceeding with the error bounds for the dual variable we have to pay special attention to the final condition of the dual problem. The primal error at time zero, $e^{\rm pr}(\mu, t^0)$, vanishes (for our zero initial conditions) and therefore does not contribute to the error bound. For the dual problem, however, the error at the final time t^{K+1}, $e^{\rm du}(\mu, t^{K+1}) \equiv \Psi(\mu, t^{K+1}) - \Psi_N(\mu, t^{K+1})$ is – in general – nonzero since $\Psi(\mu, t^{K+1})$ is not necessarily a member of $W^{\rm du}_{N_{\rm du}}$. Instead, we obtain from (4.13) that $e^{\rm du}(\mu, t^{K+1})$ satisfies

$$m(v, e^{\rm du}(\mu, t^{K+1}); \mu) = R^{\Psi_f}(v; \mu), \quad \forall\, v \in Y, \tag{4.74}$$

where

$$R^{\Psi_f}(v; \mu) \equiv \ell(v) - m(v, \Psi_N(\mu, t^{K+1}); \mu), \quad \forall\, v \in Y, \tag{4.75}$$

is the residual associated to the final condition. It can be shown that $e^{\rm du}(\mu, t^{K+1})$ satisfies the following bound [91, 123].

Lemma 7. *The error $e^{\rm du}(\mu, t^{K+1}) \equiv \Psi(\mu, t^{K+1}) - \Psi_N(\mu, t^{K+1})$ is bounded by*

$$\|e^{\rm du}(\mu, t^{K+1})\|_X \leq \Delta^{\Psi_f}_{N_{\rm du}}(\mu) \equiv \frac{\varepsilon^{\Psi_f}_{N_{\rm du}}(\mu)}{\hat{\alpha}_m(\mu)}\ , \tag{4.76}$$

where

$$\varepsilon^{\Psi_f}_{N_{\rm du}}(\mu) \equiv \sup_{v \in Y} \frac{R^{\Psi_f}(v; \mu)}{\|v\|_X} \tag{4.77}$$

is the dual norm of the residual associated to the final condition. □

It directly follows from Lemma 7 and (4.74) that

$$\begin{aligned} m(e^{\mathrm{du}}(\mu,t^{K+1}),e^{\mathrm{du}}(\mu,t^{K+1});\mu) &= R^{\Psi_f}(e^{\mathrm{du}}(\mu,t^{K+1});\mu) \le \varepsilon^{\Psi_f}_{N_{\mathrm{du}}}(\mu)\, \|e^{\mathrm{du}}(\mu,t^{K+1})\|_X \\ &\le \hat{\alpha}_m(\mu)\, {\Delta^{\Psi_f}_{N_{\mathrm{du}}}(\mu)}^2 . \end{aligned} \tag{4.78}$$

Note that for the special case in which the bilinear form m is parameter-independent, we can guarantee that $\Psi(\mu,t^{K+1})$ is a member of $W^{\mathrm{du}}_{N_{\mathrm{du}}}$ and thus $e^{\mathrm{du}}(\mu,t^{K+1})$ is identically zero.

We are now ready to prove the bounding property for the dual problem.

Proposition 7. *Let $e^{\mathrm{du}}(\mu,t^k) \equiv \Psi(\mu,t^k) - \Psi_N(\mu,t^k)$ be the error in the dual variable and define*

$$|||v(\mu,t^k)|||^{\mathrm{du}} \equiv \left(m(v(\mu,t^k),v(\mu,t^k);\mu) + \sum_{k'=k}^{K} a(v(\mu,t^{k'}),v(\mu,t^{k'});\mu)\ \Delta t \right)^{\frac{1}{2}} . \tag{4.79}$$

The error in the dual variable is then bounded by

$$|||e^{\mathrm{du}}(\mu,t^k)|||^{\mathrm{du}} \le \Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{4.80}$$

where the error bound $\Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k)$ is defined as

$$\Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k) \equiv \left(\frac{\Delta t}{\hat{\alpha}_a(\mu)} \sum_{k'=k}^{K} {\varepsilon^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^{k'})}^2 + \hat{\alpha}_m(\mu)\, {\Delta^{\Psi_f}_{N_{\mathrm{du}}}(\mu)}^2 \right)^{\frac{1}{2}} , \tag{4.81}$$

and $\varepsilon^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k)$ is the dual norm of the dual residual defined in (4.60).

Proof. We immediately derive from (4.12) and (4.61) that $e^{\mathrm{du}}(\mu,t^k) = \Psi(\mu,t^k) - \Psi_N(\mu,t^k)$ satisfies

$$m(v,e^{\mathrm{du}}(\mu,t^k);\mu) + \Delta t\ a(v,e^{\mathrm{du}}(\mu,t^k);\mu) = m(v,e^{\mathrm{du}}(\mu,t^{k+1});\mu) + \Delta t\ R^{\mathrm{du}}(v;\mu,t^k), \quad \forall v \in Y,\ \forall k \in \mathbb{K}, \tag{4.82}$$

with final condition $m(v,e^{\mathrm{du}}(\mu,t^{K+1});\mu) = R^{\Psi_f}(v;\mu)$, $\forall\, v \in Y$. Choosing $v = e^{\mathrm{du}}(\mu,t^k)$, invoking the Cauchy-Schwarz inequality, and applying (4.60) we obtain

$$\begin{aligned} & m(e^{\mathrm{du}}(\mu,t^k),e^{\mathrm{du}}(\mu,t^k);\mu) + \Delta t\ a(e^{\mathrm{du}}(\mu,t^k),e^{\mathrm{du}}(\mu,t^k);\mu) \\ &\quad \le m^{\frac{1}{2}}(e^{\mathrm{du}}(\mu,t^{k+1}),e^{\mathrm{du}}(\mu,t^{k+1});\mu)\ m^{\frac{1}{2}}(e^{\mathrm{du}}(\mu,t^k),e^{\mathrm{du}}(\mu,t^k);\mu) \\ &\qquad + \Delta t\ \varepsilon^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k)\ \|e^{\mathrm{du}}(\mu,t^k)\|_Y, \forall\, k \in \mathbb{K}. \end{aligned} \tag{4.83}$$

We now apply (4.69) twice: first, with

$$\begin{aligned} c &= m^{\frac{1}{2}}(e^{\mathrm{du}}(\mu,t^{k+1}),e^{\mathrm{du}}(\mu,t^{k+1});\mu), \\ d &= m^{\frac{1}{2}}(e^{\mathrm{du}}(\mu,t^k),e^{\mathrm{du}}(\mu,t^k);\mu), \end{aligned}$$

and $\rho = 1$; and second, with $c = \varepsilon^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k)$, $d = \|e^{\mathrm{du}}(\mu,t^k)\|_Y$, and $\rho = \hat{\alpha}_a(\mu)^{\frac{1}{2}}$. Invoking (4.7) and (4.57), we arrive at

$$m(e^{\mathrm{du}}(\mu,t^k),e^{\mathrm{du}}(\mu,t^k);\mu) - m(e^{\mathrm{du}}(\mu,t^{k+1}),e^{\mathrm{du}}(\mu,t^{k+1});\mu) \\ + \Delta t\ a(e^{\mathrm{du}}(\mu,t^k),e^{\mathrm{du}}(\mu,t^k);\mu) \le \frac{\Delta t}{\hat{\alpha}_a(\mu)}\ \varepsilon^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k)^2, \quad \forall\, k \in \mathbb{K}, \quad (4.84)$$

We now perform the sum from $k' = k$ to K and invoke (4.78) to obtain

$$m(e^{\mathrm{du}}(\mu,t^k),e^{\mathrm{du}}(\mu,t^k);\mu) + \sum_{k'=k}^{K} \Delta t\ a(e^{\mathrm{du}}(\mu,t^{k'}),e^{\mathrm{du}}(\mu,t^{k'});\mu) \\ \le \frac{\Delta t}{\hat{\alpha}_a(\mu)} \sum_{k'=k}^{K} \varepsilon^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^{k'})^2 + \hat{\alpha}_m(\mu)\,\Delta^{\Psi_f}_{N_{\mathrm{du}}}(\mu)^2, \quad \forall\, k \in \mathbb{K}, \quad (4.85)$$

which is the result stated in Proposition 7. □

Output Bound

Finally, the error bound for the output estimate is given in the following proposition.

Proposition 8. *Let the output of interest, $s(\mu,t^k)$, and the reduced-basis output estimate, $s_N(\mu,t^k)$, be given by*

$$s(\mu,t^k) = \ell(y(\mu,t^k)), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \quad (4.86)$$

and

$$s_N(\mu,t^k) = \ell(y_N(\mu,t^k)) + \sum_{k'=1}^{k} R^{\mathrm{pr}}(\Psi_N(\mu,t^{K-k+k'});\mu,t^{k'})\,\Delta t, \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \quad (4.87)$$

respectively. The error in the output of interest is then bounded by

$$|s(\mu,t^k) - s_N(\mu,t^k)| \le \Delta^s_N(\mu,t^k), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \quad (4.88)$$

where the output bound $\Delta^s_N(\mu,t^k)$ is defined as

$$\Delta^s_N(\mu,t^k) \equiv \Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^k)\ \Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^{K-k+1}), \quad (4.89)$$

and $\Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^k)$ and $\Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k)$ are defined in Propositions 6 and 7, respectively.

Proof. To begin, we recall the definition of the dual problem for the output at time t^L, $L \in \mathbb{K}$, given by

$$m(v,\psi_L(\mu,t^k);\mu) + \Delta t\ a(v,\psi_L(\mu,t^k);\mu) = m(v,\psi_L(\mu,t^{k+1});\mu), \\ \forall\, v \in Y,\ (K \ge)L \ge k \ge 1, \quad (4.90)$$

with final condition $m(v,\psi_L(\mu,t^{L+1});\mu) \equiv \ell(v)$, $\forall\, v \in Y$. We now choose $v = e^{\mathrm{pr}}(\mu,t^k) =$

$y(\mu, t^k) - y_N(\mu, t^k)$ in (4.90) and sum from $k = 1$ to L, to obtain

$$\sum_{k'=1}^{L} m(e^{\mathrm{pr}}(\mu, t^{k'}), \psi_L(\mu, t^{k'}) - \psi_L(\mu, t^{k'+1}); \mu) + \sum_{k'=1}^{L} \Delta t \; a(e^{\mathrm{pr}}(\mu, t^{k'}), \psi_L(\mu, t^{k'}); \mu) = 0. \tag{4.91}$$

This equation can be rewritten in the form

$$\sum_{k'=1}^{L} m(e^{\mathrm{pr}}(\mu, t^{k'}) - e^{\mathrm{pr}}(\mu, t^{k'-1}), \psi_L(\mu, t^{k'}); \mu) - m(e^{\mathrm{pr}}(\mu, t^{L}), \psi_L(\mu, t^{L+1}); \mu) + \sum_{k'=1}^{L} \Delta t \; a(e^{\mathrm{pr}}(\mu, t^{k'}), \psi_L(\mu, t^{k'}); \mu) = 0, \tag{4.92}$$

where we used the fact that $e^{\mathrm{pr}}(\mu, t^0) = 0$. We now note from the final condition of the dual problem that $m(e^{\mathrm{pr}}(\mu, t^L), \psi_L(\mu, t^{L+1}); \mu) = \ell(e^{\mathrm{pr}}(\mu, t^L))$ to obtain

$$\ell(e^{\mathrm{pr}}(\mu, t^L)) = \sum_{k'=1}^{L} m(e^{\mathrm{pr}}(\mu, t^{k'}) - e^{\mathrm{pr}}(\mu, t^{k'-1}), \psi_L(\mu, t^{k'}); \mu) + \sum_{k'=1}^{L} \Delta t \; a(e^{\mathrm{pr}}(\mu, t^{k'}), \psi_L(\mu, t^{k'}); \mu). \tag{4.93}$$

We next choose $v = \psi_L(\mu, t^k)$ in the error equation for the primal variable, (4.67), and sum from $k = 1$ to L, to find

$$\sum_{k'=1}^{L} m(e^{\mathrm{pr}}(\mu, t^{k'}) - e^{\mathrm{pr}}(\mu, t^{k'-1}), \psi_L(\mu, t^{k'}); \mu) + \sum_{k'=1}^{L} \Delta t \; a(e^{\mathrm{pr}}(\mu, t^{k'}), \psi_L(\mu, t^{k'}); \mu) = \sum_{k'=1}^{L} R^{\mathrm{pr}}(\psi_L(\mu, t^{k'}); \mu, t^{k'}) \; \Delta t. \tag{4.94}$$

From (4.93) and (4.94) we thus obtain

$$\begin{aligned} \ell(e^{\mathrm{pr}}(\mu, t^L)) &= \sum_{k'=1}^{L} R^{\mathrm{pr}}(\psi_L(\mu, t^{k'}); \mu, t^{k'}) \; \Delta t. && (4.95) \\ &= \sum_{k'=1}^{L} R^{\mathrm{pr}}(\Psi(\mu, t^{K-L+k'}); \mu, t^{k'}) \; \Delta t. && (4.96) \end{aligned}$$

From the definition of $s(\mu, t^k)$ and $s_N(\mu, t^k)$, and (4.96) we now obtain

$$s(\mu, t^k) - s_N(\mu, t^k) = \ell(e^{\mathrm{pr}}(\mu, t^k)) - \sum_{k'=1}^{k} R^{\mathrm{pr}}(\Psi_N(\mu, t^{K-k+k'}); \mu, t^{k'})\ \Delta t \tag{4.97}$$

$$= \sum_{k'=1}^{k} R^{\mathrm{pr}}(\Psi(\mu, t^{K-k+k'}) - \Psi_N(\mu, t^{K-k+k'}); \mu, t^{k'})\ \Delta t \tag{4.98}$$

$$= \sum_{k'=1}^{k} R^{\mathrm{pr}}(e^{\mathrm{du}}(\mu, t^{K-k+k'}); \mu, t^{k'})\ \Delta t. \tag{4.99}$$

Invoking (4.59) and the Cauchy-Schwarz inequality we arrive at

$$|s(\mu, t^k) - s_N(\mu, t^k)| \leq \sum_{k'=1}^{k} \varepsilon^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu, t^{k'})\ \|e^{\mathrm{du}}(\mu, t^{K-k+k'})\|_Y\ \Delta t \tag{4.100}$$

$$\leq \left(\sum_{k'=1}^{k} \varepsilon^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu, t^{k'})^2\ \Delta t\right)^{\frac{1}{2}} \left(\sum_{k'=1}^{k} \|e^{\mathrm{du}}(\mu, t^{K-k+k'})\|^2_Y\ \Delta t\right)^{\frac{1}{2}} . \tag{4.101}$$

Let us first bound the second term on the right hand side. From (4.7) and the fact that $\hat{\alpha}_a(\mu) \leq \alpha_a(\mu),\ \forall\, \mu \in \mathcal{D}$, we obtain

$$\|e^{\mathrm{du}}(\mu, t^{K-k+k'})\|^2_Y \leq \frac{1}{\hat{\alpha}_a(\mu)} a(e^{\mathrm{du}}(\mu, t^{K-k+k'}), e^{\mathrm{du}}(\mu, t^{K-k+k'}); \mu), \qquad \forall\, \mu \in \mathcal{D}. \tag{4.102}$$

Performing the sum from $k' = 1$ to k leads to

$$\sum_{k'=1}^{k} \|e^{\mathrm{du}}(\mu, t^{K-k+k'})\|^2_Y\ \Delta t \leq \frac{1}{\hat{\alpha}_a(\mu)} \sum_{k'=1}^{k} a(e^{\mathrm{du}}(\mu, t^{K-k+k'}), e^{\mathrm{du}}(\mu, t^{K-k+k'}); \mu)\ \Delta t \tag{4.103}$$

$$= \frac{1}{\hat{\alpha}_a(\mu)} \sum_{k'=K-k+1}^{K} a(e^{\mathrm{du}}(\mu, t^{k'}), e^{\mathrm{du}}(\mu, t^{k'}); \mu)\ \Delta t \tag{4.104}$$

$$\leq \frac{1}{\hat{\alpha}_a(\mu)} \left(\sum_{k'=K-k+1}^{K} a(e^{\mathrm{du}}(\mu, t^{k'}), e^{\mathrm{du}}(\mu, t^{k'}); \mu)\ \Delta t \right. \tag{4.105}$$

$$\left. +\ m(e^{\mathrm{du}}(\mu, t^{K-k+1}), e^{\mathrm{du}}(\mu, t^{K-k+1}); \mu) \right) \tag{4.106}$$

$$= \frac{1}{\hat{\alpha}_a(\mu)} \left(|||e^{\mathrm{du}}(\mu, t^{K-k+1})|||^{\mathrm{du}} \right)^2, \tag{4.107}$$

where the second inequality follows from the coercivity of $m(\cdot, \cdot; \mu)$ and the last equality from the definition (4.79) of the $|||\cdot|||^{\mathrm{du}}$-norm. Finally, inserting (4.107) into (4.101) and invoking (4.80) and (4.66), we obtain

$$|s(\mu, t^k) - s_N(\mu, t^k)| \leq \Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu, t^k)\ \Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu, t^{K-k+1}), \tag{4.108}$$

which is the result stated in Proposition 8. □

4.4.3 An Alternative (Simpler) Output Bound

We note from (4.89) that, using the dual formulation, we obtain a square effect in the output bound since $\Delta_N^s(\mu, t^k)$ is the product of the primal and dual error bounds. We will observe this effect also in the numerical results in Section 4.6.2. This desirable effect, however, comes with the additional complexity and computational effort of the dual formulation[1]. To avoid this additional effort, we can also define a "simple" output approximation, $\hat{s}_N(\mu, t^k)$, and corresponding output bound, $\hat{\Delta}_N^s(\mu, t^k)$, which does not require the dual formulation and — in certain cases — still results in a satisfactory convergence rate of the output approximation and sharpness of the output bound. We state the result in

Proposition 9. *Let the output of interest, $s(\mu, t^k)$, and the (simple) output estimate, $\hat{s}_N(\mu, t^k)$, be given by*

$$s(\mu, t^k) = \ell(y(\mu, t^k)), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{4.109}$$

and

$$\hat{s}_N(\mu, t^k) = \ell(y_N(\mu, t^k)), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{4.110}$$

respectively. The error in the output of interest is then bounded by

$$|s(\mu, t^k) - \hat{s}_N(\mu, t^k)| \le \hat{\Delta}_N^s(\mu, t^k), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{4.111}$$

where the (simple) output bound $\hat{\Delta}_N^s(\mu, t^k)$ is defined as

$$\hat{\Delta}_N^s(\mu, t^k) \equiv \frac{1}{\hat{\alpha}_m(\mu)} \sup_{v \in Y} \frac{\ell(v)}{\|v\|_X}\ \Delta_{N_{\rm pr}}^{\rm pr}(\mu, t^k), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{4.112}$$

and $\hat{\alpha}_m(\mu)$ is the lower bound for the coercivity constant $\alpha_m(\mu)$ defined in (4.58).

Proof. From (4.109) and (4.110) we obtain

$$\begin{aligned} |s(\mu, t^k) - \hat{s}_N(\mu, t^k)| &= |\ell(y(\mu, t^k)) - \ell(y_{N,M}(\mu, t^k))| \\ &= |\ell(e^{\rm pr}(\mu, t^k))| \le \sup_{v \in Y} \frac{\ell(v)}{\|v\|_X} \|e^{\rm pr}(\mu, t^k)\|_X, \end{aligned} \tag{4.113}$$

from which the result immediately follows since $\hat{\alpha}_m(\mu)\, \|e^{\rm pr}(\mu, t^k)\|_X \le m(e^{\rm pr}(\mu, t^k), e^{\rm pr}(\mu, t^k); \mu)^{1/2} \le \Delta_{N_{\rm pr}}^{\rm pr}(\mu, t^k)$, $\forall\, \mu \in \mathcal{D}$, $\forall\, k \in \mathbb{K}$. We note that we can only bound the X-norm of the error $e^{\rm pr}(\mu, t^k)$ by the error bound $\Delta_{N_{\rm pr}}^{\rm pr}(\mu, t^k)$ at any given time t^k. We thus obtain the requirement that $\|\ell\|_{X'}$ has to be bounded, otherwise the upper bound in (4.113) does not exist. □

We note that — given $\underline{y}_N(\mu, t^k)$ — the computational cost to evaluate $\hat{s}_N(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, is only $O(KN_{\rm pr})$ whereas the computational cost to evaluate $s_N(\mu, t^k)$, $\forall k \in \mathbb{K}$, using the primal-dual formulation is $O(K(K+1)N_{\rm pr}N_{du})$. If K is of the order of $N_{\rm pr}$ or $N_{\rm du}$ or if several outputs of interest have to be evaluated — which is often the case in practice — employing the simple bound

[1] In actual practice, of course, the primal and dual problem should be solved in parallel; they only "interact" through the residual correction term for the output estimate.

is computationally more efficient; we return to this discussion, and a comparison of both output estimates and bounding properties, in Sections 4.7.2 and 4.8.5.

4.4.4 Offline-Online Computational Procedure

We now turn to the development of offline-online computational procedures for the calculation of $\Delta^{\rm pr}_{N_{\rm pr}}(\mu, t^k)$, $\Delta^{\rm du}_{N_{\rm du}}(\mu, t^k)$, and $\Delta^s_N(\mu, t^k)$. The necessary computations for the offline and online stages — by construction rather similar to the elliptic case [91] — are detailed in Appendix A. Here, we only summarize the computational costs involved.

The computational cost in the offline stage is (to leading order) $O((N_{\rm pr,max}+N_{\rm du,max})(Q_a+Q_m))$ solutions of the underlying "truth" finite element approximation and $O((N^2_{\rm pr,max} + N^2_{\rm du,max})(Q^2_a + Q_aQ_m + Q^2_m))$ $\mathcal{N}$-inner products; the storage requirement is $O((N^2_{\rm pr,max} + N^2_{\rm du,max})(Q^2_a + Q_aQ_m + Q^2_m))$. In the online stage — given a new parameter value μ and associated reduced-basis solutions $\underline{y}_N(\mu, t^k)$ and $\underline{\Psi}_N(\mu, t^k)$, $\forall\, k \in \mathbb{K}$ — the computational cost to evaluate $\Delta^s_N(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, is $O(K(N^2_{\rm pr} + N^2_{\rm du})(Q^2_a + Q_aQ_m + Q^2_m))$. Thus, all online calculations needed are *independent* of $\mathcal{N}$.

4.5 Adaptive Sampling Procedure

Our error estimation procedures not only allow us to determine the accuracy of the output estimate but also to pursue a more rational construction of the sampling set $S^{\rm pr}_{N_{\rm pr}}$ (and $S^{\rm du}_{N_{\rm du}}$) and associated reduced-basis space $W^{\rm pr}_{N_{\rm pr}}$ (and $W^{\rm du}_{N_{\rm du}}$). The crucial point is that the error bound $\Delta^{\rm pr}_{N_{\rm pr}}(\mu, t^k)$ (respectively, $\Delta^{\rm du}_{N_{\rm du}}(\mu, t^k)$) is an accurate surrogate for the true error $|||y(\mu, t^k) - y_N(\mu, t^k)|||^{\rm pr}$ (respectively, $|||\Psi(\mu, t^k) - \Psi_N(\mu, t^k)|||^{\rm du}$) that can be very efficiently calculated in the limit of many queries. We may thus perform an exhaustive search over the parameter-time space to find the best sample sets $S^{\rm pr}_{N_{\rm pr}}$ (and $S^{\rm du}_{N_{\rm du}}$): in essence, a snapshot procedure in which only the snapshots retained must actually be evaluated.

The sampling procedure for the primal and dual problem is very similar; we thus focus on the primal problem and comment only briefly on the dual problem. Also recall that the control input sequence $u(t^k)$ is assumed to be known — either a prescribed function or the impulse (see Section 4.2.3).

4.5.1 Greedy Algorithm

To begin, we assume that we are given a sample set $S^{\rm pr}_{N_{\rm pr}}$ and associated reduced-basis space $W^{\rm pr}_{N_{\rm pr}}$. We then choose the next sampling point based on the following two steps: first, we search in parameter space and select the parameter value μ^* for which $\Delta^{\rm pr}_{N_{\rm pr}}(\mu, t^K)$ is maximized,[2]

$$\mu^* = \arg \max_{\mu \in \Xi_F} \Delta^{\rm pr}_{N_{\rm pr}}(\mu, t^K); \tag{4.114}$$

we then select the timestep t^{k^*} for which the temporal rate of change of $\Delta^{\rm pr}_{N_{\rm pr}}(\mu, t^k)$ is largest,

$$t^{k^*} = \arg \max_{t^k \in \mathbb{I}} (\Delta^{\rm pr}_{N_{\rm pr}}(\mu^*, t^k) - \Delta^{\rm pr}_{N_{\rm pr}}(\mu^*, t^{k-1})). \tag{4.115}$$

[2] Note that $\Delta^{\rm pr}_{N_{\rm pr}}(\mu, t^k)$ is a nondecreasing sequence in k and the maximum therefore always occurs at $k = K$.

Here, $\Xi_{\mathrm{F}} \in (\mathcal{D})^{n_F}$ is a random parameter test sample of size n_F; since the marginal cost to evaluate $\Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu, t^K)$ is small, the random sample can be very large, i.e., $n_F \gg 1$. We then append $\tilde{\mu}^* = (\mu^*, t^{k^*})$ to $S^{\mathrm{pr}}_{N_{\mathrm{pr}}}$ to form $S^{\mathrm{pr}}_{N_{\mathrm{pr}}+1}$, and hence $W^{\mathrm{pr}}_{N_{\mathrm{pr}}+1}$, and update the reduced-basis approximation and error estimation procedure accordingly. We repeat this process until the maximum error bound at the final time t^K over Ξ_F is less than a desired (most stringent anticipated) error tolerance $\epsilon_{\mathrm{tol,min}}$: this determines $N_{\mathrm{pr,max}}$.

We note that our sample selection process is not truly optimal: given the prescribed error tolerance $\epsilon_{\mathrm{tol,min}}$, there are undoubtedly parameter samples with fewer than $N_{\mathrm{pr,max}}$ points that suffice. Unfortunately, the latter can only be identified by prohibitively (combinatorially) expensive calculation, and thus we must resort to heuristic approaches. Our particular heuristic, described above, is of the "greedy" [21] variety: we focus on just the next sample point and just the currently largest error with no regard to more global objectives. In actual practice, as we shall see in Section 4.6.1, this *carpe diem* philosophy indeed leads to good samples; but we are not able to characterize the degree of sub-optimality relative to truly optimal samples.

We elaborate on three refinements. First, we invoke a *normalized* error bound for the sampling procedure to avoid dependence on the magnitude of the forcing term (the control input): in particular, we normalize with respect to $|||y_N(\mu, t^K)|||^{\mathrm{pr}}$, which can be calculated online in only $O(KN^2_{\mathrm{pr}})$ operations. Second, we are careful to orthonormalize the basis functions ζ^{pr}_n with respect to the $(\cdot,\cdot)_Y$ inner product by (say) Gram-Schmidt: this guarantees, for example, that the condition number of the reduced-basis matrix $A_N(\mu)$ is bounded from above by $\frac{\gamma^0_a}{\alpha^0_a}$ for all N. Third, as regards initialization, we simply set $\mu_1 = \mu_{\min}$ and choose $\zeta^{\mathrm{pr}}_1 = y(\mu_1, t^k) \neq 0$ for some small k, i.e., we select $\zeta^{\mathrm{pr}}_1 = y(\mu_1, t^1)$ for $u(t^1) \neq 0$. This choice has a simple justification: the adaptive sampling procedure is likely to select samples corresponding to transient behaviour which, in most cases — and certainly for the impulse input — occurs during the first few timesteps (also see the numerical results in Figure 4-6).

The Dual Problem

From our previous discussion in Section 4.4.2 we know that the dual error at time t^{K+1}, $e^{\mathrm{du}}(\mu, t^{K+1})$, may not necessarily be zero if the bilinear form m depends on μ. We thus need to guarantee that the final condition, $\Psi(\mu, t^{K+1})$, is sufficiently represented in $W^{\mathrm{du}}_{N_{\mathrm{du}}}$ for all $\mu \in \mathcal{D}$. We can guarantee this by either (i) considering only the "elliptic" problem 4.13 and generating a basis for $\Psi(\mu, t^{K+1})$ *first* before proceeding with the greedy algorithm described above, i.e., until $\Delta^{\Psi_f}_{N_{\mathrm{du}}}(\mu) \leq \epsilon^{\mathrm{ell}}_{\mathrm{tol,min}}$, where $\epsilon^{\mathrm{ell}}_{\mathrm{tol,min}}$ is a desired error tolerance; or (ii) simply combining these two procedures, i.e., for each new μ^* selected, we first check if $\Delta^{\Psi_f}_{N_{\mathrm{du}}}(\mu) \leq \epsilon^{\mathrm{ell}}_{\mathrm{tol,min}}$ is satisfied — if it is satisfied, we proceed with 4.115, if it is not satisfied, we append (μ^*, t^{K+1}) to $S^{\mathrm{du}}_{N_{\mathrm{du}}}$ and continue with the search for the next sample.

4.5.2 Extensions

The extension of the adaptive procedure to the case of multiple control inputs is straightforward. If the control inputs are given, the sampling algorithm can directly be applied; however, if the control inputs are unknown, e.g., in the optimal control context, we can simply adjust the impulse approach. We begin with an impulse in the first control input — all other control inputs are set to zero — and generate the basis using the standard algorithm. When the adaptive procedure terminates, we set the first control input to zero and the second control input to the impulse and restart the adaptive sampling — initialized to the already existing sample set $S_{N_{\text{pr}}}^{\text{pr}}$ and associated reduced-basis space $W_{N_{\text{pr}}}^{\text{pr}}$. In effect, the multiple control input scenario simply adds an "outer loop" to the standard algorithm.

We may also consider nonzero initial conditions. In the case of a parameter-independent nonzero initial condition, we simply set $\zeta_1^{\text{pr}} = y_0$ and apply the standard algorithm. For (affinely) parameter-dependent initial conditions $y_0(\mu)$ we may write

$$y_0(\mu) = \sum_{q=1}^{Q_y} \Theta_y^q(\mu)\; y_0^q, \quad \forall\, \mu \in \mathcal{D}, \tag{4.116}$$

where the $y_0^q \in Y$, $1 \leq q \leq Q_y$, are given members of Y; only the functions $\Theta_y^q(\mu) : \mathcal{D} \rightarrow \boldsymbol{R}$, $1 \leq q \leq Q_y$ depend on μ. In this case we initialize $W_{N_{\text{pr}}}^{\text{pr}}$ to $\text{span}_{1 \leq q \leq Q_y}\{y_0^q\}$, and then apply the standard sampling algorithm of Section 4.5.1 (with initial condition $y_0(\mu)$). In both these cases we retain the condition $e^{\text{pr}}(\mu, t^0) = 0$.

Note that the case of multiple control inputs with nonzero initial conditions is a straightforward combination of the previous two cases. We first generate a reduced-basis for the nonzero initial condition (with zero control input); given this basis, we then further adapt to the control inputs using the impulse approach (for zero initial condition).

4.5.3 Backup Procedure

At this point we need to clarify that our proposed adaptive sampling procedure is not foolproof, i.e., the method can fail by possibly selecting a new sample point $\tilde{\mu}^* = (\mu^*, t^{k^*})$ which *already* is a member of $S_{N_{\text{pr}}}^{\text{pr}}$. In this case we cannot, of course, append $\tilde{\mu}^*$ to $S_{N_{\text{pr}}}^{\text{pr}}$ since this would directly result in a singular reduced-basis system. We note that this problem does not appear for elliptic problems because the true error and the error bound are (theoretically) zero for all $\mu \in S_N$. For parabolic problems, however, this is no longer the case due to the time-dependence. The reason for this is twofold: first, even the actual error $|||e^{\text{pr}}(\mu, t^k)|||^{\text{pr}}$ is nonzero for $\tilde{\mu} = (\mu, t^k) \in S_{N_{\text{pr}}}^{\text{pr}}$; and second, in our adaptive sampling procedure we base our decision on the error bound, $\Delta_{N_{\text{pr}}}^{\text{pr}}(\mu, t^k)$, which serves as a surrogate for the actual error — even if the temporal rate of change of $|||e^{\text{pr}}(\mu, t^k)|||^{\text{pr}}$ is small, the temporal rate of change of $\Delta_{N_{\text{pr}}}^{\text{pr}}(\mu, t^k)$ itself can be much larger and result in a "false" timestep decision t^{k^*} in (4.115).

If our standard procedure fails, we thus revert to a backup routine to select the timestep t^{k^*}. We first introduce a small random sample set $\Xi_{\text{test}}^{\mathbb{I}}$ in time. Then, for *each* t_{test}^j in $\Xi_{\text{test}}^{\mathbb{I}}$ we append $\tilde{\mu}_{\text{test}}^j = (\mu^*, t_{\text{test}}^j)$ to $S_{N_{\text{pr}}}^{\text{pr}}$ to form $S_{N_{\text{pr}}}^{\text{pr}\,j}$, and hence $W_{N_{\text{pr}}}^{\text{pr}\,j}$, update the reduced-basis approximation and error estimation procedure accordingly, and calculate the error bound $\Delta_{N_{\text{pr}}}^{\text{pr}\,j}(\mu^*, t^K)$ at the final time t^K. We then choose the timestep t_{test}^j that results in the largest reduction of the error bound, i.e., we set

$$t^{k^*} = \arg \min_{t^j \in \Xi_{\text{test}}^{\mathbb{I}}} \Delta_{N_{\text{pr}}}^{\text{pr}\,j}(\mu, t^K). \tag{4.117}$$

The idea behind this approach is simple: were we to append $\tilde{\mu}_{\text{test}}^j$ to $S_{N_{\text{pr}}}^{\text{pr}}$, the new error bound at the final time would be $\Delta_{N_{\text{pr}}}^{\text{pr}\,j}(\mu, t^K)$ — we simply select that timestep t_{test}^j in $\Xi_{\text{test}}^{\mathbb{I}}$ that reduces the error bound at t^K by the largest amount.

4.6 Results for Numerical Exercise 2

We now present numerical results for the example introduced in Section 4.2.4. We first define the inner product $(w,v)_Y \equiv \int_\Omega \nabla w \, \nabla v + 0.01 \int_{\partial\Omega_{\rm out}} w\, v + 0.001 \int_{\partial\Omega_{\rm in}} w\, v$, corresponding to (4.62) for $\mu_{\rm ref} = (0.01, 0.001)$; from the bilinear form a in Section 4.2.4 it follows that we may choose $\hat{\alpha}_a(\mu) = 1$ in (4.57). Note that the bilinear form m happens to be parameter-independent in this example, and thus $e^{\rm du}(\mu, t^{K+1}) = 0$ here. We thus have no (computational) need for $(\cdot,\cdot)_X$. We recall that the time interval is $\bar{I} = [0, 20]$, the timestep $\Delta t = 0.2$, and $K = 100$.

4.6.1 Adaptive Sampling Procedure

Before discussing the convergence properties we present numerical results for our adaptive sampling procedure. For purposes of illustration, we construct a reduced-basis space for the (one-)parameter set $\mathcal{D}_1 \equiv [0.01] \times [0.001, 0.1]$, i.e., we assume $\mu_{(1)} = 0.01$ is fixed. We initialize the procedure with $S_1^{\rm pr} = (\mu_{(2),\rm min} = 0.001, t^1)$ and set the desired error tolerance (for the primal energy norm) to $\epsilon_{\rm tol,min} = 1\,\mathrm{E}{-3}$. We plot and tabulate the resulting sample set $S_{N_{\rm pr}}^{\rm pr}$ in $\mu_{(2)}$-t^k space in Figure 4-6 — we need $N_{\rm pr} = 15$ basis functions to obtain the desired accuracy. We note that for this problem the adaptive sampling procedure selects all the samples on the $\mu_{(2)} = 0.001$ axis before selecting any other samples. Also, samples taken from only near the extreme parameter values (minimum and maximum) in $\mathcal{D}_1$ are sufficient to guarantee the desired tolerance everywhere in $\mathcal{D}_1$; in general, this is not the case.

4.6.2 Convergence Results

We now present convergence results for the full two-parameter numerical example. The primal and dual samples in $\tilde{\mathcal{D}} = \mathcal{D} \times \mathbb{I}$ are constructed according to the adaptive sampling procedure in Section 4.5; we obtain $N_{\rm pr,max} = 22$ and $N_{\rm du,max} = 21$ for $\epsilon_{\rm tol,min} = 1\,\mathrm{E}{-3}$. We first define the effectivity associated to the primal and dual error bounds as

$$\eta^{\rm pr}(\mu, t^k) \equiv \frac{\Delta_{N_{\rm pr}}^{\rm pr}(\mu, t^k)}{|||e^{\rm pr}(\mu, t^k)|||^{\rm pr}} \tag{4.118}$$

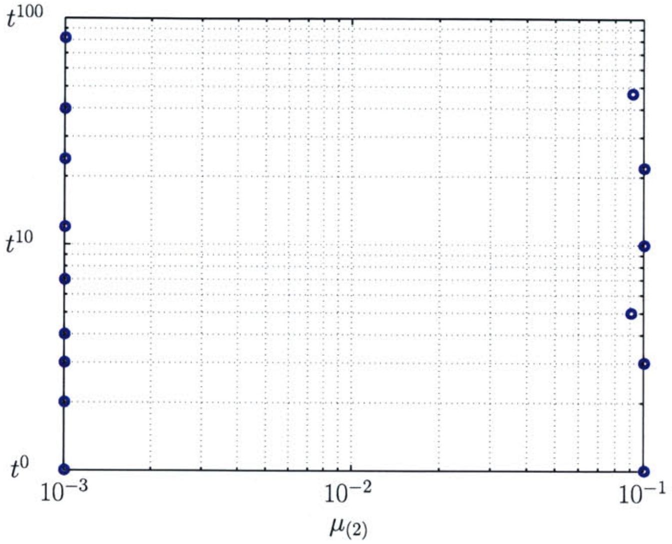
and

$$\eta^{\rm du}(\mu, t^k) \equiv \frac{\Delta_{N_{\rm du}}^{\rm du}(\mu, t^k)}{|||e^{\rm du}(\mu, t^k)|||^{\rm du}}, \tag{4.119}$$

respectively. Similarly, the effectivity for the output bound is defined as

$$\eta^{s}(\mu, t^k) \equiv \frac{\Delta_N^s(\mu, t^k)}{|s(\mu, t^k) - s_N(\mu, t^k)|}. \tag{4.120}$$

The effectivity serves as a measure of rigour and sharpness of the error bounds: we have $\eta^{\rm pr}(\mu, t^k) \geq 1$, $\forall\, \mu \in \mathcal{D}$, since $\Delta^{\rm pr}(\mu, t^k)$ is a true upper bound to the error in the $|||\cdot|||^{\rm pr}$-norm; and ideally we would like $\eta^{\rm pr}(\mu, t^k) \approx 1$, $\forall\, \mu \in \mathcal{D}$, so as to obtain a sharp bound for the error. (Similar arguments apply to the dual and to the output.)



n	μ_n^{pr}	k_n^{pr}
1	0.001	1
2	0.001	2
3	0.001	3
4	0.001	4
5	0.001	7
6	0.001	12
7	0.001	24
8	0.001	40
9	0.001	82
10	0.100	1
11	0.100	3
12	0.100	10
13	0.100	22
14	0.090	5
15	0.091	47

Figure 4-6: NE 2: Sample set $S_{N_{\text{pr}}}^{\text{pr}}$ for $\mathcal{D}_1 \equiv [0.01] \times [0.001, 0.1]$ and $N_{\text{pr}} = 15$.

In Table 4.1(a) we present, as a function of $N_{\rm pr}$ $(= N_{\rm du})$,$\epsilon^{\rm pr}_{\rm max,rel}$, $\Delta^{\rm pr}_{\rm max,rel}$, and $\overline{\eta}^{\rm pr}$: ,$\epsilon^{\rm pr}_{\rm max,rel}$ is the maximum over $\Xi_{\rm Test}$ of $|||y(\mu,t^k) - y_N(\mu,t^K)|||/|||y(\mu_y,t^K)|||$, $\Delta^{\rm pr}_{\rm max,rel}$ is the maximum over $\Xi_{\rm Test}$ of $\Delta^{\rm pr}_{N_{\rm pr}}(\mu,t^K)/|||y(\mu_y,t^K)|||$, $\overline{\eta}^{\rm pr}$ is the average over $\Xi_{\rm Test} \times \mathbb{I}$ of $\Delta^{\rm pr}_{N_{\rm pr}}(\mu,t^k)/|||y(\mu,t^k) - y_N(\mu,t^k)|||$. Here $\Xi_{\rm Test} \in (\mathcal{D})^{400}$ is a random input sample of size 400; $\mu_y \equiv \arg\max_{\mu\in\Xi_{\rm Test}} |||y(\mu,t^K)|||$. We also present in 4.1(b), as a function of $N_{\rm pr}$ $(= N_{\rm du})$,$\epsilon^{s}_{\rm max,rel}$, $\Delta^{s}_{\rm max,rel}$, and $\overline{\eta}^{s}$: $\epsilon^{s}_{\rm max,rel}$ is the maximum over $\Xi_{\rm Test}$ of $|s(\mu,t^K) - s_N(\mu,t^K)|/|s(\mu_s,t^K)|$, $\Delta^{s}_{\rm max,rel}$ is the maximum over $\Xi_{\rm Test}$ of $\Delta^s_N(\mu,t^K)/|s(\mu_s,t^K)|$, and $\overline{\eta}^s$ is the average over $\Xi_{\rm Test}$ of $\Delta^s_N(\mu,t_\eta(\mu))/|s(\mu,t_\eta(\mu)) - s_N(\mu,t_\eta(\mu))|$. Here $\mu_s \equiv \arg\max_{\mu\in\Xi_{\rm Test}} |s(\mu,t^K)|$ (note the output grows with time), and $t_\eta(\mu) \equiv \arg\max_{t^k\in\mathbb{I}}$ $|s(\mu,t^k) - s_N(\mu,t^k)|$. We observe very rapid convergence of the reduced-basis approximation. Furthermore, as we may expect, $\Delta^s_N(\mu,t^k)$ converges roughly as the square of $\Delta^{\rm pr}_{N_{\rm pr}}(\mu,t^k)$; we see that for only $N_{\rm pr} = N_{\rm du} = 8$ the error in the output is less than one percent. Also, the effectivities are very good: $O(1)$ for the primal error bound, and $O(10)$ for the output bound; note the latter are worse than the former as our bound cannot take into account any correlation between the primal and dual error. (We do not at present have good *a priori* upper bounds for the effectivities; see [91] for treatment of the elliptic case.)

In Table 4.2 we present, as a function of $N_{\rm pr}(= N_{\rm du})$, the online computational times to calculate $s_N(\mu,t^k)$ and $\Delta^s_N(\mu,t^k)$, $\forall\, k \in \mathbb{K}$. The values are normalized with respect to the computational time for the direct calculation of the truth approximation output $s(\mu,t^k) = \ell(u(\mu,t^k))$, $\forall\, k \in \mathbb{K}$. We note that even for the largest value of $N_{\rm pr}(= N_{\rm du})$ the calculation of $s_N(\mu,t^k)$ and $\Delta^s_N(\mu,t^k)$ is approximately 100 times faster than the direct calculation of $s(\mu,t^k)$. The actual average run-time to compute $s_N(\mu,t^k)$ and $\Delta^s_N(\mu,t^k)$ in MATLAB 6.5 on a 750 MHz Pentium III varies from 0.041 sec. (for $N_{\rm pr} = N_{\rm du} = 4$) to 0.061 sec. (for $N_{\rm pr} = N_{\rm du} = 20$). (The growth with $N_{\rm pr}$ is less than expected due to memory-access issues.) We emphasize that the reduced-basis entry does

not include the extensive offline computations — and is thus only meaningful in the real-time or many-query contexts.

We can now define lower and upper output bounds

$$s^-_N(\mu,t^k) \equiv s_N(\mu,t^k) - \Delta^s_N(\mu,t^k) \le s(\mu,t^k) \le s_N(\mu,t^k) + \Delta^s_N(\mu,t^k) \equiv s^+_N(\mu,t^k). \qquad (4.121)$$

We know that $s^+_N(\mu,t^k)$ (respectively, $s^-_N(\mu,t^k)$) are *certifiably* upper (respectively, lower) bounds for the true output $s(\mu,t^k)$ — see Proposition 8; that these bounds are accurate — see Table 4.1; and that these bounds may be evaluated very fast online — see Table 4.2. The bounds may thus serve to ensure a feasible design[3], a "good" design, and a fast design process or real-time decision [79].

4.7 AP I: Nondestructive Evaluation of Delamination

With the theory developed thus far we are ready to consider the reduced-basis approximation for the delamination problem introduced in Section 1.1.1. The sketch of the FRP reinforced concrete slab is shown in Figure 1-1. We first exploit symmetry and consider only the half-width of the slab ($x_1 \ge 0$) for our truth approximation. We also note that the geometry of the slab depends on the delamination width $w_{\rm del}$; we treat the geometric variation in an indirect way by performing an affine geometric mapping (see [120] for a detailed discussion of affine mappings) from the parameter

[3] For example, to honor an optimal-control constraint of the form $s(\mu,t^k) \le T_{\rm max}$ we may conservatively impose $s^+_N(\mu,t^k) \le T_{\rm max}$.

N_{pr}	$\epsilon^{\text{pr}}_{\text{max,rel}}$	$\Delta^{\text{pr}}_{\text{max,rel}}$	$\overline{\eta}^{\text{pr}}$
4	3.19 E−01	1.37 E+00	5.44
8	4.71 E−02	5.93 E−02	1.84
12	1.12 E−02	1.15 E−02	1.04
16	1.23 E−03	1.24 E−03	1.02
20	1.60 E−04	1.62 E−04	1.04

(a)

N_{pr}	$\epsilon^{s}_{\text{max,rel}}$	$\Delta^{s}_{\text{max,rel}}$	$\overline{\eta}^{s}$
4	1.40 E−02	2.15 E+00	97.5
8	1.19 E−03	1.07 E−02	49.9
12	1.59 E−04	7.42 E−04	5.95
16	1.73 E−06	1.22 E−05	16.8
20	3.10 E−08	1.30 E−07	22.5

(b)

Table 4.1: NE 2: Convergence rate and effectivities for the output, $N_{\text{pr}} = N_{\text{du}}$.

N_{pr}	$s_N(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$\Delta^s_N(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$s(\mu, t^k), \forall\, k \in \mathbb{K}$
4	3.00 E−03	3.11 E−03	1
8	3.59 E−03	3.20 E−03	1
12	4.19 E−03	3.28 E−03	1
16	4.77 E−03	3.36 E−03	1
20	5.57 E−03	3.48 E−03	1

Table 4.2: NE 2: Online computational times (normalized with respect to the time to solve for $s(\mu, t^k), \forall\, k \in \mathbb{K}$).

dependent solution domain to a fixed reference domain Ω with $\mu_{1,\text{ref}} = 5$, shown in Figure 4-7. The reference domain Ω, a typical point in which is (x_1, x_2), is then given by $\Omega \equiv [0, 30] \times [0, 11]$. By dividing Ω into 10 subdomains, Ω^i, $1 \le i \le 10$, we only have to consider the geometric variations in regions Ω^2, Ω^3, Ω^7, and Ω^8 — the remaining regions do not vary with the delamination width w_{del}. We define the outputs, i.e., the surface temperatures at the two measurement points, to be the average temperatures over the "fictitious" regions Ω_6 and Ω_{10} of size $|\Omega_6| = 0.125$ and $|\Omega_{10}| = 0.5$, respectively. We note that these regions are introduced for easier reference only, the affine mapping does not require a distinction between the regions Ω_9 and Ω_{10} and the regions Ω_5 and Ω_6, respectively. The delamination is indicated by the magenta horizontal line, Γ_{del}, between the domains Ω^1, Ω^2 and Ω^5, Ω^7, respectively. Note that the delamination is modeled as having zero width and homogeneous Neumann boundary conditions on the surface. We assume homogeneous Neumann boundary conditions on Γ_{N} and homogeneous Dirichlet boundary conditions on Γ_{D}.

We next identify the input parameter set[4] $\mu = (\mu_1, \mu_2) \equiv (w_{\text{del}}/2, \varkappa) \in \mathcal{D} \equiv [1, 10] \times [0.4, 1.8] \subset \mathbb{R}^{P=2}$, where $\varkappa \equiv k_{\text{FRP}}/k_{\text{C}}$, and derive the time-discrete weak form of the governing equations (1.1)-(1.10). The temperature distribution $T(\mu, t^k) \in Y$ in the FRP-concrete slab then satisfies (4.3) with initial condition $T(\mu, t^0) = 0$, where $Y \subset Y^e \equiv \{v | v \in H^1(\Omega), v = 0|_{\Gamma_{\text{D}}}\}$ is a linear finite element truth approximation subspace of dimension $\mathcal{N} = 5603$. The truth approximation mesh and a zoom on the delamination are shown in Figure 4-8. It can be shown that the bilinear and linear forms a, m, and b admit the affine representation (4.9)-(4.11) with $Q_a = 10$, $Q_m = 3$, $Q_b = 3$, respectively; we summarize these terms in detail in Appendix C. We also define the inner product $(w, v)_Y \equiv a(w, v; \mu_{\text{ref}})$ and $(w, v)_X \equiv m(w, v; \mu_{\text{ref}})$, corresponding to (4.62) and (4.63) for $\mu_{\text{ref}} = (5, 1.2)$, respectively. We immediately observe from the definitions of the a^q, m^q, and

[4] Because of symmetry, we consider only the half-width of the delamination as our parameter

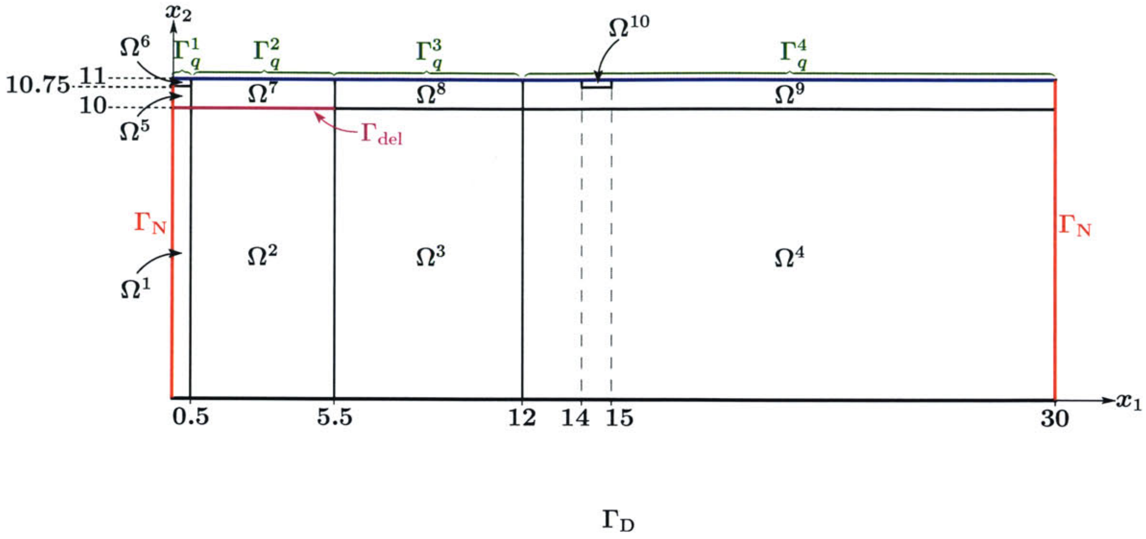
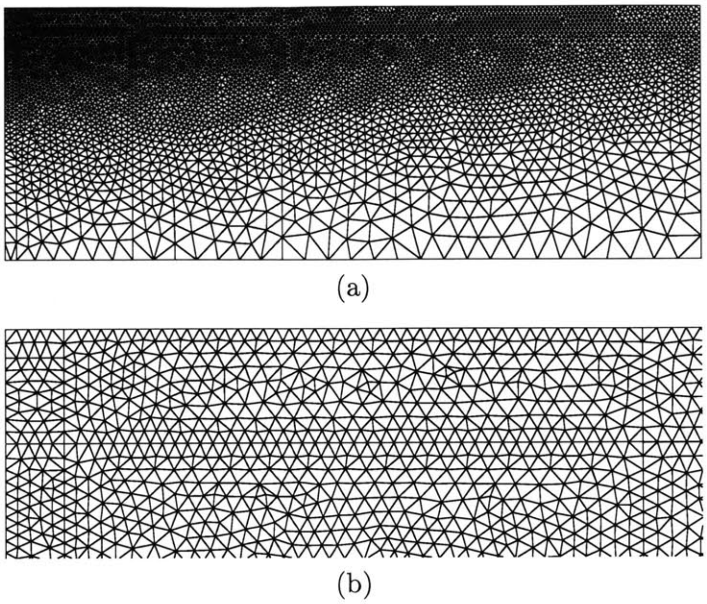


Figure 4-7: AP I: Reference domain $\Omega = \bigcup_{1\leq i\leq 10} \Omega^i$.

$\theta^q_{a,m}(\mu)$ listed in Appendix C that the necessary conditions from Lemma 6 are satisfied; we may thus choose $\hat{\alpha}_a(\mu) = \min_{1\leq q\leq Q_a}(\Theta^q_a(\mu)/\Theta^q_a(\mu_{\rm ref}))$ and $\hat{\alpha}_m(\mu) = \min_{1\leq q\leq Q_m}(\Theta^q_m(\mu)/\Theta^q_m(\mu_{\rm ref}))$ in (4.57) and (4.58), respectively. The two outputs, $s_1(\mu, t^k) = \ell_1(T(\mu, t^k))$, $\forall\, k \in \mathbb{K}$ and $s_2(\mu, t^k) = \ell_2(T(\mu, t^k))$, $\forall\, k \in \mathbb{K}$, can be written in the form (4.4), where $\ell_1(v) = |\Omega^6|^{-1}\int_{\Omega^6} v$ and $\ell_2(v) = |\Omega^{10}|^{-1}\int_{\Omega^{10}} v$. We shall consider the time interval $\bar{I} = [0, 10]$ and a timestep $\Delta t = 5\,\text{E}-2$; we thus have $K = 200$.

We briefly return to our previous discussion concerning the set of admissible output functionals ℓ. We remarked in the proof of the simple output bound (4.112) that the output functional ℓ has to be bounded with respect to $\|\cdot\|_X$. This requirement is the reason why we define the output here to be the average temperature over a small patch: as $\mathcal{N} \to \infty$, the dual norms of the outputs functionals $\|\ell_1\|_{X'}$ and $\|\ell_2\|_{X'}$ tend to $|\Omega_6|^{-\frac{1}{2}}$ and $|\Omega_{10}|^{-\frac{1}{2}}$, respectively — the dual norms thus blow up as the size of the regions Ω_6 and Ω_{10} goes to zero. Since our methods must remain stable as $\mathcal{N} \to \infty$ we do have to choose a (small) finite area for our output measurement.

We first present numerical results for the truth approximation. We assume that the surface is exposed to the heat source $u(t^k) = q(t^k) = 1$ for $1 \leq k \leq 10$ and $u(t^k) = q(t^k) = 0$ for $k \geq 11$, corresponding to the heat being applied for $t \in [0, 0.5]$. We show in Figure 4-9 snapshots of the temperature distribution over Ω at four timesteps for $\mu = (5, 1)$. The temperature reaches its peak on the top surface for t^{10}. At this point the heat turns off and the temperature evens out in the structure; we observe that the FRP cools down much slower on top of the delamination. We also present, in Figure 4-20(a), the thermal signal $s_1(\mu, t^k) - s_2(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, for $\mu_2 = 1$ fixed as a function of μ_1. The thermal signal is defined as the difference between the surface temperature on top of the delamination and the surface temperature on top of the undamaged structure. In actual practice, the thermal signal should be normalized with respect to some measure of either $s_1(\mu, t^k)$ or $s_2(\mu, t^k)$, e.g., the maximum output $\max_{t^k\in\mathbb{I}} s_{1,2}(\mu, t^k)$. We then avoid the dependence on the magnitude of the heat input which may be hard to determine in practice — and thus cause significant difficulty. We note that the thermal signal is more pronounced with increasing width of the delamination. We also plot in Figure 4-20(b) the thermal signal for $\mu_1 = 3$ fixed and varying


(a)

(b)

Figure 4-8: AP I: (a) Finite element truth approximation mesh; and (b) zoom on the delamination shown in magenta.

μ_2; as expected, the thermal response is faster for larger ratios of $\varkappa = k_{\rm FRP}/k_{\rm C}$.

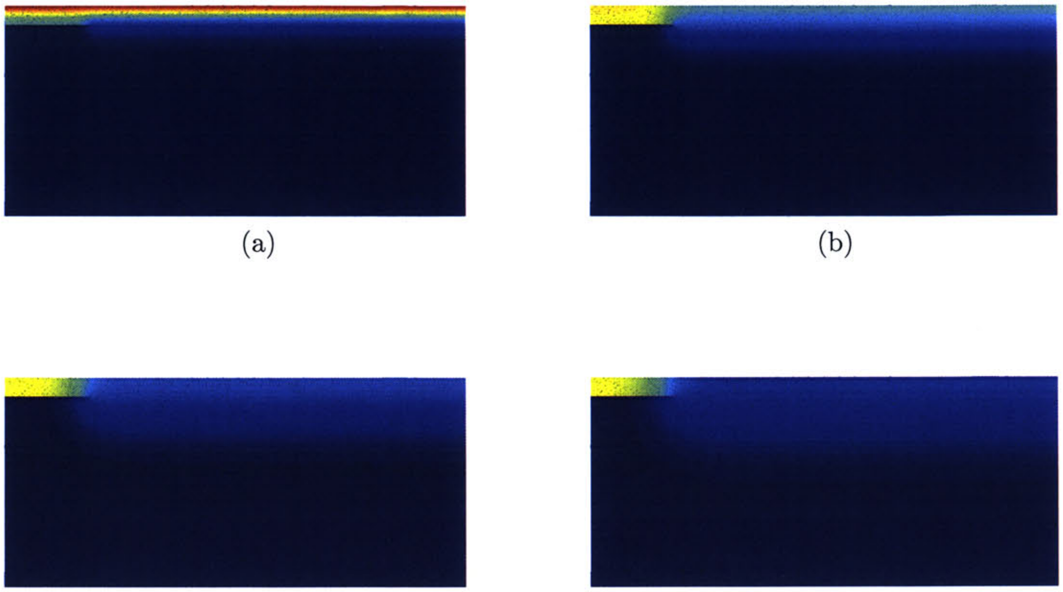
4.7.1 Reduced-Basis Approximation

We next generate the sample set $S^{\rm pr}_{N_{\rm pr}}$ and associated reduced basis space $W^{\rm pr}_{N_{\rm pr}}$ according to the adaptive sampling procedure described in Section 4.5. We initialize the procedure with $\mu_1^{\rm pr} = (5, 1.2)$ and $t^{k_1^{\rm pr}} = 1\Delta t$ and set the desired error tolerance (for the relative error in the energy norm) to $\epsilon_{\rm tol,min} = 1\,\mathrm{E}{-}4$. We sample on a random parameter test sample $\Xi_{\rm F} \in (\mathcal{D})^{400}$ of size 400 — we need $N_{\rm pr,max} = 217$ basis functions to obtain the desired accuracy. We also generate the sample sets $S^{\rm du,1}_{N_{\rm du,1}}$ and $S^{\rm du,2}_{N_{\rm du,2}}$ and associated reduced-basis spaces $W^{\rm du,1}_{N_{\rm du,1}}$ and $W^{\rm du,2}_{N_{\rm du,2}}$ for the two dual problems corresponding to the two output functionals $\ell_1(v)$ and $\ell_2(v)$, respectively; we obtain, for $\epsilon_{\rm tol,min} = 1\,\mathrm{E}{-}4$, $\mu_1^{\rm du} = (5, 1.2)$, and $t^{k_1^{\rm du}} = 201\Delta t$: $N_{\rm du,1,max} = 169$ and $N_{\rm du,2,max} = 135$.

We plot the sample sets $S^{\rm pr}_{N_{\rm pr}}$ and $S^{\rm du,1}_{N_{\rm du,1}}$ in $\mu - t^k$-space in Figure 4-11(a) and (b), respectively (the color of the sample points is associated with the magnitude of μ_2 so as to better identify the location in three-dimensional parameter-time space). The sample sets reflect the transients occurring at small k for the primal problem and at large k for the dual problem. We also note that the samples are largely located along the "boundary" of the parameter domain $\mathcal{D}$.

4.7.2 Numerical Results

We now present convergence results and error bounds for the primal problem, the two dual problems, and the output estimates. In Table 4.3 we present, as a function of $N_{\rm pr}$, the maximum relative error in the energy norm $\epsilon^{\rm pr}_{\rm max,rel}$, the maximum relative error bound $\Delta^{\rm pr}_{\rm max,rel}$, and the average effectivity $\overline{\eta}^{\rm pr}$: $\epsilon^{\rm pr}_{\rm max,rel}$ is the maximum over $\Xi_{\rm Test}$ of $|||e^{\rm pr}(\mu, t^K)|||^{\rm pr}/|||y(\mu_y, t^K)|||$, $\Delta^{\rm pr}_{\rm max,rel}$ is

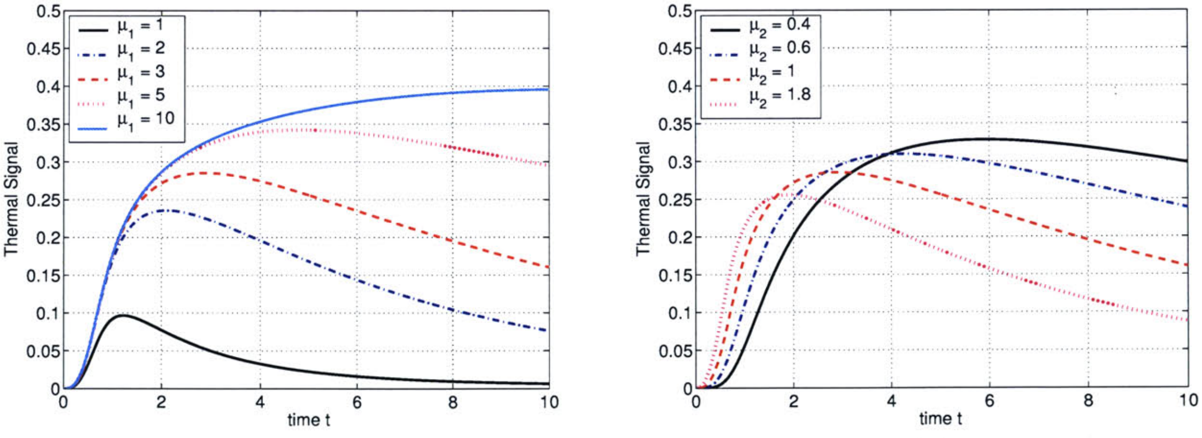

(a)

(b)

(c)

(d)

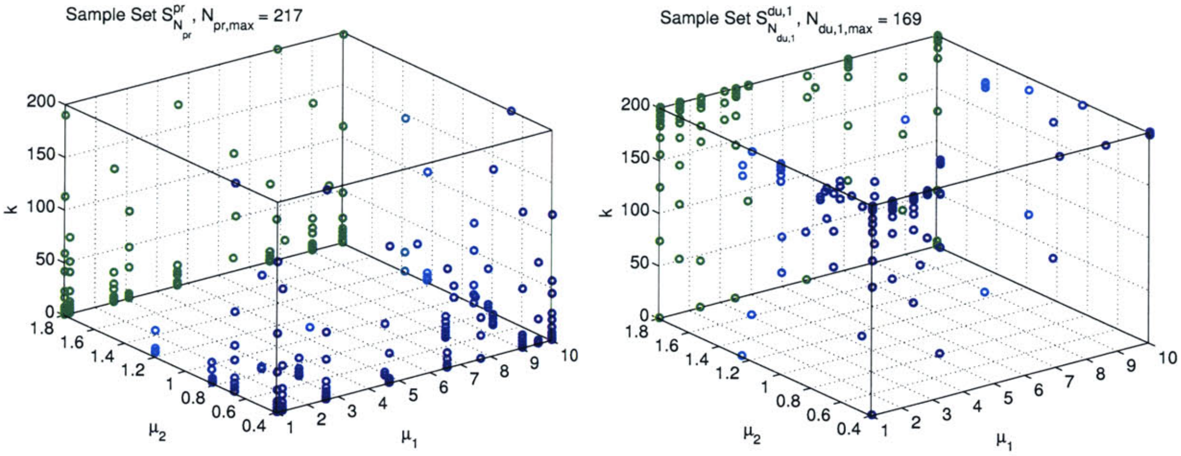
Figure 4-9: AP I: Temperature distribution for $\mu = (5, 1)$ at (a) $t = t^{10}$, (b) t^{20}, (c) t^{40}, and (d) t^{60}.



(a)

(b)

Figure 4-10: AP I: Thermal Signal for (a) $\mu_2 = 1$ as a function of μ_1 and (b) for $\mu_1 = 3$ as a function of μ_2.



(a)

(b)

Figure 4-11: AP II: (a) Sample set $S^{\text{pr}}_{N_{\text{pr}}}$ with $N_{\text{pr,max}} = 217$, and (b) sample set $S^{\text{du},1}_{N_{\text{du},1}}$ for the first output, s_1, with $N_{\text{du},1,\text{max}} = 169$.

the maximum over Ξ_{Test} of $\Delta^{\text{pr}}_{N_{\text{pr}}}(\mu, t^K)/|||y(\mu_y, t^K)|||$, and $\overline{\eta}^{\text{pr}}$ is the average over $\Xi_{\text{Test}} \times \mathbb{I}$ of $\Delta^{\text{pr}}_{N_{\text{pr}}}(\mu, t^k)/|||y(\mu, t^k) - y_N(\mu, t^k)|||$, where $\mu_y \equiv \arg\max_{\mu \in \Xi_{\text{Test}}} |||y(\mu, t^K)|||$. Here $\Xi_{\text{Test}} \in (\mathcal{D})^{121}$ is an input sample of size 121 (a regular 11 × 11 grid). We also present in Table 4.4(a) and (b) the corresponding results for the two dual problem: we tabulate, as a function of N_{du}, $\epsilon^{\text{du}}_{\text{max,rel}}$, $\Delta^{\text{du}}_{\text{max,rel}}$, and $\overline{\eta}^{\text{du}}$ (the definitions of these quantities are similar to the ones for the primal problem). We observe that the primal and dual error converge rapidly and that the bounds are very sharp; the effectivities are $O(1)$ for all values of N_{pr} and N_{du}. We also note that both dual problems require less basis functions than the primal problem to obtain the desired accuracy. Furthermore, we also obtain a smaller $N_{\text{du,max}}$ for the dual problem corresponding to the second output, s_2, because the influence of the parameter μ_1 on the dual problem is smaller for this output functional.

We next present in Table 4.5(a) and (b), as a function of $N_{\text{pr}} = N_{\text{du}}$, the convergence rates

N_{pr}	$\epsilon^{\text{pr}}_{\text{max,rel}}$	$\Delta^{\text{pr}}_{\text{max,rel}}$	$\overline{\eta}^{\text{pr}}$
20	8.09 E−02	3.18 E−01	2.74
40	2.71 E−02	8.01 E−02	2.77
60	1.02 E−02	2.01 E−02	2.58
80	5.02 E−03	8.40 E−03	2.83
100	1.68 E−03	2.91 E−03	2.50
120	7.40 E−04	1.71 E−03	2.45
140	4.37 E−04	8.56 E−04	2.32
160	2.13 E−04	4.84 E−04	2.21
180	1.30 E−04	3.16 E−04	2.18
200	9.55 E−05	2.70 E−04	2.20

Table 4.3: AP I: Convergence rate and effectivities for primal problem.

$N_{\rm du}$	$\epsilon^{\rm du}_{\rm max,rel}$	$\Delta^{\rm du}_{\rm max,rel}$	$\overline{\eta}^{\rm du}$
20	2.04E-01	7.46E-01	2.62
40	5.23E-02	9.69E-02	2.41
60	1.36E-02	2.23E-02	2.56
80	3.30E-03	5.39E-03	2.61
100	1.74E-03	2.27E-03	2.29
120	6.45E-04	9.00E-04	2.25
140	1.51E-04	3.77E-04	2.13
160	8.16E-05	1.41E-04	2.09

(a)

$N_{\rm du}$	$\epsilon^{\rm du}_{\rm max,rel}$	$\Delta^{\rm du}_{\rm max,rel}$	$\overline{\eta}^{\rm du}$
20	5.10E-02	2.03E-01	3.05
40	7.66E-03	1.73E-02	2.65
60	1.93E-03	3.57E-03	2.37
80	5.75E-04	1.03E-03	2.39
100	1.89E-04	3.18E-04	2.36
120	7.97E-05	1.25E-04	2.25

(b)

Table 4.4: AP I: Convergence rate and effectivities for dual problem corresponding to (a) output 1 and (b) output 2.

and error bounds for the two outputs. To this end, we define the maximum relative output error $\epsilon^s_{\rm max,rel}$, the maximum relative output bound $\Delta^s_{\rm max,rel}$, and the average output effectivity $\overline{\eta}^s$: $\epsilon^s_{\rm max,rel}$ is the maximum over $\Xi_{\rm Test}$ of $|s(\mu, t_\eta(\mu)) - s_N(\mu, t_\eta(\mu))|/s_{\rm max}$, $\Delta^s_{\rm max,rel}$ is the maximum over $\Xi_{\rm Test}$ of $\Delta^s_N(\mu, t^K)/|s_{\rm max}|$, and $\overline{\eta}^s$ is the average over $\Xi_{\rm Test}$ of $\Delta^s_N(\mu, t_\eta(\mu))/|s(\mu, t_\eta(\mu)) - s_N(\mu, t_\eta(\mu))|$; here $t_\eta(\mu) \equiv \arg\max_{t^k \in \mathbb{I}} |s(\mu, t^k) - s_N(\mu, t^k)|$ and $s_{\rm max} \equiv \max_{t^k \in \mathbb{I}} \max_{\mu \in \Xi_{\rm Test}} |s(\mu, t^k)|$. We also plot $\epsilon^s_{\rm max,rel}$ and $\Delta^s_{\rm max,rel}$ as a function of $N_{\rm pr}$ and $N_{\rm du}$ for output 1 and 2 in Figures 4-12 and 4-13, respectively. We first observe that the output error and output bound converges roughly as the square of the primal and dual errors and error bounds, respectively. We only need $N_{\rm pr} = N_{\rm du} = 50$ to obtain an accuracy in the output bound for output 1 of approximately 1%; for output 2 even $N_{\rm pr} = N_{\rm du} = 30$ is sufficient. However, the output effectivities are considerably large, $O(100)$, for all values of $N_{\rm pr}$ and $N_{\rm du}$ — in fact, the effectivities are much larger than we may anticipate. We note, however, that the output effectivities are generally worse than the effectivities for the energy bound because our output bound cannot take into account any correlation between the primal and dual error (see [74] for a discussion in the elliptic context).

We further see from Figure 4-15 that the error (and bound) decreases for fixed $N_{\rm du}$ as $N_{\rm pr}$ increases; similarly, for fixed $N_{\rm pr}$ the error (and bound) decreases as $N_{\rm du}$ increases. Note that we can obtain a specific desired accuracy in the output bound for different combinations of $N_{\rm pr}$ and $N_{\rm du}$. We may thus select values for $N_{\rm pr}$ and $N_{\rm du}$ so as to minimize the computation cost involved to obtain a desired accuracy.

We now consider numerical results for the simple bound of Proposition 9. We present in Table 4.6(a) and (b), as a function of $N_{\rm pr}$, $\epsilon^{\hat{s}}_{\rm max,rel}$, $\Delta^{\hat{s}}_{\rm max,rel}$, and $\overline{\eta}^{\hat{s}}$; these quantities are defined with respect to the simple output estimate, $\hat{s}_N(\mu, t^k)$, and simple output bound, $\hat{\Delta}_N(\mu, t^k)$. The convergence rate of the output error and output bound is now "only" of the order of $\epsilon^{\rm pr}_{\rm max,rel}$ and $\Delta^{\rm pr}_{\rm max,rel}$. The output effectivities are also worse than for the primal-dual formulation. For an accuracy of 1% in the output bound we would now require $N_{\rm pr} = 180$.

We note that, if our interest lies in sharp output bounds and small effectivities, we could also adopt the approach taken in [121], i.e., we define the output estimate

$$\tilde{s}_N(\mu, t^k) = \ell(y_N(\mu, t^k)), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{4.122}$$

N_{pr}	N_{du}	$\epsilon^{s}_{\text{max,rel}}$	$\Delta^{s}_{\text{max,rel}}$	$\overline{\eta}^{s}$
20	20	1.78 E−02	1.23 E+00	174
40	40	1.75 E−03	3.85 E−02	260
60	60	1.67 E−04	2.24 E−03	189
80	80	7.57 E−06	2.43 E−04	268
100	100	6.21 E−07	3.21 E−05	222
120	120	1.34 E−07	6.84 E−06	212
140	140	3.36 E−08	1.82 E−06	210
160	160	8.64 E−09	4.14 E−07	384

(a)

N_{pr}	N_{du}	$\epsilon^{s}_{\text{max,rel}}$	$\Delta^{s}_{\text{max,rel}}$	$\overline{\eta}^{s}$
20	20	7.53 E−04	2.72 E−01	817
40	40	8.32 E−05	3.59 E−03	636
60	60	7.82 E−06	2.35 E−04	242
80	80	1.07 E−06	2.21 E−05	324
100	100	7.96 E−08	3.34 E−06	274
120	120	6.02 E−09	8.30 E−07	258

(b)

Table 4.5: AP I: Convergence rates and effectivities for (a) output 1 and (b) output 2.

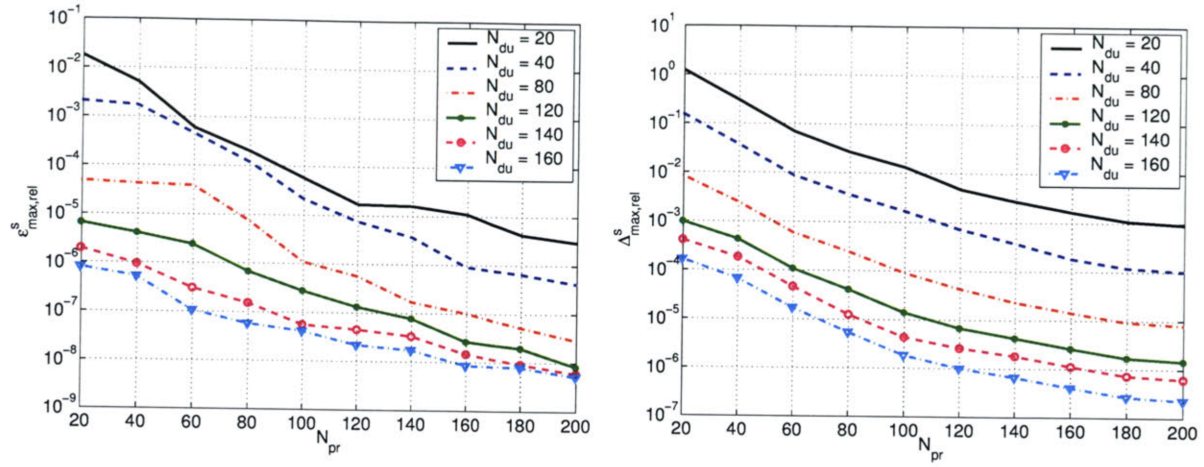
N_{pr}	$\epsilon^{\hat{s}}_{\text{max,rel}}$	$\Delta^{\hat{s}}_{\text{max,rel}}$	$\overline{\eta}^{\hat{s}}$
20	6.76 E−02	2.58 E+01	211
40	1.44 E−02	6.24 E+00	341
60	3.34 E−03	1.46 E+00	363
80	1.43 E−03	4.73 E−01	379
100	3.71 E−04	2.77 E−01	445
120	9.81 E−05	1.24 E−01	604
140	4.59 E−05	6.33 E−02	573
160	2.34 E−05	2.88 E−02	674
180	1.03 E−05	1.08 E−02	1002
200	6.02 E−06	9.18 E−03	1117

(a)

N_{pr}	$\epsilon^{\hat{s}}_{\text{max,rel}}$	$\Delta^{\hat{s}}_{\text{max,rel}}$	$\overline{\eta}^{\hat{s}}$
20	9.22 E−03	1.81 E+01	1379
40	2.56 E−03	4.38 E+00	1185
60	1.55 E−03	1.02 E+00	1220
80	9.79 E−04	3.32 E−01	1321
100	1.95 E−04	1.95 E−01	813
120	1.99 E−04	8.66 E−02	619
140	9.37 E−05	4.44 E−02	479
160	3.45 E−05	2.02 E−02	459
180	2.33 E−05	7.55 E−03	391
200	1.56 E−05	6.44 E−03	780

(b)

Table 4.6: AP I: Convergence rate and effectivities using simple bounds for (a) output 1 and (b) output 2.



(a)

(b)

Figure 4-12: AP I: (a) Maximum relative output error $\varepsilon^s_{\mathrm{max,rel}}$ and (b) maximum relative output bound $\Delta^s_{\mathrm{max,rel}}$ for output 1.

and corresponding output bound

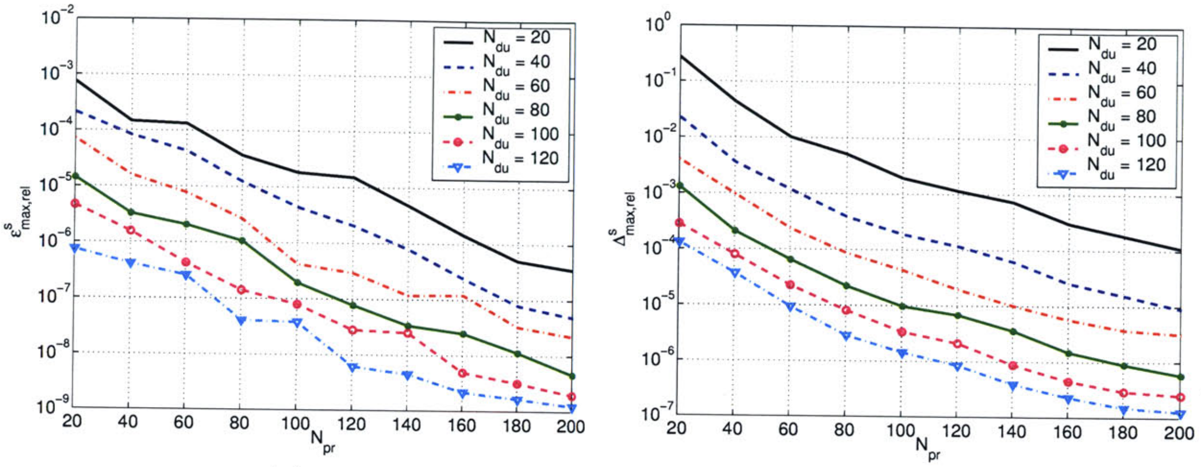
$$\tilde{\Delta}^s_N(\mu, t^k) \equiv \Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu, t^k)\ \Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu, t^{K-k+1}) + \left| \sum_{k'=1}^{k} R^{\mathrm{pr}}(\Psi_N(\mu, t^{K-k+k'}); \mu, t^{k'})\, \Delta t \right|, \tag{4.123}$$

where $\Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu, t^k)$ and $\Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu, t^k)$ are defined in Propositions 6 and 7, respectively. It is then easy to show that

$$|s(\mu, t^k) - \tilde{s}_N(\mu, t^k)| \leq \tilde{\Delta}^s_N(\mu, t^k), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}. \tag{4.124}$$

Note that we also require the dual problem following this approach. However, we employ the residual correction term to increase the sharpness of the output bound instead of the accuracy of the output estimate. We present in Table 4.7, as a function of N_{pr} and N_{du}, the maximum relative output error $\epsilon^{\tilde{s}}_{\mathrm{max,rel}}$, maximum relative output bound $\Delta^{\tilde{s}}_{\mathrm{max,rel}}$, and average output effectivity $\overline{\eta}^{\tilde{s}}$ obtained for this formulation. We observe that the convergence rate of the output error and bound is slower than for the standard primal-dual formulation, but the output effectivities are extremely good, $O(1)$ almost throughout. Due to the slower convergence rate, however, we still require approximately $N_{\mathrm{pr}} = N_{\mathrm{du}} = 50$ to obtain an accuracy of 1% in the error bound for output 1 and $N_{\mathrm{pr}} = N_{\mathrm{du}} = 30$ for output 2. Thus, in terms of the size of N_{pr} and N_{du} there is no gain in choosing this formulation over the standard primal-dual approach.

Finally, we present in Table 4.8, as a function of $N_{\mathrm{pr}}(= N_{\mathrm{du}})$, the online computational time to calculate $s_N(\mu, t^k)$ and $\Delta^s_N(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, for output 1. The values are normalized with respect to the computational time for the direct calculation of the truth approximation output $s(\mu, t^k) = \ell(u(\mu, t^k))$, $\forall\, k \in \mathbb{K}$. We note that the time to compute $s_N(\mu, t^k)$ dominates the time to compute $\Delta^s_N(\mu, t^k)$ for small values of $N_{\mathrm{pr}} = N_{\mathrm{du}}$. This is due to the K^2-complexity of evaluating the residual correction term. Since K is not too large in the present example we still obtain computational savings of a factor of 125 for $N_{\mathrm{pr}} = N_{\mathrm{du}} = 50$ (corresponding to an accuracy in the output bound of 1%). The average actual run-time for the output estimate and output bound in



(a)

(b)

Figure 4-13: AP I: (a) Maximum relative output error $\varepsilon^s_{\mathrm{max,rel}}$ and (b) maximum relative output bound $\Delta^s_{\mathrm{max,rel}}$ for output 2.

MATLAB 6.5 on a 750 MHz Pentium III is 0.5 sec.

We should note, however, that a large number of timesteps often results because of the requirement that the (discrete) time integration be accurate — and thus Δt be small. For the time history of the output estimate $s_N(\mu, t^k)$ itself, on the other side, a coarser time grid often suffices. Evaluating the output estimate, $s_N(\mu, t^k)$, at only every (say) 10th timestep, decreases the computational cost by a factor of 10, i.e., the complexity is $O(K(K+1)N_{\mathrm{pr}}N_{\mathrm{du}}/10)$.

We thus define $\overline{\mathbb{K}} = \{10, 20, 30, \ldots, K\}$ and present in Table 4.9 the online computational times to calculate $s_N(\mu, t^k)$, $\forall\, k \in \overline{\mathbb{K}}$, and $\Delta^s_N(\mu, t^k)$, $\forall\, k \in \overline{\mathbb{K}}$, i.e., we evaluate the output estimate, $s_N(\mu, t^k)$, only at every 10th timestep. The online time to calculate $\Delta^s_N(\mu, t^k)$ remains unchanged, but we observe — especially for small $N_{\mathrm{pr}} = N_{\mathrm{du}}$ — up to $O(5)$ reduction in computational effort to evaluate $s_N(\mu, t^k)$, $\forall\, k \in \overline{\mathbb{K}}$. (The savings are not quite $O(10)$ here because the size of K is moderate; in Section 4.8.5 we will consider an example where K is larger resulting in savings of very close to $O(10)$.) We recall that we require $N_{\mathrm{pr}} = N_{\mathrm{du}} = 50$ for an accuracy in the output bound of 1%: in this case, the overall savings compared to the direct calculation of the truth approximation $s(\mu, t^k), \forall\, k \in \overline{\mathbb{K}}$, are now almost a factor of 800. Also, the average actual run-time in MATLAB 6.5 on a 750 MHz Pentium III now decreases to only 0.18 sec.

Finally, we compare these results with the online computational times to calculate the simple output estimate and output bound, $\hat{s}_N(\mu, t^k)$ and $\hat{\Delta}^s_N(\mu, t^k)$, $\forall\, k \in \mathbb{K}$. For the same value of N_{pr} these values are smaller since the computation does not involve the solution of the dual problem. In order to make a valid comparison, however, we need to compare the computational efficiency at a fixed accuracy: we thus recall that we required $N_{\mathrm{pr}} = 180$ to obtain an accuracy of 1% in the output bound (corresponding to $N_{\mathrm{pr}} = N_{\mathrm{du}} = 50$) — the resulting computational savings are then only of the order of 30, the average run-time is 1.87 sec.

At this point we should also recall that, in the case of multiple outputs, each output requires a separate dual problem. The computational complexity of the primal-dual formulation thus increases with the number of outputs, while the computational complexity of the primal-only approach is (effectively) independent of the number of outputs. In applications with a large number of outputs

N_{pr}	N_{du}	$\epsilon^s_{\text{max,rel}}$	$\Delta^s_{\text{max,rel}}$	$\overline{\eta}^s$
20	20	6.76 E−02	1.24 E+00	12.3
40	40	1.44 E−02	4.23 E−02	3.88
60	60	3.34 E−03	5.08 E−03	1.68
80	80	1.43 E−03	1.57 E−03	1.18
100	100	3.71 E−04	3.84 E−04	1.07
120	120	9.81 E−05	9.98 E−05	1.05
140	140	4.59 E−05	4.61 E−05	1.01
160	160	2.34 E−05	2.35 E−05	1.01

(a)

N_{pr}	N_{du}	$\epsilon^s_{\text{max,rel}}$	$\Delta^s_{\text{max,rel}}$	$\overline{\eta}^s$
20	20	9.22 E−03	2.82 E−01	1.92
40	40	2.56 E−03	4.61 E−03	2.40
60	60	1.55 E−03	1.64 E−03	1.30
80	80	9.79 E−04	9.87 E−04	1.06
100	100	1.95 E−04	1.97 E−04	1.02
120	120	1.99 E−04	1.99 E−04	1.00

(b)

Table 4.7: AP I: Convergence rate and effectivities for output estimate and output bound defined in (4.122) and (4.123), respectively: (a) output 1 and (b) output 2.

$N_{\text{pr}} = N_{\text{du}}$	$s_N(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$\Delta^s_N(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$s(\mu, t^k), \forall\, k \in \mathbb{K}$
20	3.11 E−03	9.78 E−04	1
40	5.22 E−03	1.54 E−03	1
60	7.90 E−03	2.34 E−03	1
80	9.49 E−03	3.88 E−03	1
100	1.48 E−02	9.98 E−03	1
120	2.01 E−02	1.74 E−02	1
140	2.55 E−02	3.21 E−02	1
160	3.10 E−02	4.36 E−02	1

Table 4.8: AP I: Online computational times to calculate $s_N(\mu, t^k)$ and $\Delta^s_N(\mu, t^k)$ for all $k \in \mathbb{K}$ (normalized with respect to the time to solve for $s(\mu, t^k), \forall\, k \in \mathbb{K}$).

the primal-only approach may therefore be advantageous, despite the larger value of N_{pr} required to obtain a desired accuracy.

4.8 Nonsymmetric Problems: Convection-Diffusion Equation

We now relax the assumption of symmetry on the bilinear form $a(\cdot, \cdot; \mu)$, permitting treatment of a wider class of problems. A representative example is the unsteady convection-diffusion equation, where the presence of the convective term renders the operator nonsymmetric. We will see that most results directly carry over from the symmetric to the nonsymmetric case; we therefore focus on the differences and refer back to the symmetric case whenever possible.

We note that the time-discretization deserves more attention when solving convection-diffusion equations. The Euler-Backward scheme, although employed in this section for simplicity, can introduce too much numerical diffusion for certain problems and should not be used for pure advection problems solved over long timescales. However, it also has desirable "stability" advantages, e.g., errors are always damped. The Crank-Nicolson scheme, on the other side, is neutrally stable and introduces no numerical damping. The disadvantage of this method is that perturbations or

$N_{\mathrm{pr}} = N_{\mathrm{du}}$	$s_N(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$\Delta_N^s(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$s(\mu, t^k), \forall\, k \in \mathbb{K}$
20	6.90E-04	9.78E-04	1
40	9.70E-04	1.54E-03	1
60	1.31E-03	2.34E-03	1
80	1.82E-03	3.88E-03	1
100	2.97E-03	9.98E-03	1
120	5.59E-03	1.74E-02	1
140	9.28E-03	3.21E-02	1
160	1.23E-02	4.36E-02	1

Table 4.9: AP I: Online computational times to calculate $s_N(\mu, t^k)$ and $\Delta_N^s(\mu, t^k)$ for all $k \in \overline{\mathbb{K}}$ (normalized with respect to the time to solve for $s(\mu, t^k), \forall\, k \in \mathbb{K}$).

N_{pr}	$\hat{s}_N(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$\hat{\Delta}_N^s(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$s(\mu, t^k), \forall\, k \in \mathbb{K}$
20	2.10E-04	4.52E-04	1
40	3.97E-04	6.36E-04	1
60	6.73E-04	8.75E-04	1
80	1.08E-03	1.33E-03	1
100	2.05E-03	3.70E-03	1
120	4.37E-03	6.20E-03	1
140	6.44E-03	1.20E-02	1
160	8.24E-03	1.65E-02	1
180	1.06E-02	2.09E-02	1
200	1.41E-02	2.68E-02	1

Table 4.10: AP I: Online computational times to calculate $\hat{s}_N(\mu, t^k)$ and $\hat{\Delta}_N^s(\mu, t^k)$ (normalized with respect to the time to solve for $s(\mu, t^k), \forall\, k \in \mathbb{K}$).

round-off errors are not damped, and too large a timestep can result in a phenomenon referred to as "ringing." The time-discretization scheme chosen thus depends on the specific problem at hand, for a further discussion see [42].

4.8.1 Abstract Formulation

We directly consider the "truth" approximation here: given a parameter $\mu \in \mathcal{D} \subset \mathbb{R}^P$, we evaluate the output of interest

$$s(\mu, t^k) = \ell(y(\mu, t^k)), \quad \forall\, k \in \mathbb{K}. \tag{4.125}$$

where the field variable, $y(\mu, t^k) \in Y$, satisfies

$$m(y(\mu,t^k), v; \mu) + \Delta t\, a^{\mathrm{CD}}(y(\mu, t^k), v; \mu) = m(y(\mu, t^{k-1}), v; \mu) + \Delta t\, b(v; \mu)\, u(t^k), \qquad \forall v \in Y,\ \forall k \in \mathbb{K}, \tag{4.126}$$

with initial condition (say) $y(\mu, t^0) = y_0(\mu) = 0$. Here, $a^{\mathrm{CD}}(\cdot, \cdot; \mu)$ and $b(\cdot; \mu)$ are Y-continuous bilinear and linear forms, respectively; $m(\cdot, \cdot; \mu)$ and $\ell(\cdot)$ are X-continuous bounded bilinear and linear forms, respectively; and $u(t^k)$ denotes the control input at time t^k.

We shall make the following assumption. First, we assume that the bilinear form $m(\cdot, \cdot; \mu)$ is symmetric, $m(v, w; \mu) = m(w, v; \mu)$, $\forall w, v \in X$, $\forall \mu \in \mathcal{D}$, and satisfies the continuity and coercivity conditions (4.6) and (4.8), respectively. We also require that the linear forms $b(\cdot; \mu) : Y \to \boldsymbol{R}$ and $\ell(\cdot) : Y \to \boldsymbol{R}$ be bounded with respect to $\|\cdot\|_Y$ and $\|\cdot\|_X$, respectively. We also assume that the bilinear form $a^{\mathrm{CD}}(\cdot, \cdot; \mu)$ is continuous,

$$a^{\mathrm{CD}}(w, v; \mu) \le \gamma_a(\mu) \|w\|_Y \|v\|_Y \ \le\ \gamma_a^0 \|w\|_Y \|v\|_Y, \quad \forall\, w, v \in Y,\ \forall\, \mu \in \mathcal{D}, \tag{4.127}$$

and coercive,

$$0 < \alpha_a^0 \ \le\ \alpha_a(\mu) \equiv \inf_{v \in Y} \frac{a^{\mathrm{CD}}(v, v; \mu)}{\|v\|_Y^2}, \quad \forall\, \mu \in \mathcal{D}. \tag{4.128}$$

(We (plausibly) suppose that γ_a^0, α_a^0, may be chosen independent of $\mathcal{N}$.) It thus follows that a solution to (4.126) exists and is unique [94].

We note that we *do not* require the bilinear form a^{CD} to be symmetric anymore. However, we point out that a^{CD} can always be written in the form

$$a^{\mathrm{CD}}(w, v; \mu) = a^{\mathrm{D}}(w, v; \mu) + a^{\mathrm{C}}(w, v; \mu), \quad \forall\, w, v \in Y,\ \forall\, \mu \in \mathcal{D}, \tag{4.129}$$

where $a^{\mathrm{D}}(\cdot, \cdot; \mu)$ is symmetric, $a^{\mathrm{D}}(v, w; \mu) = a^{\mathrm{D}}(w, v; \mu)$, $\forall\, w, v \in X$, $\forall\, \mu \in \mathcal{D}$, and $a^{\mathrm{C}}(\cdot, \cdot; \mu)$ is skew-symmetric, $a^{\mathrm{C}}(v, v; \mu) = 0$, $\forall\, v \in X$, $\forall\, \mu \in \mathcal{D}$. It directly follows that

$$a^{\mathrm{CD}}(v, v; \mu) = a^{\mathrm{D}}(v, v; \mu) + a^{\mathrm{C}}(v, v; \mu) = a^{\mathrm{D}}(v, v; \mu), \quad \forall\, v \in Y,\ \forall\, \mu \in \mathcal{D}; \tag{4.130}$$

thus, only the symmetric part a^{D} will enter into the coercivity (4.128). This results will be useful for our *a posteriori* error estimation procedure to follow.

We shall assume that a^{CD}, m, and b depend affinely on the parameter μ, i.e., m and b can be expressed in the form of (4.10) and (4.11), respectively; and a^{CD} can be written as

$$a^{\mathrm{CD}}(w, v; \mu) = \sum_{q=1}^{Q_{a^{\mathrm{CD}}}} \Theta^q_{a^{\mathrm{CD}}}(\mu)\, a^{\mathrm{CD}\,q}(w, v), \quad \forall\, w, v \in Y,\ \forall\, \mu \in \mathcal{D}, \tag{4.131}$$

for some (preferably) small integer $Q_{a^{\mathrm{CD}}}$. Here, the function $\Theta^q_{a^{\mathrm{CD}}}(\mu) : \mathcal{D} \rightarrow \boldsymbol{R}$ depends on μ, but the continuous forms $a^{\mathrm{CD}\,\mathrm{q}}$ do *not* depend on μ. Finally, we also require that all linear and bilinear forms are independent of time — the system is thus linear time-invariant (LTI).

Dual Problem

To ensure rapid convergence of the reduced-basis output approximation we also introduce a dual problem in the nonsymmetric case — which shall evolve backward in time [20]. Invoking the LTI property we can express the adjoint for the output at time t^L, $1 \leq L \leq K$, as $\psi_L(\mu, t^k) = \Psi(\mu, t^{K-L+k})$, $1 \leq k \leq L$, where $\Psi(\mu, t^k) \in Y$ satisfies

$$m(v, \Psi(\mu, t^k); \mu) + \Delta t\ a^{\mathrm{CD}}(v, \Psi(\mu, t^k); \mu) = m(v, \Psi(\mu, t^{k+1}); \mu), \qquad \forall\, v \in Y,\ \forall\, k \in \mathbb{K}, \tag{4.132}$$

with final condition

$$m(v, \Psi(\mu, t^{K+1}); \mu) \equiv \ell(v), \qquad \forall\, v \in Y. \tag{4.133}$$

Again, to obtain $\psi_L(\mu, t^k)$, $1 \leq k \leq L$, $\forall\, L \in \mathbb{K}$, we solve *once* for $\Psi(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, and then appropriately shift the result — we do not need to solve K separate dual problems.

Conservation Form

Before we introduce the reduced-basis approximation we have to comment on the specific choice of the weak form for convection-diffusion equations — these problems are an important class of applications since they widely appear in fluid dynamics; we already mentioned one such example in the Introduction in (1.11). To this end, we first consider the "exact" problem

$$\frac{y^{\mathrm{e}}(t^k) - y^{\mathrm{e}}(t^{k-1})}{\Delta t} + \mathbf{U} \cdot \nabla y^{\mathrm{e}}(t^k) = \frac{1}{\mathrm{Pe}}\, y^{\mathrm{e}}(t^{k-1}) + g(x)\, u(t^k), \qquad \forall\, k \in \mathbb{K}, \tag{4.134}$$

with initial condition (say) $y^{\mathrm{e}}(t^0) = 0$. Here, $y^{\mathrm{e}} \in Y^{\mathrm{e}}$ is the field variable and Y^{e} is an appropriate Hilbert space; $g(x)\, u(t^k)$ represent the source term; $\mathbf{U}$ is the velocity field; and Pe is the Peclet number. The Peclet number, $\mathrm{Pe} = \mathrm{U}_0 L_c / \kappa$, represents the ratio between the strength of the advective and diffusive process, where U_0 is the average velocity, L_c is the characteristic length, and κ the diffusivity.

We shall require that the (exact) velocity field is incompressible, i.e, $\nabla \cdot \mathbf{U} = 0$. However, in actual practice (and for our truth finite element discretization) we introduce a (piecewise-polynomial) approximation, $\tilde{\mathbf{U}}$, to the exact velocity field $\mathbf{U}$ which may only be approximately divergence-free. In general, we *do not* obtain $\nabla \cdot \tilde{\mathbf{U}} = 0$ pointwise, and thus the specific form of the weak form of (4.134) is important in determining the lower bound for the coercivity constant $\alpha_a(\mu)$ and guaranteeing that (4.130) is satisfied [42]. More specifically, we require the convective term, $\tilde{\mathbf{U}} \cdot \nabla w$, be written in conservation form, that is

$$a^{\mathrm{C}}(w, v; \tilde{\mathbf{U}}) = \int_\Omega v\, (\tilde{\mathbf{U}} \cdot \nabla w) + \frac{1}{2} \int_\Omega v\, w\, \left(\nabla \cdot \tilde{\mathbf{U}}\right). \tag{4.135}$$

It is then easy to show that the bilinear form $a^{\mathrm{C}}(\cdot, \cdot; \mathbf{U})$ is skew-symmetric for (i) a contained flow, that is $(\mathbf{n} \cdot \tilde{\mathbf{U}}) = 0$ everywhere on the boundary, where $\mathbf{n}$ is the unit outward normal, or (ii) a flow with Dirichlet boundary conditions. We have

$$\begin{aligned} a^{\mathrm{C}}(v, v; \tilde{\mathbf{U}}) &= \int_\Omega v\,(\tilde{\mathbf{U}} \cdot \nabla v) + \frac{1}{2}\int_\Omega v^2\,(\nabla \cdot \tilde{\mathbf{U}}) \\ &= \frac{1}{2}\int_\Omega \tilde{\mathbf{U}} \cdot \nabla v^2 + \frac{1}{2}\int_\Omega v^2\,(\nabla \cdot \tilde{\mathbf{U}}) \\ &= \frac{1}{2}\int_\Omega \nabla \cdot (\tilde{\mathbf{U}}\, v^2) - \frac{1}{2}\int_\Omega v^2\,(\nabla \cdot \tilde{\mathbf{U}}) + \frac{1}{2}\int_\Omega v^2\,(\nabla \cdot \tilde{\mathbf{U}}) \\ &= \frac{1}{2}\int_\Gamma (\mathbf{n} \cdot \tilde{\mathbf{U}})\, v^2 = 0, \quad \forall\, v \in Y. \end{aligned} \tag{4.136}$$

We will thus employ the conservation form (4.135) in our numerical examples to follow. Note that we implicitly assume here (and in the examples to follow) that all quadratures are performed exactly.

We remarked earlier that the Peclet number, Pe, represents the ratio between the strength of the advective and diffusive process: for Pe $\ll 1$ the flow is diffusion-dominated, whereas for Pe $\gg 1$ the flow is advection-dominated. In the applications considered in this thesis the Peclet number varies in the range $1 \le \mathrm{Pe} \le 100$; for certain ranges of the parameter values, small diffusivities (usually), the problems we consider are thus advection-dominated — this is also the reason why Crank-Nicolson is the preferred time-integration scheme. We do not, however, consider stabilization methods, such as bubble functions, for the convection terms here.

4.8.2 Reduced-Basis Approximation

We first introduce the nested sample sets $S^{\mathrm{pr}}_{N_{\mathrm{pr}}} = \{\tilde{\mu}^{\mathrm{pr}}_1 \in \tilde{\mathcal{D}}, \ldots, \tilde{\mu}^{\mathrm{pr}}_{N_{\mathrm{pr}}} \in \tilde{\mathcal{D}}\}$, $1 \le N_{\mathrm{pr}} \le N_{\mathrm{pr,max}}$, and $S^{\mathrm{du}}_{N_{\mathrm{du}}} = \{\tilde{\mu}^{\mathrm{du}}_1 \in \tilde{\mathcal{D}}, \ldots, \tilde{\mu}^{\mathrm{du}}_{N_{\mathrm{du}}} \in \tilde{\mathcal{D}}\}$, $1 \le N_{\mathrm{du}} \le N_{\mathrm{du,max}}$, where $\tilde{\mu} \equiv (\mu, t^k)$ and $\tilde{\mathcal{D}} \equiv \mathcal{D} \times \mathbb{I}$. We then define the associated nested Lagrangian [85] reduced-basis spaces

$$W^{\mathrm{pr}}_{N_{\mathrm{pr}}} = \operatorname{span}\{\zeta^{\mathrm{pr}}_n \equiv y(\tilde{\mu}^{\mathrm{pr}}_n),\ 1 \le n \le N_{\mathrm{pr}}\}, \quad 1 \le N_{\mathrm{pr}} \le N_{\mathrm{pr,max}}, \tag{4.137}$$

and

$$W^{\mathrm{du}}_{N_{\mathrm{du}}} = \operatorname{span}\{\zeta^{\mathrm{du}}_n \equiv \Psi(\tilde{\mu}^{\mathrm{du}}_n),\ 1 \le n \le N_{\mathrm{du}}\}, \quad 1 \le N_{\mathrm{du}} \le N_{\mathrm{du,max}}, \tag{4.138}$$

where $y(\tilde{\mu}^{\mathrm{pr}}_n)$ is the solution of (4.126) at time $t = t^{k^{\mathrm{pr}}_n}$ for $\mu = \mu^{\mathrm{pr}}_n$ and $\Psi(\tilde{\mu}^{\mathrm{du}}_n)$ is the solution of (4.132) at time $t = t^{k^{\mathrm{du}}_n}$ for $\mu = \mu^{\mathrm{du}}_n$.

Our reduced-basis approximation $y_N(\mu, t^k)$ to $y(\mu, t^k)$ is then obtained by a standard Galerkin projection: given $\mu \in \mathcal{D}$, $y_N(\mu, t^k) \in W^{\mathrm{pr}}_{N_{\mathrm{pr}}}$ satisfies

$$m(y_N(\mu, t^k), v; \mu) + \Delta t\ a^{\mathrm{CD}}(y_N(\mu, t^k), v; \mu) = m(y_N(\mu, t^{k-1}), v; \mu) + \Delta t\ b(v; \mu)\ u(t^k), \quad \forall\, v \in W^{\mathrm{pr}}_{N_{\mathrm{pr}}},\ \forall\, k \in \mathbb{K}, \tag{4.139}$$

with initial condition $y_N(\mu, t^0) = 0$. Similarly, we obtain the reduced-basis approximation $\Psi_N(\mu, t^k) \in W^{\mathrm{du}}_{N_{\mathrm{du}}}$ to $\Psi(\mu, t^k)$ as the solution of

$$m(v, \Psi_N(\mu, t^k); \mu) + \Delta t\ a^{\mathrm{CD}}(v, \Psi_N(\mu, t^k); \mu) = m(v, \Psi_N(\mu, t^{k+1}); \mu), \quad \forall\, v \in W^{\mathrm{du}}_{N_{\mathrm{du}}},\ \forall\, k \in \mathbb{K}, \tag{4.140}$$

with final condition

$$m(v, \Psi_N(\mu, t^{K+1}); \mu) \equiv \ell(v), \quad \forall\, v \in W^{\mathrm{du}}_{N_{\mathrm{du}}}. \tag{4.141}$$

Finally, we evaluate the output estimate, $s_N(\mu, t^k)$, from

$$s_N(\mu, t^k) \equiv \ell(y_N(\mu, t^k)) + \sum_{k'=1}^{k} R^{\mathrm{pr}}(\Psi_N(\mu, t^{K-k+k'}); \mu, t^{k'})\ \Delta t, \quad \forall\, k \in \mathbb{K}, \tag{4.142}$$

where

$$R^{\mathrm{pr}}(v; \mu, t^k) \equiv b(v;\mu)\ u(t^k) - a^{\mathrm{CD}}(y_N(\mu, t^k), v; \mu) - \frac{1}{\Delta t} m(y_N(\mu, t^k) - y_N(\mu, t^{k-1}), v; \mu), \qquad \forall\, v \in Y,\ \forall\, k \in \mathbb{K}, \tag{4.143}$$

is the primal residual.

A Priori Convergence Theory

We consider here the rate at which $y_N(\mu, t^k)$ converges to $y(\mu, t^k)$. The proof for the dual variable is very similar and therefore omitted.

Proposition 10. *Assume that the "truth" solution $y(\mu, t^k)$ and the corresponding reduced-basis solution $y_N(\mu, t^k)$ satisfy (4.126) and (4.139), respectively. The error, $e^{\mathrm{pr}}(\mu, t^k) = y(\mu, t^k) - y_N(\mu, t^k)$, is bounded by*

$$\begin{aligned} &\alpha_m(\mu)\ \|e(\mu, t^k)\|_X^2 + \alpha_a(\mu)\ \Delta t\ \sum_{k'=1}^{k} \|e(\mu, t^{k'})\|_Y^2 \\ &\quad \le \inf_{w_N(t^k) \in W_N} \left\{ \gamma_m(\mu)\ \|y(\mu, t^k) - w_N(t^k)\|_X^2 + \frac{\gamma_a(\mu)^2}{\alpha_a(\mu)}\ \Delta t\ \|y(\mu, t^k) - w_N(t^k)\|_Y^2 \right\} \\ &\qquad + \sum_{k'=1}^{k} \inf_{w_N(t^{k'}) \in W_N} \frac{\gamma_a(\mu)^2}{\alpha_a(\mu)}\ \Delta t\ \|y(\mu, t^{k'}) - w_N(t^{k'})\|_Y^2. \end{aligned} \tag{4.144}$$

Before turning to the proof, we note that we obtain the additional factor $\gamma_a(\mu)/\alpha_a(\mu)$ in the nonsymmetric case as compared to the symmetric case of Proposition 5. Since this factor is greater than one, we expect a slower convergence of our reduced-basis approximation (and, in turn, a larger N_{pr} (and N_{du}) required to obtain a desired accuracy) for nonsymmetric problems. We confirm this observation numerically in Section 4.8.5 (see Figure 4-23).

Proof. It directly follows from (4.126) and (4.139) that $e^{\mathrm{pr}}(\mu, t^k) = y(\mu, t^k) - y_N(\mu, t^k)$ satisfies

$$m(e^{\mathrm{pr}}(\mu, t^k), v; \mu) + \Delta t\ a^{\mathrm{CD}}(e^{\mathrm{pr}}(\mu, t^k), v; \mu) = m(e^{\mathrm{pr}}(\mu, t^{k-1}), v; \mu), \quad \forall\, v \in W^{\mathrm{pr}}_{N_{\mathrm{pr}}}, \tag{4.145}$$

with initial condition $e^{\mathrm{pr}}(\mu, t^0) = y(\mu, t^0) - y_N(\mu, t^0) = 0$, since $y(\mu, t^0) = y_N(\mu, t^0) = 0$ by assumption. We next let $w_N(t^k) \in W^{\mathrm{pr}}_{N_{\mathrm{pr}}}$ be the projection of $y(\mu, t^k)$ with respect to the "m" scalar product and choose $v \equiv w_N(t^k) - y_N(\mu, t^k) = e^{\mathrm{pr}}(\mu, t^k) - (y(\mu, t^k) - w_N(t^k))$ in (4.145). The treatment of the bilinear form m follows directly the proof of Proposition 5 for the symmetric case.

For the bilinear form $a^{\rm CD}$ we obtain

$$\begin{align}
& 2\ \Delta t\ a^{\rm CD}(e^{\rm pr}(\mu, t^k), y(\mu, t^k) - w_N(t^k); \mu) \tag{4.146}\\
& \qquad \leq\ 2\ \Delta t\ \gamma_a(\mu)\ \|e^{\rm pr}(\mu, t^k)\|_Y\ \|y(\mu, t^k) - w_N(t^k)\|_Y \tag{4.147}\\
& \qquad \leq\ \frac{\gamma_a(\mu)^2}{\alpha_a(\mu)}\ \Delta t\ \|y(\mu, t^k) - w_N(t^k)\|_Y^2 + \alpha_a(\mu)\ \Delta t\ \|e^{\rm pr}(\mu, t^{k'})\|_Y^2, \tag{4.148}
\end{align}$$

where we used the continuity of $a^{\rm CD}$ and invoked (4.30) with $c = \gamma_a(\mu)\ \|y(\mu, t^k) - w_N(t^k)\|_Y$, $d = \|e^{\rm pr}(\mu, t^k)\|_Y$, and $\rho = \alpha_a(\mu)$. We can then write

$$\begin{align}
& m(e^{\rm pr}(\mu, t^k), e^{\rm pr}(\mu, t^k); \mu) - m(e^{\rm pr}(\mu, t^{k-1}), e^{\rm pr}(\mu, t^{k-1}); \mu) + 2\ \Delta t\ a^{\rm CD}(e^{\rm pr}(\mu, t^k), e^{\rm pr}(\mu, t^k); \mu) \nonumber\\
& \qquad \leq\ m(y(\mu, t^k) - w_N(t^k), y(\mu, t^k) - w_N(t^k); \mu) \nonumber\\
& \qquad\quad -m(y(\mu, t^{k-1}) - w_N(t^{k-1}), y(\mu, t^{k-1}) - w_N(t^{k-1}); \mu) \nonumber\\
& \qquad\quad +\frac{\gamma_a(\mu)^2}{\alpha_a(\mu)}\ \Delta t\ \|y(\mu, t^k) - w_N(t^k)\|_Y^2 + \alpha_a(\mu)\ \Delta t\ \|e^{\rm pr}(\mu, t^{k'})\|_Y^2. \tag{4.149}
\end{align}$$

The desired results directly follows by summing from 1 to k and invoking the coercivity and continuity of the bilinear forms $a^{\rm CD}$ and m.

□

As in Section 4.3.2, we can also prove the boundedness of $y_\mu(\mu, t^k)$ (and $y_{\mu\mu}(\mu, t^k)$) for the nonsymmetric problem (4.126). The proof follows the same lines as in the symmetric case and is therefore omitted (we invoke the coercivity, continuity, and affine decomposition for the bilinear form $a^{\rm CD}$ and require a condition similar to (4.40) for the bilinear forms $a^{\rm CD\,q}$, $1 \leq q \leq Q_{a^{\rm CD}}$).

Offline-Online Computational Procedure

The offline-online computation decomposition and the corresponding operation counts for the primal and dual problems is equivalent to the procedure discussed in Section 4.3.3. We only recall that — given a new parameter value μ — the online cost to evaluate the output estimate $s_N(\mu, t^k)$, $\forall\, k \in \mathbb{K}$ is $O(N_{\rm pr}^3 + N_{\rm du}^3 + K(N_{\rm pr}^2 + N_{\rm du}^2) + K(K+1)N_{\rm pr}N_{\rm du})$ and thus *independent* of $\mathcal{N}$.

4.8.3 *A Posteriori* Error Estimation

We now turn to the *a posteriori* error estimation for the nonsymmetric problem. To begin, we define — similar to (4.57) and (4.58) — positive lower bounds $\hat{\alpha}_a(\mu) : \mathcal{D} \rightarrow \boldsymbol{R}_+$ and $\hat{\alpha}_m(\mu) : \mathcal{D} \rightarrow \boldsymbol{R}_+$ for the coercivity constants $\alpha_a(\mu)$ in (4.128) and $\alpha_m(\mu)$ in (4.8), respectively. We also introduce the dual norm of the primal residual (4.143)

$$\varepsilon_{N_{\rm pr}}^{\rm pr}(\mu, t^k) \equiv \sup_{v \in Y} \frac{R^{\rm pr}(v; \mu, t^k)}{\|v\|_Y}, \quad \forall\, k \in \mathbb{K}, \tag{4.150}$$

and the dual norm of the dual residual

$$\varepsilon_{N_{\rm du}}^{\rm du}(\mu, t^k) \equiv \sup_{v \in Y} \frac{R^{\rm du}(v; \mu, t^k)}{\|v\|_Y}, \quad \forall\, k \in \mathbb{K}, \tag{4.151}$$

where

$$R^{\mathrm{du}}(v;\mu,t^k) \equiv -a^{\mathrm{CD}}(v,\Psi_N(\mu,t^k);\mu) - \frac{1}{\Delta t} m(v,\Psi_N(\mu,t^k) - \Psi_N(\mu,t^{k+1});\mu), \quad \forall\, v \in Y,\ \forall\, k \in \mathbb{K}, \tag{4.152}$$

is the dual residual in the nonsymmetric case. We also specify the inner products

$$(v,w)_Y \equiv a^{\mathrm{D}}(v,w;\mu_{\mathrm{ref(s)}}), \quad \forall\, v,w \in Y, \tag{4.153}$$

and

$$(v,w)_X \equiv m(v,w;\mu_{\mathrm{ref(s)}}), \quad \forall\, v,w \in Y, \tag{4.154}$$

for some constant reference value(s) $\mu_{\mathrm{ref(s)}}$, and recall that $\|\cdot\|_Y = (\cdot,\cdot)_Y^{1/2}$, $\|\cdot\|_X = (\cdot,\cdot)_X^{1/2}$.

We note that, because of the skew-symmetry of a^{C}, the definition of the Y-norm and the lower bound for the coercivity constant only refer to the symmetric part a^{D}. Thus, only the affine decomposition of $a^{\mathrm{D}}(v,w;\mu)$ is important for the choice of our bound conditioner in Lemma 6.

We now present the bounding properties for the errors in the primal variable, the dual variable, and the output estimate. The error bounds are indeed equivalent — taking into account that the primal and dual residuals are different now — to the results presented in Section 4.4.2. Throughout this section we assume that the "truth" solutions $y(\mu,t^k)$ and $\Psi(\mu,t^k)$ satisfy (4.126) and (4.132), respectively, and the corresponding reduced-basis approximations $y_N(\mu,t^k)$ and $\Psi_N(\mu,t^k)$ satisfy (4.139) and (4.140), respectively.

Proposition 11. *Let $e^{\mathrm{pr}}(\mu,t^k) \equiv y(\mu,t^k) - y_N(\mu,t^k)$ be the error in the primal variable and define the "spatio-temporal" energy norm*

$$|||v(\mu,t^k)|||^{\mathrm{pr}} \equiv \left(m(v(\mu,t^k),v(\mu,t^k);\mu) + \sum_{k'=1}^{k} a^{\mathrm{D}}(v(\mu,t^{k'}),v(\mu,t^{k'});\mu)\ \Delta t \right)^{\frac{1}{2}}, \quad \forall v \in Y. \tag{4.155}$$

The error in the primal variable is then bounded by

$$|||e^{\mathrm{pr}}(\mu,t^k)|||^{\mathrm{pr}} \le \Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^k), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{4.156}$$

where the error bound $\Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^k)$ is defined as

$$\Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^k) \equiv \left(\frac{\Delta t}{\hat{\alpha}_a(\mu)} \sum_{k'=1}^{k} \varepsilon^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^{k'})^2 \right)^{\frac{1}{2}}, \tag{4.157}$$

and $\varepsilon^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^k)$ is the dual norm of the primal residual defined in (4.150).

Proof. The proof directly follows from Proposition 6 and (4.130).

□

Proposition 12. *Let $e^{\mathrm{du}}(\mu,t^k) \equiv \Psi(\mu,t^k) - \Psi_N(\mu,t^k)$ be the error in the dual variable and define*

$$|||v(\mu,t^k)|||^{\mathrm{du}} \equiv \left(m(v(\mu,t^k),v(\mu,t^k);\mu) + \sum_{k'=k}^{K} a^{\mathrm{D}}(v(\mu,t^{k'}),v(\mu,t^{k'});\mu)\ \Delta t \right)^{\frac{1}{2}}. \tag{4.158}$$

The error in the dual variable is then bounded by

$$|||e^{\mathrm{du}}(\mu,t^k)|||^{\mathrm{du}} \le \Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{4.159}$$

where the error bound $\Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k)$ *is defined as*

$$\Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k) \equiv \left(\frac{\Delta t}{\hat{\alpha}_a(\mu)} \sum_{k'=k}^{K} \varepsilon^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^{k'})^2 + \hat{\alpha}_m(\mu)\, \Delta^{\Psi_f}_{N_{\mathrm{du}}}(\mu)^2 \right)^{\frac{1}{2}}, \tag{4.160}$$

and $\varepsilon^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k)$ *is the dual norm of the dual residual defined in (4.151).*

Proof. The proof directly follows from Proposition 7 and (4.130).

□

For the output bound we obtain the following results.

Proposition 13. *Let the output of interest,* $s(\mu,t^k)$*, and the reduced-basis output estimate,* $s_N(\mu,t^k)$*, be given by*

$$s(\mu,t^k) = \ell(y(\mu,t^k)), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{4.161}$$

and

$$s_N(\mu,t^k) = \ell(y_N(\mu,t^k)) + \sum_{k'=1}^{k} R^{\mathrm{pr}}(\Psi_N(\mu,t^{K-k+k'});\mu,t^{k'})\,\Delta t, \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{4.162}$$

respectively. The error in the output of interest is then bounded by

$$|s(\mu,t^k) - s_N(\mu,t^k)| \le \Delta^s_N(\mu,t^k), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{4.163}$$

where the output bound $\Delta^s_N(\mu,t^k)$ *is defined as*

$$\Delta^s_N(\mu,t^k) \equiv \Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^k)\ \Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^{K-k+1}), \tag{4.164}$$

and $\Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^k)$ *and* $\Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k)$ *are defined in Propositions 11 and 12, respectively.*

Proof. The proof follows directly from Proposition 8 and (4.130).

□

Finally, we note that the simple output approximation, $\hat{s}_N(\mu,t^k)$, and corresponding bound, $\hat{\Delta}^s_N(\mu,t^k)$, according to Proposition 9 also holds for the nonsymmetric problem.

Offline-Online Computational Procedure

The offline-online computational decomposition directly follows our previous discussion in Section 4.4.4 and in Appendix A; also see Appendix B for the necessary computations if the Crank-Nicolson scheme is used. We therefore only summarize the computational costs involved in the online stage; that is — given a new parameter value μ and associated reduced-basis solutions $\underline{y}_N(\mu,t^k)$ and $\underline{\Psi}_N(\mu,t^k)$, $\forall\, k \in \mathbb{K}$ — the computational cost to evaluate $\Delta^s_N(\mu,t^k)$, $\forall\, k \in \mathbb{K}$, is $O(K(N^2_{\mathrm{pr}} + N^2_{\mathrm{du}})(Q^2_a + Q_aQ_m + Q^2_m))$. Again, all online calculations needed are *independent* of $\mathcal{N}$.

4.8.4 Numerical Exercise 3: Banks and Kunisch

We now turn to a particular numerical example of a convection-diffusion equation. We consider a one-dimensional model for brain transport discussed in [14], where the authors use this problem as a test case for their parameter estimation algorithms. In this section, we focus on generating the reduced-basis approximation for this model and discuss the sampling procedure and convergence results. In Section 7.4 we return to this example and employ our reduced-basis approximation to solve the parameter estimation problem.

The transport system, defined on the one-dimensional domain $\Omega = [0, 1]$, is given by

$$y_t = q_1\, y_{xx} + q_2\, y_x - q_2, \tag{4.165}$$

with initial conditions

$$y(0, x) = -2x^2 + 2x, \tag{4.166}$$

and homogeneous Dirichlet boundary conditions, $y(t, 0) = y(t, 1) = 0$. The outputs of interest is the average values of y over a small domain centered at the points $x = 0.25$, 0.5, 0.75 as a function of time t, where $t \in \bar{I} = [0, 1]$.

The system response is influenced by two parameters, the diffusivity q_1 and the velocity (and forcing) q_2. We assume that q_1 varies in the range $0.1 \leq q_1 \leq 1$, and that q_2 satisfies $0.5 \leq q_2 \leq 5$. Our input parameter is hence $\mu \equiv (\mu_1, \mu_2) \equiv (q_1, q_2) \in \mathcal{D} \equiv [0.1, 1] \times [0.5, 5] \subset \mathbb{R}^{P=2}$.

We next consider the time-discrete "truth" approximation of (4.165). The weak form of the governing equation for $y(\mu, t^k) \in Y$ is (4.126), where $Y \subset Y^{\mathrm{e}} \equiv H_0^1(\Omega)$ is a linear finite element truth approximation subspace of dimension $\mathcal{N} = 800$, and $u(t^k) = 1$, $\forall\, k \in \mathbb{K}$. The bilinear forms are given by $m(w, v) = \int_\Omega w\, v$, $a^{\mathrm{CD}}(w, v; \mu) = \mu_1 \int_\Omega w_x\, v_x - \mu_2 \int_\Omega w_x\, v$, and $b(v; \mu) = -\mu_2 \int_\Omega v$; the bilinear forms admit the obvious affine representations (4.10), (4.11), and (4.131) with $Q_m = 1$, $Q_b = 1$, and $Q_{a^{\mathrm{CD}}} = 2$. Also note that the bilinear a^{CD} satisfies (4.129) with $a^{\mathrm{D}}(w, v; \mu) = \mu_1 \int_\Omega w_x\, v_x$ and $a^{\mathrm{C}}(w, v; \mu) = \mu_2 \int_\Omega w_x\, v$. We also define the inner products $(w, v)_X \equiv \int_\Omega wv$ and $(w, v)_Y \equiv \int_\Omega w_x\, v_x$, corresponding to (4.153) for $\mu_1 = 1$; we may hence choose (see Lemma 6) $\hat{\alpha}_a(\mu) = \mu_1$ in (4.57). Note that the bilinear form m happens to be parameter-independent in this example, and thus $e^{\mathrm{du}}(\mu, t^{K+1}) = 0$ here. We have three outputs, $s_q(\mu, t^k) = \ell_q(y(\mu, t^k))$, $1 \leq q \leq 3$, can be written in the form (4.125), where $\ell_q(v) = |\Omega^{\ell_q}|^{-1} \int_{\Omega^{\ell_q}} v$, $1 \leq q \leq 3$, and $\Omega^{\ell_1} = [0.245, 0.255]$, $\Omega^{\ell_2} = [0.495, 0.505]$, and $\Omega^{\ell_1} = [0.745, 0.755]$. We note that $\ell_q(v) \in X$, $1 \leq q \leq 3$, since the outputs are integrals over small regions. We choose a discrete timestep $\Delta t = 0.01$ for the time interval $\bar{I} = [0, 1]$; we thus have $K = 100$.

Reduced-Basis Approximation

We generate the sample set $S_{N_{\mathrm{pr}}}^{\mathrm{pr}}$ and associated reduced basis space $W_{N_{\mathrm{pr}}}^{\mathrm{pr}}$ according to the adaptive sampling procedure described in Section 4.5. Since the initial condition is nonzero, we initialize the procedure with $\zeta_1^{\mathrm{pr}} = y_0$ and set the desired error tolerance (for the relative error in the energy norm) to $\epsilon_{\mathrm{tol,min}} = 1\,\mathrm{E}{-4}$. We sample on a random parameter test sample $\Xi_{\mathrm{F}} \in (\mathcal{D})^{1600}$ of size 1600. We plot and tabulate the resulting sample set $S_{N_{\mathrm{pr}}}^{\mathrm{pr}}$ in $\mu_1 - \mu_2 - t^k$-space in Figure 4-14 — we need $N_{\mathrm{pr,max}} = 16$ basis functions to obtain the desired accuracy. We note that although $K = 100$, only basis functions within the first 30 timesteps are selected. Furthermore, we observe that the samples are not distributed uniformly over $\mathcal{D} \times \mathbb{I}$. Instead, more samples are (adaptively) chosen in the "difficult" parameter range — for small diffusivities μ_1 and large velocities μ_2.

N_{pr}	$\epsilon^{y}_{\text{max,rel}}$	$\Delta^{\text{pr}}_{\text{max,rel}}$	$\overline{\eta}^{\text{pr}}$
3	4.54 E−01	7.39 E−01	1.31
6	5.22 E−02	6.08 E−02	1.15
9	2.20 E−03	2.69 E−03	1.12
12	2.85 E−04	3.76 E−04	1.08
15	8.51 E−05	1.09 E−04	1.24

Table 4.11: NE 3: Convergence rate and effectivities for primal problem.

N_{du}	$\epsilon^{\Psi}_{\text{max,rel}}$	$\Delta^{\text{du}}_{\text{max,rel}}$	$\overline{\eta}^{\text{du}}$
3	6.03 E−01	1.78 E+00	2.22
6	3.97 E−01	1.18 E+00	1.66
9	1.31 E−01	2.26 E−01	1.26
12	8.88 E−02	1.10 E−01	1.12
15	7.45 E−03	8.67 E−03	1.08
18	3.16 E−03	3.99 E−03	1.08
21	2.72 E−04	3.50 E−04	1.04
24	5.47 E−05	6.54 E−05	1.03

Table 4.12: NE 3: Convergence rate and effectivities for dual problem corresponding to output 1.

We also generate reduced-basis spaces for the three dual problems corresponding to the three output functionals $\ell_q(v)$, $1 \leq q \leq 3$; we obtain (for $\epsilon_{\text{tol,min}} = 1\,\text{E}-4$): $N^1_{\text{du,max}} = 24$, $N^2_{\text{du,max}} = 24$, and $N^3_{\text{du,max}} = 22$.

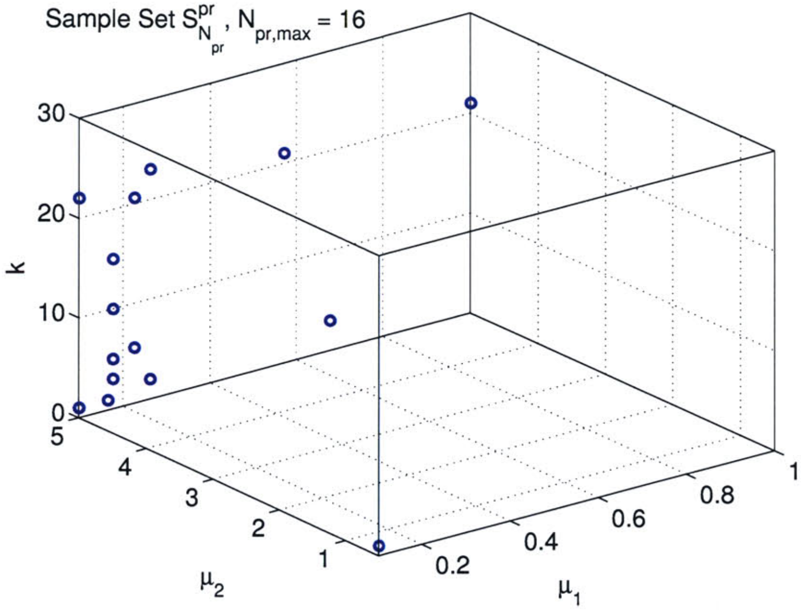
Numerical Results

We now discuss the convergence results and effectivities for the primal problem, the dual problem, and the output estimate; we only consider the dual corresponding to output 1, $s_1(\mu, t^k) = \ell_1(y(\mu, t^k))$ (the results for the other two outputs are almost identical).

In Table 4.11 we present, as a function of N_{pr}, $\epsilon^{\text{pr}}_{\text{max,rel}}$, $\Delta^{\text{pr}}_{\text{max,rel}}$, and $\overline{\eta}^{\text{pr}}$: for the definitions of these quantities see Section 4.7.2. Here $\Xi_{\text{Test}} \in (\mathcal{D})^{400}$ is a random input sample of size 400. The convergence results for the dual problem, $\epsilon^{\text{du}}_{\text{max,rel}}$, $\Delta^{\text{du}}_{\text{max,rel}}$, and $\overline{\eta}^{\text{du}}$ as a function of N_{du} are presented in Table 4.12. We observe a rapid convergence of the primal and dual reduced-basis approximation and that the error bounds are very sharp. We also note that the error for the dual problem converges slower than the error for the primal problem. This is related to the output functional which reflects as a very "rough" initial condition for the dual; the initial condition for the primal, on the other side, is much smoother.

We next present in Table 4.13(a) and (b) the convergence rate for the output using the dual formulation of Proposition 13 and the simple bound of Proposition 9[5], respectively. To this end, we define $\epsilon^{s}_{\text{max,rel}}$, $\Delta^{s}_{\text{max,rel}}$, and $\overline{\eta}^{s}$: $\epsilon^{s}_{\text{max,rel}}$ is the maximum over Ξ_{Test} of $|s(\mu, t_\eta(\mu)) - s_N(\mu, t_\eta(\mu))|/s_{\text{max}}$, $\Delta^{s}_{\text{max,rel}}$ is the maximum over Ξ_{Test} of $\Delta^{s}_{N}(\mu, t^K)/|s_{\text{max}}|$, and $\overline{\eta}^{s}$ is the average over Ξ_{Test} of

[5] Since m is parameter independent here, we can simply choose $\hat{\alpha}_m(\mu) = 1$ for the output bound



n	μ_n^{pr}	k_n^{pr}
1	$y_0(x)$	
2	$(0.100, 5.000)$	1
3	$(0.100, 5.000)$	22
4	$(0.262, 5.000)$	2
5	$(0.262, 5.000)$	23
6	$(0.678, 5.000)$	3
7	$(0.571, 5.000)$	21
8	$(0.117, 4.294)$	9
9	$(0.117, 4.294)$	24
10	$(1.000, 5.000)$	21
11	$(0.166, 5.000)$	1
12	$(0.100, 0.500)$	1
13	$(0.177, 5.000)$	10
14	$(0.177, 5.000)$	3
15	$(0.177, 5.000)$	5
16	$(0.177, 5.000)$	15

Figure 4-14: NE 3: Sample set $S^{\mathrm{pr}}_{N_{\mathrm{pr}}}$ for $\mathcal{D} \equiv [0.1, 1] \times [0.5, 5.0]$ and $N_{\mathrm{pr}} = 16$.

$N_{\rm pr}$	$N_{\rm du}$	$\epsilon^s_{\rm max,rel}$	$\Delta^s_{\rm max,rel}$	$\overline{\eta}^s$
3	12	4.92 E−02	9.38 E−01	17.9
6	15	1.06 E−04	3.68 E−03	28.0
9	18	1.11 E−05	9.86 E−05	5.52
12	21	1.29 E−08	6.48 E−07	17.3
15	24	5.75 E−10	2.27 E−08	10.1

(a)

$N_{\rm pr}$	$\epsilon^{\hat{s}}_{\rm max,rel}$	$\Delta^{\hat{s}}_{\rm max,rel}$	$\overline{\eta}^{\hat{s}}$
3	1.85 E−01	2.92 E+01	31.5
6	1.36 E−02	2.40 E+00	88.3
9	1.10 E−03	1.06 E−01	98.6
12	1.38 E−04	1.48 E−02	42.1
15	8.95 E−06	4.28 E−03	111

(b)

Table 4.13: NE 3: Convergence rate and effectivities for output 1 using (a) dual formulation and (b) simple output bound.

$\Delta^s_N(\mu, t_\eta(\mu))/|s(\mu, t_\eta(\mu)) - s_N(\mu, t_\eta(\mu))|$; here $t_\eta(\mu) \equiv \arg\max_{t^k \in \mathbb{I}} |s(\mu, t^k) - s_N(\mu, t^k)|$ and $s_{\rm max} \equiv \max_{t^k\in\mathbb{I}} \max_{\mu\in\Xi_{\rm Test}} |s(\mu, t^k)|$ (the quantities $\epsilon^{\hat{s}}_{\rm max,rel}$, $\Delta^{\hat{s}}_{\rm max,rel}$, and $\overline{\eta}^{\hat{s}}$ are defined similarly). We also plot $\epsilon^s_{\rm max,rel}$ and $\Delta^s_{\rm max,rel}$ as a function of $N_{\rm pr}$ and $N_{\rm du}$ in Figure 4-15(a) and (b), respectively.

We first note that, because of the slower convergence of the dual bound, we need to choose $N_{\rm du} > N_{\rm pr}$ to observe the square effect in the output error and output bound: however, for our choice of $N_{\rm pr}$ and $N_{\rm du}$, $\epsilon^s_{\rm max,rel}$ (respectively, $\Delta^s_N(\mu, t^k)$) does converge roughly as the square of $\epsilon^y_{\rm max,rel}$ (respectively, $\Delta^{\rm pr}_{N_{\rm pr}}(\mu, t^k)$). The output effectivities are $O(10)$ for the output bound in Table 4.13(a) and $O(10 - 100)$, for the simple output bound in Table 4.13(b). We observe from Figure 4-15 that increasing $N_{\rm du}$ while keeping $N_{\rm pr}$ constant results in a more accurate output estimate and output bound. Similarly, for constant $N_{\rm du}$ the accuracy of the output estimate and bound increases as $N_{\rm pr}$ increases. We note that we need approximately $N_{\rm pr} = 7$ and $N_{\rm du} = 12$ to obtain an accuracy in the output bound of 1%; using the simple output bound we obtain the same accuracy for $N_{\rm pr} = 12$. It thus follows that introducing the dual formulation in this problem does not pay off — the simple output bound converges fast enough and is thus sufficiently accurate even for small N. We also pointed out already that we need to introduce a separate dual problem for *each* output — thus, for problems with many outputs (also see the next section), the simple bound is certainly advantageous.

4.8.5 AP II: Dispersion of Pollutants

We now return to example AP II — the dispersion of a pollutant in a two-dimensional flow — introduced in Section 1.1.1. We assume here that the location of source term is known. A sketch of the flow field $\tilde{\mathbf{U}}$ with the source location and the eight measurement sensors is shown in Figure 4-16. The domain Ω, a typical point in which is (x_1, x_2), is given by $\Omega \equiv [0, 4] \times [0, 1]$. We assume that the concentration at the left boundary, $\Gamma_{\rm D}$, is zero and that the remaining boundaries, $\Gamma_{\rm N}$, are impermeable. The diffusivity κ is assumed to vary in the range $0.01 \le \kappa \le 1$; our input parameter is hence $\mu \equiv (\mu_1) \equiv (\kappa) \in \mathcal{D} \equiv [0.01, 1] \subset \mathbb{R}^{P=1}$.

The time-discrete weak form of the governing equation (1.11) for the concentration $c(\mu, t^k) \in Y$ is thus (B.2) (we use the Crank-Nicolson scheme for the time-integration here[6]) with initial condition $c(\mu, t^0) = 0$, where $Y \subset Y^e \equiv \{v \,|\, v \in H^1(\Omega), v = 0|_{\Gamma_{\rm D}}\}$ is a linear finite element truth

[6] Here, we have $\mathrm{U}_0 = 1$, $L_c = 1$, and $\kappa \in [0.01, 1]$; the Peclet number thus varies in the range $1 \le \mathrm{Pe} \le 100$. For small diffusivities the flow is clearly advection-dominated and thus Crank-Nicolson is the preferred time-integration scheme.

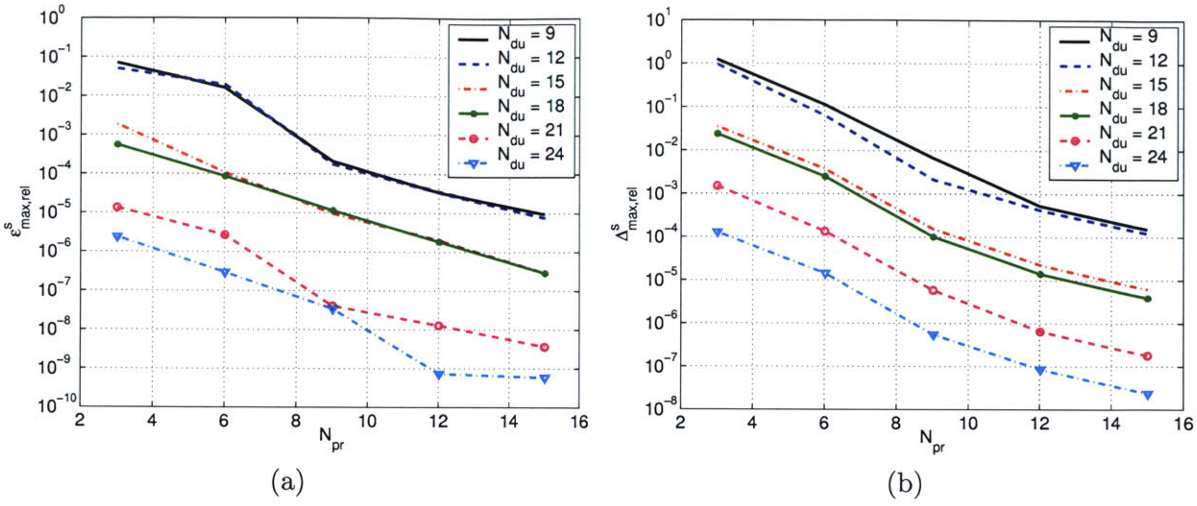


Figure 4-15: NE 3: (a) Maximum relative error $\varepsilon^{s}_{\max,\text{rel}}$ and (b) maximum relative output bound $\Delta^{s}_{\max,\text{rel}}$ for output 1 using the dual formulation.

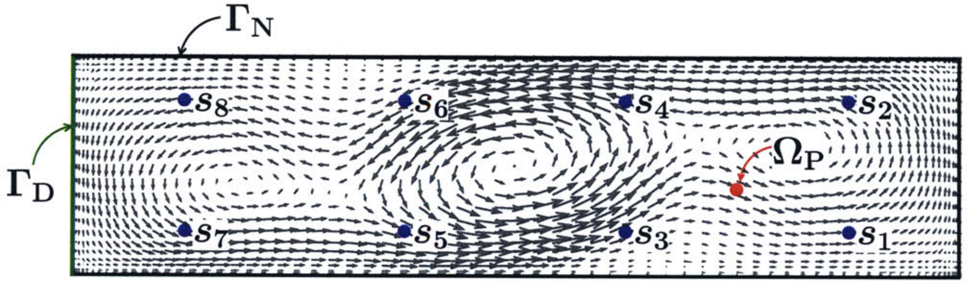


Figure 4-16: AP II: Velocity field with pollution source and measurement locations.

approximation subspace of dimension $\mathcal{N} = 3720$. The truth approximation mesh is shown in Figure 4-17. The bilinear and linear forms are given by

$$m(w,v) \equiv \int_{\Omega} w\, v, \tag{4.167}$$

$$a^{\text{CD}}(w,v;\mu) \equiv \mu_1 \int_{\Omega} \nabla w \cdot \nabla v + \int_{\Omega} v\, (\tilde{\mathbf{U}} \cdot \nabla w) + \frac{1}{2} \int_{\Omega} v\, w \left(\nabla \cdot \tilde{\mathbf{U}}\right), \tag{4.168}$$

$$b(v) \equiv \int_{\Omega} g^{\text{PS}}(x)\, v, \tag{4.169}$$

where the source term $g^{\text{PS}}(x)$ is defined as

$$g^{\text{PS}}(x) = \frac{1}{2\pi(\sigma^{\text{PS}})^2}\, e^{-\frac{(x_1 - x_1^{\text{PS}})^2 + (x_2 - x_2^{\text{PS}})^2}{2(\sigma^{\text{PS}})^2}}, \tag{4.170}$$

with the source location $x^s = (x_1^{\text{PS}}, x_2^{\text{PS}}) = (3, 0.4)$ and standard deviation $\sigma^{\text{PS}} = 0.05$. The velocity field, $\tilde{\mathbf{U}}$, is a Natural Convection (Navier Stokes) Flow with Gr $= 10^5$ and Pr $= 0$ taken from [121].

We note that the bilinear and linear forms m and b are parameter independent, that a^{CD} admits the affine representation (4.131) with $Q_a = 2$, and that the convective term a^{C} is written in conservation form. We also define the inner products $(w, v)_X \equiv \int_\Omega wv$ and $(w, v)_Y \equiv \int_\Omega \nabla w \cdot \nabla v$, corresponding to (4.153) for $\mu_1 = 1$; we may thus choose $\hat{\alpha}_a(\mu) = \mu_1$ in (4.57). Again, m is parameter independent here and thus $e^{\mathrm{du}}(\mu, t^{K+1}) = 0$. The sensor measurements, or outputs, are given by

$$s_q(\mu, t^k) = \ell(y(\mu, t^k); l_q(x)) \equiv \int_\Omega l_q(x)\, y(\mu, t^k), \quad 1 \le q \le 8, \ \forall\, k \in \mathbb{K} \tag{4.171}$$

where the spatial sensitivity is modeled as

$$l_q(x) = \frac{1}{2\pi\sigma_l^2} e^{-\frac{(x_1 - x_1^{l_q})^2 + (x_2 - x_2^{l_q})^2}{2\sigma_l^2}} \tag{4.172}$$

with standard deviation $\sigma_l = 0.05$. We have $\ell(y(\mu, t^k); l_q(x)) \in X$, $1 \le q \le 8$, since the $l_q(x)$ are (smooth) Gaussian functions. The sensor locations, $x^{l_q} = (x_1^{l_q}, x_2^{l_q})$, are given by $x^{l_1} = (3.5, 0.2)$, $x^{l_2} = (3.5, 0.8)$, $x^{l_3} = (2.5, 0.2)$, $x^{l_4} = (2.5, 0.8)$, $x^{l_5} = (1.5, 0.2)$, $x^{l_6} = (1.5, 0.8)$, $x^{l_7} = (0.5, 0.2)$, and $x^{l_8} = (0.5, 0.8)$. We shall consider the time interval $\bar{I} = [0, 2]$ and a timestep $\Delta t = 2.5\,\mathrm{E}{-3}$; we thus have $K = 800$.

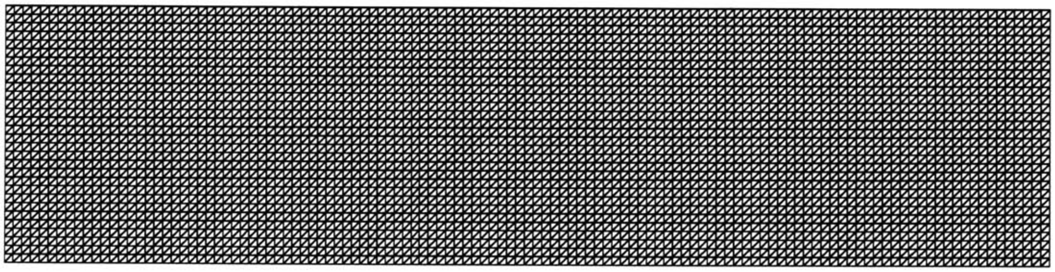


Figure 4-17: AP II: Finite element truth approximation mesh.

We present in Figures 4-18 and 4-19 snapshots of the concentration over Ω at eight timesteps for the two extreme values of the diffusivity — $\mu = 0.01$ and $\mu = 1.00$ — in the parameter set $\mathcal{D}$, respectively. For $\mu = 0.01$ the flow is clearly advection dominated, whereas for $\mu = 1.00$ the diffusive terms dominate. We also show, in Figures 4-20 and 4-21, the outputs, i.e., concentration readings at the 8 measurement locations, for the two parameter values $\mu = 0.01$ and $\mu = 1.00$, respectively. We first note that the measured concentrations are considerably lower for $\mu = 1.00$ because of the diffusive effects. On the other side, as already seen in Figure 4-18, the pollution "cloud" stays more compact for $\mu = 0.01$ thus resulting in higher concentration outputs. Comparing Figures 4-18 and 4-20 we can in effect track the pollution cloud as it moves through the domain — the pollution is first registered at sensor 1, then sensor 2, sensor 4, and so on. However, due to the small amount of diffusion present in the system the peak concentration in the outputs does decrease over time.

Reduced-Basis Approximation

We generate the sample set $S_{N_{\mathrm{pr}}}^{\mathrm{pr}}$ and associated reduced basis space $W_{N_{\mathrm{pr}}}^{\mathrm{pr}}$ according to the adaptive sampling procedure described in Section 4.5. We initialize the procedure with $\tilde{\mu}_1^{\mathrm{pr}} = (\mu_1^{\mathrm{pr}}, t^{k_1^{\mathrm{pr}}}) =$

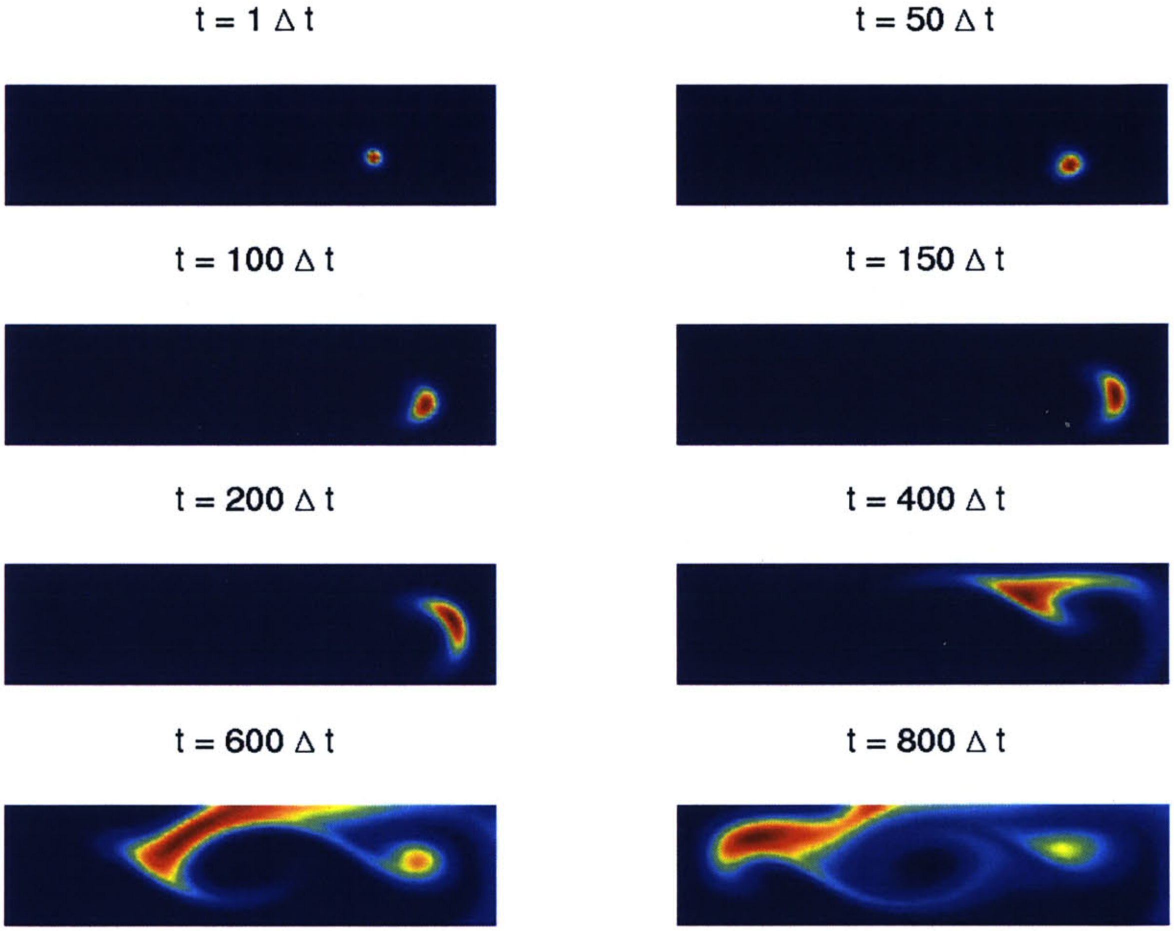


Figure 4-18: AP II: Concentration $c(\mu = 0.01, t^k)$ at $t = t^1$, t^{50}, t^{100}, t^{150}, t^{200}, t^{400}, t^{600}, t^{800}.

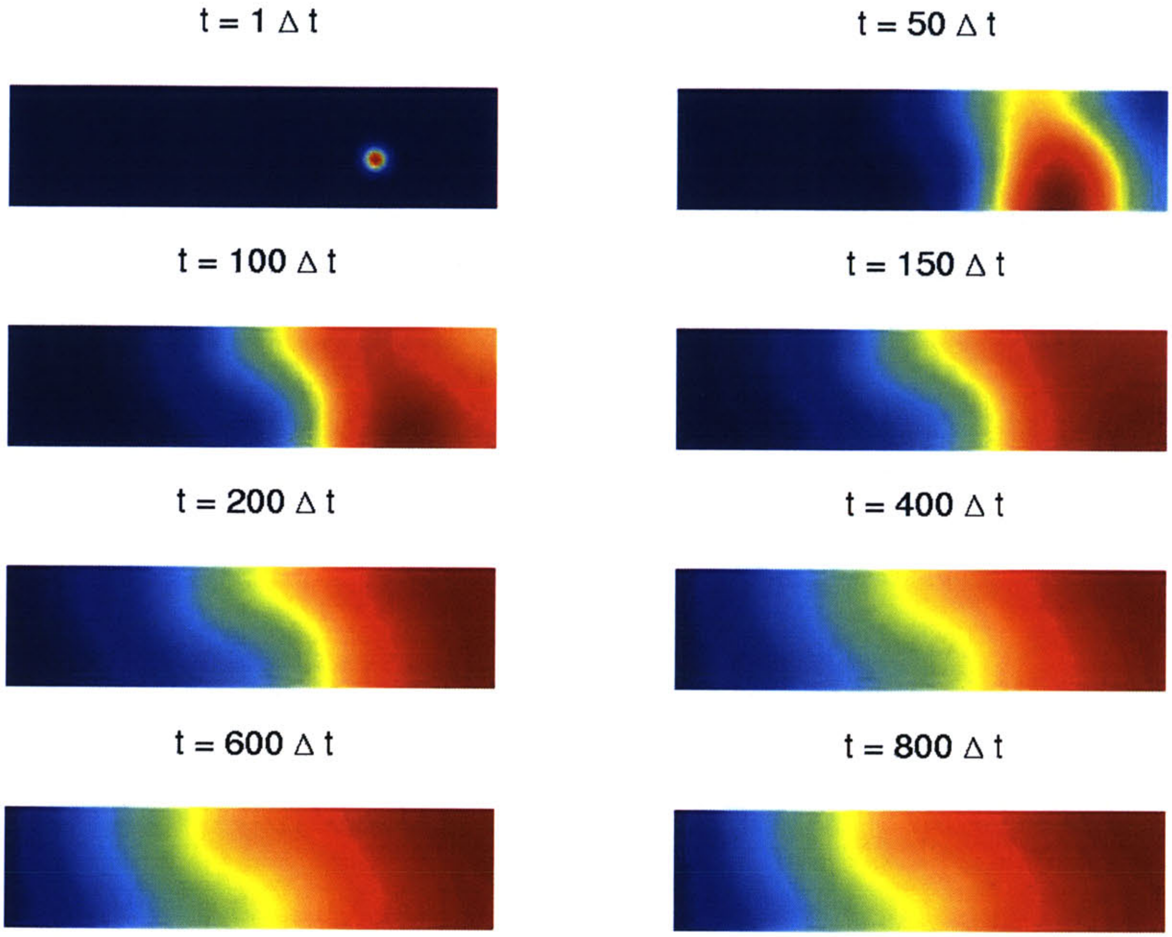


Figure 4-19: AP II: Concentration $c(\mu = 1.00, t^k)$ at $t = t^1$, t^{50}, t^{100}, t^{150}, t^{200}, t^{400}, t^{600}, t^{800}.

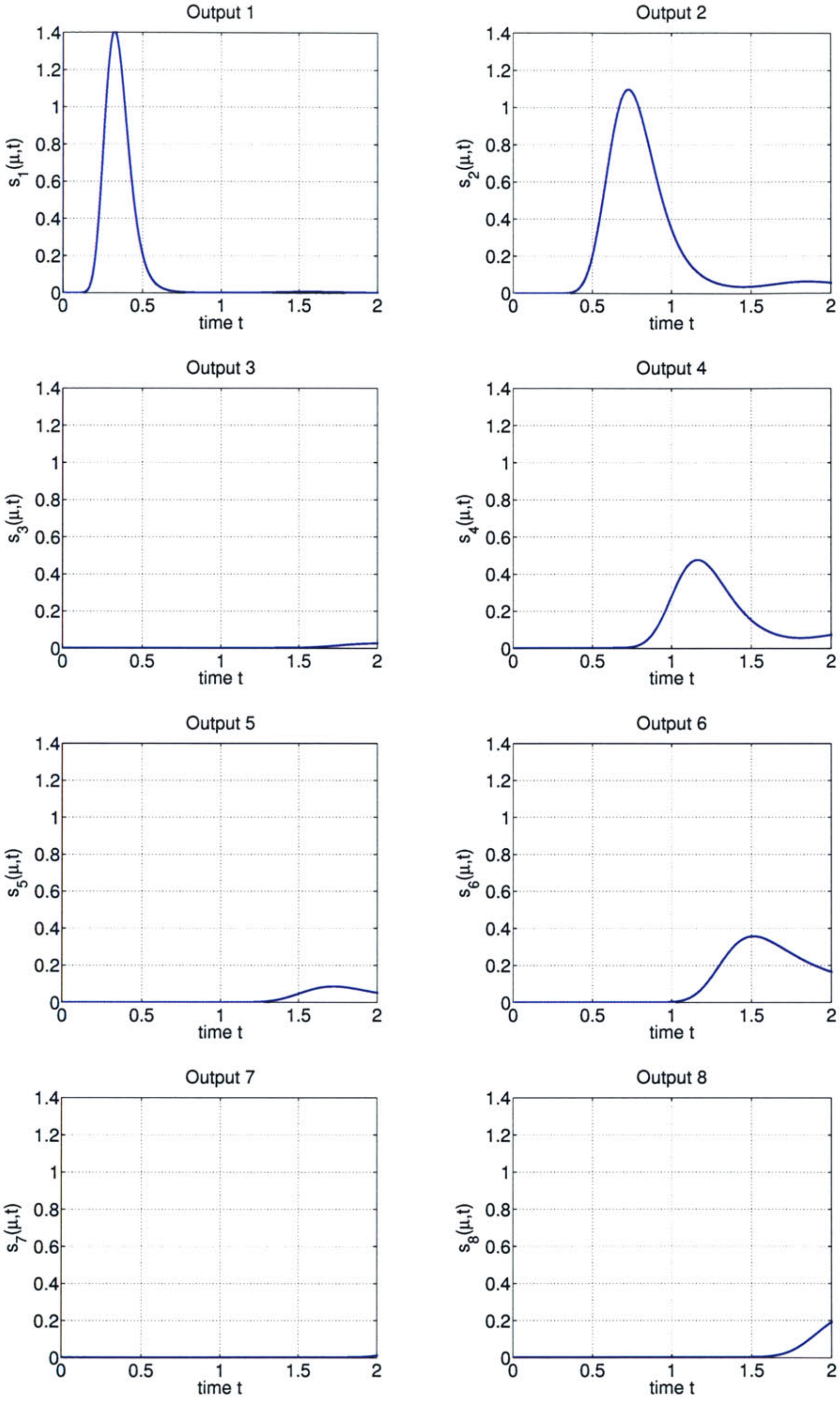


Figure 4-20: AP II: Outputs $s_q(\mu = 0.01, t^k)$, $1 \leq q \leq 8$, as a function of time.

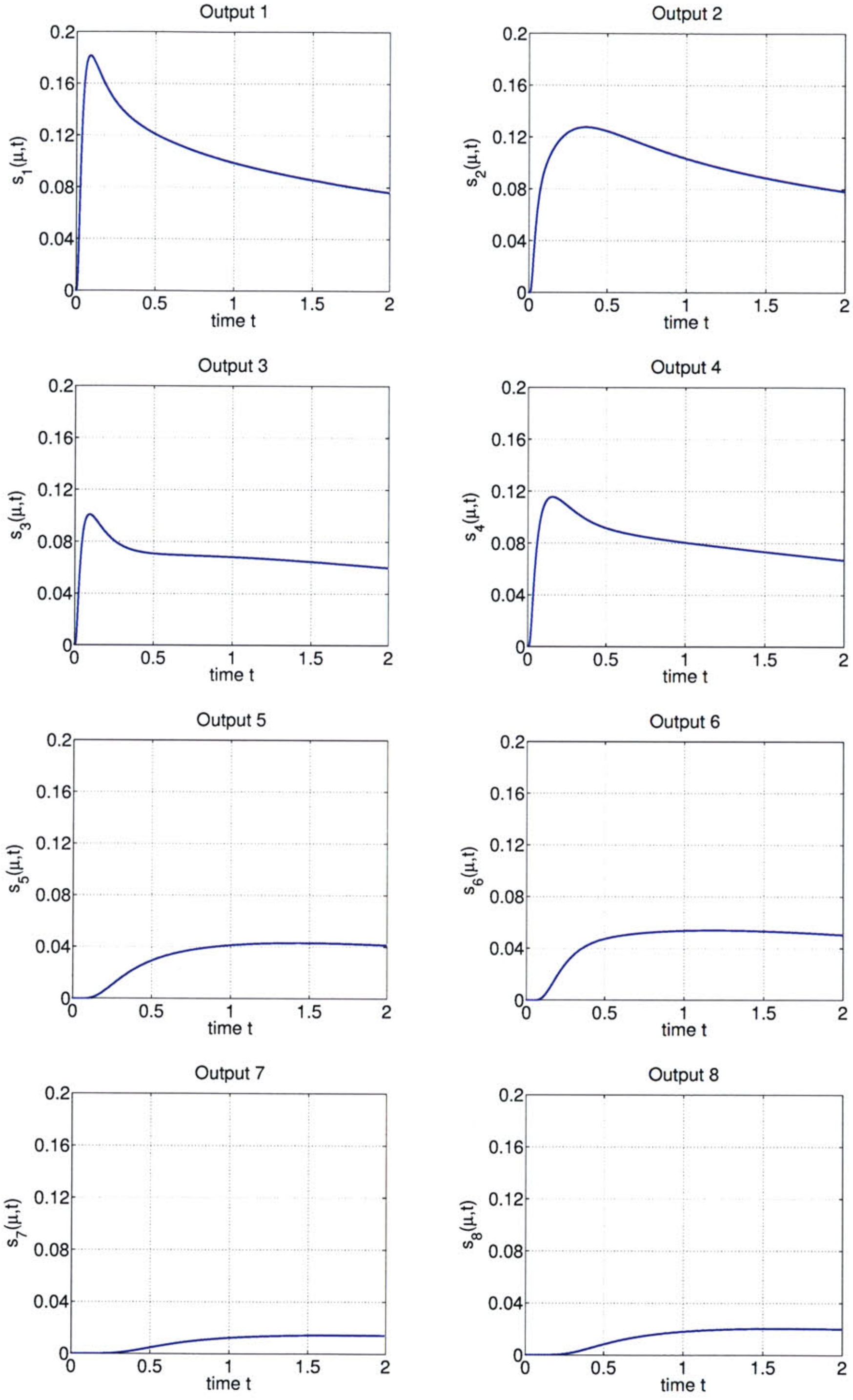


Figure 4-21: AP II: Outputs $s_q(\mu = 1.00, t^k)$, $1 \leq q \leq 8$, as a function of time.

$(0.01, 1\Delta t)$ and set the desired error tolerance (for the relative error in the energy norm) to $\epsilon_{\text{tol,min}} = 1\,\text{E}-4$. We sample on a log-random parameter test sample $\Xi_{\text{F}} \in (\mathcal{D})^{60}$ of size 60 — we need $N_{\text{pr,max}} = 207$ basis functions to obtain the desired accuracy. We also generate the sample sets $S^{\text{du},1}_{N_{\text{du},1}}$ and $S^{\text{du},2}_{N_{\text{du},2}}$ and corresponding reduced-basis spaces $W^{\text{du},1}_{N_{\text{du},1}}$ and $W^{\text{du},2}_{N_{\text{du},2}}$ for the two dual problems corresponding to the two output functionals $\ell_q(v)$, $q = 1, 2$, respectively; we obtain (for $\epsilon_{\text{tol,min}} = 1\,\text{E}-4$ and $\tilde{\mu}^{\text{du}}_1 = (\mu^{\text{du}}_1, t^{k^{\text{du}}_1}) = (0.01, 801\Delta t)$): $N_{\text{du,1,max}} = 225$, $N_{\text{du,2,max}} = 200$ [7].

We plot the sample sets $S^{\text{pr}}_{N_{\text{pr}}}$ and $S^{\text{du},1}_{N_{\text{du},1}}$ in $\mu - t^k$-space (note the logarithmic scale in μ) in Figure 4-22(a) and (b), respectively. The plots reflect the fact that the primal evolves forward in time whereas the dual evolves backward in time — $S^{\text{pr}}_{N_{\text{pr}}}$ and $S^{\text{du},1}_{N_{\text{du},1}}$ are biased towards the beginning and end of the time interval, respectively. We also note that the primal and dual sample sets are biased towards smaller μ values — convection dominates diffusion in this parameter range (see Figures 4-18 and 4-19) and the problem is thus more complicated. We could *a priori* expect this behavior from Propositions 5 and 10: a more "nonsymmetric" (i.e., complicated) problem results in a slower convergence rate of the reduced-basis approximation and thus more samples are required to obtain the desired accuracy. We also confirm this result numerically in Figure 4-23: we generate two separate reduced-basis approximations for the time histories at two *fixed* parameter values, $\mu = 0.01$ and $\mu = 1.00$, i.e., we consider a problem where time is the only varying parameter (as is usually the case in POD). We plot the convergence rates of the relative errors, $|||e(\mu, t^k)|||^{\text{pr}} / |||y(\mu, t^k)|||$, as a function of N in Figure 4-23. The convergence is slower for the smaller μ-value and more samples are required to obtain a specific desired accuracy.

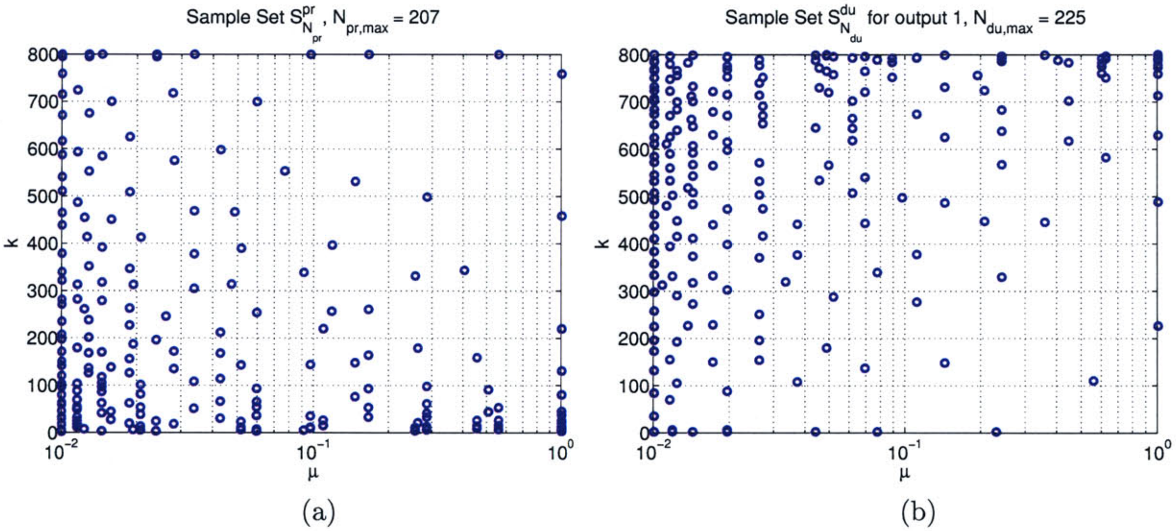


Figure 4-22: AP II: (a) Sample set $S^{\text{pr}}_{N_{\text{pr}}}$ with $N_{\text{pr,max}} = 207$, and (b) sample set $S^{\text{du}}_{N_{\text{du}}}$ for the first output, s_1, with $N^1_{\text{du,max}} = 225$.

[7] The dual problems corresponding to the remaining outputs give similar results so we restrict our attention to only the first two outputs

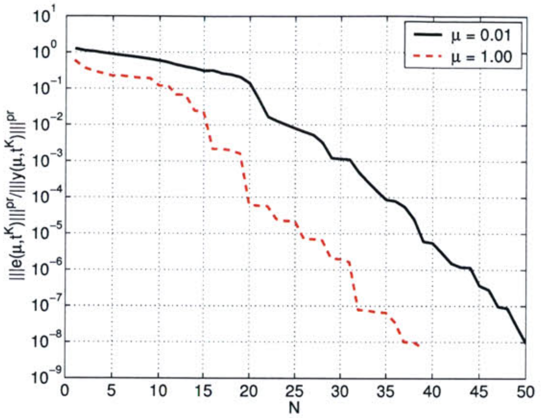


Figure 4-23: AP II: Convergence rate of the relative error $\frac{|||e(\mu,t^k)|||^{\rm pr}}{|||y(\mu,t^k)|||}$ for $\mu = 0.01$ and $\mu = 1.00$.

Numerical Results

We now present convergence results and effectivities for the primal problem, the dual problems corresponding to the first and second output, and the output estimate (for the definitions of the quantities presented see Sections 4.6.2 and 4.8.4). Here, the parameter test sample, $\Xi_{\rm Test} \in (\mathcal{D})^{50}$, is a log-random sample of size 50.

In Table 4.14 we present, as a function of $N_{\rm pr}$, the maximum relative error, $\epsilon^{\rm pr}_{\rm max,rel}$, the maximum relative error bound $\Delta^{\rm pr}_{\rm max,rel}$, and the average effectivity $\overline{\eta}^{\rm pr}$; for the dual problems 1 and 2 we tabulate $\epsilon^{\rm du}_{\rm max,rel}$, $\Delta^{\rm du}_{\rm max,rel}$, and $\overline{\eta}^{\rm du}$ as a function of $N_{\rm du}$ in Table 4.15(a) and (b), respectively. We observe that the primal and dual reduced-basis approximations converge very fast. The effectivities are very good, $O(1)$, for both the primal and dual problem. We note that the effectivities are slightly larger for small diffusivities $O(3)$ and smaller, $O(1)$, for diffusivities close to $\mu = 1$. In general, we would expect the bound conditioner to perform better for small diffusivities and worse for larger diffusivities. However, for small diffusivities the convective term is dominant and the solution is far from elliptic which usually results in larger effectivities. We would thus expect overall higher effectivities than observed here.

We next turn to the convergence of the output estimate. We plot in Figures 4-24 and 4-25(a) the maximum relative output error, $\epsilon^{s}_{\rm max,rel}$, and in Figures 4-24 and 4-25(b) the maximum relative output bound, $\Delta^{s}_{\rm max,rel}$, as a function of $N_{\rm pr}$ and $N_{\rm du}$ for output 1 and 2, respectively. As expected, the error (and bound) decreases for fixed $N_{\rm du}$ as $N_{\rm pr}$ increases; similarly, for fixed $N_{\rm pr}$ the error (and bound) decreases as $N_{\rm du}$ increases.

We also tabulate the relative maximum output error, output bound and output effectivities, as a function of $N_{\rm pr} = N_{\rm du}$, for output 1 and 2 in Tables 4.16(a) and (b), respectively. We observe the square effect in the output error and output bound: $\Delta^{s}_{N}(\mu, t^k)$ converges roughly as the square of $\Delta^{\rm pr}_{N}(\mu, t^k)$. However, the effectivities are considerably larger, $O(100)$, because our bound cannot take into account any correlation between the primal and dual error. In Table 4.17(a) and (b) we present the maximum relative output error, output bound, and effectivity for output 1 and 2 using the simple bound of Proposition 9 (for $\hat{\alpha}_m(\mu) = 1$). The convergence of the output error and bound is now only $O(\Delta^{\rm pr}_{N}(\mu, t^k))$ and thus considerably slower; the effectivities, on the other side, are only $O(10 - 100)$. To obtain an accuracy of the bound for output 1 of one percent, we require

N_{pr}	$\epsilon^{\mathrm{pr}}_{\mathrm{max,rel}}$	$\Delta^{\mathrm{pr}}_{\mathrm{max,rel}}$	$\overline{\eta}^{\mathrm{pr}}$
20	3.96 E-01	2.45 E+00	5.20
40	1.39 E-01	5.04 E-01	2.77
60	4.60 E-02	1.08 E-01	2.17
80	1.95 E-02	3.60 E-02	2.03
100	1.17 E-02	1.46 E-02	1.85
120	3.03 E-03	3.91 E-03	1.74
140	7.61 E-04	1.37 E-03	1.66
160	6.01 E-04	7.17 E-04	1.62
180	1.74 E-04	2.88 E-04	1.58
200	9.59 E-05	1.33 E-04	1.60

Table 4.14: AP II: Convergence rate and effectivities for primal problem.

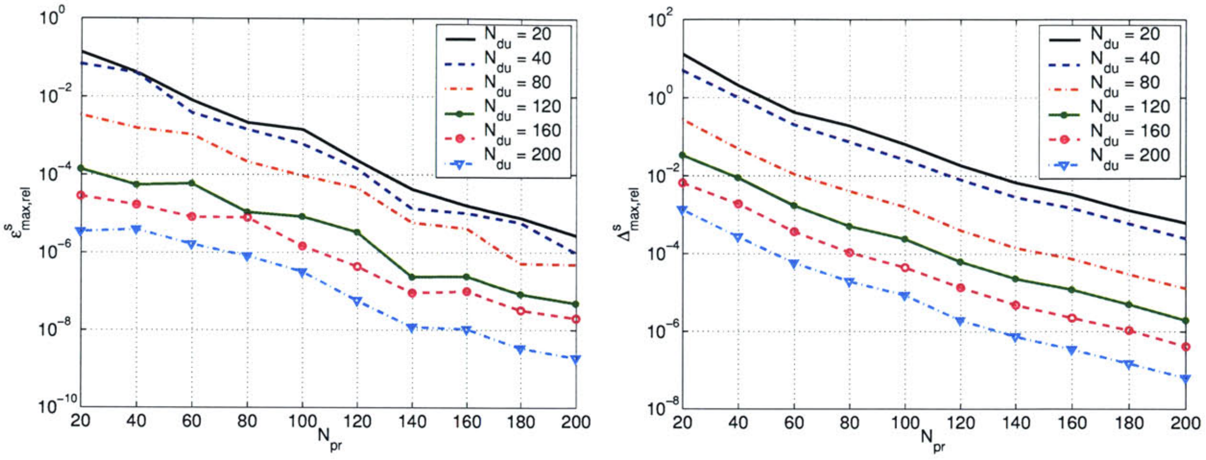
N_{du}	$\epsilon^{\mathrm{du}}_{\mathrm{max,rel}}$	$\Delta^{\mathrm{du}}_{\mathrm{max,rel}}$	$\overline{\eta}^{\mathrm{du}}$
20	4.23 E-01	1.85 E+00	4.58
40	1.75 E-01	7.51 E-01	3.93
60	7.66 E-02	1.55 E-01	2.43
80	3.43 E-02	4.38 E-02	2.06
100	1.22 E-02	1.54 E-02	1.90
120	4.54 E-03	6.35 E-03	1.82
140	1.53 E-03	2.04 E-03	1.73
160	1.10 E-03	1.31 E-03	1.69
180	3.77 E-04	5.39 E-04	1.62
200	1.97 E-04	2.25 E-04	1.62

(a)

N_{du}	$\epsilon^{\mathrm{du}}_{\mathrm{max,rel}}$	$\Delta^{\mathrm{du}}_{\mathrm{max,rel}}$	$\overline{\eta}^{\mathrm{du}}$
20	4.20 E-01	2.07 E+00	5.14
40	1.88 E-01	3.92 E-01	2.84
60	6.66 E-02	1.13 E-01	2.66
80	2.08 E-02	3.41 E-02	1.95
100	7.32 E-03	9.86 E-03	1.77
120	2.39 E-03	5.11 E-03	1.76
140	7.27 E-04	1.09 E-03	1.66
160	4.57 E-04	5.81 E-04	1.63
180	2.03 E-04	2.31 E-04	1.60
200	5.91 E-05	1.01 E-04	1.56

(b)

Table 4.15: AP II: Convergence rate and effectivities for dual problem corresponding to (a) output 1 and (b) output 2.



(a)

(b)

Figure 4-24: AP II: (a) Maximum relative error $\varepsilon^{s}_{\max,\mathrm{rel}}$ and (b) maximum relative output bound $\Delta^{s}_{\max,\mathrm{rel}}$ for output 1 using the dual formulation.

approximately $N_{\mathrm{pr}} = N_{\mathrm{du}} = 70$ for the primal-dual formulation and approximately $N_{\mathrm{pr}} = 140$ for the simple bound.

We next turn to the computational efficiency of the proposed method. We recall that the computational cost is $O(N_{\mathrm{pr}}^3 + KN_{\mathrm{pr}}^2)$ to solve for $\underline{y}_N(\mu, t^k)$, $O(2KN_{\mathrm{pr}})$ to evaluate the simple output estimate, $\hat{s}_N(\mu, t^k)$, $\forall k \in \mathbb{K}$, and $O(K(K+1)N_{\mathrm{pr}}N_{\mathrm{du}})$ to evaluate $s_N(\mu, t^k)$, $\forall k \in \mathbb{K}$, using the primal-dual formulation. The computational cost involved thus depends strongly on the total number of timesteps K: if $K \ll N_{\mathrm{pr}}, N_{\mathrm{du}}$, the $O(N_{\mathrm{pr}}^3)$-term dominates and — given a certain desired accuracy — introducing the dual problem can result in computational savings of up to $O(4)$. In most cases, however, $K \geq N_{\mathrm{pr}}, N_{\mathrm{du}}$, and the computation cost due to the residual correction term, $O(K(K+1)N_{\mathrm{pr}}N_{\mathrm{du}})$, is likely to dominate. We already observed this behaviour in Section 4.7.2 for the delamination problem for small N_{pr} and N_{du}.

We present in Table 4.19, as a function of $N_{\mathrm{pr}}(= N_{\mathrm{du}})$, the online computational times to calculate $s_N(\mu, t^k)$ and $\Delta^s_N(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, for output 1. The values are normalized with respect to the computational time for the direct calculation of the truth approximation output $s(\mu, t^k) = \ell(u(\mu, t^k))$, $\forall k \in \mathbb{K}$. We note that the computational saving are moderate due to the K^2-complexity of the residual correction term. We thus define $\overline{\mathbb{K}} = \{10, 20, 30, \ldots, K\}$ and present in Table 4.20 the online computational times to calculate $s_N(\mu, t^k)$, $\forall\, k \in \overline{\mathbb{K}}$ and $\Delta^s_N(\mu, t^k)$, $\forall\, k \in \overline{\mathbb{K}}$. Again, the online time to calculate $\Delta^s_N(\mu, t^k)$ remains unchanged. However, we observe — especially for small $N_{\mathrm{pr}} = N_{\mathrm{du}}$ — up to $O(10)$ reduction in computational effort to evaluate $s_N(\mu, t^k)$, $\forall\, k \in \overline{\mathbb{K}}$. For small $N_{\mathrm{pr}} = N_{\mathrm{du}}$ the computational effort is dominated by the residual correction term with the $O(K(K+1)N_{\mathrm{pr}}N_{\mathrm{du}}/10)$ complexity, for larger $N_{\mathrm{pr}} = N_{\mathrm{du}}$ this effect is less obvious because of the $O(N_{\mathrm{pr}}^3)$ complexity to solve for $\underline{y}_N(\mu, t^k)$.

We compare these results with the online computational times to calculate the simple output estimate and output bound, $\hat{s}_N(\mu, t^k)$ and $\hat{\Delta}^s_N(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, presented in Table 4.21. The online times are smaller since the computation does not involve the solution of the dual problem. We also note that these times are (effectively) independent off the number of outputs considered. To obtain an accuracy for the output bound of one percent, we require approximately $N_{\mathrm{pr}} = N_{\mathrm{du}} = 70$ for the

N_{pr}	N_{du}	$\epsilon^{s}_{\mathrm{max,rel}}$	$\Delta^{s}_{\mathrm{max,rel}}$	$\overline{\eta}^{s}$
20	20	1.37 E−01	1.30 E+01	1908
40	40	4.09 E−02	1.02 E+00	150
60	60	3.07 E−03	4.62 E−02	182
80	80	2.12 E−04	4.01 E−03	185
100	100	3.87 E−05	6.37 E−04	229
120	120	3.40 E−06	6.49 E−05	119
140	140	1.22 E−07	7.51 E−06	150
160	160	9.99 E−08	2.24 E−06	175
180	180	4.55 E−09	3.71 E−07	140
200	200	1.98 E−09	6.72 E−08	110

(a)

N_{pr}	N_{du}	$\epsilon^{s}_{\mathrm{max,rel}}$	$\Delta^{s}_{\mathrm{max,rel}}$	$\overline{\eta}^{s}$
20	20	1.60 E−01	1.87 E+01	3096
40	40	4.13 E−02	6.47 E−01	1722
60	60	3.45 E−03	4.30 E−02	209
80	80	3.11 E−04	3.52 E−03	160
100	100	3.58 E−05	4.77 E−04	206
120	120	1.01 E−06	6.63 E−05	160
140	140	9.77 E−08	5.14 E−06	185
160	160	3.10 E−08	1.53 E−06	174
180	180	6.14 E−09	2.18 E−07	111
200	200	5.58 E−10	4.61 E−08	180

(b)

Table 4.16: AP II: Convergence rate and effectivities using dual formulation for (a) output 1 and (b) output 2.

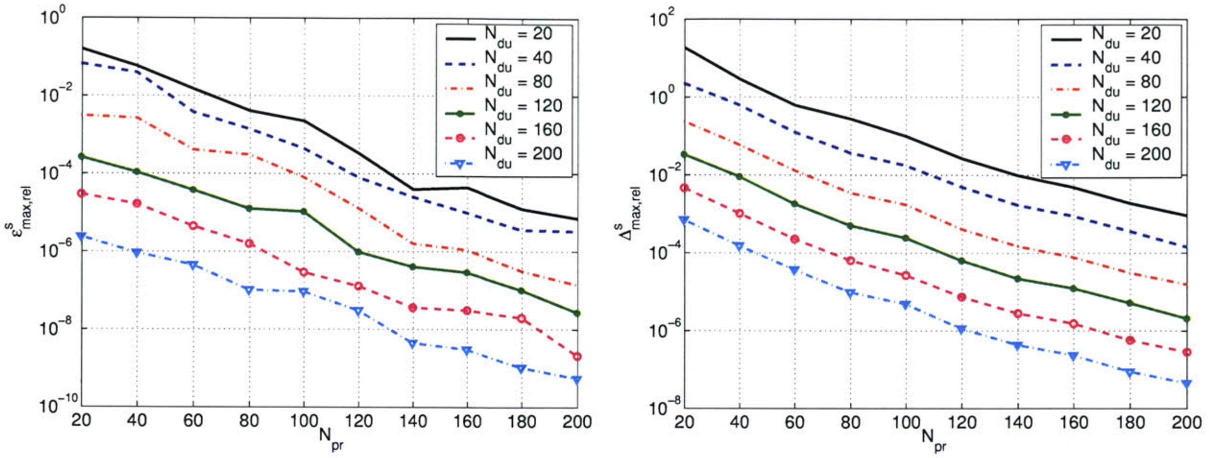
N_{pr}	$\epsilon^{\hat{s}}_{\mathrm{max,rel}}$	$\Delta^{\hat{s}}_{\mathrm{max,rel}}$	$\overline{\eta}^{\hat{s}}$
20	2.84 E−01	9.86 E+00	41.8
40	4.62 E−02	2.03 E+00	41.1
60	1.13 E−02	4.35 E−01	55.2
80	2.46 E−03	1.45 E−01	48.9
100	1.69 E−03	5.86 E−02	42.3
120	2.52 E−04	1.57 E−02	78.5
140	1.40 E−04	5.49 E−03	107
160	3.43 E−05	2.88 E−03	146
180	1.50 E−05	1.16 E−03	113
200	1.11 E−05	5.37 E−04	148

(a)

N_{pr}	$\epsilon^{\hat{s}}_{\mathrm{max,rel}}$	$\Delta^{\hat{s}}_{\mathrm{max,rel}}$	$\overline{\eta}^{\hat{s}}$
20	1.69 E−01	1.27 E+01	99.1
40	5.76 E−02	2.60 E+00	47.9
60	1.61 E−02	5.58 E−01	37.7
80	4.75 E−03	1.86 E−01	32.7
100	2.77 E−03	7.53 E−02	43.0
120	4.53 E−04	2.02 E−02	46.4
140	1.06 E−04	7.06 E−03	64.1
160	9.53 E−05	3.70 E−03	56.1
180	3.38 E−05	1.49 E−03	46.8
200	1.92 E−05	6.89 E−04	37.1

(b)

Table 4.17: AP II: Convergence rate and effectivities using simple bounds for (a) output 1 and (b) output 2.



(a)

(b)

Figure 4-25: AP II: (a) Maximum relative error $\varepsilon^s_{\mathrm{max,rel}}$ and (b) maximum relative output bound $\Delta^s_{\mathrm{max,rel}}$ for output 2 using the dual formulation.

primal-dual formulation and $N_{\mathrm{pr}} = 140$ for the simple bound. The computational savings compared to the underlying finite element truth approximation are a factor of 40 for the simple bound, a factor of 15 for the primal-dual formulation with the output evaluated for all $k \in \mathbb{K}$, and a factor of 70 for the primal-dual formulation with the output evaluated for all $k \in \overline{\mathbb{K}}$. The actual run-times to compute the output estimate and output bound in MATLAB 6.5 on a 750 MHz Pentium III are 2.59 sec. (simple bound), 6.25 sec. (primal-dual, $k \in \mathbb{K}$), and 1.35 sec. (primal-dual, $k \in \overline{\mathbb{K}}$). We note that the computational saving observed here are smaller than in Section 4.7.2 because the dimension of the truth approximation, $\mathcal{N}$, is smaller. We also note that these results are for a single output — were we to consider several outputs, the online computational time for the primal-only approach would remain the same, whereas the computational time for the primal-dual formulation would increase with the number of outputs.

Let us now assume that the acceptable accuracy in the output bound is 10% instead of 1%. We observe from Table 4.16 and 4.17 that $N_{\mathrm{pr}} = N_{\mathrm{du}} = 50$ and $N_{\mathrm{pr}} = 90$ are now sufficient for the primal-dual and primal-only approach, respectively. The computational savings are now a factor of 110 for the simple bound, a factor of 19 for the primal-dual formulation with the output evaluated for all $k \in \mathbb{K}$, and a factor of 90 for the primal-dual formulation with the output evaluated for all $k \in \overline{\mathbb{K}}$ — the simple bound is now clearly preferable. The run-times are now 0.85 sec., 5.02 sec., and 1.06 sec., respectively.

Finally, we remark that the decision about employing the primal-dual formulation for the output bound or the simple (primal-only) output bound usually depends on the specific problem. The simple bound is advantageous if (*i*) K is large, i.e., $K \gg N_{\mathrm{pr}}$, and the output estimate has to be evaluated for all $k \in \mathbb{K}$ — in this case the $O(K(K+1)N_{\mathrm{pr}}N_{\mathrm{du}})$ complexity is detriment to the computational efficiency, and (*ii*) if many outputs are required and thus as many dual problems have to be evaluated.

N_{pr}	N_{du}	$\epsilon^{s}_{\mathrm{max,rel}}$	$\Delta^{s}_{\mathrm{max,rel}}$	$\overline{\eta}^{s}$
20	20	2.84 E−01	1.30 E+01	39.4
40	40	4.62 E−02	1.03 E+00	15.7
60	60	1.13 E−02	5.27 E−02	5.18
80	80	2.46 E−03	5.65 E−03	2.18
100	100	1.69 E−03	2.25 E−03	1.32
120	120	2.52 E−04	2.90 E−04	1.18
140	140	1.40 E−04	1.41 E−04	1.07
160	160	3.43 E−05	3.57 E−05	1.08
180	180	1.50 E−05	1.53 E−05	1.02
200	200	1.11 E−05	1.11 E−05	1.01

(a)

N_{pr}	N_{du}	$\epsilon^{s}_{\mathrm{max,rel}}$	$\Delta^{s}_{\mathrm{max,rel}}$	$\overline{\eta}^{s}$
20	20	1.69 E−01	1.87 E+01	120
40	40	5.76 E−02	6.55 E−01	8.41
60	60	1.61 E−02	4.77 E−02	2.44
80	80	4.75 E−03	8.20 E−03	1.40
100	100	2.77 E−03	3.15 E−03	1.22
120	120	4.53 E−04	4.67 E−04	1.08
140	140	1.06 E−04	1.09 E−04	1.03
160	160	9.53 E−05	9.55 E−05	1.01
180	180	3.38 E−05	3.39 E−05	1.00
200	200	1.92 E−05	1.92 E−05	1.00

(b)

Table 4.18: AP II: Convergence rate and effectivities for output estimate and output bound defined in (4.122) and (4.123), respectively: (a) output 1 and (b) output 2.

$N_{\mathrm{pr}} = N_{\mathrm{du}}$	$s_N(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$\Delta^s_N(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$s(\mu, t^k), \forall\, k \in \mathbb{K}$
20	4.33 E−02	2.10 E−03	1
40	5.22 E−02	2.95 E−03	1
60	5.70 E−02	4.00 E−03	1
80	6.27 E−02	7.86 E−03	1
100	6.78 E−02	1.63 E−02	1
120	7.78 E−02	2.63 E−02	1
140	8.73 E−02	3.53 E−02	1
160	9.44 E−02	4.56 E−02	1
180	1.04 E−01	5.74 E−02	1
200	1.16 E−01	7.39 E−02	1

Table 4.19: AP II: Online computational times to calculate $s_N(\mu, t^k)$ and $\Delta^s_N(\mu, t^k)$ for all $k \in \mathbb{K}$ (normalized with respect to the time to solve for $s(\mu, t^k), \forall\, k \in \mathbb{K}$).

$N_{\text{pr}} = N_{\text{du}}$	$s_N(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$\Delta^s_N(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$s(\mu, t^k), \forall\, k \in \mathbb{K}$
20	5.34 E-03	2.10 E-03	1
40	7.00 E-03	2.95 E-03	1
60	8.53 E-03	4.00 E-03	1
80	1.03 E-02	7.86 E-03	1
100	1.35 E-02	1.63 E-02	1
120	1.96 E-02	2.63 E-02	1
140	2.73 E-02	3.53 E-02	1
160	3.36 E-02	4.56 E-02	1
180	4.14 E-02	5.74 E-02	1
200	5.10 E-02	7.39 E-02	1

Table 4.20: AP II: Online computational times to calculate $s_N(\mu, t^k)$ and $\Delta^s_N(\mu, t^k)$ for all $k \in \overline{\mathbb{K}}$ (normalized with respect to the time to solve for $s(\mu, t^k), \forall\, k \in \mathbb{K}$).

N_{pr}	$\hat{s}_N(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$\hat{\Delta}^s_N(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$s(\mu, t^k), \forall\, k \in \mathbb{K}$
20	8.63 E-04	9.30 E-04	1
40	1.72 E-03	1.13 E-03	1
60	2.82 E-03	1.28 E-03	1
80	4.13 E-03	2.10 E-03	1
100	7.02 E-03	5.38 E-03	1
120	1.23 E-02	7.91 E-03	1
140	1.69 E-02	1.00 E-02	1
160	2.10 E-02	1.30 E-02	1
180	2.66 E-02	1.59 E-02	1
200	3.45 E-02	1.99 E-02	1

Table 4.21: AP II: Online computational times to calculate $\hat{s}_N(\mu, t^k)$ and $\hat{\Delta}^s_N(\mu, t^k)$ (normalized with respect to the time to solve for $s(\mu, t^k), \forall\, k \in \mathbb{K}$).

Chapter 5

Nonaffine Linear Parabolic Equations

5.1 Introduction

In Chapter 4 we developed the reduced-basis method and associated *a posteriori* error estimation for linear parabolic problems with affine parameter dependence. Based on the affine assumption we introduced very efficient offline-online computational procedure relevant to the many query or real-time context.

Unfortunately, if the affine parameter dependence is not met, this computational strategy breaks down; the online complexity will still depend on $\mathcal{N}$. For example, for *general* $g(x;\mu)$ (here $x \in \Omega$ and $\mu \in \mathcal{D}$) the bilinear form

$$a(w, v; g(x;\mu)) \equiv \int_\Omega \nabla w \cdot \nabla v + \int_\Omega g(x;\mu)\, w\, v \tag{5.1}$$

will not admit an efficient, i.e., online $\mathcal{N}$-independent, computational decomposition. In a recent note Barrault *et al.* [15] introduce a technique that recovers the efficient offline-online decomposition even in the presence of nonaffine parameter dependence — we briefly reviewed the empirical interpolation method in Section 2.4. In this approach, the authors develop a "collateral" reduced-basis expansion $g_M(x;\mu)$ for $g(x;\mu)$ and then replace $g(x;\mu)$ in (5.1) with the (necessarily) affine approximation $g_M(x;\mu) = \sum_{m=1}^{M} \varphi_{M\,m}(\mu) q_m(x)$. The essential ingredients are (*i*) a "good" collateral reduced-basis approximation space, $W_M^g = \text{span}\{q_m(x), 1 \leq m \leq M\}$, of dimension M, (*ii*) a stable and inexpensive interpolation procedure by which to determine the $\varphi_{M\,m}(\mu), 1 \leq m \leq M$, and (*iii*) an effective *a posteriori* estimator with which to quantify the newly introduced error terms.

We now apply this technique and extend the results of the previous chapter to parabolic problems with nonaffine parameter dependence, i.e., where g is a nonaffine function of the parameter μ and spatial coordinate x; we will consider nonlinear problems in Chapter 6. Since the primary focus here is the treatment of the nonaffine terms, we do not consider adjoint formulations in this chapter.

5.2 Abstract Formulation

We directly consider a time-discrete framework associated to the time interval $I \equiv]0, t_f]$. We recall that $\bar{I}$ is divided into K subintervals of equal length $\Delta t = \frac{t_f}{K}$, that t^k is defined by $t^k \equiv k\Delta t$, $0 \leq k \leq K \equiv \frac{t_f}{\Delta t}$; furthermore, $\mathbb{I} \equiv \{t^0, \ldots, t^k\}$ and $\mathbb{K} \equiv \{1, \ldots, K\}$. We shall consider Euler-Backward for the time integration. We also recall our reference (or "truth") finite element approximation space Y of very large dimension $\mathcal{N}$. Clearly, our results must be stable as $\Delta t \to 0$, $K \to \infty$, and $\mathcal{N} \to \infty$.

We may now directly consider our "truth" finite element approximation: given a parameter $\mu \in \mathcal{D}$, we evaluate the (here, single) output of interest

$$s(\mu, t^k) = \ell(y(\mu, t^k)), \quad \forall\, k \in \mathbb{K}, \tag{5.2}$$

where the field variable $y(\mu, t^k) \in Y$, $\forall\, k \in \mathbb{K}$, satisfies the nonaffine parabolic partial differential equation

$$m(y(\mu, t^k), v) + \Delta t\ a(y(\mu, t^k), v; g(x;\mu)) = m(y(\mu, t^{k-1}), v) + \Delta t\ b(v; h(x;\mu))\ u(t^k), \quad \forall\, v \in Y, \tag{5.3}$$

with initial condition (say) $y(\mu, t^0) = y_0(\mu) = 0$. Here, μ and $\mathcal{D}$ are the input and input domain; $m(\cdot,\cdot)$ is a X-continuous bilinear form; $a(\cdot,\cdot; g(x;\mu))$ is a Y-continuous linear operator; $b(\cdot; h(x;\mu))$ and $\ell(\cdot)$ are X-continuous linear forms; and $u(t^k)$ denotes the (here, single) "control input" at time $t = t^k$. Note that a and b depend on $g(\cdot;\mu) \in L^\infty(\Omega)$ and $h(\cdot;\mu) \in L^\infty(\Omega)$; we further assume that these functions are continuous in the closed domain $\bar{\Omega}$ and sufficiently smooth with respect to $\mu \in \mathcal{D}$. We shall suppose that a is of the form

$$a(w, v; g(x;\mu)) = a_0(w, v) + a_1(w, v, g(x;\mu)), \tag{5.4}$$

where $a_0(\cdot,\cdot)$ is a continuous (and, for simplicity, parameter-independent) bilinear form and $a_1(\cdot,\cdot,g(\cdot))$ is a trilinear form. For simplicity of exposition, we assume here that $h(x;\mu) = g(x;\mu)$ and also that m and ℓ do not depend on the parameter.

We shall make the following assumptions. We assume that $a(\cdot,\cdot; g(x;\mu))$ and $m(\cdot,\cdot)$ are continuous

$$\begin{aligned} a(w, v; g(x;\mu)) &\leq \gamma_a(\mu)\|w\|_Y\|v\|_Y \leq \gamma_a^0\|w\|_Y\|v\|_Y, \quad \forall\, w, v \in Y,\ \forall\, \mu \in \mathcal{D}, && (5.5)\\ m(w, v) &\leq \gamma_m\|w\|_X\|v\|_X, \quad \forall\, w, v \in Y; && (5.6) \end{aligned}$$

coercive,

$$\begin{aligned} 0 &< \alpha_a^0 \leq \alpha_a(\mu) \equiv \inf_{w \in X} \frac{a(w, w; g(x;\mu))}{\|w\|_X^2}, \quad \forall\, \mu \in \mathcal{D}, && (5.7)\\ 0 &< \alpha_m \equiv \inf_{v \in Y} \frac{m(v, v)}{\|v\|_X^2}; && (5.8) \end{aligned}$$

and symmetric, $a(v, w; g(x;\mu)) = a(w, v; g(x;\mu))$, $\forall\, v, w \in Y$, $\forall\, \mu \in \mathcal{D}$, and $m(v, w) = m(w, v)$, $\forall\, w, v \in Y$, $\forall\, \mu \in \mathcal{D}$. (We (plausibly) suppose that γ_a^0, γ_m, α_a^0, α_m may be chosen independent of

$\mathcal{N}$.) We also assume that the trilinear form a_1 satisfies

$$a_1(w,v,z) \leq \gamma_{a_1} \|w\|_X \, \|v\|_X \, \|z\|_{L^\infty(\Omega)}, \quad \forall\, w,v \in Y. \tag{5.9}$$

Next, we require that the linear forms $b(\cdot\,; h(x;\mu)) : Y \to \mathbb{R}$ and $\ell(\cdot) : Y \to \mathbb{R}$ be bounded with respect to $\|\cdot\|_X$. And finally, we require that all linear and bilinear forms are independent of time — the system is thus linear time-invariant (LTI). It follows, given that $g(\cdot\,;\mu) \in L^\infty(\Omega)$, that a solution to (5.3) exists and is unique [94].

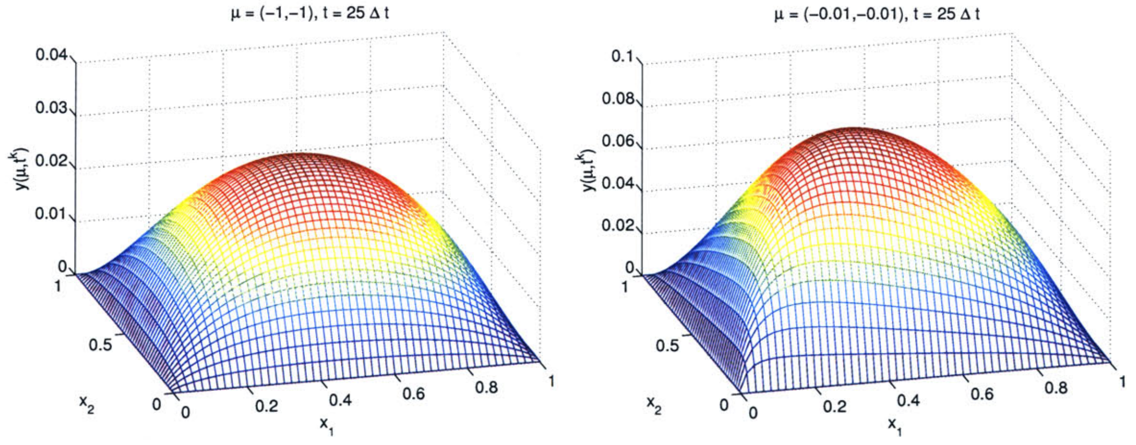
5.2.1 Numerical Exercise 4: A Nonaffine Diffusion Problem

As a numerical example we consider the following nonaffine diffusion problem defined on the unit square, $\Omega =]0,1[^2 \in \mathbb{R}^2$: Given $\mu \equiv (\mu_1,\mu_2) \in \mathcal{D} \equiv [-1,-0.01]^2 \subset \mathbb{R}^{P=2}$, we evaluate $y(\mu,t^k) \in Y$ from (5.3), where $Y \subset Y^{\mathrm{e}} \equiv H_0^1(\Omega)$ is a linear finite element truth approximation subspace of dimension $\mathcal{N} = 2601$,

$$m(w,v) \equiv \int_\Omega w\,v, \quad a_0(w,v) \equiv \int_\Omega \nabla w \cdot \nabla v, \quad a_1(w,v,z) \equiv \int_\Omega z\,w\,v, \quad b(v;z) \equiv \int_\Omega z\,v, \tag{5.10}$$

and $z = G(x;\mu)$ is the (nonaffine) function defined in (2.36). The output can be written in the form (5.2), $s(\mu,t^k) = \ell(y(\mu,t^k))$, $\forall\, k \in \mathbb{K}$, where $\ell(v) \equiv |\Omega|^{-1} \int_\Omega v$ — clearly a very smooth functional. We shall consider the time interval $\bar{I} = [0,2]$ and a timestep $\Delta t = 0.01$; we thus have $K = 200$. We also presume the periodic control input $u(t^k) = \sin(2\pi t^k)$, $t^k \in \mathbb{I}$.

Two snapshots of the solution $y(\mu,t^k)$ at time $t^k = 25\Delta t$ are shown in Figures 5-1(a) and (b) for $\mu = (-1,-1)$ and $\mu = (-0.01,-0.01)$, respectively. The solution oscillates in time and the peak is offset towards $x = (0,0)$ for μ near the "corner" $(-0.01,-0.01)$.



(a) (b)

Figure 5-1: Solution $y(\mu,t^k)$ at $t^k = 25\Delta t$ for (a) $\mu = (-1,-1)$ and (b) $\mu = (-0.01,-0.01)$.

5.3 Reduced-Basis Approximation

5.3.1 Formulation

We first introduce the nested sample sets $S^y_N = \{\tilde{\mu}^y_1 \in \tilde{\mathcal{D}}, \ldots, \tilde{\mu}^y_N \in \tilde{\mathcal{D}}\}$, $1 \leq N \leq N_{\max}$, where $\tilde{\mu} \equiv (\mu, t^k)$ and $\tilde{\mathcal{D}} \equiv \mathcal{D} \times \mathbb{I}$. We then define the associated nested Lagrangian [85] reduced-basis space

$$W^y_N = \operatorname{span}\{\zeta_n \equiv y(\tilde{\mu}^y_n),\ 1 \leq n \leq N\}, \quad 1 \leq N \leq N_{\max}, \tag{5.11}$$

where $y(\tilde{\mu}^y_n)$ is the solution of (5.3) at time $t = t^{k^y_n}$ for $\mu = \mu^y_n$.

Before we proceed, let us first motivate the need for the empirical interpolation approach in dealing with nonaffine problems: were we to follow the classical recipe, the reduced-basis approximation would be obtained by a standard Galerkin projection: given $\mu \in \mathcal{D}$, we evaluate

$$s_N(\mu, t^k) = \ell(y_N(\mu, t^k)), \quad \forall\, k \in \mathbb{K}, \tag{5.12}$$

where $y_N(\mu, t^k) \in W^y_N$ satisfies

$$\begin{aligned} m(y_N(\mu, t^k), v) &+ \Delta t\ a(y_N(\mu, t^k), v; g(x; \mu)) \\ &= m(y_N(\mu, t^{k-1}), v) + \Delta t\ b(v; g(x; \mu))\ u(t^k), \quad \forall\, v \in W^y_N,\ \forall\, k \in \mathbb{K}, \end{aligned} \tag{5.13}$$

with initial condition $y_N(\mu, t^0) = 0$. We now express $y_N(\mu, t^k) = \sum_{j=1}^N y_{N\,j}(\mu, t^k)\ \zeta_j$ and choose as test functions $v = \zeta_n$, $1 \leq n \leq N$ in (5.3) to obtain, $\forall\, k \in \mathbb{K}$,

$$\begin{aligned} \sum_{j=1}^N \left\{ m(\zeta_i, \zeta_j) + \Delta t\ \left(a_0(\zeta_i, \zeta_j) + a_1(\zeta_i, \zeta_j, g(x; \mu))\right)\right\} y_{N\,j}(\mu, t^k) \\ = \sum_{j=1}^N m(\zeta_i, \zeta_j)\ y_{N\,j}(\mu, t^{k-1}) + \Delta t\ b(\zeta_i; g(x; \mu))\ u(t^k), \quad 1 \leq i \leq N. \end{aligned} \tag{5.14}$$

We observe that while $m(\zeta_i, \zeta_j)$ and $a_0(\zeta_i, \zeta_j)$ are parameter-independent and can thus be pre-computed offline, $b(\zeta_i; g(x; \mu))$ and $a_1(\zeta_i, \zeta_j, g(x; \mu))$ depend on $g(x; \mu)$ and must therefore be evaluated online for every new parameter value μ. The operation count for the online stage will thus scale as $O(N^2\mathcal{N})$, where $\mathcal{N}$ is the dimension of the underlying truth finite element approximation space: the reduction in marginal cost gain obtained in moving from the truth finite element approximation space to the reduced-basis space will be quite modest regardless of the dimension reduction.

To recover online $\mathcal{N}$-independence, we appeal to the empirical interpolation method discussed in Section 2.4. We simply replace $g(x; \mu)$ in (5.13) with the (necessarily) affine approximation $g_M(x; \mu) = \sum_{m=1}^M \varphi_{M\,m}(\mu) q_m(x)$ from (5). We thus construct the nested samples $S^g_M = \{\mu^g_1 \in \mathcal{D}, \cdots, \mu^g_M \in \mathcal{D}\}$, $1 \leq M \leq M_{\max}$, associated nested approximation spaces $W^g_M = \operatorname{span}\{\xi_m \equiv g(\mu^g_m),\ 1 \leq m \leq M\} = \operatorname{span}\{q_m,\ 1 \leq m \leq M\}$, $1 \leq M \leq M_{\max}$, and nested sets of interpolation points $T_M = \{t_1, \ldots, t_M\}$, $1 \leq M \leq M_{\max}$, following the procedure of Section 2.4. Our reduced-basis approximation $y_{N,M}(\mu, t^k)$ to $y(\mu, t^k)$ is then: given $\mu \in \mathcal{D}$, $y_{N,M}(\mu, t^k) \in W_N^y$ satisfies

$$\begin{aligned} m(y_{N,M}(\mu,t^k), v) &+ \Delta t\ a(y_{N,M}(\mu,t^k), v; g_M(x;\mu)) \\ &= m(y_{N,M}(\mu,t^{k-1}), v) + \Delta t\ b(v; g_M(x;\mu))\ u(t^k), \quad \forall\, v \in W_N^y,\ \forall\, k \in \mathbb{K}, \end{aligned} \tag{5.15}$$

with initial condition $y_{N,M}(\mu, t^0) = 0$. We then evaluate the output estimate, $s_{N,M}(\mu, t^k)$, from

$$s_{N,M}(\mu, t^k) \equiv \ell(y_{N,M}(\mu, t^k)), \quad \forall\, k \in \mathbb{K}. \tag{5.16}$$

We now express $y_{N,M}(\mu, t^k) = \sum_{n=1}^{N} y_{N,M\,n}(\mu, t^k)\ \zeta_n$, choose as test functions $v = \zeta_n$, $1 \le n \le N$, and invoke (2.29) to obtain

$$\begin{aligned} \sum_{j=1}^{N} \left\{ m(\zeta_i, \zeta_j) + \Delta t \left(a_0(\zeta_i, \zeta_j) + + \sum_{m=1}^{M} \varphi_{M\,m}(\mu)\, a_1(\zeta_i, \zeta_j, q_m) \right) \right\}\ y_{N,M\,j}(\mu, t^k) \\ = \sum_{j=1}^{N} m(\zeta_i, \zeta_j)\ y_{N,M\,j}(\mu, t^{k-1}) + \Delta t \sum_{m=1}^{M} \varphi_{M\,m}(\mu)\, b(\zeta_i; q_m)\ u(t^k), \quad 1 \le i \le N. \end{aligned} \tag{5.17}$$

where $\varphi_{M\,m}(\mu)$, $1 \le m \le M$, is determined from (2.30). We indeed recover the online $\mathcal{N}$-independence: the quantities $m(\zeta_i, \zeta_j)$, $a_0(\zeta_i, \zeta_j)$, $a_1(\zeta_i, \zeta_j, q_m)$, and $b(\zeta_i; q_m)$ are all *parameter independent* and can thus be pre-computed offline, as discussed further in Section 5.3.3.

Note that we construct the parameter-time sample set S_N^y and associated reduced-basis space W_N^y using the adaptive sampling procedure described in Section 4.5. During the sampling process we shall use the "best" possible approximation $g_M(x;\mu)$ of $g(x;\mu)$ so as to minimize the error induced by the empirical interpolation procedure, i.e., we set $M = M_{\max}$.

5.3.2 *A Priori* Convergence Theory

We consider here the rate at which $y_{N,M}(\mu, t^k)$ converges to $y(\mu, t^k)$. To this end, we first define

$$\psi_1(\mu) \equiv \frac{1}{\varepsilon_M(\mu)} \sup_{v \in Y} \frac{b(v; g(\cdot;\mu) - g_M(\cdot;\mu))}{\|v\|_Y}, \tag{5.18}$$

$$\psi_2(\mu) \equiv \frac{1}{\varepsilon_M(\mu)} \sup_{w \in Y} \sup_{v \in Y} \frac{a(w, v; g(\cdot;\mu) - g_M(\cdot;\mu))}{\|w\|_Y \|v\|_Y}, \tag{5.19}$$

$$\psi_3(\mu) \equiv \sup_{v \in Y} \frac{b(w, v; g_M(\cdot;\mu))}{\|v\|_Y}. \tag{5.20}$$

where $\varepsilon_M(\mu)$ is the interpolation error defined in (2.31). We can prove

Proposition 14. *For $\varepsilon_M(\mu)$ of (2.31) satisfying $\varepsilon_M(\mu) < \alpha_a(\mu)/(4\,\psi_2(\mu))$ (say), the error $e(\mu, t^k) \equiv$*

$y(\mu, t^k) - y_{N,M}(\mu, t^k)$ *satisfies*

$$\alpha_m \, \|e(\mu, t^k)\|_X^2 + \frac{\alpha_a(\mu)}{2} \, \Delta t \sum_{k'=1}^{k} \|e(\mu, t^{k'})\|_Y^2 \le \Upsilon(\mu) \, \Delta t \sum_{k'=1}^{k} u(t^{k'})^2$$
$$+ \inf_{w_N(t^k) \in W_N^y} \left\{ \gamma_m \, \|y(\mu, t^k) - w_N(t^k)\|_X^2 + \Delta t \, (\gamma_a(\mu) + 2\,\alpha_a(\mu)) \, \|y(\mu, t^k) - w_N(t^k)\|_Y^2 \right\}$$
$$+ \Delta t \, (\gamma_a(\mu) + 2\,\alpha_a(\mu)) \sum_{k'=1}^{k} \inf_{w_N(t^{k'}) \in W_N^y} \|y(\mu, t^{k'}) - w_N(t^{k'})\|_Y^2, \quad (5.21)$$

where

$$\Upsilon(\mu) = \frac{5}{\alpha_a(\mu)} \, \varepsilon_M(\mu)^2 \left(\psi_1(\mu)^2 + \psi_2(\mu)^2 \frac{2\,\psi_3(\mu)^2}{\alpha_a(\mu)^2} \right).$$

Proof. To begin, we note from (5.3) and (5.15) that

$$m(e(\mu, t^k) - e(\mu, t^{k-1}), v) + \Delta t \; a(e(\mu, t^k), v; g(x; \mu))$$
$$= \Delta t \left(b(v; g(x; \mu) - g_M(x; \mu)) \; u(t^k) - a(y_{N,M}(\mu, t^k), v; g(x; \mu) - g_M(x; \mu)) \right), \; \forall \, v \in W_N^y. \quad (5.22)$$

with initial condition $e(\mu, t^0) = 0$, since $y(\mu, t^0) = y_{N,M}(\mu, t^0) = 0$ by assumption. Let $w_N(t^k) \in W_N^y$ be the projection of $y(\mu, t^k)$ with respect to the "m" scalar product and choose $v \equiv w_N(t^k) - y_N(\mu, t^k) = e(\mu, t^k) - (y(\mu, t^k) - w_N(t^k))$ in (5.22). Following the same steps as in the proof of Proposition 5 we obtain

$$\begin{aligned} & m(e(\mu, t^k), e(\mu, t^k)) - m(e(\mu, t^{k-1}), e(\mu, t^{k-1})) + \Delta t \; a(e(\mu, t^k), e(\mu, t^k); g(x; \mu)) \\ & \quad \le \; m(y(\mu, t^k) - w_N(t^k), y(\mu, t^k) - w_N(t^k)) \\ & \qquad - m(y(\mu, t^{k-1}) - w_N(t^{k-1}), y(\mu, t^{k-1}) - w_N(t^{k-1})) \\ & \qquad + \Delta t \; a(y(\mu, t^k) - w_N(t^k), y(\mu, t^k) - w_N(t^k); g(x; \mu)) \\ & \qquad + 2\,\Delta t \left(b(v; g(x; \mu) - g_M(x; \mu)) \; u(t^k) - a(y_{N,M}(\mu, t^k), v; g(x; \mu) - g_M(x; \mu)) \right), \quad (5.23) \end{aligned}$$

which after summing from $k' = 1$ to k leads to

$$\begin{aligned} & m(e(\mu, t^k), e(\mu, t^k)) + \Delta t \sum_{k'=1}^{k} a(e(\mu, t^{k'}), e(\mu, t^{k'}); g(x; \mu)) \\ & \quad \le \; m(y(\mu, t^k) - w_N(t^k), y(\mu, t^k) - w_N(t^k)) \\ & \qquad + \Delta t \sum_{k'=1}^{k} a(y(\mu, t^{k'}) - w_N(t^{k'}), y(\mu, t^{k'}) - w_N(t^{k'}); g(x; \mu)) \\ & \qquad + 2\,\Delta t \sum_{k'=1}^{k} \left(\psi_1(\mu) \, |u(t^{k'})| + \psi_2(\mu) \, \|y_{N,M}(\mu, t^{k'})\|_Y \right) \varepsilon_M(\mu) \, \|v\|_Y, \quad (5.24) \end{aligned}$$

where the last inequality follows from (5.18) and (5.19). We now note that $\|v\|_Y \le \|y(\mu, t^k) - w_N\|_Y + \|e(\mu, t^k)\|_Y$ and recall the identity (4.30) which we apply four times: first, with $c =$

$\varepsilon_M(\mu)\,\psi_1(\mu)\,|u(t^k)|$, $d = \|y(\mu,t^k) - w_N(t^k)\|_Y$, and $\rho^2 = \alpha_a(\mu)$; second, with $c = \varepsilon_M(\mu)\,\psi_1(\mu)\,|u(t^k)|$, $d = \|e(\mu,t^k)\|_Y$, and $\rho^2 = \alpha_a(\mu)/4$; third, with $c = \varepsilon_M(\mu)\,\psi_2(\mu)\,\|y_{N,M}(\mu,t^k)\|_Y$, $d = \|y(\mu,t^k) - w_N(t^k)\|_Y$, and $\rho^2 = \alpha_a(\mu)$; and fourth, with $c = \varepsilon_M(\mu)\,\psi_2(\mu)\,\|y_{N,M}(\mu,t^k)\|_Y$, $d = \|e(\mu,t^k)\|_Y$, and $\rho^2 = \alpha_a(\mu)/4$. We can then bound the last term of (5.24) by

$$\begin{aligned}
2\,\Delta t \sum_{k'=1}^{k} &\left(\psi_1(\mu)\,|u(t^{k'})| + \psi_2(\mu)\,\|y_{N,M}(\mu,t^{k'})\|_Y\right) \varepsilon_M(\mu)\,\|v\|_Y \\
&\leq\; \varepsilon_M(\mu)^2\,\frac{5}{\alpha_a(\mu)}\left(\psi_1(\mu)^2\,\Delta t \sum_{k'=1}^{k} u(t^{k'})^2 + \psi_2(\mu)^2 \Delta t \sum_{k'=1}^{k} \|y_{N,M}(\mu,t^{k'})\|_Y^2\right) \\
&\qquad +2\,\Delta t\,\alpha_a(\mu) \sum_{k'=1}^{k} \|y(\mu,t^k) - w_N(t^k)\|_X^2 + \Delta t \frac{\alpha_a(\mu)}{2} \sum_{k'=1}^{k} \|e(\mu,t^{k'})\|_Y^2. \qquad (5.25)
\end{aligned}$$

We next obtain from (5.15) with $v = y_{N,M}(\mu,t^k)$, invoking the Cauchy-Schwarz inequality for the cross-term $m(y_{N,M}(\mu,t^k), y_{N,M}(\mu,t^{k-1}))$ and applying (4.30) with $c = m^{1/2}(y_{N,M}(\mu,t^k), y_{N,M}(\mu,t^k))$, $d = m^{1/2}(y_{N,M}(\mu,t^{k-1}), y_{N,M}(\mu,t^{k-1}))$, and $\rho = 1$, that

$$\begin{aligned}
m(y_{N,M}(\mu,t^k), &y_{N,M}(\mu,t^k)) - m(y_{N,M}(\mu,t^{k-1}), y_{N,M}(\mu,t^{k-1})) \\
&+2\,\Delta t\, a(y_{N,M}(\mu,t^k), y_{N,M}(\mu,t^k); g(x;\mu)) \\
&\qquad\leq\; 2\,\Delta t\, b(y_{N,M}(\mu,t^k); g_M(x;\mu))\, u(t^k) \\
&\qquad\qquad +2\,\Delta t\, a(y_{N,M}(\mu,t^k), y_{N,M}(\mu,t^k); g(x;\mu) - g_M(x;\mu)) \\
&\qquad\leq\; 2\,\Delta t\,\psi_3(\mu)\,\|y_{N,M}(\mu,t^k)\|_Y\,|u(t^k)| + 2\,\Delta t\,\varepsilon_M(\mu)\,\psi_2(\mu)\,\|y_{N,M}(\mu,t^k)\|_Y^2 \\
&\qquad\leq\; \frac{\Delta t}{\alpha_a(\mu) - 2\,\psi_2(\mu)\,\varepsilon_M(\mu)}\psi_3(\mu)^2\, u(t^k)^2 + \Delta t\,\alpha_a(\mu)\,\|y_{N,M}(\mu,t^k)\|_Y^2, \qquad (5.26)
\end{aligned}$$

where the second inequality follows from (5.19) and (5.20), and the last inequality from (4.30) with $c = \psi_3(\mu)\,u(t^k)$, $d = \|y_{N,M}(\mu,t^k)\|_Y$, and $\rho = \alpha_a(\mu) - 2\,\psi_2(\mu)\,\varepsilon_M(\mu)$; note that $\rho > 0$ from our assumption on $\varepsilon_M(\mu)$. Invoking (5.5) and summing from $k' = 1$ to k we obtain

$$\begin{aligned}
m(y_{N,M}(\mu,t^k), y_{N,M}(\mu,t^k)) + \Delta t \sum_{k'=1}^{k} a(y_{N,M}(\mu,t^{k'}), y_{N,M}(\mu,t^{k'}); g(x;\mu)) \\
\leq \frac{\psi_3(\mu)^2}{\alpha_a(\mu) - 2\,\psi_2(\mu)\,\varepsilon_M(\mu)}\,\Delta t \sum_{k'=1}^{k} u(t^k)^2. \qquad (5.27)
\end{aligned}$$

From the coercivity of m and a, and our assumption on $\varepsilon_M(\mu)$ it then directly follows that

$$\begin{aligned}
\Delta t \sum_{k'=1}^{k} \|y_{N,M}(\mu,t^{k'})\|_X^2 \;&\leq\; \frac{\psi_3(\mu)^2}{\alpha_a(\mu)(\alpha_a(\mu) - 2\,\psi_2(\mu)\,\varepsilon_M(\mu))}\,\Delta t \sum_{k'=1}^{k} u(t^k)^2 \qquad (5.28) \\
&\leq\; \frac{2\,\psi_3(\mu)^2}{\alpha_a(\mu)^2}\,\Delta t \sum_{k'=1}^{k} u(t^k)^2. \qquad (5.29)
\end{aligned}$$

From (5.24) and invoking (5.25) and (5.29) we obtain

$$
\begin{aligned}
m(e(\mu,t^k),e(\mu,t^k)) + \Delta t \sum_{k'=1}^{k} a(e(\mu,t^{k'}),e(\mu,t^{k'});g(x;\mu)) \\
\le \; & m(y(\mu,t^k)-w_N(t^k),y(\mu,t^k)-w_N(t^k)) \\
& +\Delta t \sum_{k'=1}^{k} a(y(\mu,t^{k'})-w_N(t^{k'}),y(\mu,t^{k'})-w_N(t^{k'});g(x;\mu)) \\
& +2\,\Delta t\,\alpha_a(\mu) \sum_{k'=1}^{k} \|y(\mu,t^{k'})-w_N(t^{k'})\|_X^2 + \Delta t \frac{\alpha_a(\mu)}{2} \sum_{k'=1}^{k} \|e(\mu,t^{k'})\|_Y^2 \\
& +\Upsilon(\mu)\,\Delta t \sum_{k'=1}^{k} u(t^{k'})^2, \qquad (5.30)
\end{aligned}
$$

where

$$
\Upsilon(\mu) = \frac{5}{\alpha_a(\mu)}\,\varepsilon_M(\mu)^2 \left(\psi_1(\mu)^2 + \psi_2(\mu)^2 \frac{2\,\psi_3(\mu)^2}{\alpha_a(\mu)^2} \right).
$$

The desired result then directly follows from the continuity and coercivity of m and a.

□

We note from Proposition 14 that M should be chosen such that $\epsilon_M(\mu)$ is of the same order as the best-fit error, otherwise the term $\Upsilon(\mu)$ may limit the convergence of the reduced-basis approximation. We will observe a similar requirement for M and the nonaffine function approximation error $\epsilon_M(\mu)$ when we discuss *a posteriori* error estimation in Section 5.4.

As regards the best approximation, we note that W_N^y comprises "snapshots" on the ($P+1$ dimensional) manifold $\mathcal{M}^y \equiv \{y(\mu,t^k) | \forall\, (\mu,t^k) \in \tilde{\mathcal{D}}\}$ induced by the parametric and temporal dependence. The critical observation is that $\mathcal{M}^y$ is very *low-dimensional* and *smooth* under our hypotheses on stability and continuity — the proof follows the same lines as in Section 4.3.2 for the affine case. We thus expect that the best approximation will converge to $y(\mu,t^k)$ very rapidly, and hence that N may be chosen small.

5.3.3 Offline-Online Computational Procedure

In this section we develop the offline-online computational procedure in order to fully exploit the dimension reduction of the problem [9, 48, 60, 91]. We first express $y_{N,M}(\mu,t^k)$ as

$$
y_{N,M}(\mu,t^k) = \sum_{n=1}^{N} y_{N,M\,n}(\mu,t^k)\;\zeta_n, \qquad (5.31)
$$

and choose as test functions $v=\zeta_n$, $1\le n\le N$ in (5.15). (We prefer Galerkin over Petrov-Galerkin for purposes of stability.) It then follows from (5.17) that $\underline{y}_{N,M}(\mu,t^k) = [y_{N,M\,1}(\mu,t^k)\;\; y_{N,M\,2}(\mu,t^k)\ldots y_{N,M\,N}(\mu,t^k)]^T \in \mathbb{R}^N$ satisfies

$$
(M_N + \Delta t\; A_N(\mu))\;\underline{y}_{N,M}(\mu,t^k) = M_N\;\underline{y}_{N,M}(\mu,t^{k-1}) + \Delta t\;F_N(\mu)\;u(t^k), \quad \forall\, k \in \mathbb{K}, \qquad (5.32)
$$

with initial condition $y_{N,M\,n}(\mu, t^0) = 0$, $1 \leq n \leq N$. Given $\underline{y}_{N,M}(\mu, t^k), \forall\, k \in \mathbb{K}$, we finally evaluate the output estimate from

$$s_{N,M}(\mu, t^k) = L_N^T\, \underline{y}_{N,M}(\mu, t^k), \quad \forall\, k \in \mathbb{K}. \tag{5.33}$$

Here, $M_N \in \mathbb{R}^{N \times N}$ is a *parameter-independent* SPD matrix with entries

$$M_{N\,i,j} = m(\zeta_i, \zeta_j), \quad 1 \leq i,j \leq N. \tag{5.34}$$

Furthermore, we obtain from (2.29) and (5.4) that $A_N(\mu) \in \mathbb{R}^{N \times N}$ and $F_N(\mu) \in \mathbb{R}^N$ can be expressed as

$$\begin{aligned} A_N(\mu) &= A_{0,N} + \sum_{m=1}^{M} \varphi_{M\,m}(\mu)\; A_{1,N}^m, &(5.35)\\ F_N(\mu) &= \sum_{m=1}^{M} \varphi_{M\,m}(\mu)\; F_N^m, &(5.36)\end{aligned}$$

where $\varphi_{M\,m}(\mu)$, $1 \leq m \leq M$, is calculated from (2.30), and the *parameter-independent* quantities $A_{0,N} \in \mathbb{R}^{N \times N}$, $A_{1,N}^m \in \mathbb{R}^{N \times N}$, and $F_N^m \in \mathbb{R}^N$ are given by

$$\begin{aligned} A_{0,N\,i,j} &= a_0(\zeta_i, \zeta_j), & &1 \leq i,j \leq N,\\ A_{1,N\,i,j}^m &= a_1(\zeta_i, \zeta_j, q_m), & &1 \leq i,j \leq N,\ 1 \leq m \leq M,\\ F_{N\,i}^m &= b(\zeta_i; q_m), & &1 \leq i \leq N,\ 1 \leq m \leq M, \end{aligned} \tag{5.37}$$

respectively. Finally, $L_N \in \mathbb{R}^N$ is the output vector with entries $L_{N\,i} = \ell(\zeta_i)$, $1 \leq i \leq N$.

The offline-online decomposition is now clear. In the offline stage — performed only *once* — we first construct the nested approximation spaces W_M^g and sets of interpolation points T_M, $1 \leq M \leq M_{\max}$; we then solve for the ζ_n, $1 \leq n \leq N_{\max}$ and compute and store the μ-independent quantities in (5.34), (5.37), and L_N. The computational cost — without taking into account the construction of W_M^g and T_M — is therefore $O(KN_{\max})$ solutions of the underlying $\mathcal{N}$-dimensional "truth" finite element approximation and $O(M_{\max}N_{\max}^2)$ $\mathcal{N}$-inner products; the storage requirements are also $O(M_{\max}N_{\max}^2)$. In the online stage — performed many times, for each new parameter value μ — we first compute $\varphi_M(\mu)$ from (2.30) at cost $O(M^2)$ by multiplying the pre-computed inverse of B^M with the vector $g(t_i; \mu)$, $1 \leq i \leq M$; we then assemble the reduced-basis matrix (5.35) and vector (5.36); this requires $O(MN^2)$ operations. We then solve (5.32) for $\underline{y}_{N,M}(\mu, t^k)$; since the reduced-basis matrices are in general full, the operation count (based on LU factorization and our LTI assumption) is $O(N^3 + KN^2)$. Finally, given $\underline{y}_{N,M}(\mu, t^k)$ we evaluate the output estimate $s_{N,M}(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, from (5.33) at a cost of $O(KN)$.

Hence, as required in the many-query or real-time contexts, the online complexity is *independent* of $\mathcal{N}$, the dimension of the underlying "truth" finite element approximation space. Since $N, M \ll \mathcal{N}$ we expect significant computational savings in the online stage relative to classical discretization and solution approaches (and relative to standard reduced-basis approaches built upon (5.13)).

5.3.4 Implementation Issues

At this point we need to comment on an important issue concerning the actual numerical implementation of our proposed method which, if not addressed properly, can lead to erroneous results. We first note that solving the "truth" finite element approximation (5.3) for $y(\mu, t^k)$ necessitates the integration of terms of the form $\int_\Omega v\, g(x;\mu)$ (and $\int_\Omega w\, v\, g(x;\mu)$) — which (usually) have to be evaluated by (say) Gaussian quadrature:

$$\int_\Omega v\ g(x;\mu) \approx \sum_{j=1}^{\mathcal{N}_{\mathrm{QP}}} \omega_j\, v(x_j^{\mathrm{QP}})\ g(x_j^{\mathrm{QP}};\mu), \tag{5.38}$$

where the ω_j are the elemental Gauss-Legendre quadrature weights, x_j^{QP} are the corresponding elemental quadrature points, and $\mathcal{N}_{\mathrm{QP}}$ is the total number of quadrature points. Similarly, the reduced-basis approximation procedure requires, during the offline stage, the evaluation of (say) $\int_\Omega \zeta_i\, q_m$. For consistency, the term $\int_\Omega \zeta_i\, q_m$ should be evaluated using the same quadrature rule (5.38) that was used to develop the "truth" finite element approximation,

$$\int_\Omega \zeta_i\ q_m \approx \sum_{j=1}^{\mathcal{N}_{\mathrm{QP}}} \omega_j\, \zeta_i(x_j^{\mathrm{QP}})\ q_m(x_j^{\mathrm{QP}}); \tag{5.39}$$

absent this consistency, $u_{N,M}(\mu)$ will not converge to $u(\mu)$ as $N, M \to \infty$.

From the construction of the interpolation points t_i, $1 \le i \le M_{\max}$, we note that the q_m, $1 \le m \le M_{\max}$, can be written as a linear combination of the basis function $\xi_i = g(x;\mu_i^g)$, $1 \le i \le M_{\max}$, obtained from our greedy adaptive procedure in Section 2.4: $\xi_i = T_{im} q_m$, $1 \le i, m \le M_{\max}$, where $T \in \mathbb{R}^{M_{\max} \times M_{\max}}$ is the corresponding transformation matrix . Unfortunately, it turns out that T is badly conditioned and the resulting q_m required in (5.39) susceptible to large round-off errors. To avoid this problem we follow a different route. First, while generating the basis functions $\xi_i = g(x;\mu_i^g) \in \mathbb{R}^{\mathcal{N}}$, $1 \le i \le M_{\max}$, we also generate a corresponding set of functions ξ_i^{QP} evaluated at the quadrature points x_j^{QP}, $1 \le j \le \mathcal{N}_{\mathrm{QP}}$, that is $\xi_i^{\mathrm{QP}}(x_j^{\mathrm{QP}}) = g(x_j^{\mathrm{QP}};\mu_i^g)$, $1 \le j \le \mathcal{N}_{\mathrm{QP}}$, $1 \le i \le M_{\max}$. Next, we construct the set of interpolation points t_i and functions q_i from the ξ_i according to the procedure of Section 2.4. During this procedure, we also construct vectors of quadrature-point values, $q_m^{\mathrm{QP}} \in \mathbb{R}^{\mathcal{N}_{\mathrm{QP}}}$, $1 \le m \le M_{\max}$, from the ξ_i^{QP}: starting with $q_1^{\mathrm{QP}} = \xi_1^{\mathrm{QP}}(x)/\xi_1(t_1)$ and then setting $r_M^{\mathrm{QP}}(x) = \xi_M^{\mathrm{QP}}(x) - \sum_{i=1}^{M-1} \sigma_i^{M-1}\, q_i^{\mathrm{QP}}(x)$, $q_M^{\mathrm{QP}}(x) = r_M^{\mathrm{QP}}(x)/r_M(t_M)$, $2 \le M \le M_{\max}$, where the σ_j^{M-1} are determined during the construction of the q_i. Note that q_m^{QP} is simply the "basis" function corresponding to q_m, but evaluated at the quadrature points, i.e., $q_m^{\mathrm{QP}}(x_j^{\mathrm{QP}}) = q_m(x_j^{\mathrm{QP}})$, $1 \le j \le \mathcal{N}_{\mathrm{QP}}$, $1 \le m \le M_{\max}$. Given the q_m^{QP}, we can then directly evaluate the integral $\int_\Omega \zeta_i\, q_m$ from

$$\int_\Omega \zeta_i\ q_m \approx \sum_{j=1}^{\mathcal{N}_{\mathrm{QP}}} \omega_j\ \zeta_i(x_j^{\mathrm{QP}})\ q_m^{\mathrm{QP}}(x_j^{\mathrm{QP}}). \tag{5.40}$$

Using this approach during the numerical implementation we can avoid the round-off errors that resulted from the conditioning of the transformation matrix T.

5.4 *A Posteriori* Error Estimation

5.4.1 Preliminaries

We now turn to the development of our *a posteriori* error estimator. To begin, we recall the definition of $\hat{\alpha}_a(\mu) : \mathcal{D} \rightarrow \mathbb{R}^+$ as a lower bound for the coercivity constant $\alpha_a(\mu)$. We next introduce the dual norm of the residual

$$\varepsilon_{N,M}(\mu, t^k) \equiv \sup_{v \in Y} \frac{R(v; \mu, t^k)}{\|v\|_Y}, \quad \forall\, k \in \mathbb{K}, \tag{5.41}$$

where

$$\begin{aligned} R(v; \mu, t^k) \equiv b(v; g_M(x; \mu))\ u(t^k) - a(y_{N,M}(\mu, t^k), v; g_M(x; \mu)) \\ - \frac{1}{\Delta t} m(y_{N,M}(\mu, t^k) - y_{N,M}(\mu, t^{k-1}), v), \quad \forall\, v \in Y,\ \forall\, k \in \mathbb{K}, \end{aligned} \tag{5.42}$$

is the residual. We also introduce the dual norm

$$\Phi_M^{\text{na}}(\mu, t^k) \equiv \sup_{v \in Y} \frac{b(v; q_{M+1})\ u(t^k) - a_1(y_{N,M}(\mu, t^k), v, q_{M+1})}{\|v\|_Y}, \quad \forall\, k \in \mathbb{K}, \tag{5.43}$$

which reflects the contribution of the nonaffine terms. Finally, we specify the inner products $(v, w)_Y \equiv a_0(v, w)$, $\forall\, v, w \in Y$ and $(v, w)_X \equiv m(v, w)$, $\forall\, v, w \in Y$, and recall the definition $\hat{\varepsilon}_M(\mu) = |g(t_{M+1}; \mu) - g_M(t_{M+1}; \mu)|$ from Section 2.4.2.

We now present and prove the bounding properties for the errors in the field variable and the output estimate. Throughout this section we assume that the "truth" solution $y(\mu, t^k)$ satisfies (5.3) and the corresponding reduced-basis approximation $y_{N,M}(\mu, t^k)$ satisfies (5.15).

5.4.2 Error Bound Formulation

Primal Variable

We obtain the following result for the error bound.

Proposition 15. *Suppose that $g(x; \mu) \in W_{M+1}^g$. Let $e(\mu, t^k) \equiv y(\mu, t^k) - y_{N,M}(\mu, t^k)$ be the error in the field variable and define the "spatio-temporal" energy norm*

$$|||v(\mu, t^k)||| \equiv \left(m(v(\mu, t^k), v(\mu, t^k)) + \sum_{k'=1}^{k} a(v(\mu, t^{k'}), v(\mu, t^{k'}); g(x; \mu))\ \Delta t \right)^{\frac{1}{2}}, \quad \forall\, v \in Y. \tag{5.44}$$

The error is then bounded by

$$|||e(\mu, t^k)||| \leq \Delta_{N,M}^y(\mu, t^k), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{5.45}$$

where the error bound $\Delta^y_{N,M}(\mu, t^k)$ *is defined as*

$$\Delta^y_{N,M}(\mu, t^k) \equiv \left(\frac{2\Delta t}{\hat{\alpha}_a(\mu)} \sum_{k'=1}^{k} \varepsilon_{N,M}(\mu, t^{k'})^2 + \frac{2\Delta t}{\hat{\alpha}_a(\mu)} \hat{\varepsilon}^2_M(\mu) \sum_{k'=1}^{k} \Phi^{\mathrm{na}}_M(\mu, t^{k'})^2 \right)^{\frac{1}{2}} . \tag{5.46}$$

Proof. We immediately derive from (5.3) and (5.42) that $e(\mu, t^k) = y(\mu, t^k) - y_{N,M}(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, satisfies

$$\begin{aligned} m(e(\mu, t^k), v) + \Delta t\; a(e(\mu, t^k), v; g(x;\mu)) &= m(e(\mu, t^{k-1}), v) + \Delta t\; R(v; \mu, t^k) \\ &+ \Delta t\; \Big(b(v; g(x;\mu) - g_M(x;\mu))\; u(t^k) - a_1(y_{N,M}(\mu, t^k), v, g(x;\mu) - g_M(x;\mu)) \Big), \\ &\qquad \forall\, v \in Y, \end{aligned} \tag{5.47}$$

where $e(\mu, t^0) = 0$ since $y(\mu, t^0) = y_{N,M}(\mu, t^0) = 0$ by assumption. We now choose $v = e(\mu, t^k)$, invoke the Cauchy-Schwarz inequality for the cross term $m(e(\mu, t^{k-1}), e(\mu, t^k))$, and apply (5.41) to obtain, $\forall\, k \in \mathbb{K}$,

$$\begin{aligned} & m(e(\mu, t^k), e(\mu, t^k)) + \Delta t\; a(e(\mu, t^k), e(\mu, t^k); g(x;\mu)) \\ & \quad \le\; m^{\frac{1}{2}}(e(\mu, t^k), e(\mu, t^k))\; m^{\frac{1}{2}}(e(\mu, t^{k-1}), e(\mu, t^{k-1})) + \Delta t\; \varepsilon_{N,M}(\mu, t^k)\; \|e(\mu, t^k)\|_X \\ & \qquad + \Delta t\; \Big(b(e(\mu, t^k); g(x;\mu) - g_M(x;\mu))\; u(t^k) - a_1(y_{N,M}(\mu, t^k), e(\mu, t^k), g(x;\mu) - g_M(x;\mu)) \Big). \end{aligned} \tag{5.48}$$

From our assumption, $g(x;\mu) \in W^g_{M+1}$, Proposition 2, and (5.43) it directly follows that

$$\begin{aligned} & b(e(\mu, t^k); g(x;\mu) - g_M(x;\mu))\; u(t^k) - a_1(y_{N,M}(\mu, t^k), e(\mu, t^k), g(x;\mu) - g_M(x;\mu)) \\ & \qquad \le\; \hat{\varepsilon}_M(\mu)\; \sup_{v \in Y} \frac{b(v; q_{M+1})\; u(t^k) - a_1(y_{N,M}(\mu, t^k), v, q_{M+1})}{\|v\|_Y}\; \|e(\mu, t^k)\|_Y \\ & \qquad \le\; \hat{\varepsilon}_M(\mu)\; \Phi^{\mathrm{na}}_M(\mu, t^k)\; \|e(\mu, t^k)\|_Y. \end{aligned} \tag{5.49}$$

We will now apply (4.30) thrice: first, choosing $c = m^{\frac{1}{2}}(e(\mu, t^k), e(\mu, t^k))$, $d = m^{\frac{1}{2}}(e(\mu, t^{k-1}), e(\mu, t^{k-1}))$, and $\rho = 1$, we obtain

$$\begin{aligned} & 2\; m^{\frac{1}{2}}(e(\mu, t^k), e(\mu, t^k))\; m^{\frac{1}{2}}(e(\mu, t^{k-1}), e(\mu, t^{k-1})) \\ & \qquad \le m(e(\mu, t^{k-1}), e(\mu, t^{k-1})) + m(e(\mu, t^k), e(\mu, t^k)); \end{aligned} \tag{5.50}$$

second, choosing $c = \varepsilon_{N,M}(\mu, t^k)$, $d = \|e(\mu, t^k)\|_Y$, and $\rho = (\hat{\alpha}_a(\mu)/2)^{\frac{1}{2}}$ we have

$$2\; \varepsilon_N(\mu, t^k)\; \|e(\mu, t^k)\|_Y \le \frac{2}{\hat{\alpha}_a(\mu)}\; \varepsilon_N(\mu, t^k)^2 + \frac{\hat{\alpha}_a(\mu)}{2}\; \|e(\mu, t^k)\|^2_Y; \tag{5.51}$$

and third, choosing $c = \hat{\varepsilon}_M(\mu)\; \Phi^{\mathrm{na}}_M(\mu, t^k)$, $d = \|e(\mu, t^k)\|_Y$, and $\rho = (\hat{\alpha}_a(\mu)/2)^{\frac{1}{2}}$ gives

$$2\; \hat{\varepsilon}_M(\mu)\; \Phi^{\mathrm{na}}_M(\mu, t^k)\; \|e(\mu, t^k)\|_X \le \frac{2}{\hat{\alpha}_a(\mu)}\; \hat{\varepsilon}^2_M(\mu)\; \Phi^{\mathrm{na}}_M(\mu, t^k)^2 + \frac{\hat{\alpha}_a(\mu)}{2}\; \|e(\mu, t^k)\|^2_Y; \tag{5.52}$$

Combining (5.48) and (5.49), and invoking (5.7) and (5.50)-(5.52), we obtain

$$m(e(\mu,t^k),e(\mu,t^k)) - m(e(\mu,t^{k-1}),e(\mu,t^{k-1})) + \Delta t\ a(e(\mu,t^k),e(\mu,t^k);g(x;\mu)) \\ \leq \frac{2\Delta t}{\hat{\alpha}_a(\mu)} \left(\varepsilon_{N,M}(\mu,t^k)^2 + \hat{\varepsilon}_M^2(\mu)\ \Phi_M^{\text{na}}(\mu,t^k)^2\right), \quad \forall\, k \in \mathbb{K}, \quad (5.53)$$

where we used the fact that $\hat{\alpha}_a(\mu) \leq \alpha_a(\mu)$, $\forall\, \mu \in \mathcal{D}$. We now perform the sum from $k' = 1$ to k and recall that $e(\mu,t^0) = 0$, leading to

$$m(e(\mu,t^k),e(\mu,t^k)) + \sum_{k'=1}^{k} \Delta t\ a(e(\mu,t^{k'}),e(\mu,t^{k'});g(x;\mu)) \\ \leq \frac{2\Delta t}{\hat{\alpha}_a(\mu)} \sum_{k'=1}^{k} \left(\varepsilon_{N,M}(\mu,t^{k'})^2 + \hat{\varepsilon}_M^2(\mu)\ \Phi_M^{\text{na}}(\mu,t^{k'})^2\right), \quad \forall\, k \in \mathbb{K}, \quad (5.54)$$

which is the result stated in Proposition 15. □

We note from (5.46) that our error bound comprises the affine as well as the nonaffine error contributions. We may thus choose N and M such that both contributions balance, i.e., neither N nor M should be chosen unnecessarily high. We also recall that our (crucial) assumption $g(x;\mu) \in W_{M+1}^g$ cannot be confirmed in actual practice — in fact, we generally have $g(x;\mu) \notin W_{M+1}^g$ and hence our error bound (5.46) is *not* completely rigorous, since $\hat{\varepsilon}_M(\mu) \leq \epsilon_M(\mu)$. We comment on both of these issue again in detail in Section 5.5 when discussing numerical results.

Output Bound

We can now define the (simple) output bound in

Proposition 16. *Suppose that $g(x;\mu) \in W_{M+1}^g$. Let the output, $s(\mu,t^k)$, and the output estimate, $s_{N,M}(\mu,t^k)$, be given by (5.2) and (5.16), respectively. The error in the output of interest is then bounded by*

$$|s(\mu,t^k) - s_{N,M}(\mu,t^k)| \leq \Delta_{N,M}^s(\mu,t^k), \quad \forall\, k \in \mathbb{K},\ \forall\, \mu \in \mathcal{D}, \quad (5.55)$$

where the output bound $\Delta_{N,M}^s(\mu,t^k)$ is defined as

$$\Delta_{N,M}^s(\mu,t^k) \equiv \sup_{v \in Y} \frac{\ell(v)}{\|v\|_X} \Delta_{N,M}^y(\mu,t^k)\,. \quad (5.56)$$

Proof. From (5.2) and (5.16) we obtain

$$\begin{aligned} |s(\mu,t^k) - s_{N,M}(\mu,t^k)| &= |\ell(y(\mu,t^k)) - \ell(y_{N,M}(\mu,t^k))| \\ &= |\ell(e(\mu,t^k))| \\ &\leq \sup_{v \in Y} \frac{\ell(v)}{\|v\|_X} \|e(\mu,t^k)\|_X \end{aligned}$$

from which the result immediately follows since $\|e(\mu,t^k)\|_X \leq \Delta_{N,M}^y(\mu,t^k)$, $\forall\, \mu \in \mathcal{D}$, $\forall\, k \in \mathbb{K}$. □

Note that m is *parameter-independent* here; we thus have no need for the bound conditioner $\hat{\alpha}_m(\mu)$ required for the simple bound in (4.112).

5.4.3 Offline-Online Computational Procedure

We now turn to the development of offline-online computational procedures for the calculation of $\Delta^y_{N,M}(\mu, t^k)$ and $\Delta^s_{N,M}(\mu, t^k)$. We first note from standard duality arguments that

$$\begin{aligned} \varepsilon_{N,M}(\mu, t^k) &\equiv \sup_{v\in X} \frac{R(v;\mu,t^k)}{\|v\|_Y} && (5.57)\\ &= \|\hat{e}(\mu,t^k)\|_Y, && (5.58)\end{aligned}$$

where $\hat{e}(\mu, t^k) \in X$ is given by

$$(\hat{e}(\mu,t^k), v)_Y = R(v;\mu,t^k), \qquad \forall\, v \in X; \tag{5.59}$$

(5.59) is effectively a Poisson problem for each $t^k \in \mathbb{I}$.

From (5.42), (5.4), and (5) it thus follows that $\hat{e}(\mu, t^k)$ satisfies

$$\begin{aligned}(\hat{e}(\mu,t^k),v)_X = \sum_{m-1}^{M} \varphi_{M\,m}(\mu)\, u(v;q_m)\, y(t^k) - \sum_{n=1}^{N}\Bigg\{ \frac{1}{\Delta t}\left(y_{N,M\,n}(\mu,t^k) - y_{N,M\,n}(\mu,t^{k-1})\right) m(\zeta_n, v)\\ + y_{N,M\,n}(\mu,t^k)\, a_0(\zeta_n,v) + \sum_{m=1}^{M} \varphi_{M\,m}(\mu)\, y_{N,M\,n}(\mu,t^k)\, a_1(\zeta_n,v,q_m)\Bigg\}, \ \forall\, v\in X. \quad (5.60)\end{aligned}$$

It is clear from linear superposition that we can express $\hat{e}(\mu, t^k)$ as

$$\begin{aligned}\hat{e}(\mu) = \sum_{q=m}^{M} \varphi_{M\,m}(\mu)\, y(t^k)\, \mathcal{F}_m - \sum_{n=1}^{N}\Bigg\{\frac{1}{\Delta t}\left(y_{N,M\,n}(\mu,t^k) - y_{N,M\,n}(\mu,t^{k-1})\right) \mathcal{M}_n\\ + \left(\mathcal{A}^0_n + \sum_{m=1}^{M} \varphi_{M\,m}(\mu)\, \mathcal{A}^1_{m,n}\right) y_{N,M\,n}(\mu,t^k)\Bigg\}, \quad (5.61)\end{aligned}$$

where we calculate $\mathcal{F}_m \in X$, $\mathcal{A}^0_n \in X$, $\mathcal{A}^1_{m,n} \in X$, and $\mathcal{M}_n \in X$ from

$$\begin{aligned}(\mathcal{F}_m, v)_Y &= b(v;q_m), && \forall\, v \in X,\ 1 \le m \le M_{\max},\\ (\mathcal{A}^0_n, v)_Y &= a_0(\zeta_n, v), && \forall\, v \in X,\ 1 \le n \le N_{\max},\\ (\mathcal{A}^1_{m,n}, v)_Y &= a_1(\zeta_n, v, q_m), && \forall\, v \in X,\ 1 \le n \le N_{\max},\ 1 \le m \le M_{\max},\\ (\mathcal{M}_n, v)_Y &= m(\zeta_n, v), && \forall\, v \in X,\ 1 \le n \le N_{\max};\end{aligned} \tag{5.62}$$

note $\mathcal{B}$, $\mathcal{A}^{0,1}$, and $\mathcal{M}$ are parameter independent.

From (5.58) and (5.62) it follows that

$$
\begin{aligned}
\varepsilon_{N,M}(\mu,t^k)^2 \;=\; & \sum_{m,m'=1}^{M} \varphi_{M\,m}(\mu)\,\varphi_{m'}(\mu)\,u(t^k)\,u(t^k)\,\Lambda^{ff}_{mm'} \\
+\; & \sum_{m=1}^{M}\sum_{n=1}^{N} \varphi_{M\,m}(\mu)\,u(t^k)\left(\left(\Lambda^{a_0f}_{mn} + \sum_{m'=1}^{M} \varphi_{m'}(\mu)\,\Lambda^{a_1f}_{mnm'}\right) y_{N,M\,n}(\mu,t^k)\right. \\
& \left. + \left(y_{N,M\,n}(\mu,t^k) - y_{N,M\,n}(\mu,t^{k-1})\right)\,\Lambda^{mf}_{mn}\right) \\
+\; & \sum_{n,n'=1}^{N} \left\{ \left(y_{N,M\,n}(\mu,t^k) - y_{N,M\,n}(\mu,t^{k-1})\right)\,\left(y_{N,M\,n'}(\mu,t^k) - y_{N,M\,n'}(\mu,t^{k-1})\right)\,\Lambda^{mm}_{nn'}\right. \\
& + y_{N,M\,n}(\mu,t^k)\,\left(y_{N,M\,n'}(\mu,t^k) - y_{N,M\,n'}(\mu,t^{k-1})\right)\,\left(\Lambda^{a_0m}_{nn'} + \sum_{m=1}^{M} \varphi_{M\,m}(\mu)\Lambda^{a_1m}_{nn'm}\right) \\
& + y_{N,M\,n}(\mu,t^k)\,y_{N,M\,n'}(\mu,t^k)\left(\Lambda^{a_0a_0}_{nn'} + \sum_{m=1}^{M} \varphi_{M\,m}(\mu)\Lambda^{a_0a_1}_{nn'm}\right) \qquad (5.63) \\
& \left. + \sum_{m,m'=1}^{M} \varphi_{M\,m}(\mu)\,\varphi_{m'}(\mu)\,y_{N,M\,n}(\mu,t^k)\,y_{N,M\,n'}(\mu,t^k)\,\Lambda^{a_1a_1}_{nn'mm'}\right\}, \qquad (5.64)
\end{aligned}
$$

where the parameter-independent quantities Λ are defined as

$$
\begin{aligned}
\Lambda^{ff}_{mm'} &= (\mathcal{F}_m, \mathcal{F}_{m'})_Y, && 1 \le m, m' \le M_{\max}; \\
\Lambda^{a_0f}_{mn} &= -2\;(\mathcal{F}_m, \mathcal{A}^0_n)_Y, && 1 \le m \le M_{\max},\ 1 \le n \le N_{\max}; \\
\Lambda^{a_1f}_{mnm'} &= -2\;(\mathcal{F}_m, \mathcal{A}^1_{m',n})_Y, && 1 \le m, m' \le M_{\max},\ 1 \le n \le N_{\max}; \\
\Lambda^{mf}_{mn} &= -\frac{2}{\Delta t}\;(\mathcal{F}_m, \mathcal{M}_n)_Y, && 1 \le m \le M_{\max},\ 1 \le n \le N_{\max}; \\
\Lambda^{a_0a_0}_{nn'} &= (\mathcal{A}^0_n, \mathcal{A}^0_{n'})_Y, && 1 \le n, n' \le N_{\max}; \\
\Lambda^{a_0a_1}_{nn'm} &= 2(\mathcal{A}^0_n, \mathcal{A}^1_{m,n'})_Y, && 1 \le m \le M_{\max},\ 1 \le n, n' \le N_{\max}; \\
\Lambda^{a_1a_1}_{nn'mm'} &= (\mathcal{A}^1_{m,n}, \mathcal{A}^1_{m',n'})_Y, && 1 \le m, m' \le M_{\max},\ 1 \le n, n' \le N_{\max}; \\
\Lambda^{a_0m}_{nn'} &= \frac{2}{\Delta t}\;(\mathcal{A}^0_n, \mathcal{M}_{n'})_Y, && 1 \le n, n' \le N_{\max}; \\
\Lambda^{a_1m}_{nn'm} &= \frac{2}{\Delta t}\;(\mathcal{A}^1_{m,n}, \mathcal{M}_{n'})_Y, && 1 \le m \le M_{\max},\ 1 \le n, n' \le N_{\max}; \\
\Lambda^{mm}_{nn'} &= \frac{1}{\Delta t^2}(\mathcal{M}_n, \mathcal{M}_{n'})_Y, && 1 \le n, n' \le N_{\max}.
\end{aligned}
\qquad (5.65)
$$

The evaluation of $\Phi^{\mathrm{na}}_M(\mu,t^k)$ is very similar; to this end, we first calculate $\mathcal{F}_{M+1} \in X$ and $\mathcal{A}^1_{M+1,n} \in$

X from

$$\begin{aligned}(\mathcal{F}_{M+1}, v)_Y &= b(v; q_{M+1}), \qquad \forall\, v \in X,\\ (\mathcal{A}^1_{M+1,n}, v)_Y &= a_1(\zeta_n, v; q_{M+1}), \quad \forall\, v \in X,\ 1 \le n \le N_{\max};\end{aligned} \tag{5.66}$$

It then follows from (5.43) and standard duality arguments that

$$\begin{aligned}\Phi^{\mathrm{na}}_M(\mu, t^k)^2 = {} & y(t^k)^2\, \Lambda^{ff}_{M+1\,M+1}\\ & + \sum_{n=1}^{N} y_{N,M\,n}(\mu, t^k) \left\{ y(t^k)\, \Lambda^{a_1 f}_{n\,M+1\,M+1} + \sum_{n'=1}^{N} y_{N,M\,n'}(\mu, t^k)\, \Lambda^{a_1 a_1}_{nn'\,M+1\,M+1} \right\}\end{aligned}$$

where the parameter-independent quantities Λ are defined as

$$\begin{aligned}\Lambda^{ff}_{M+1\,M+1} &= (\mathcal{F}_{M+1}, \mathcal{F}_{M+1})_Y;\\ \Lambda^{a_1 f}_{n\,M+1\,M+1} &= -2\,(\mathcal{F}_{M+1}, \mathcal{A}^1_{M+1,n})_Y, \quad 1 \le n \le N_{\max};\\ \Lambda^{a_1 a_1}_{nn'\,M+1\,M+1} &= (\mathcal{A}^1_{M+1,n}, \mathcal{A}^1_{M+1,n'})_Y, \quad 1 \le n, n' \le N_{\max}.\end{aligned} \tag{5.67}$$

The offline-online decomposition is now clear.

In the offline stage we first compute the quantities $\mathcal{F}$, $\mathcal{A}^{0,1}$, and $\mathcal{M}$ from (5.62) and (5.66) and then evaluate the Λ from (5.65) and (5.67); this requires (to leading order) $O(M_{\max} N_{\max})$ expensive "truth" finite element solutions, and $O(M^2_{\max} N^2_{\max})$ $\mathcal{N}$-inner products. In the online stage — given a new parameter value μ and associated reduced-basis solution $\underline{y}_{N,M}(\mu, t^k)$, $\forall\, k \in \mathbb{K}$ — the computational cost to evaluate $\Delta^y_{N,M}(\mu, t^k)$ and $\Delta^s_{N,M}(\mu, t^k)$, $\forall k \in \mathbb{K}$, is $O(KM^2N^2)$. Thus, all online calculations needed are *independent* of $\mathcal{N}$.

5.5 Results for Numerical Exercise 4

We now return to the numerical example introduced in Section 5.2.1. We first employ the empirical interpolation method of Section 2.4 to construct the approximation to the nonaffine function $G(x; \mu)$ defined in (2.36). The function $G(x; \mu)$ already served as our test problem in Numerical Exercise 1, we can thus directly use the sample set S^g_M and associated basis W^g_M — and hence T_M and B_M — constructed in Section 2.4.3 (we use the $L^\infty(\Omega)$-norm results here).

We next generate the sample set S^y_N and associated reduced basis space W^y_N according to the adaptive sampling procedure described in Section 4.5. We initialize the procedure with $\mu^y_1 = (-0.01, -0.01)$ and $t^{k^y_1} = 1\Delta t$ and set the desired error tolerance (for the relative error in the energy norm) to $\epsilon_{\mathrm{tol,min}} = 1\,\mathrm{E}{-6}$. We sample on a deterministic parameter test sample $\Xi_{\mathrm{F}} \in (\mathcal{D})^{1600}$ of size 1600 (Ξ_F is the same sample set used in Section 2.4.3 for the construction of S^g_M) — we need $N_{\max} = 54$ basis functions to obtain the desired accuracy.

We plot the sample sets S^y_N in $\mu - t^k$-space in Figure 5-2. We note that (i) more samples are selected in the difficult parameter range near the corner $\mu = (-0.01, -0.01)$; and (ii) although $K = 200$ here, the samples are only selected within the first 40 timesteps because of the periodic control input. Were we to follow a naïve approach and select the samples on a regular grid in time (without using the adaptive sampling procedure), our reduced-basis system could easily become ill-conditioned.

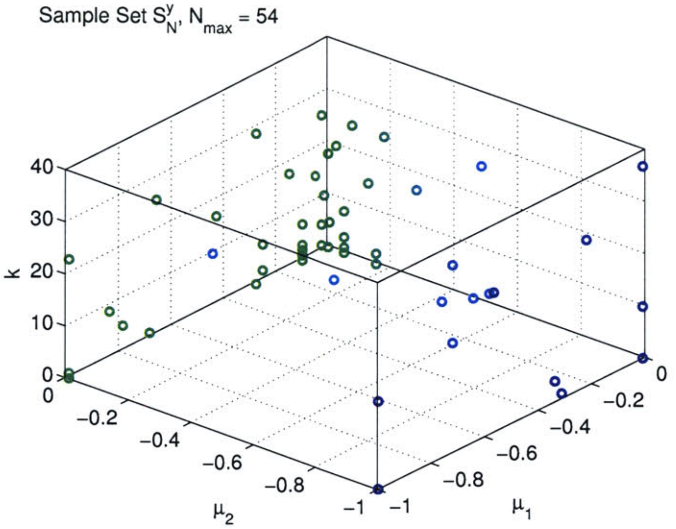


Figure 5-2: NE 4: Sample set S^y_N.

We now turn to the convergence results. In Figure 5-3 we plot, as a function of N and M, the maximum relative error in the energy norm $\epsilon^y_{N,M,\mathrm{max,rel}}$; here, $\epsilon^y_{N,M,\mathrm{max,rel}}$ is the maximum over Ξ_{Test} of $|||e(\mu, t^K)|||/|||y(\mu_y, t^K)|||$, where $\mu_y \equiv \arg\max_{\mu\in\Xi_{\mathrm{Test}}} |||y(\mu, t^K)|||$, and $\Xi_{\mathrm{Test}} \in (\mathcal{D})^{225}$ is a input sample of size 225 (a regular 15×15 grid). We observe that the error levels off at smaller and smaller values as we increase M reflecting the trade-off influence between the reduced-basis approximation and the coefficient function approximation contribution to the error: for fixed M the error in the coefficient function approximation $g_M(x; \mu)$ to $g(x; \mu)$ will ultimately dominate for large N; increasing M renders the coefficient function approximation more accurate, which in turn leads to the drops in the error. We further note that the separation points reflect a balanced contribution of both approximations to the error: increasing either N or M while keeping the other one fixed has a very small effect on the error; to reduced the error both N and M have to be increased.

In Table 5.1 we present, as a function of N and M, the maximum relative error in the energy norm $\epsilon^y_{N,M,\mathrm{max,rel}}$, the maximum relative error bound $\Delta^y_{N,M,\mathrm{max,rel}}$, and the average effectivity $\overline{\eta}^y$: $\Delta^y_{N,M,\mathrm{max,rel}}$ is the maximum over Ξ_{Test} of $\Delta^y_{N,M}(\mu, t^K)/|||y(\mu_y, t^K)|||$, and $\overline{\eta}^y$ is the average over $\Xi_{\mathrm{Test}} \times \mathbb{I}$ of $\Delta^y_{N,M}(\mu, t^k)/|||y(\mu, t^k) - y_N(\mu, t^k)|||$. The specific N, M combinations presented roughly correspond to the separation points of the convergence curves in Figure 5-3. We observe that the reduced-basis approximation converges very fast and the bounds are very sharp — actually too sharp. For the first three N, M combinations we obtain effectivities less than one: our error bound is *not* an upper bound for the true error. We recall that our proof for the bounding property is based on the assumption that $g(x; \mu) \in W^g_{M+1}$.In general, however, this assumption is not satisfied and our error estimators may not be rigorous upper bounds since $\hat{\varepsilon}_M(\mu) \leq \varepsilon_M(\mu)$ if $g(x; \mu) \notin W^g_{M+1}$ (also see Section 2.4.2). To regain the bounding property, we shall select N and M such that the contribution to the error bound due to the nonaffine function approximation is much smaller than the contribution of the affine terms. Although this approach does not guarantee that $g(x; \mu) \in W^g_{M+1}$, we can hope to absorb the lower bound property of $\hat{\varepsilon}_M(\mu)$ in the rigorous

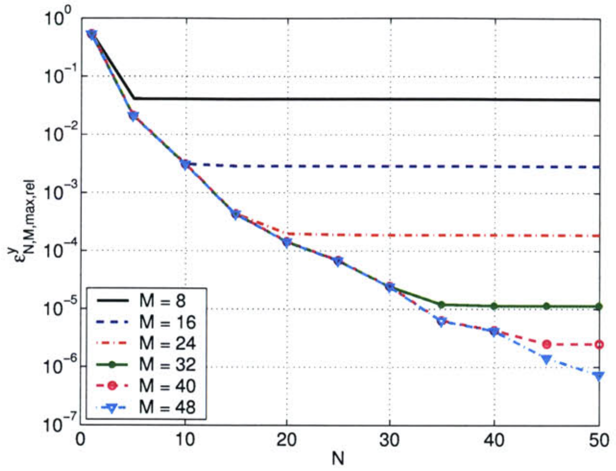


Figure 5-3: NE 4: Maximum relative error in the energy norm.

N	M	$\epsilon^{y}_{N,M,\max,\mathrm{rel}}$	$\Delta^{y}_{N,M,\max,\mathrm{rel}}$	$\overline{\eta}^{y}$
5	8	4.12 E−02	2.87 E−02	0.50
10	16	3.12 E−03	4.40 E−03	0.96
20	24	1.97 E−04	2.05 E−04	0.87
30	32	2.46 E−05	3.49 E−05	1.19
40	40	4.27 E−06	6.03 E−06	1.19
50	48	7.48 E−07	1.06 E−06	1.38

Table 5.1: NE 4: Convergence rate and effectivities as a function of N and M.

error bound for the affine terms. If we thus increase M while keeping N fixed, we indeed obtain effectivities larger than one even for small values of N, M, as shown in Table 5.2.

We next turn to the output estimate. In Table 5.3 we present, as a function of N and M, the maximum relative output error $\epsilon^{s}_{N,M,\max,\mathrm{rel}}$, the maximum relative output bound $\Delta^{s}_{N,M,\max,\mathrm{rel}}$, and the average effectivity $\overline{\eta}^{s}_{N,M}$: $\epsilon^{s}_{N,M,\max,\mathrm{rel}}$ is the maximum over Ξ_{Test} of $|s(\mu, t_\eta(\mu)) - s_N(\mu, t_\eta(\mu))|/ s_{\max}$, $\Delta^{s}_{N,M,\max,\mathrm{rel}}$ is the maximum over Ξ_{Test} of $\Delta^{s}_{N,M}(\mu, t^K)/|s_{\max}|$ and $\overline{\eta}^{s}$ is the average over Ξ_{Test} of $\Delta^{s}_{N,M}(\mu, t_\eta(\mu))/|s(\mu, t_\eta(\mu)) - s_{N,M}(\mu, t_\eta(\mu))|$. Here $t_\eta(\mu) \equiv \arg\max_{t^k \in \mathbb{I}} |s(\mu, t^k) - s_N(\mu, t^k)|$ and $s_{\max} \equiv \max_{t^k \in \mathbb{I}} \max_{\mu \in \Xi_{\mathrm{Test}}} |s(\mu, t^k)|$. We note that, to calculate the average output effectivity $\overline{\eta}^{s}$, we exclude parameter values from the sample Ξ_{Test} where $\epsilon^{s}_{N,M,\max,\mathrm{rel}} \leq 1\,\mathrm{E}{-8}$ so as to avoid contamination by round-off errors. We see that the output error and output bound converges very fast — for only $N = 20$ and $M = 24$ our error bound for the error in the output — even for our simple bound — is quite close to 0.1%. Also, the effectivities are, of course, not as good as for the energy norm bound, but still acceptable for the simple bound.

In Table 5.4 we present, as a function of N and M, the online computational times to calculate $s_{N,M}(\mu, t^k)$ and $\Delta^{s}_{N,M}(\mu, t^k)$, $\forall k \in \mathbb{K}$. The values are normalized with respect to the computational time for the direct calculation of the truth approximation output $s(\mu, t^k) = \ell(y(\mu, t^k))$, $\forall\, k \in \mathbb{K}$. (The growth with N and M is less than expected due to memory-access issues.) The computational

N	M	$\epsilon^y_{N,M,\mathrm{max,rel}}$	$\Delta^y_{N,M,\mathrm{max,rel}}$	$\bar{\eta}^y$
5	16	2.09E-02	2.98E-02	1.34
10	24	3.09E-03	4.38E-03	1.42
20	32	1.45E-04	2.05E-04	1.42
30	40	2.46E-05	3.48E-05	1.41
40	48	4.26E-06	6.04E-06	1.42

Table 5.2: NE 4: Convergence rate and effectivities as a function of N and M.

N	M	$\epsilon^s_{N,M,\mathrm{max,rel}}$	$\Delta^s_{N,M,\mathrm{max,rel}}$	$\bar{\eta}^s_{N,M}$
5	8	4.23E-02	1.67E-01	7.70
10	16	3.03E-03	2.56E-02	21.2
20	24	1.79E-04	1.19E-03	159
30	32	7.65E-06	2.03E-04	52.2
40	40	2.21E-06	3.52E-05	20.5
50	48	1.29E-07	6.17E-06	33.2

Table 5.3: NE 4: Maximum relative output error, output bound, and effectivities.

saving for an accuracy of close to 0.1 percent ($N = 20$, $M = 24$) in the output bound are close to a factor of 100. The actual run-time to compute the output estimate and output bound in MATLAB 6.5 on a 750 MHz Pentium III is 0.11 sec. (for $N = 20$, $M = 24$). We also note that the time to calculate $\Delta^s_{N,M}(\mu, t^k)$ exceeds that of calculating $s_N(\mu, t^k)$ considerably — this is due to the higher computational cost, $O(KM^2N^2)$, to evaluate $\Delta^y_{N,M}(\mu, t^k)$. Thus, although the theory suggest to choose M large so that the error due to the nonaffine function approximation is small, we should choose M as small as possible to retain the computational efficiency of our method. We emphasize that the reduced-basis entry does *not* include the extensive offline computations — and is thus only meaningful in the real-time or many-query contexts.

N	M	$s_{N,M}(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$\Delta^s_{N,M}(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$s(\mu, t^k),\ \forall\, k \in \mathbb{K}$
5	8	6.96E-04	4.29E-03	1
10	16	7.61E-04	6.51E-03	1
20	24	1.05E-03	1.09E-02	1
30	32	1.25E-03	1.87E-02	1
40	40	1.68E-03	3.30E-02	1
50	48	2.06E-03	5.32E-02	1

Table 5.4: NE 4:Online computational times (normalized with respect to the time to solve for $s(\mu, t^k),\ \forall\, k \in \mathbb{K}$).

5.6 AP II: Dispersion of Pollutants

We now return to the pollutant dispersion problem introduced in Section 1.1.1. In the last chapter we discussed the reduced basis approximation and associated *a posteriori* error estimation for the case of affine parameter dependence — we assumed that the location of the source term is fixed and the diffusivity is the only varying parameter. In this section we take the next step and assume that the source location can also vary in a certain range. The source term $g^{\mathrm{PS}}(x;\mu)$, defined in (4.170), is given by

$$g^{\mathrm{PS}}(x;\mu) = \frac{1}{2\pi\sigma^{\mathrm{PS}^2}}\, e^{-\frac{(x_1-x_1^{\mathrm{PS}})^2+(x_2-x_2^{\mathrm{PS}})^2}{2\sigma^{\mathrm{PS}^2}}}, \tag{5.68}$$

where $x^{\mathrm{PS}} = (x_1^{\mathrm{PS}}, x_2^{\mathrm{PS}})$ denotes the source location and σ^{PS} is the standard deviation; note that we use the notation $g^{\mathrm{PS}}(\cdot;\mu)$ now to signify that x^{PS} is an input parameter. We immediately recognize that the parameter dependence is nonaffine — we thus require the theory developed for nonaffine problems in this chapter.

For easier reference, we repeat the sketch of the flow field $\mathbf{U}$ with the source location and the measurement sensors in Figure 5-4. The domain Ω, a typical point in which is (x_1, x_2), is given by $\Omega \equiv [0,4] \times [0,1]$. We shall assume that the concentration at the left boundary, Γ_{D}, is zero and that the remaining boundaries, Γ_{N}, are impermeable. We shall also assume that the diffusivity κ varies in the (now, smaller) range $0.05 \leq \kappa \leq 0.5$ and that the source location, x^{PS}, satisfies $2.9 \leq x_1^{\mathrm{PS}} \leq 3.1$ and $0.3 \leq x_2^{\mathrm{PS}} \leq 0.5$; our input parameter is hence $\mu \equiv \{\mu_1, \mu_2, \mu_3\} \equiv \{\kappa, x_1^{\mathrm{PS}}, x_2^{\mathrm{PS}}\} \in \mathcal{D} \equiv [0.05, 0.5] \times [2.9, 3.1] \times [0.3, 0.5] \subset \mathbb{R}^{P=3}$; the standard deviation $\sigma^{\mathrm{PS}} = 0.1$ is assumed fixed. For notational convenience we also define $\Omega^{\mathrm{PS}} \equiv [2.9, 3.1] \times [0.3, 0.5]$, i.e., the spatial domain in which the source is located.

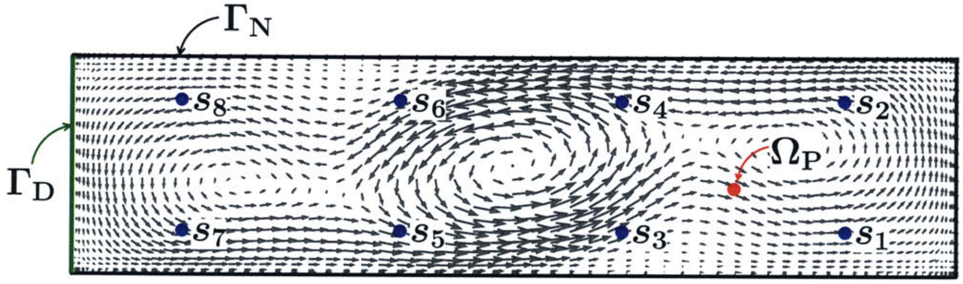


Figure 5-4: AP II: Velocity field with pollution source and measurement locations.

The time-discrete weak form of the governing equation (1.11) for the concentration $c(\mu, t^k) \in Y$ is again (B.2) with initial condition $c(\mu, t^0) = 0$, where $Y \subset Y^{\mathrm{e}} \equiv \{v | v \in H^1(\Omega), v = 0|_{\Gamma_{\mathrm{D}}}\}$ is a linear finite element truth approximation subspace of dimension $\mathcal{N} = 3720$ (the finite element mesh is the same as in Figure 4-17). The bilinear and linear forms m, a^{CD} and b are defined in (4.167), (4.168) and (4.169), respectively, where $g^{\mathrm{PS}}(x;\mu)$ is now given by (5.68). We note that the parameter dependence of $b(v; g(x;\mu))$ is now *nonaffine*, m is parameter independent and a^{CD} admits the affine representation (4.131) with $Q_a = 2$. We also define the inner products $(w,v)_X \equiv \int_\Omega w\, v = m(w,v)$ and $(w,v)_Y \equiv \int_\Omega \nabla w \cdot \nabla v$, corresponding to (4.153) for $\mu_1 = 1$; we may thus choose $\hat{\alpha}_a(\mu) = \mu_1$ in (4.57). The outputs, $s_q(\mu, t^k)$, $1 \leq q \leq 4$, are evaluated from (4.171) with $l_q(x)$ defined in (4.172); the sensor locations x^{l_q} are the same as in Section 4.8.5. We shall now consider the time interval $\bar{I} = [0,1]$ and a timestep $\Delta t = 5\,\mathrm{E}{-3}$; we thus have $K = 200$. We note that, since we consider a shorter time interval here, we restrict our attention mainly to the first four outputs located in the immediate vicinity of the source.

We point out that the admissible spatial variation of the source location is fairly small compared to the domain Ω; we would require 100 patches of size Ω^{PS} to completely cover Ω. The reason for choosing a small region Ω^{PS} is twofold: first, the flow field, $\mathbf{U}$, is quite complex, i.e., sources that are close to each other can result in very different pollutant distributions. We plot in Figures 5-5 and 5-6 the field variable $c(\mu, t^k)$ at six timesteps for $\mu = (0.05, 2.9, 0.3)$ and $\mu = (0.05, 3.1, 0.5)$, respectively. These parameter values correspond to the source being located at two opposite corners of Ω^{PS}. Although the two sources are close to each other, the resulting dispersion patterns are already decidedly different. This difference would be far more evident for source locations close to separation points in the flow. We also plot, in Figures 5-7 and 5-8, the first four outputs $s_q(\mu, t^k)$, $1 \leq q \leq 4$, as a function of time for $\mu = (0.05, 2.9, 0.3)$ and $\mu = (0.05, 3.1, 0.5)$, respectively. We observe that the difference in the field variable is also reflected in the outputs. Second, the minimum admissible diffusivity considered is small, and thus convection plays a dominant role in the problem solution. We recall from Section 4.8.5 that more samples are required for "more" non-symmetric problems — even the fixed source location with $\kappa \in [0.01, 1]$ resulted in a basis with $N_{\mathrm{pr,max}} = 207$. Thus, choosing Ω^{PS} larger would result in a reduced-basis approximation which is too high-dimensional with associated detriment to the computational efficiency (we will see that the computational savings for the given problem are already modest).

5.6.1 Reduced-Basis Approximation

We first consider the approximation to the nonaffine function $g^{\mathrm{PS}}(x;\mu)$ defined in (5.68). We choose for Ξ^g a deterministic grid of 41×41 parameter points over $[2.9, 3.1] \times [0.3, 0.5]$ and we choose $(\mu^g_{2,1}, \mu^g_{3,1}) = (3, 0.4)$. Note that $g^{\mathrm{PS}}(x;\mu)$ does not depend on μ_1. Next, we pursue the empirical interpolation method of Section 2.4 (using the $L^\infty(\Omega)$-norm) to construct S^g_M, W^g_M, T_M, and B^M, $1 \leq M \leq M_{\max}$, for $M_{\max} = 66$. We plot in Figures 5-9(a) and (b) the sample set S^g_M and the set of interpolation points T_M, respectively. The samples are (almost) symmetric around the center $(3, 0.4)$ and the interpolation points are, of course, centered around $x = (3, 0.4)$.

We next generate the sample set S^y_N and associated reduced basis space W^y_N according to the adaptive sampling procedure described in Section 4.5 with $M = M_{\max}$ for the nonaffine function approximation. We initialize the procedure with $\mu^y_1 = (0.05, 3, 0.4)$ and $t^{k^y_1} = 1\Delta t$ and set the desired error tolerance (for the relative error in the energy norm) to $\epsilon_{\mathrm{tol,min}} = 1\,\mathrm{E}{-4}$. We sample on a parameter test sample $\Xi_{\mathrm{F}} \in (\mathcal{D})^{2420}$ of size 2420 (Ξ_F is log-random in μ_1 and deterministic in μ_2 and μ_3); we require $N_{\max} = 256$ basis functions to obtain the desired accuracy.

We next present the sample set S^y_N: since we obtain a four-dimensional parameter-time space now, we first project the sample set S^y_N down onto (a) the μ_1-t^k-space, and (b) the μ_2-μ_3-t^k-space and plot the result in Figure 5-10(a) and (b), respectively. We observe, as in Section 4.8.5, that the samples are biased towards smaller diffusivities μ_1 (note the logarithmic scale). Furthermore, it is very interesting to note that almost all samples are located on the boundary of the mu_2-μ_3-space, i.e., the physical domain in which the source location lies.

5.6.2 Numerical Results

To begin, we present convergence results for the nonaffine function approximation. We present in Table 5.5 $\varepsilon^*_{M,\max}$, $\overline{\rho}_M$, Λ_M, $\overline{\eta}_M$, and $\varkappa_M$ as a function of M (see Section 2.4.3 for the definitions

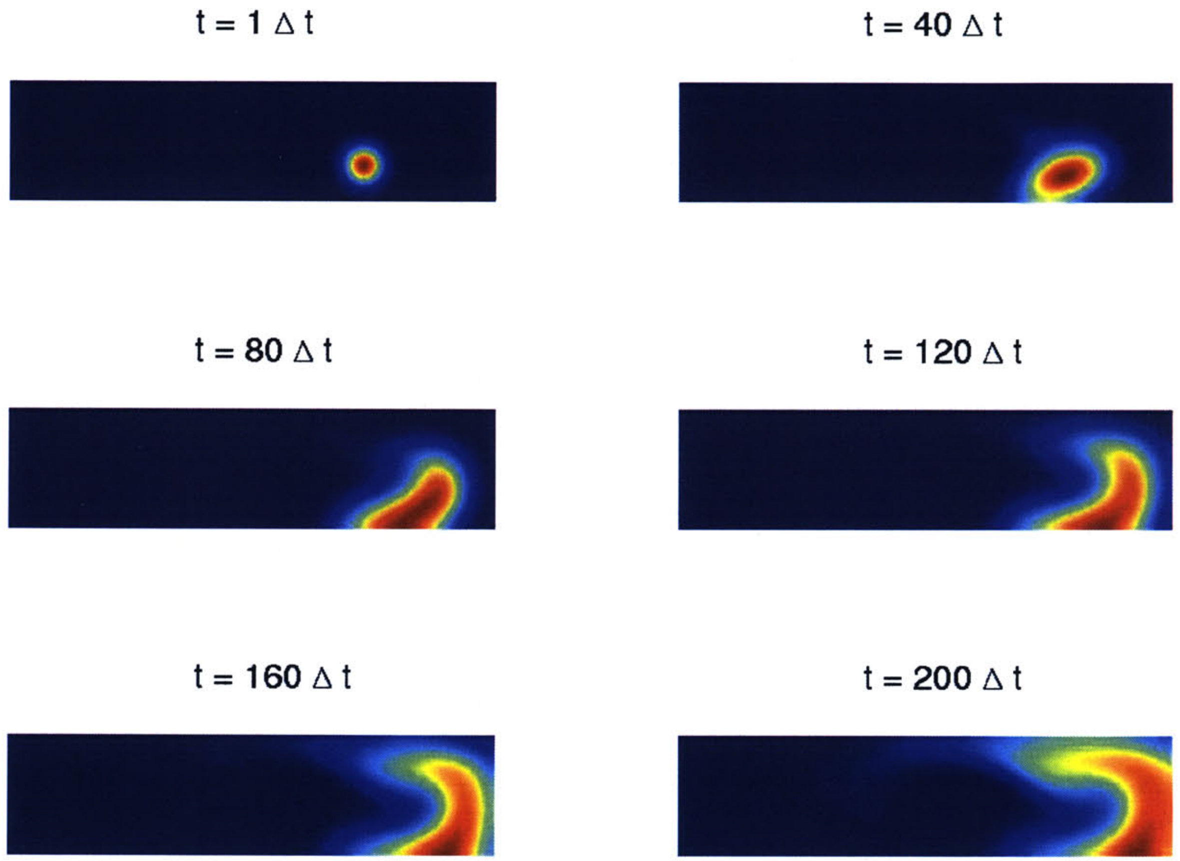


Figure 5-5: AP II: Concentration $c(\mu, t^k)$ for $\mu = (0.05, 2.9, 0.3)$ at $t = t^1$, t^{40}, t^{80}, t^{120}, t^{160}, t^{200}.

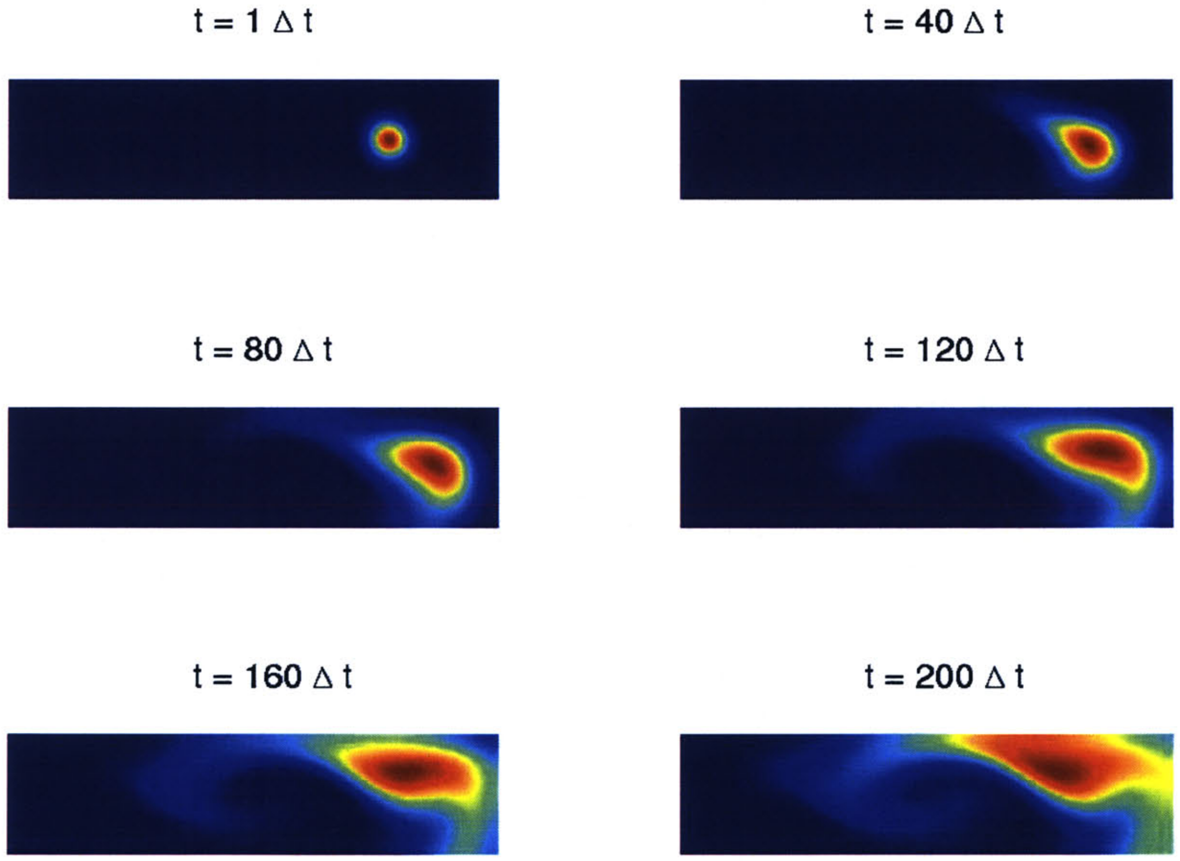


Figure 5-6: AP II: Concentration $c(\mu, t^k)$ for $\mu = (0.05, 3.1, 0.5)$ at $t = t^1$, t^{40}, t^{80}, t^{120}, t^{160}, t^{200}.

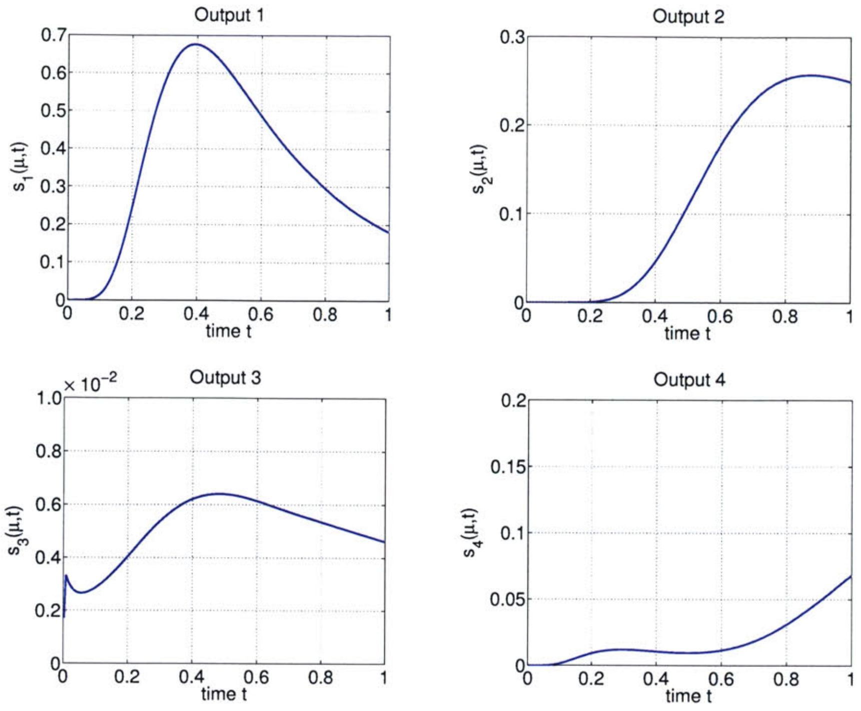


Figure 5-7: AP II: Outputs $s_q(\mu, t^k)$, $1 \leq q \leq 8$, for $\mu = (0.05, 2.9, 0.3)$ as a function of time.

of these quantities; here Ξ^g_{Test} is a test sample of size 225). We observe that the maximum error $\varepsilon^*_{M,\max}$ for the best approximation converges rapidly with M; that the Lebesgue constant provides a reasonably sharp measure of the interpolation-induced error; that the Lebesgue constant grows slowly with M but deteriorates slightly for $M = 40$; that the error estimator effectivity is reasonably close to unity (recall that $\hat{\varepsilon}_M(\mu) \leq \varepsilon_M(\mu)$, $1 \leq M \leq M_{\max}$); and that B^M is well-conditioned for our choice of basis.

We now turn to the convergence results and error bounds for the reduced-basis approximation. In Figure 5-3(a) and (b) we plot, as a function of N and M, the maximum relative error in the energy norm $\epsilon^y_{N,M,\max,\text{rel}}$ and the maximum relative error bound $\Delta^y_{N,M,\max,\text{rel}}$; here, $\epsilon^y_{N,M,\max,\text{rel}}$ is the maximum over Ξ_{Test} of $|||e(\mu, t^K)|||/|||y(\mu_y, t^K)|||$ and $\Delta^y_{N,M,\max,\text{rel}}$ is the maximum over Ξ_{Test} of $\Delta^y_{N,M}(\mu, t^K)/|||y(\mu_y, t^K)|||$, where $\Xi_{\text{Test}} \subset (\mathcal{D})^{1000}$ is an input sample of size 1000 (a $10 \times 10 \times 10$ grid log-random in μ_1 and random in μ_2 and μ_3), and $\mu_y \equiv \arg\max_{\mu \in \Xi_{\text{Test}}} |||y(\mu, t^K)|||$.

We first note that N is much larger than in the previous example because the problem is much more complex. We observe the same convergence behavior as in the previous numerical example. For increasing N and fixed M the interpolation induced error starts to dominate at one point and the error does not further decrease — only by simultaneously increasing N and M can we obtain a smaller error and hence a better approximation. We also note that the error bound shows the same behavior as the actual error. However, there are differences in the (N, M) asymptotes which may reflect in our choice of N and M: concerning the actual error, we may select $M = 30$ for $N = 160$ without being limited by the interpolation induced error. The $M = 30$ asymptote for the error bound, on the other side, already levels off for approximately $N = 100$ — we thus need

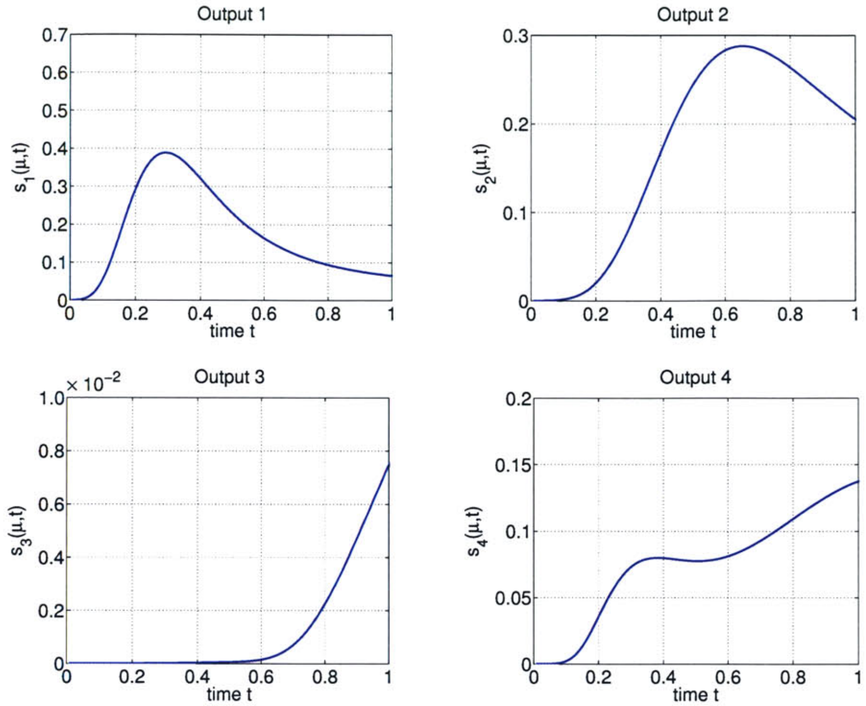
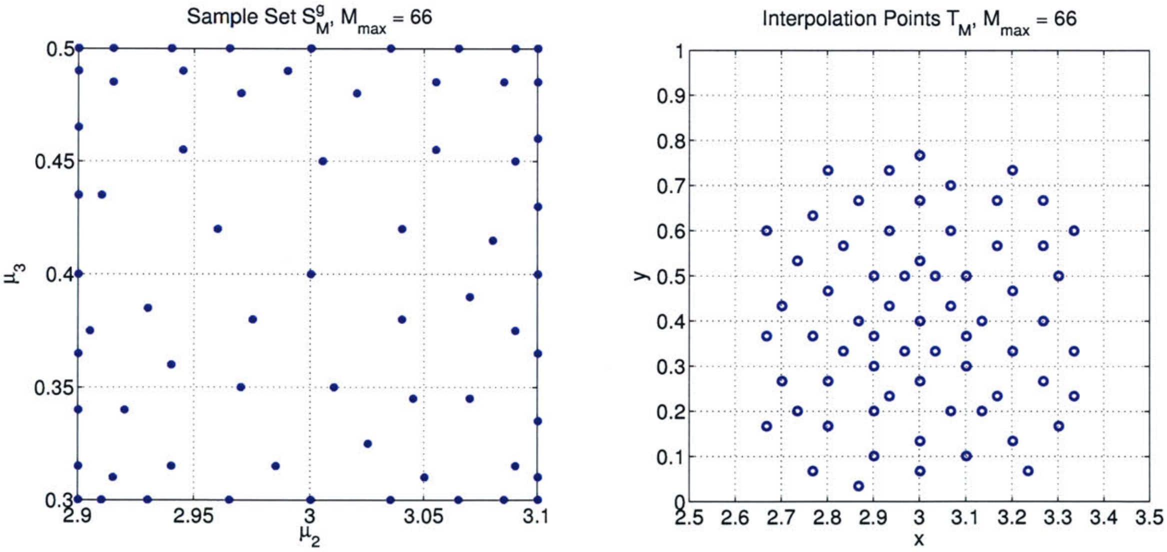


Figure 5-8: AP II: Outputs $s_q(\mu, t^k)$, $1 \leq q \leq 8$, for $\mu = (0.05, 3.1, 0.5)$ as a function of time.



(a)

(b)

Figure 5-9: AP II: (a) Parameter sample set S^g_M, $M_{\max} = 66$, and (b) interpolation points t_m, $1 \leq m \leq M_{\max}$, for the nonaffine pollution source (5.68).

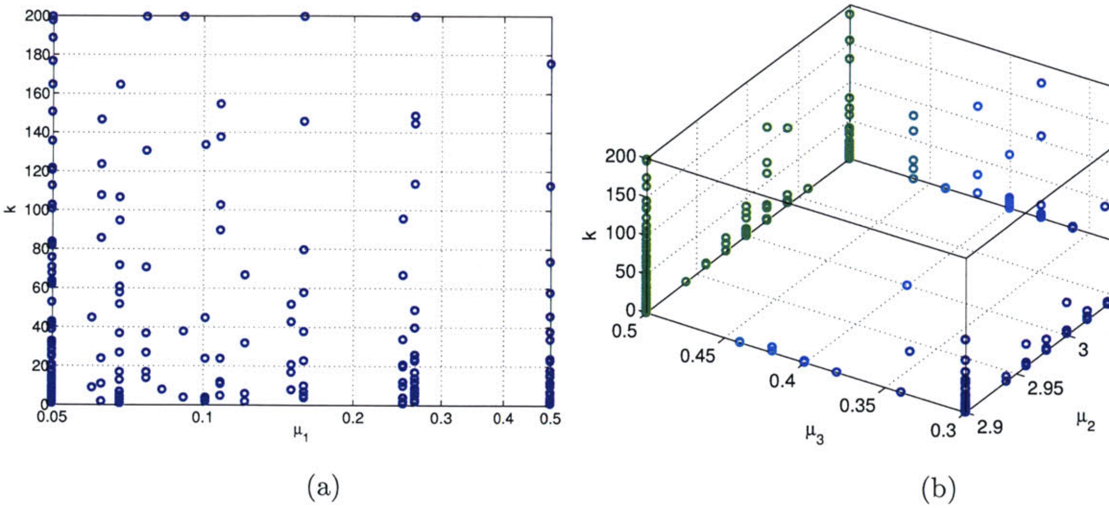


Figure 5-10: AP II: (a) Parameter sample set S_N^y, $N_{\max} = 256$, (a) projected onto μ_1-t^k-space; and (b) projected onto μ_2-μ_3-t^k-space .

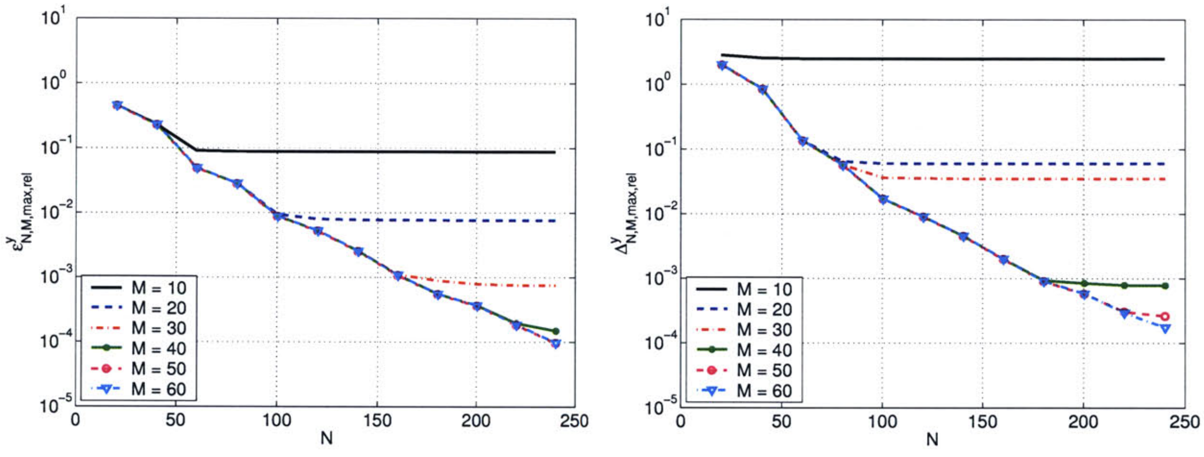
M	$\varepsilon^*_{M,\max}$	$\overline{\rho}_M$	Λ_M	$\overline{\eta}_M$	$\varkappa_M$
5	1.96 E−01	0.561	1.53	0.381	2.11
10	4.17 E−02	0.377	3.92	0.455	4.66
20	3.12 E−03	0.334	5.70	0.432	11.5
30	2.23 E−04	0.351	9.97	0.497	14.3
40	3.26 E−05	0.289	14.8	0.544	21.1
50	4.43 E−06	0.258	12.2	0.337	30.8
60	6.74 E−07	0.178	40.2	0.595	34.6

Table 5.5: AP II: Nonaffine function approximation; $\varepsilon^*_{M,\max}$, $\overline{\rho}_M$, Λ_M, $\overline{\eta}_M$, and $\varkappa_M$ as a function of M.

N	M	$\epsilon^y_{N,M,\max,\text{rel}}$	$\Delta^y_{N,M,\max,\text{rel}}$	$\overline{\eta}^y$
40	20	2.26 E−01	8.38 E−01	3.59
80	30	2.78 E−02	5.70 E−02	2.40
120	40	5.18 E−03	8.87 E−03	1.89
160	40	1.07 E−03	1.97 E−03	1.79
200	50	3.60 E−04	5.70 E−04	1.72
240	60	9.55 E−05	1.71 E−04	1.65

Table 5.6: AP II: Convergence rate and effectivities as a function of N and M.

to choose $M = 40$ so as not to limit the convergence of the error bound. Also, we observe that the $M = 50$ and $M = 60$ curves for the actual error coincide (the $M = 40$ curve deviates slightly only for $N \gtrapprox 220$); thus choosing $M = 50$ is sufficient even for $N = N_{\max}$. For the error bound the $M = 40$ asymptote separates earlier, and there is a slight difference between the $M = 50$ and $M = 60$ asymptotes for $N \gtrapprox 220$.



(a) (b)

Figure 5-11: AP II: (a) Maximum relative error in the energy norm and (b) error bound.

In Table 5.6 we present, as a function of N and M, $\epsilon^y_{N,M,\max,\text{rel}}$, $\Delta^y_{N,M,\max,\text{rel}}$, and the average effectivity $\overline{\eta}^y$, where $\overline{\eta}^y$ is the average over $\Xi_{\text{Test}} \times \mathbb{I}$ of $\Delta^y_{N,M}(\mu, t^k)/|||y(\mu, t^k) - y_N(\mu, t^k)|||$. Here, we select M, for a given N, such that the nonaffine function approximation does not limit the convergence of the error bound. We confirm the fast convergence already observed in the convergence plots and note that the effectivities are very good throughout.

We next present in Tables 5.7 and 5.8 the maximum relative output error $\epsilon^s_{N,M,\max,\text{rel}}$, the maximum relative output bound $\Delta^s_{N,M,\max,\text{rel}}$, and the average effectivity $\overline{\eta}^s_{N,M}$ as a function of N and M for output 1 and 2, respectively (the results for the third and fourth output are similar to the results presented). Here, $\epsilon^s_{N,M,\max,\text{rel}}$ is the maximum over Ξ_{Test} of $|s(\mu, t_\eta(\mu)) - s_N(\mu, t_\eta(\mu))|/s_{\max}$, $\Delta^s_{N,M,\max,\text{rel}}$ is the maximum over Ξ_{Test} of $\Delta^s_{N,M}(\mu, t^K)/|s_{\max}|$ and $\overline{\eta}^s$ is the average over Ξ_{Test} of $\Delta^s_{N,M}(\mu, t_\eta(\mu))/|s(\mu, t_\eta(\mu)) - s_{N,M}(\mu, t_\eta(\mu))|$, where $t_\eta(\mu) \equiv \arg\max_{t^k \in \mathbb{I}} |s(\mu, t^k) - s_N(\mu, t^k)|$

N	M	$\epsilon^s_{N,M,\max,\text{rel}}$	$\Delta^s_{N,M,\max,\text{rel}}$	$\bar{\eta}^s_{N,M}$
40	20	2.82 E-02	2.69 E+00	109
80	30	3.91 E-03	1.83 E-01	70.7
120	40	5.64 E-04	2.85 E-02	66.2
160	40	1.03 E-04	6.33 E-03	66.7
200	50	1.77 E-05	1.83 E-03	142
240	60	4.27 E-06	5.49 E-04	134

Table 5.7: AP II: Maximum relative output error, output bound, and effectivities for output 1.

N	M	$\epsilon^s_{N,M,\max,\text{rel}}$	$\Delta^s_{N,M,\max,\text{rel}}$	$\bar{\eta}^s_{N,M}$
40	20	1.64 E-01	7.34 E+00	50.5
80	30	1.47 E-02	4.99 E-01	57.4
120	40	2.18 E-03	7.76 E-02	53.1
160	40	5.36 E-04	1.72 E-02	68.4
200	50	1.33 E-04	4.99 E-03	85.2
240	60	1.94 E-05	1.49 E-03	77.0

Table 5.8: AP II: Maximum relative output error, output bound, and effectivities for output 2.

and $s_{\max} \equiv \max_{t^k \in \mathbb{I}} \max_{\mu \in \Xi_{\text{Test}}} |s(\mu, t^k)|$. The error in the output converges fast and the output effectivities are still acceptable for the simple output bound. However, for an accuracy of the *output error bound* of 1%, we require approximately $N = 160$ and $M = 40$. Although the true error is almost two magnitudes smaller, in actual practice we can only guarantee the accuracy as determined from our error bound. Better effectivities, i.e., sharper bounds, would therefore be advantageous since the required values for N and M to obtain the desired accuracy decrease — if the effectivities would be close to one, $N = 80$ and $M = 30$ would suffice for a 1% accuracy in the output bound. We could probably achieve these numbers by introducing the dual problem (see the results in Section 4.8.5).

Finally, in Table 5.9 we present, as a function of N and M, the online computational times to calculate $s_{N,M}(\mu, t^k)$ and $\Delta^s_{N,M}(\mu, t^k)$, $\forall\, k \in \mathbb{K}$. The values are normalized with respect to the computational time for the direct calculation of the truth approximation output $s(\mu, t^k) = \ell(y(\mu, t^k))$, $\forall\, k \in \mathbb{K}$. The computational savings for $N = 160$ and $M = 40$ — corresponding to an output bound with 1% accuracy — are approximately a factor of 30. The corresponding run-time in MATLAB 6.5 on a 750 MHz Pentium III is 0.78 sec. We note that, despite the $O(KM^2N^2)$ complexity to calculate the output bound, the time to calculate $s_N(\mu, t^k)$ ultimately dominates because of the large N. The savings here are pretty small because (i) the dimension $\mathcal{N}$ of the underlying truth approximation is not very large, and (ii) we have to choose N large to obtain the desired accuracy.

N	M	$s_{N,M}(\mu, t^k)$, $\forall\, k \in \mathbb{K}$	$\Delta^s_{N,M}(\mu, t^k)$, $\forall\, k \in \mathbb{K}$	$s(\mu, t^k)$, $\forall\, k \in \mathbb{K}$
40	20	1.04 E-03	2.34 E-03	1
80	30	2.78 E-03	5.19 E-03	1
120	40	1.04 E-02	1.19 E-02	1
160	40	1.85 E-02	1.75 E-02	1
200	50	3.20 E-02	2.65 E-02	1
240	60	4.85 E-02	3.62 E-02	1

Table 5.9: AP II: Online computational times (normalized with respect to the time to solve for $s(\mu, t^k)$, $\forall\, k \in \mathbb{K}$).

Chapter 6

Nonlinear Parabolic Equations

6.1 Introduction

In the last two chapters we developed reduced-basis methods and associated *a posteriori* error estimation procedures for *linear* parabolic partial differential equations with *affine* and *nonaffine* parameter dependence. We now extend our methodology to treat certain classes of *nonaffine nonlinear* parabolic partial differential equations. More specifically, we consider problems where $a(w, v; g)$ can be written as

$$a(w, v; g(w; x; \mu)) \equiv a^L(w, v) + \int_\Omega g(w; x; \mu)\, v. \tag{6.1}$$

Here, $x \in \Omega$ is the spatial coordinate, $\mu \in \mathcal{D}$ is the input parameter, $a^L(w, v)$ is a bounded bilinear form, and $g(w; x; \mu)$ is a nonaffine nonlinear function which is monotonically increasing in w. Similar to the previous chapter, we will introduce a "collateral" reduced-basis expansion for $g(w; x; \mu)$ and employ the empirical interpolation method to determine the coefficients for the approximation to $g(w; x; \mu)$.

The nonlinear dependence on the field variable, however, introduces new numerical difficulties: first, our greedy choice of basis functions ensures good approximation properties, but is very expensive for the nonlinear time-dependent case; second, since the field variable is not known in advance, it is difficult to generate an explicit affine approximation of $g(w; x; \mu)$; and third, it is challenging to ensure that the online complexity remains independent of $\mathcal{N}$ even in the presence of highly nonlinear terms.

In the first part of this chapter we develop the necessary theory and also present a numerical example to test and confirm our approach. In Section 6.6 we consider the application to a specific problem in the class of reaction-diffusion systems. Since the focus here is the treatment of the nonlinear term, we do not consider adjoint formulations (see [121] for an application of adjoint techniques in the reduced-basis context to the steady Navier-Stokes equation).

6.2 Abstract Formulation

As in the last two chapters, we directly consider a time-discrete framework associated to the time interval $I \equiv]0, t_f]$. We recall that $\bar{I}$ is divided into K subintervals of equal length $\Delta t = \frac{t_f}{K}$, that t^k is defined by $t^k \equiv k\Delta t$, $0 \leq k \leq K \equiv \frac{t_f}{\Delta t}$; furthermore, $\mathbb{I} \equiv \{t^0, \dots, t^k\}$ and $\mathbb{K} \equiv \{1, \dots, K\}$. We shall consider Euler-Backward for the time integration. We also recall our reference (or "truth") finite element approximation space Y of very large dimension $\mathcal{N}$. Clearly, our results must be stable as $\Delta t \to 0$, $K \to \infty$, and $\mathcal{N} \to \infty$.

We directly consider the "truth" approximation here: Given a parameter $\mu \in \mathcal{D}$, we evaluate the (here, single) output of interest

$$s(\mu, t^k) = \ell(y(\mu, t^k)), \quad \forall\, k \in \mathbb{K} \tag{6.2}$$

where the field variable $y(\mu, t^k) \in Y$, $\forall\, k \in \mathbb{K}$, satisfies the weak form of the nonlinear parabolic partial differential equation

$$\begin{aligned} m(y(\mu, t^k), v) + \Delta t\ a^L(y(\mu, t^k), v) + \Delta t \int_\Omega g(y(\mu, t^k); x; \mu)\, v \\ = m(y(\mu, t^{k-1}), v) + \Delta t\ b(v)\ u(t^k), \quad \forall\, v \in Y, \end{aligned} \tag{6.3}$$

with initial condition (say) $y(\mu, t^0) = 0$. Here, μ and $\mathcal{D}$ are the input and input domain; $m(\cdot,\cdot)$ and $b(\cdot)$, $\ell(\cdot)$ are X-continuous bilinear and linear forms, respectively; $a^L(\cdot,\cdot)$ is a Y-continuous bilinear form; $u(t^k)$ denotes the (here, single) control input; and $g(w; x; \mu) \in L^2(\Omega)$, $\forall\, w \in Y$ is a nonlinear nonaffine function of the field variable $y(\mu, t^k)$, the spatial coordinate x, and the parameter μ, which is monotonically increasing in w for all $\mu \in \mathcal{D}$. We note that the field variable, $y(\mu, t^k)$, is of course also a function of the spatial coordinate x. In the sequel we will use the notation $y(x; \mu, t^k)$ to signify this dependence whenever it is crucial.

We shall make the following assumptions. We assume that $a^L(\cdot,\cdot)$ and $m(\cdot,\cdot)$ are continuous

$$a^L(w, v) \;\leq\; \gamma_a \|w\|_Y \|v\|_Y, \quad \forall\, w, v \in Y, \tag{6.4}$$

$$m(w, v) \;\leq\; \gamma_m \|w\|_X \|v\|_X, \quad \forall\, w, v \in Y; \tag{6.5}$$

coercive,

$$0 \;<\; \alpha_a \equiv \inf_{w \in X} \frac{a^L(w, w)}{\|w\|_X^2}, , \tag{6.6}$$

$$0 \;<\; \alpha_m \equiv \inf_{v \in Y} \frac{m(v, v)}{\|v\|_X^2}; \tag{6.7}$$

and symmetric, $a^L(v, w) = a^L(w, v)$, $\forall\, v, w \in Y$, and $m(v, w) = m(w, v)$, $\forall\, w, v \in X$. (We (plausibly) suppose that γ_a, γ_m, α_a, α_m may be chosen independent of $\mathcal{N}$.) We also require that the linear forms $b(\cdot) : Y \to \mathbb{R}$ and $\ell(\cdot) : Y \to \mathbb{R}$ be bounded with respect to $\|\cdot\|_X$. And finally, we require that all linear and bilinear forms are independent of time — the system is thus linear time-invariant (LTI).

Since the focus of this section is the treatment of the nonlinearity $g(w; x; \mu)$ we assume that the bilinear and linear forms m, a^L and b, ℓ are parameter independent; a parameter dependence of either form is readily admitted. Note also that our results presented here directly carry over to the case where g is also an explicit function of (discrete) time t^k.

6.2.1 Numerical Exercise 5: A Nonlinear Diffusion Problem

We now turn to a numerical example. We consider the following nonlinear diffusion problem defined on the unit square, $\Omega =]0,1[^2 \in \mathbb{R}^2$: Given $\mu = (\mu_1, \mu_2) \in \mathcal{D}^\mu \equiv [0.01, 10]^2$, we evaluate $y(\mu, t^k) \in Y$ from (6.3), where $Y \subset Y^{\mathrm{e}} \equiv H_0^1(\Omega)$ is a linear finite element truth approximation subspace of dimension $\mathcal{N} = 2601$,

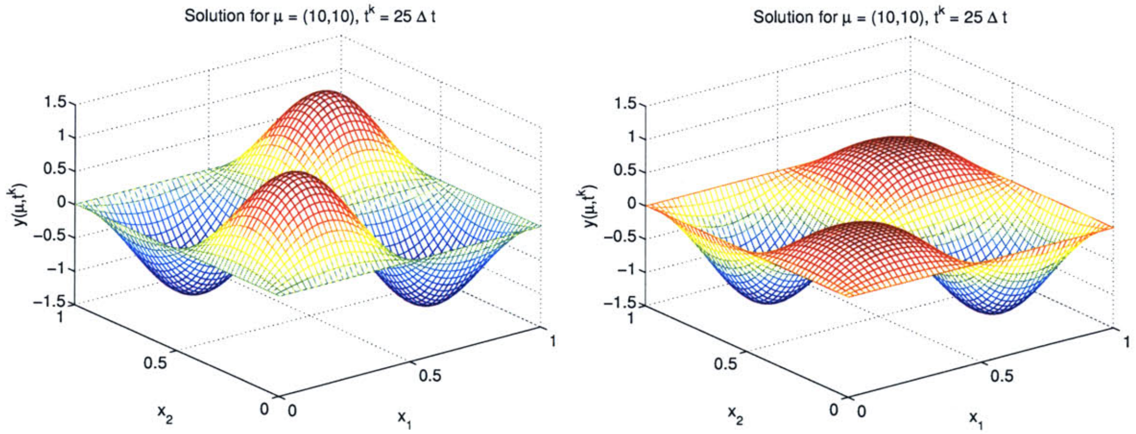
$$m(w,v) \equiv \int_\Omega w\, v, \qquad a^L(w,v) \equiv \int_\Omega \nabla w \cdot \nabla v, \qquad b(v) \equiv 100 \int_\Omega v \, \sin(2\pi x_1)\, \cos(2\pi x_2), \tag{6.8}$$

and the nonlinearity is given by

$$g(y(\mu, t^k); \mu) = \mu_1 \, \frac{e^{\mu_2\, y(\mu, t^k)} - 1}{\mu_2}. \tag{6.9}$$

The output $s(\mu, t^k)$ is evaluated from (6.2) with $\ell(v) = \int_\Omega v$. We presume the periodic control input $u(t^k) = \sin(2\pi t^k)$, $t^k \in \mathbb{I}$. We shall consider the time interval $\bar{I} = [0, 2]$ and a timestep $\Delta t = 0.01$; we thus have $K = 200$.

We note that μ_2 represent the strength of the nonlinearity whereas μ_1 represents strength of the sink term in (6.9); as $\mu_2 \to 0$ we have $g(w;\mu) \to \mu_1 w$. The solution thus tends to the solution for the linear problem as μ_2 tends to zero. We believe that, because of the monotonicity of g, it can be proven that the problem is well-posed. Two snapshots of the solution $y(\mu, t^k)$ at time $t^k = 25\Delta t$ are shown for $\mu = (0.01, 0.01)$ and $\mu = (10, 10)$ in Figures 6-1(a) and (b), respectively. We observe that the solution has two negative peaks and two positive peaks with similar height for $\mu = (0.01, 0.01)$ (which oscillate back and forth in time). As μ_2 increases, the height of the negative peaks remains largely unchanged, while the positive peaks get rectified as shown in Figure 6-1(b). The exponential nonlinearity has a damping effect on the positive part of $y(\mu, t^k)$, but has (almost) no effect on the negative part. Note that the solution for $\mu = (10, 10)$, of course, also oscillates in time — with the positive peaks always being smaller than the negative peaks.

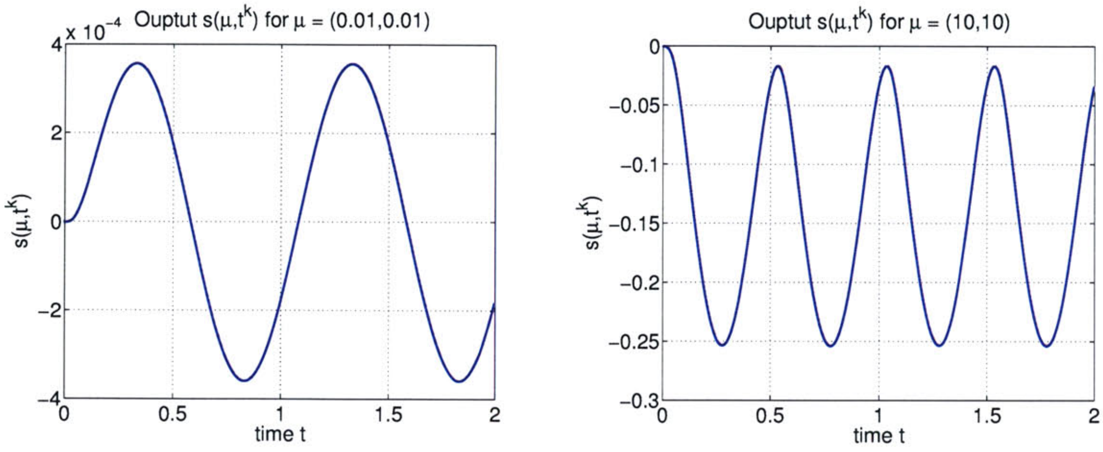


(a) (b)

Figure 6-1: NE 5: Solution $y(\mu, t^k)$ at $t^k = 25\Delta t$ for (a) $\mu = (0.01, 0.01)$ and (b) $\mu = (10, 10)$.

The influence of the nonlinearity is also evident in the output $s(\mu, t^k)$ — the average of $y(\mu, t^k)$

over Ω — plotted in Figure 6-2(a) and (b) for $\mu = (0.01, 0.01)$ and $\mu = (10, 10)$, respectively. Note the *very* different scaling of the output for the two parameters: $s(\mu, t^k)$ is close to zero for $\mu = (0.01, 0.01)$ since the positive and negative peaks of $y(\mu, t^k)$ have almost the same hight. For $\mu = (10, 10)$ the output is well below zero for all times and oscillates with approximately twice the frequency.



(a)

(b)

Figure 6-2: NE 5: Output $s(\mu, t^k)$ for (a) $\mu = (0.01, 0.01)$ and (b) $\mu = (10, 10)$.

6.3 Reduced-Basis Approximation

6.3.1 Formulation

We first introduce the nested sample sets $S_N^y = \{\tilde{\mu}_1^y \in \tilde{\mathcal{D}}, \ldots, \tilde{\mu}_N^y \in \tilde{\mathcal{D}}\}$, $1 \leq N \leq N_{\max}$, and $S_M^g = \{\tilde{\mu}_1^g \in \tilde{\mathcal{D}}, \ldots, \tilde{\mu}_M^g \in \tilde{\mathcal{D}}\}$, $1 \leq M \leq M_{\max}$ where $\tilde{\mu} \equiv (\mu, t^k)$ and $\tilde{\mathcal{D}} \equiv \mathcal{D} \times \mathbb{T}$. Note that, since $g(\cdot; x; \mu)$ is a function of the field variable $y(\mu, t^k)$, the sample set S_M^g must now also reside in *parameter-time* space $\tilde{\mathcal{D}}$; in general, $S_N^y \neq S_M^g$ and in fact $N \neq M$. We then define the associated nested Lagrangian [85] reduced-basis space

$$W_N^y = \operatorname{span}\{\zeta_n \equiv y(\tilde{\mu}_n^y),\ 1 \leq n \leq N\}, \quad 1 \leq N \leq N_{\max}, \tag{6.10}$$

where $y(\tilde{\mu}_n^y)$ is the solution of (6.2) at time $t = t^{k_n^y}$ for $\mu = \mu_n^y$. We also define the nested collateral reduced-basis space

$$W_M^g = \operatorname{span}\{\xi_n \equiv g(y(\tilde{\mu}_n^g); x; \mu),\ 1 \leq n \leq M\} = \operatorname{span}\{q_1, \ldots,\ q_M\}, \quad 1 \leq M \leq M_{\max}, \tag{6.11}$$

and nested set of interpolation points $T_M = \{t_1, \ldots, t_M\}$, $1 \leq M \leq M_{\max}$.

Let us first assume that we do not have access to the empirical interpolation method and instead follow the standard approach. We would then our reduced-basis approximation by a standard Galerkin projection: given $\mu \in \mathcal{D}$, the reduced-basis approximation $y_N(\mu, t^k) \in W_N^y$ to $y(\mu, t^k)$

satisfies

$$m(y_N(\mu,t^k),v) + \Delta t\ a^L(y_N(\mu,t^k),v) + \Delta t \int_\Omega g(y_N(\mu,t^k);x;\mu)\ v \\ = m(y_N(\mu,t^{k-1}),v) + \Delta t\ b(v)\ u(t^k), \quad \forall\, v \in W_N^y,\ \forall\, k \in \mathbb{K}, \quad (6.12)$$

We may now express $y_N(\mu,t^k) = \sum_{j=1}^N y_{Nj}(\mu,t^k)\zeta_j$ and choose as test functions $v = \zeta_i,\ 1 \le i \le N$, in (6.12) to obtain, $\forall\, k \in \mathbb{K}$,

$$\sum_{j=1}^N \left\{m(\zeta_j,\zeta_i) + \Delta t\ a^L(\zeta_j,\zeta_i)\right\}\ y_{Nj}(\mu,t^k) + \Delta t \int_\Omega g\Big(\sum_{j=1}^N y_{Nj}(\mu,t^k)\zeta_j;x;\mu\Big)\ \zeta_i \\ = \sum_{j=1}^N m(\zeta_j,\zeta_i)\ y_{Nj}(\mu,t^{k-1}) + \Delta t\ b(\zeta_i)\ u(t^k), \quad 1 \le i \le N. \quad (6.13)$$

We may now apply a (say) Newton iterative scheme to solve (6.13) at each timestep for $y_{Nj}(\mu,t^k),\ 1 \le j \le N$: given the solution at the previous timestep, $y_{Nj}(\mu,t^{k-1}),\ 1 \le j \le N$ and a current iterate $\bar{y}_{Nj}(\mu,t^k),\ 1 \le j \le N$, we find an increment $\delta y_{Nj},\ 1 \le j \le N$, such that

$$\sum_{j=1}^N \left\{ m(\zeta_j,\zeta_i) + \Delta t\ a^L(\zeta_j,\zeta_i) + \Delta t \int_\Omega g_1\Big(\sum_{n=1}^N \bar{y}_{Nn}(\mu,t^k)\zeta_n;x;\mu\Big)\ \zeta_j\ \zeta_i \right\}\ \delta y_{Nj} \\ = \sum_{j=1}^N m(\zeta_j,\zeta_i)\ y_{Nj}(\mu,t^{k-1}) + \Delta t\ b(\zeta_i)\ u(t^k) - \sum_{j=1}^N \left\{m(\zeta_j,\zeta_i)\right. \\ \left. + \Delta t\ a^L(\zeta_j,\zeta_i)\right\}\ \bar{y}_{Nj}(\mu,t^k) - \Delta t \int_\Omega g\Big(\sum_{j=1}^N \bar{y}_{Nj}(\mu,t^k)\zeta_j;x;\mu\Big)\ \zeta_i, \quad 1 \le i \le N(6.14)$$

where g_1 is the partial derivative with respect to the first argument.

We note that if g is a low-order (at most quadratically) polynomial nonlinearity in $y(\mu,t^k)$, we can expand the nonlinear terms $g(\sum_{j=1}^N \bar{y}_{Nj}(\mu,t^k)\zeta_j;x;\mu)$ and $g_1(\sum_{n=1}^N \bar{y}_{Nn}(\mu,t^k)\zeta_n;x;\mu)$ into their power series and develop an efficient, i.e., online $\mathcal{N}$-independent, offline-online computational decomposition [122, 121]. Unfortunately, for high-order polynomial or non-polynomial nonlinearities this trick cannot be used. Hence, $g(\sum_{j=1}^N \bar{y}_{Nj}(\mu,t^k)\zeta_j;x;\mu)\zeta_i$ and $g_1(\sum_{n=1}^N \bar{y}_{Nn}(\mu,t^k)\zeta_n;x;\mu)\zeta_j\zeta_i$ must be evaluated online at every Newton iteration with $\mathcal{N}$-dependent cost.

To recover online $\mathcal{N}$-independent cost, we again appeal to the empirical interpolation method. We replace the nonlinearity $g(y_N(\mu,t^k);x;\mu)$ in (6.12) by the affine approximation $g_M^{y_N,M}(x;\mu,t^k)$ given by

$$g_M^{y_N,M}(x;\mu,t^k) = \sum_{m=1}^M \varphi_{Mm}(\mu,t^k)\ q_m(x) \quad (6.15)$$

where the coefficients $\varphi_{Mm}(\mu, t^k)$ are determined from

$$\sum_{j=1}^{M} B_{ij}^M \ \varphi_{Mj}(\mu, t^k) = g(y_{N,M}(t_i; \mu, t^k); t_i; \mu), \quad 1 \le i \le M, \tag{6.16}$$

and $B_{ij}^M = q_j(t_i)$, $1 \le i, j \le M$. Note that, contrary to the nonaffine case, $\varphi_M(\mu, t^k)$ now also depends on time. Our reduced-basis approximation $y_{N,M}(\mu, t^k)$ to $y(\mu, t^k)$ is then obtained by a standard Galerkin projection: given $\mu \in \mathcal{D}$, $y_{N,M}(\mu, t^k) \in W_N^y$ satisfies

$$\begin{aligned} m(y_{N,M}(\mu, t^k), v) + \Delta t \ a^L(y_{N,M}(\mu, t^k), v) + \Delta t \int_\Omega g_M^{y_{N,M}}(x; \mu, t^k) \ v \\ = m(y_{N,M}(\mu, t^{k-1}), v) + \Delta t \ b(v) \ u(t^k), \quad \forall\, v \in W_N^y, \ \forall\, k \in \mathbb{K}, \end{aligned} \tag{6.17}$$

with initial condition $y_{N,M}(\mu, t^0) = 0$. We will show in the next section that (6.17) indeed allows an efficient offline-online decomposition. Finally, we evaluate the output from

$$s_{N,M}(\mu, t^k) = \ell(y_{N,M}(\mu, t^k)), \quad \forall\, k \in \mathbb{K}. \tag{6.18}$$

At this point we should remark that our current approach of constructing the sample set S_M^g and associated reduced-basis space W_M^g in the nonlinear parabolic case is computationally prohibitively profligate. The reason, related to our greedy adaptive sampling procedure proposed in Section 2.4, is twofold. First, we need to calculate and store the "truth" solution $y(\mu, t^k)$ at all times $t^k \in \mathbb{I}$ on the grid Ξ^g in parameter space. For our numerical example in Section 6.2.1 Ξ^g is of size 144 — we thus need to solve (6.3) 144 times and store 144 × 200 "truth" solutions $y(\mu, t^k)$! And second, as pointed out in Section 2.4, determining the next sample point $\tilde{\mu}_n^g$ in $\tilde{\Xi}^g \equiv \Xi^g \times \mathbb{I}$ requires the solution of a linear program for all $\mu \in \tilde{\Xi}^g$ if the function g is time-varying, as is inherently the case in the nonlinear context.[1] We note, however, that both of these reasons are due to the fact that we do not have an error estimator for $y(\mu, t^k)$ without knowing the approximation for g (which depends on $y(\mu, t^k)$). Since this computation is too expensive in our current implementation, we revert to the least squares surrogate in this chapter. In choosing this approach we in fact rely on our numerical comparison in Section 2.4 which shows that we can expect similar results.

6.3.2 Offline-Online Computational Procedure

In this section we develop the offline-online computational decomposition to recover online $\mathcal{N}$-independence even in the nonlinear case. We first express $y_{N,M}(\mu, t^k)$ as

$$y_{N,M}(\mu, t^k) = \sum_{n=1}^{N} y_{N,M\,n}(\mu, t^k) \ \zeta_n, \tag{6.19}$$

and choose as test functions $v = \zeta_n$, $1 \le n \le N$, in (6.17).

It then follows from (6.15) that $\underline{y}_{N,M}(\mu, t^k) = [y_{N,M\,1}(\mu, t^k) \ \ y_{N,M\,2}(\mu, t^k) \ \ldots \ y_{N,M\,N}(\mu, t^k)]^T \in$

[1] Note that in the linear nonaffine parabolic case the function g depends only on x and μ and *not* on time.

$\mathbb{R}^N$, $\forall\, k \in \mathbb{K}$, satisfies

$$(M_N + \Delta t\ A_N)\ \underline{y}_{N,M}(\mu, t^k) + \Delta t\ C^{N,M}\ \varphi_M(\mu, t^k) = M_N\ \underline{y}_{N,M}(\mu, t^{k-1}) + \Delta t\ B_N\ u(t^k), \quad (6.20)$$

with initial condition $y_{N,M\,n}(\mu, t^0) = 0,\ 1 \le n \le N$. Here, the coefficients $\varphi_M(\mu, t^k) = [\varphi_{M\,1}(\mu, t^k)\ \varphi_{M\,2}(\mu, t^k)\ \dots\ \varphi_{M\,M}(\mu, t^k)]^T \in \mathbb{R}^M$ are determined from (6.16); $M_N \in \mathbb{R}^{N \times N}$, $A_N \in \mathbb{R}^{N \times N}$, and $C^{N,M} \in \mathbb{R}^{N \times M}$, are *parameter-independent* matrices with entries $M_{N\,i,j} = m(\zeta_i, \zeta_j),\ 1 \le i, j \le N$, $A_{N\,i,j} = a(\zeta_i, \zeta_j),\ 1 \le i, j \le N$, and $C^{N,M}_{i,j} = \int_\Omega \zeta_i\, q_j,\ 1 \le i \le N,\ 1 \le j \le M$, respectively; and $B_N \in \mathbb{R}^N$ is a *parameter independent* vector with entries $B_{N\,i} = b(\zeta_i),\ 1 \le i \le N$ [2]

We can now substitute $\varphi_{M\,m}(\mu, t^k)$ from (6.16) into (6.20) to obtain the nonlinear algebraic system

$$\begin{aligned}(M_N + \Delta t\ A_N)\ \underline{y}_{N,M}(\mu, t^k) + \Delta t\ D^{N,M}\ g(Z^{N,M}\, \underline{y}_{N,M}(\mu, t^k); \underline{t}_M; \mu)& \\ = M_N\ \underline{y}_{N,M}(\mu, t^{k-1}) + \Delta t\ B_N\ u(t^k), \quad \forall\, k \in \mathbb{K},& \quad (6.21)\end{aligned}$$

where $D^{N,M} = C^{N,M}(B^M)^{-1} \in \mathbb{R}^{N \times M}$, $Z^{N,M} \in \mathbb{R}^{M \times N}$ is a *parameter-independent* matrix with entries $Z^{N,M}_{i,j} = \zeta_j(t_i),\ 1 \le i \le M,\ 1 \le j \le N$, and $\underline{t}_M = [t_i\ \dots t_M]^T \in \mathbb{R}^M$ is the set of interpolation points. We now solve for $\underline{y}_{N,M}(\mu, t^k)$ at each timestep using a Newton iterative scheme: given the solution for the previous timestep, $\underline{y}_{N,M}(\mu, t^{k-1})$, and a current iterate $\bar{\underline{y}}_{N,M}(\mu, t^k)$, we find an increment $\delta\underline{y}_{N,M}$ such that

$$\begin{aligned}\left(M_N + \Delta t\ A_N + \Delta t \bar{E}^N\right)\ \delta\underline{y}_{N,M}& \\ = M_N\ \underline{y}_{N,M}(\mu, t^{k-1}) + \Delta t\ B_N(\mu)\ u(t^k) - (M_N + \Delta t\ A_N)\ \bar{\underline{y}}_{N,M}(\mu, t^k)& \\ - \Delta t\ D^{N,M}\ g(Z^{N,M}\, \bar{\underline{y}}_{N,M}(\mu, t^k); \underline{t}_M; \mu),& \quad (6.22)\end{aligned}$$

where $\bar{E}^N \in \mathbb{R}^{N \times N}$ must be calculated at every Newton iteration from

$$\bar{E}^N_{i,j} = \sum_{m=1}^{M} D^{N,M}_{i,m}\, g_1\!\left(\sum_{n=1}^{N} \bar{y}_{N,M\,n}(\mu, t^k) \zeta_n(t_m); t_m; \mu \right) \zeta_j(t_m), \quad 1 \le i, j \le N. \quad (6.23)$$

Finally, we evaluate the output estimate from

$$s_{N,M}(\mu, t^k) = L_N^T\ \underline{y}_{N,M}(\mu, t^k), \quad \forall\, k \in \mathbb{K}, \quad (6.24)$$

where $L_N \in \mathbb{R}^N$ is the output vector with entries $L_{N\,i} = \ell(\zeta_i),\ 1 \le i \le N$.

The offline-online decomposition is now clear. In the offline stage — performed only *once* — we first construct the nested approximation spaces W^g_M and sets of interpolation points T_M, $1 \le M \le M_{\max}$; we then solve for the ζ_n, $1 \le n \le N_{\max}$ and compute and store the μ-independent quantities M_N, A_N, B^M, $D^{N,M}$, B_N, and $Z^{N,M}$. In the online stage — performed many times, for each new parameter value μ — we solve (6.22) for $\underline{y}_{N,M}(\mu, t^k)$ and evaluate the output estimate $s_{N,M}(\mu, t^k)$ from (6.24). The operation count is dominated by the Newton update at each timestep: we first

[2] We follow the notation based on the previous chapters here. It should be straightforward to distinguish between the control vector, B_N, and the "interpolation" matrix, B^M in (6.16).

assemble $\bar{E}^N$ from (6.23) at cost $O(MN^2)$ — note that we perform the sum in the parenthesis of (6.23) first before performing the outer sum — and then invert the left hand side of (6.22) at cost $O(N^3)$. The operation count in the online stage is thus $O(\bar{\kappa}K(MN^2 + N^3))$, where $\bar{\kappa}$ is the average number of Newton steps per timestep. We thus recover $\mathcal{N}$-independence in the online stage.

We remark that, in actual practice, M can be quite large — and in fact much larger than N. In this case it is straightforward to reduce M without sacrificing accuracy by splitting the time interval $\mathbb{I}$ into several smaller subintervals $\mathbb{I}_1, \ldots, \mathbb{I}_\mathcal{I}$ such that $\mathbb{I} = \bigcup_{i=1,\mathcal{I}} \mathbb{I}_i$. We then construct, in the offline stage, $\mathcal{I}$ separate samples sets $S^g_{M\,i}$, $1 \le i \le \mathcal{I}$ and associated reduced-basis spaces $W^g_{M\,i}$, $1 \le i \le \mathcal{I}$ on each interval $\mathbb{I}_i$, $1 \le i \le \mathcal{I}$. In the online stage we simply "switch" to the corresponding sample — and hence T_M, B^M, and $D^{N,M}$ — as time progresses. This approach renders the offline computation more expensive, but can increases the online efficiency considerably while retaining the desired accuracy.

6.4 *A Posteriori* Error Estimation

6.4.1 Preliminaries

We now turn to the development of our *a posteriori* error estimator; by construction rather similar to the nonaffine parabolic case in Chapter 5. To begin, we recall that the bilinear form a^L is assumed to be parameter independent here; we can thus use the coercivity constant α_a and have no need for the lower bound $\hat{\alpha}_a(\mu)$ required earlier. We next introduce the dual norm of the residual

$$\varepsilon_{N,M}(\mu, t^k) \equiv \sup_{v \in Y} \frac{R(v; \mu, t^k)}{\|v\|_Y}, \quad \forall\, k \in \mathbb{K}, \tag{6.25}$$

where

$$R(v; \mu, t^k) \equiv b(v)\ u(t^k) - \frac{1}{\Delta t} m(y_{N,M}(\mu, t^k) - y_{N,M}(\mu, t^{k-1}), v) \\ - a^L(y_{N,M}(\mu, t^k), v) - \int_\Omega g_M^{y_{N,M}}(x; \mu, t^k)\ v, \quad \forall\, v \in Y,\ \forall\, k \in \mathbb{K}, \tag{6.26}$$

is the residual associated to the nonlinear parabolic problem. We also require the dual norm

$$\vartheta^q_M \equiv \sup_{v \in Y} \frac{\int_\Omega q_{M+1}\ v}{\|v\|_Y}. \tag{6.27}$$

and the error bound $\hat{\varepsilon}_M(\mu, t^k)$ for the nonlinear function approximation given by

$$\hat{\varepsilon}_M(\mu, t^k) \equiv |g(y_{N,M}(t_{M+1}; \mu, t^k); t_{M+1}; \mu) - g_M(y_{N,M}(t_{M+1}; \mu, t^k); t_{M+1}; \mu)|. \tag{6.28}$$

We note that, contrary to the nonaffine case, the error bound $\hat{\varepsilon}_M(\mu, t^k)$ is now also a function of (discrete) time.

The bounding properties for the errors in the field variable and the output estimate are stated below. Throughout this section we assume that the "truth" solution $y(\mu, t^k)$ satisfy (6.3) and the corresponding reduced-basis approximation $y_{N,M}(\mu, t^k)$ satisfies (6.17), respectively.

6.4.2 Error Bound Formulation

Primal Variable

We obtain the following result for the error in the energy norm.

Proposition 17. *Suppose that $g(y_{N,M}(\mu,t^k);\mu) \in W^g_{M+1}$, $\forall\, k \in \mathbb{K}$. Let $e(\mu,t^k) \equiv y(\mu,t^k) - y_{N,M}(\mu,t^k)$ be the error in the field variable and define the "spatio-temporal" energy norm*

$$|||v(\mu,t^k)||| \equiv \left(m(v(\mu,t^k),v(\mu,t^k)) + \sum_{k'=1}^{k} a^L(v(\mu,t^{k'}),v(\mu,t^{k'}))\ \Delta t \right)^{\frac{1}{2}}, \quad \forall\, v \in Y. \tag{6.29}$$

The error is then bounded by

$$|||e(\mu,t^k)||| \leq \Delta^y_{N,M}(\mu,t^k), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{6.30}$$

where the error bound $\Delta^y_{N,M}(\mu,t^k)$ is defined as

$$\Delta^y_{N,M}(\mu,t^k) \equiv \left(\frac{2\Delta t}{\alpha_a} \sum_{k'=1}^{k} \varepsilon_{N,M}(\mu,t^{k'})^2 + \frac{2\Delta t}{\alpha_a}\, {\vartheta^q_M}^2 \sum_{k'=1}^{k} \hat{\varepsilon}_M(\mu,t^{k'})^2 \right)^{\frac{1}{2}}. \tag{6.31}$$

Proof. We immediately derive from (6.3) and (6.26) that $e(\mu,t^k) = y(\mu,t^k) - y_{N,M}(\mu,t^k)$, $\forall\, k \in \mathbb{K}$, satisfies

$$\begin{aligned} &m(e(\mu,t^k),v) + \Delta t\, a^L(e(\mu,t^k),v) + \Delta t \int_\Omega \Big(g(y(\mu,t^k);x;\mu) - g(y_{N,M}(\mu,t^k);x;\mu) \Big)\ v \\ &\quad = m(e(\mu,t^{k-1}),v) + \Delta t\, R(v;\mu,t^k) + \Delta t \int_\Omega \Big(g^{y_{N,M}}_M(x;\mu,t^k) - g(y_{N,M}(\mu,t^k);x;\mu) \Big)\ v,\ \forall\, v \in Y, \end{aligned} \tag{6.32}$$

where $e(\mu,t^0) = 0$ since $y(\mu,t^0) = y_{N,M}(\mu,t^0) = 0$ by assumption. We now choose $v = e(\mu,t^k)$ in (6.32), immediately note from the monotonicity of g that

$$\int_\Omega \Big(g(y(\mu,t^k);x;\mu) - g(y_{N,M}(\mu,t^k);x;\mu) \Big)\ e(\mu,t^k) \geq 0; \tag{6.33}$$

invoke (6.25) and the Cauchy-Schwarz inequality for the cross term $m(e(\mu,t^{k-1}),e(\mu,t^k))$ to obtain, $\forall\, k \in \mathbb{K}$,

$$\begin{aligned} m(e(\mu,t^k),e(\mu,t^k)) + \Delta t\ a^L(e(\mu,t^k),e(\mu,t^k)) &\leq m^{\frac{1}{2}}(e(\mu,t^{k-1}),e(\mu,t^{k-1}))\, m^{\frac{1}{2}}(e(\mu,t^k),e(\mu,t^k)) \\ + \Delta t\, \varepsilon_{N,M}(\mu,t^k)\, \|e(\mu,t^k)\|_Y &+ \Delta t \int_\Omega \Big(g^{y_{N,M}}_M(x;\mu,t^k) - g(y_{N,M}(\mu,t^k);x;\mu) \Big)\ e(\mu,t^k). \end{aligned} \tag{6.34}$$

We will now apply (4.30) twice: first, choosing $c = m^{\frac{1}{2}}(e(\mu,t^k),e(\mu,t^k))$, $d = m^{\frac{1}{2}}(e(\mu,t^{k-1}),e(\mu,t^{k-1}))$, and $\rho = 1$, we obtain

$$2\, m^{\frac{1}{2}}(e(\mu,t^k),e(\mu,t^k))\; m^{\frac{1}{2}}(e(\mu,t^{k-1}),e(\mu,t^{k-1})) \leq m(e(\mu,t^{k-1}),e(\mu,t^{k-1}))+m(e(\mu,t^k),e(\mu,t^k)); \tag{6.35}$$

and second, choosing $c = \varepsilon_{N,M}(\mu,t^k)$, $d = \|e(\mu,t^k)\|_Y$, and $\rho = (\alpha_a/2)^{\frac{1}{2}}$ we have

$$2\,\varepsilon_N(\mu,t^k)\;\|e(\mu,t^k)\|_Y \leq \frac{2}{\alpha_a}\;\varepsilon_N(\mu,t^k)^2 + \frac{\alpha_a}{2}\;\|e(\mu,t^k)\|_Y^2. \tag{6.36}$$

We now note from our assumption $g(y_{N,M}(\mu,t^k);x;\mu) \in W^g_{M+1}$ and Proposition 2 that

$$\left(g_M^{y_{N,M}}(x;\mu,t^k) - g(y_{N,M}(\mu,t^k);x;\mu)\right) = \hat{\varepsilon}_M(\mu,t^k)\, q_{M+1}(x); \tag{6.37}$$

it thus follows that

$$\begin{aligned}
2\int_\Omega &\left(g_M^{y_{N,M}}(x;\mu,t^k) - g(y_{N,M}(\mu,t^k);x;\mu)\right) e(\mu,t^k)\\
&\leq\;\; 2\,\sup_{v\in Y}\left\{\frac{\int_\Omega\left(g_M^{y_{N,M}}(x;\mu,t^k) - g(y_{N,M}(\mu,t^k);x;\mu)\right)v}{\|v\|_Y}\right\}\;\|e(\mu,t^k)\|_Y\\
&\leq\;\; 2\,\hat{\varepsilon}_M(\mu,t^k)\,\sup_{v\in Y}\left\{\frac{\int_\Omega q_{M+1}v}{\|v\|_Y}\right\}\;\|e(\mu,t^k)\|_Y\\
&\leq\;\; 2\,\hat{\varepsilon}_M(\mu,t^k)\;\vartheta^q_M\;\|e(\mu,t^k)\|_Y\\
&\leq\;\; \frac{2}{\alpha_a}\;\hat{\varepsilon}_M(\mu,t^k)^2\;{\vartheta^q_M}^2 + \frac{\alpha_a}{2}\;\|e(\mu,t^k)\|_Y^2,
\end{aligned} \tag{6.38}$$

where we applied (4.30) with $c = \hat{\varepsilon}_M(\mu,t^k)\;\vartheta^q_M$, $d = \|e(\mu,t^k)\|_Y$, and $\rho = (\alpha_a/2)^{\frac{1}{2}}$ in the last step. Finally, from (6.34), (6.35), (6.36), (6.38), and invoking (6.6) we obtain the bound

$$m(e(\mu,t^k),e(\mu,t^k))+\Delta t\sum_{k'=1}^{k} a(e(\mu,t^{k'}),e(\mu,t^{k'})) \leq \frac{2\,\Delta t}{\alpha_a}\sum_{k'=1}^{k}\left(\varepsilon_{N,M}(\mu,t^{k'})^2 + {\vartheta^q_M}^2\;\hat{\varepsilon}_M(\mu,t^{k'})^2\right) \tag{6.39}$$

which is the result stated in Proposition 17. □

We note from (6.31) that our error bound comprises two terms: the contribution from the linear (affine) terms and from the nonlinear (nonaffine) function approximation. Similar to the linear nonaffine case, we may thus choose N and M such that both contributions balance, i.e., neither N nor M should be chosen unnecessarily high. However, our choice should also take the rigor of the error bound into account — we comment on this issue after stating the bounding property for the output estimate.

Output Bound

We can now define the (simple) output bound

Proposition 18. *Suppose that* $g(y_{N,M}(\mu, t^k); \mu) \in W^g_{M+1}$, $\forall\, k \in \mathbb{K}$. *The error in the output is then bounded by*

$$|s(\mu, t^k) - s_{N,M}(\mu, t^k)| \leq \Delta^s_{N,M}(\mu, t^k), \quad \forall\, k \in \mathbb{K},\ \forall\, \mu \in \mathcal{D}, \tag{6.40}$$

where the output bound is defined as

$$\Delta^s_{N,M}(\mu, t^k) \equiv \sup_{v \in Y} \frac{\ell(v)}{\|v\|_X} \Delta^y_{N,M}(\mu, t^k), \quad \forall\, k \in \mathbb{K},\ \forall\, \mu \in \mathcal{D}. \tag{6.41}$$

Proof. The result directly follows from (6.2), (6.18), and the fact that the error satisfies $\|e(\mu, t^k))\|_X \leq \Delta^y_{N,M}(\mu, t^k), \forall\, k \in \mathbb{K},\ \forall\, \mu \in \mathcal{D}$. □

We note that the condition $g(y_{N,M}(\mu, t^k); \mu) \in W^g_{M+1}$, $\forall\, k \in \mathbb{K}$ is *very* unlikely to hold: first, because W^g_M is constructed based on $g(y(\mu, t^k); \mu)$ and not $g(y_{N,M}(\mu, t^k))$, and second, particularly because of the time-dependence of $g(y_{N,M}(\mu, t^k); \mu)$. Our choice of N and M is thus even more important — the contribution of the non-rigorous part, $\vartheta^q_M \hat{\varepsilon}_M(\mu, t^k)$, to the error bound $\Delta^y_{N,M}(\mu, t^k)$ should be small compared to the contribution of the rigorous part $\varepsilon_{N,M}(\mu, t^k)$.

6.4.3 Offline-Online Computational Procedure

The offline-online computational procedures for the calculation of $\Delta^y_{N,M}(\mu, t^k)$ (and $\Delta^s_{N,M}(\mu, t^k)$) are very similar to the previous discussions in Sections 4.4.4 and 5.4.3. We will therefore omit the details and only summarize the computational costs involved in the online stage. In the online stage — given a new parameter value μ and associated reduced-basis solution $\underline{y}_{N,M}(\mu, t^k)$, $\forall\, k \in \mathbb{K}$ — the computational cost to evaluate $\Delta^y_{N,M}(\mu, t^k)$ (and hence $\Delta^s_{N,M}(\mu, t^k)$) is $O(K(N+M)^2)$ and thus *independent* of $\mathcal{N}$.

6.5 Results for Numerical Exercise 5

We now return to our model problem from Section 6.2.1. We first construct the sample set S^g_M and associated reduced-basis space W^g_M — and hence T_M and B_M — using the empirical interpolation method described in Section 2.4.3; we employ the surrogate least squares approach on $\tilde{\Xi}^g = \Xi^g \times \mathbb{I}$ for the greedy procedure, where $\Xi^g \subset \mathcal{D}^{144}$ is a regular 12×12 grid.

We next generate the sample set S^y_N and associated reduced basis space W^y_N according to the adaptive sampling procedure described in Section 4.5 — since the "truth" solutions $y(\mu, t^k)$ are stored on $\tilde{\Xi}^g$, we sample on this parameter-time grid and base our choice of basis on the energy norm of the true error $e(\mu, t^k)$ and *not* on the error bound. We initialize the procedure with $\mu^y_1 = (0.01, 0.01)$ and $t^{k^y_1} = 1\Delta t$ and set the desired error tolerance (for the relative error in the energy norm) to $\epsilon_{\text{tol,min}} = 1\,\text{E}-6$. We sample on the parameter test sample Ξ^g used for the construction of S^g_N. We need $N_{\max} = 55$ basis functions to obtain the desired accuracy.

We plot the sample set S^y_N in $\mu - t^k$-space in Figure 6-3. We note that the samples are largely located in the $\mu_2 = 10$ plane where the strength of the nonlinearity is largest. Furthermore, because of the periodic control input, all samples are selected within the first 100 timesteps although $K = 200$ here.

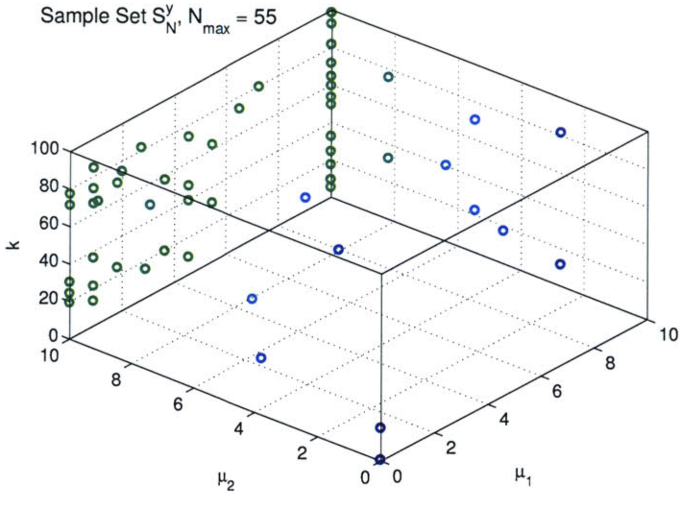


Figure 6-3: NE 5: Sample set S_N^y.

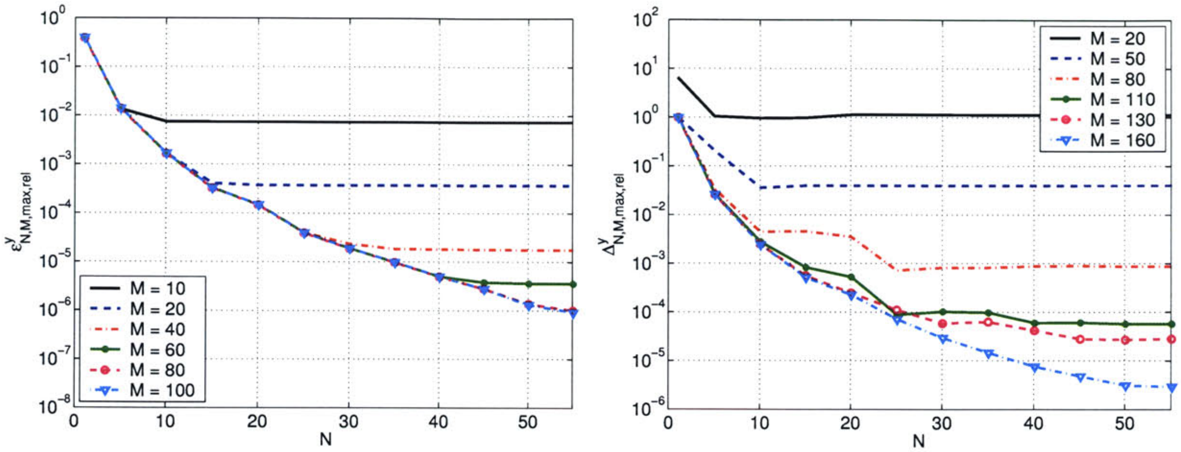
N	M	$\epsilon^y_{N,M,\max,\mathrm{rel}}$	$\Delta^y_{N,M,\max,\mathrm{rel}}$	$\overline{\eta}^y$
1	10	3.82 E−01	4.22 E+01	79.8
5	30	1.36 E−02	1.17 E+00	26.0
10	50	1.62 E−03	3.54 E−02	8.65
20	80	1.46 E−04	3.52 E−03	8.25
30	110	1.88 E−05	1.01 E−04	3.82
40	140	4.94 E−06	1.78 E−05	1.69

Table 6.1: NE 5: Convergence rate and effectivities as a function of N and M.

In Figure 6-4(a) and (b) we plot the maximum relative error $\epsilon^y_{N,M,\max,\mathrm{rel}}$ and maximum relative error bound $\Delta^y_{N,M,\max,\mathrm{rel}}$ as a function of N and M, respectively; here $\epsilon^y_{N,M,\max,\mathrm{rel}}$ is the maximum over Ξ_{Test} of $|||e(\mu, t^K)|||/|||y(\mu_y, t^K)|||$ and $\Delta^y_{N,M,\max,\mathrm{rel}}$ is the maximum over Ξ_{Test} of $\Delta^y_{N,M}(\mu, t^K)/|||y(\mu_y, t^K)|||$, where $\mu_y \equiv \arg\max_{\mu\in\Xi_{\mathrm{Test}}} |||y(\mu, t^K)|||$, and $\Xi_{\mathrm{Test}} \subset (\mathcal{D})^{225}$ is a test sample of size 225 (a regular 15×15 grid). We observe the same convergence behavior as in the nonaffine case: the curves level off at lower and lower levels as M increases. However, M is much larger now as compared to the nonaffine case due to the (implicit) time dependence of g. We also note that the true error essentially converged for $M = 80$ — the $M - 80$ and $M = 100$ asymptotes are identical. The error bound, on the other side, still decreases considerably for $M \geq 80$ (note the different M values in the two plots). To obtain small relative error bounds we should thus choose M large.

In Table 6.1 we present, as a function of N and M, $\epsilon^y_{N,M,\max,\mathrm{rel}}$, $\Delta^y_{N,M,\max,\mathrm{rel}}$, and the average effectivity $\overline{\eta}^y$: $\overline{\eta}^y$ is the average over $\Xi_{\mathrm{Test}} \times \mathbb{I}$ of $\Delta^y_{N,M}(\mu, t^k)/|||y(\mu, t^k) - y_N(\mu, t^k)|||$. We confirm the fast convergence already observed in Figure 6-4 and observe that the effectivities are very good for higher values of N and M.

We now turn to the output estimate and present the maximum relative output error $\epsilon^s_{N,M,\max,\mathrm{rel}}$,



(a) (b)

Figure 6-4: NE 5: (a) Maximum relative error in the energy norm and (b) maximum relative error bound.

N	M	$\epsilon^s_{N,M,\max,\mathrm{rel}}$	$\Delta^s_{N,M,\max,\mathrm{rel}}$	$\overline{\eta}^s_{N,M}$
1	10	1.00 E−00	9.19 E+02	494
5	30	1.91 E−02	2.55 E+01	597
10	50	1.46 E−04	7.71 E−01	1410
20	80	1.67 E−05	7.66 E−02	1357
30	110	5.16 E−06	2.21 E−03	416
40	140	1.56 E−06	3.88 E−04	200

Table 6.2: NE 5: Maximum relative output error, output bound, and effectivities.

the maximum relative output bound $\Delta^s_{N,M,\max,\mathrm{rel}}$, and the average effectivity $\overline{\eta}^s_{N,M}$ as a function of N and M in Table 6.2. Here, $\epsilon^s_{N,M,\max,\mathrm{rel}}$ is the maximum over Ξ_{Test} of $|s(\mu,t_\eta(\mu)) - s_N(\mu,t_\eta(\mu))|/s_{\max}$, $\Delta^s_{N,M,\max,\mathrm{rel}}$ is the maximum over Ξ_{Test} of $\Delta^s_{N,M}(\mu,t^K)/|s_{\max}|$ and $\overline{\eta}^s$ is the average over Ξ_{Test} of $\Delta^s_{N,M}(\mu,t_\eta(\mu))/|s(\mu,t_\eta(\mu)) - s_{N,M}(\mu,t_\eta(\mu))|$, where $t_\eta(\mu) \equiv \arg\max_{t^k \in \mathbb{I}} |s(\mu,t^k) - s_N(\mu,t^k)|$ and $s_{\max} \equiv \max_{t^k\in\mathbb{I}} \max_{\mu\in\Xi_{\mathrm{Test}}} |s(\mu,t^k)|$. The reduced-basis output estimate converges very fast: for only $N = 5$ and $M = 30$ the relative error in the output is close to 1%. However, the output bound largely overestimates the true error, which is reflected in the very large effectivities. For an accuracy of 1% in the output bound, we would thus require approximately $N = 25$ and $M = 100$. Introducing adjoint techniques [121] might be a remedy for the poor output bounds.

Finally, in Table 6.3 we present, as a function of N and M, the online computational times to calculate $s_{N,M}(\mu,t^k)$ and $\Delta^s_{N,M}(\mu,t^k)$, $\forall\, k \in \mathbb{K}$. The values are normalized with respect to the computational time for the direct calculation of the truth approximation output $s(\mu,t^k) = \ell(y(\mu,t^k))$, $\forall\, k \in \mathbb{K}$. We note that the gain in the online response time is much larger in the nonlinear case. This is mainly due to the fact that solving for the "truth" approximation (6.6) involves the matrix assembly of the nonlinear terms. The actual run-time to compute the output estimate and output bound in MATLAB 6.5 on a 750 MHz Pentium III is 1.41 (for $N = 25$,

N	M	$s_{N,M}(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$\Delta^s_{N,M}(\mu, t^k),\ \forall\, k \in \mathbb{K}$	$s(\mu, t^k),\ \forall\, k \in \mathbb{K}$
1	10	6.62E-05	8.66E-05	1
5	30	1.19E-04	7.35E-05	1
10	50	1.74E-04	9.57E-05	1
20	80	3.88E-04	1.57E-04	1
30	110	7.20E-04	2.62E-04	1
40	140	1.22E-03	4.33E-04	1

Table 6.3: Online computational times (normalized with respect to the time to solve for $s(\mu, t^k),\ \forall\, k \in \mathbb{K}$).

$M = 100$). We already pointed out, however, that the offline computations necessary in the nonlinear case are also more extensive — primarily due to the sampling procedure for S^g_M. If the many-query context, or a clear demand for real-time response of an engineering system or component in operation, can justify the offline cost, the reduced-basis methods, and in particular our approach described here, can be gainfully employed in many practical applications.

6.6 Application to Reaction-Diffusion Systems

In this section we apply our method to a specific problem belonging to the class of nonlinear reaction-diffusion systems [26]. Reaction-diffusion systems are an interesting area of application for several reasons: First of all, they appear in a large number of real-world applications: ranging from Biology [22], where reaction-diffusion equations characterize the pattern formation in morphogenesis and mutations in genetics; to Ecology, where they govern predator-prey relation and the spreading of epidemics; to Physiology, where the conduction in nerves and carbon monoxide poisoning is described by reaction-diffusion equations; to Chemistry [5, 6, 32], probably the most notable application area of reaction-diffusion equations. Furthermore, inherent to these equations and the specific application area are a large number of parameters, which, in general, have a very strong influence on the dynamic behavior of the system, e.g., such as reaction rates in chemistry. To analyze and understand the specific problem, many different parameter combinations have to be investigated. The solution of reaction-diffusion equations, however, is a very challenging task because the equations are time-dependent, and often highly nonlinear and coupled. Efficient solution techniques which can characterize many parameter combinations are therefore important. Finally, in many applications — such as chemical engineering — understanding, modeling, and simulation is often only the first step; the original goal is to control the behavior and outcome using (say) techniques from optimal control theory [58].

The general form of a reaction-diffusion equation can be written as

$$\frac{\partial \mathbf{y}(x;\mu,t)}{\partial t} = \nabla\left(D(\mu)\mathbf{y}(x;\mu,t)\right) + \mathbf{f}(\mathbf{y}(x;\mu,t);\mu). \tag{6.42}$$

Here, $x \in \Omega \subset \mathbb{R}^d$ is the spatial domain, $\mu \in \mathcal{D}$ is the parameter vector, $\mathbf{y}$ is the vector-valued field variable, e.g., containing temperatures and concentrations, $D(\mu)$ is the diffusion matrix, and $\mathbf{f}(\mathbf{y};\mu)$ is a vector-valued function containing the (non)linear reaction terms. Our particular application is the self-ignition of a coal stockpile introduced in Section 1.1.1. We note, however, that similar models are also used in combustion theory, biology, and in the description of porous catalysts.

We note that we cannot apply our *a posteriori* error estimation procedures to this problem because (*i*) the system is coupled, and (*ii*) our monotonicity assumption for the nonlinearity is not satisfied. We therefore only consider the reduced-basis approximation and do *not* discuss the *a posteriori* error estimation.

6.6.1 AP III: Self-Ignition of a Coal Stockpile

More specifically, we consider a one-dimensional non-isothermal reaction-diffusion model for the self-ignition of a coal stockpile with Arrhenius type nonlinearity [105, 103, 23], described in detail in Section 1.1.1. The parameters governing the dynamic behavior of the system are the Arrhenius number, γ, the Prater temperature, β, the Lewis number, Le, and the Thiele modulus, Φ. Here, we assume that three of those parameters — β, Le, and Φ — are fixed and only γ is varying. The values, taken from [23], are given by $\beta = 4.287$, $\Phi^2 = 70000$, and Le $= 0.233$; and γ varies in the range $12 \leq \gamma \leq 12.6$. We can thus identify the input parameter $\mu \equiv \gamma \in \mathcal{D} \equiv [12, 12.6] \subset \mathbb{R}^{P=1}$. We will see that the system exhibits a very interesting dynamical behavior in terms of complex oscillatory patterns for this parameter range.

We next derive the weak form of the governing equations (1.14) and (1.15) and discretize in time using Euler-Backward. We also introduce the linear finite truth approximation subspaces $Y_T \equiv \{v | v \in H^1(\Omega), v = 0|_{x=0,1}\}$ and $Y_c \equiv \{v | v \in H^1(\Omega), v = 0|_{x=0}\}$ of dimensions $\mathcal{N} = 800$ and $\mathcal{N} = 801$, respectively; here $\Omega \equiv [0, 1]$ is the spatial domain. We shall consider the time interval $\bar{I} = [0, 6]$ and a timestep $\Delta t = 1\,\text{E}{-}3$; we thus have $K = 6000$. Our truth approximation is thus: Given $\mu \in \mathcal{D}$, find $T(\mu, t^k) \in Y_T$ and $c(\mu, t^k) \in Y_c$, $\forall\, k \in \mathbb{K}$, such that[3]

$$m(T(\mu,t^k),v_T) + \Delta t\ a(T(\mu,t^k),v_T)$$

$$-\Delta t\ \beta\ \Phi^2 \int_\Omega (c(\mu,t^k)+1)\ e^{-\mu/(T(\mu,t^k)+1)}\ v_T = m(T(\mu,t^{k-1}),v_T), \quad \forall\, v_T \in Y_T \quad (6.43)$$

$$m(c(\mu,t^k),v_c) + \Delta t\ \text{Le}\ a(c(\mu,t^k),v_c)$$

$$+\Delta t\ \Phi^2 \int_\Omega (c(\mu,t^k)+1)\ e^{-\mu/(T(\mu,t^k)+1)}\ v_c = m(c(\mu,t^{k-1}),v_c), \quad \forall\, v_c \in Y_c \quad (6.44)$$

with inital conditions $T(\mu, t^0) = T_0,\ c(\mu, t^0) = c_0$. We then evaluate the outputs from

$$s^1(\mu,t^k) = \ell(T(\mu,t^k)), \qquad \forall\, k \in \mathbb{K}, \quad (6.45)$$

and

$$s^2(\mu,t^k) = \ell(c(\mu,t^k)), \qquad \forall\, k \in \mathbb{K}. \quad (6.46)$$

Here, $\ell(v) = \int_\Omega \delta(x - 0.2)\, v$, and the bilinear forms are given by

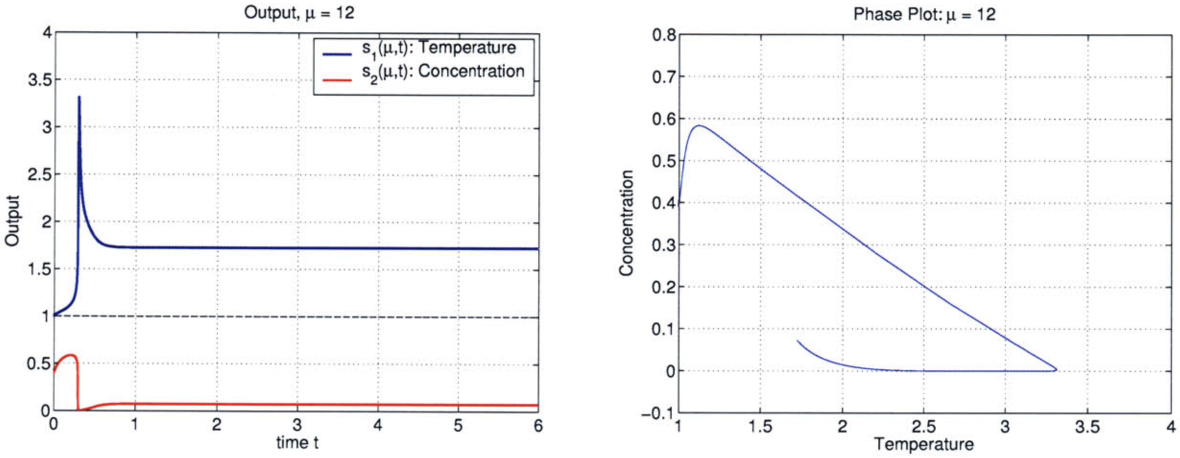
$$m(w,v) = \int_\Omega w\, v, \qquad a(w,v) = \int_\Omega \nabla w \cdot \nabla v. \quad (6.47)$$

[3] Note that we use our usual notation here: $T(\mu, t^k) = T(x; \mu, t^k)$ and $c(\mu, t^k) = c(x; \mu, t^k)$.

We also define the nonlinearity g as

$$g(c(\mu, t^k), T(\mu, t^k); \mu) = (c(\mu, t^k) + 1)\, e^{-\mu/(T(\mu,t^k)+1)}. \tag{6.48}$$

We plot in Figure 6-5(a) and (b) the outputs s^1 and s^2 for $\mu = 12.0$ over (discrete) time and in phase space $s^1 - s^2$, respectively. The sharp peak in the temperature output s^1 and corresponding drop in the concentration output s^2 indicates the ignition of the system. After the ignition, the system goes into a stable steady-state solution. In Figure 6-6(a) and (b) we show the corresponding output plots for $\mu = 12.5$; we first note that the ignition occurs later in time and that the maximum temperature reached is higher. For this parameter value the system does not return to a steady-state solution, but converges to a period 1 limit cycle. Finally, we present in Figures 6-7(a) and (b) the output plots for $\mu = 12.6$: the time of ignition occurs at a later and the maximum temperature is higher than before. Although hardly visible in the phase plot because of the initial transients, the system converges to a limit cycle with mixed mode oscillations. We show in Figures 6-8 the phase plots for the two parameter values $\mu = 12.5$ and $\mu = 12.6$ without the transient behavior. We can clearly see the period 1 limit cycle for $\mu = 12.5$; as μ is increased, a period doubling cascade occurs leading to the mixed mode oscillations for $\mu = 12.6$.



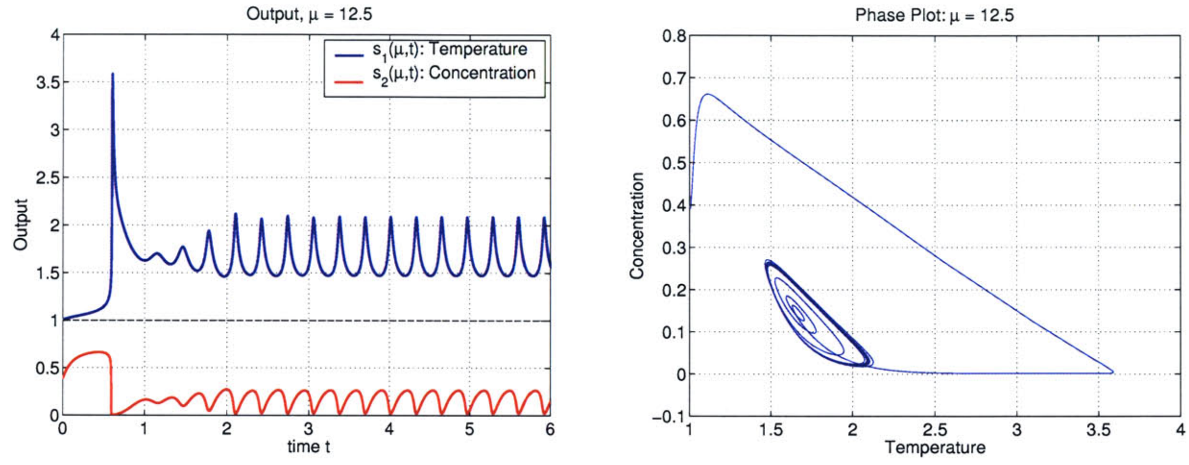
(a) (b)

Figure 6-5: AP III: Outputs $s_1(\mu, t^k)$ and $s_2(\mu, t^k)$ for $\mu = 12.0$, (a) as a function of time (b) phase plot.

6.6.2 Reduced-Basis Approximation

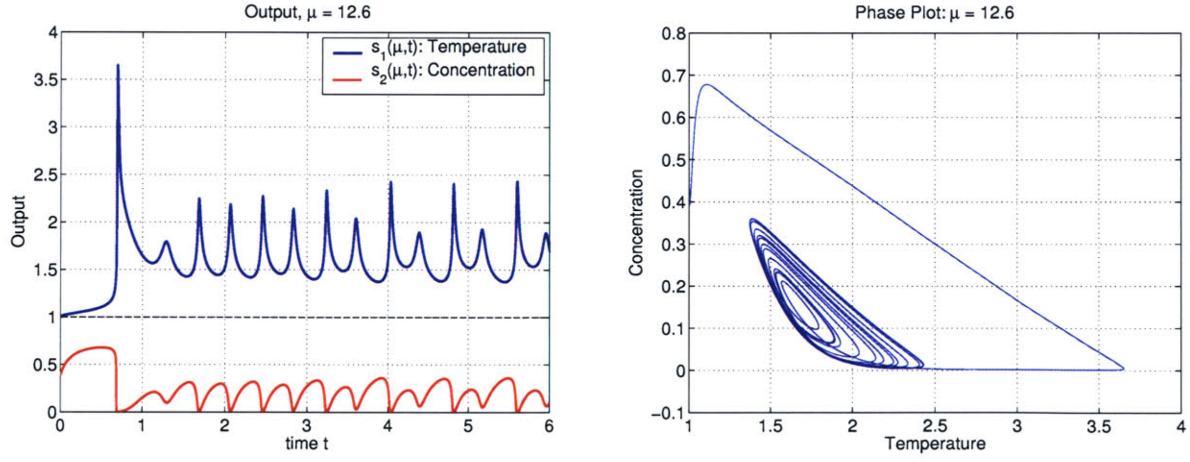
We first introduce the nested sample sets $S^T_{N_T} = \{\tilde{\mu}^T_1 \in \tilde{\mathcal{D}}, \ldots, \tilde{\mu}^T_{N_T} \in \tilde{\mathcal{D}}\}$, $1 \le N_T \le N_{T,\max}$, $S^c_{N_c} = \{\tilde{\mu}^c_1 \in \tilde{\mathcal{D}}, \ldots, \tilde{\mu}^c_{N_c} \in \tilde{\mathcal{D}}\}$, $1 \le N_c \le N_{c,\max}$, and — for the nonlinearity $S^g_M = \{\tilde{\mu}^g_1 \in \tilde{\mathcal{D}}, \ldots, \tilde{\mu}^g_M \in \tilde{\mathcal{D}}\}$, $1 \le M \le M_{\max}$, where $\tilde{\mu} \equiv (\mu, t^k)$ and $\tilde{\mathcal{D}} \equiv \mathcal{D} \times \mathbb{I}$. We then define the associated nested Lagrangian reduced-basis spaces

$$W^T_{N_T} = \text{span}\{\zeta^T_m \equiv T(\mu^T_n, t^{k^T_n}),\ 1 \le n \le N_T\}, \quad 1 \le N_T \le N_{T,\max}, \tag{6.49}$$



(a) (b)

Figure 6-6: AP III: Outputs $s_1(\mu, t^k)$ and $s_2(\mu, t^k)$ for $\mu = 12.5$, (a) as a function of time (b) phase plot.



(a) (b)

Figure 6-7: AP III: Outputs $s_1(\mu, t^k)$ and $s_2(\mu, t^k)$ for $\mu = 12.6$, (a) as a function of time (b) phase plot.

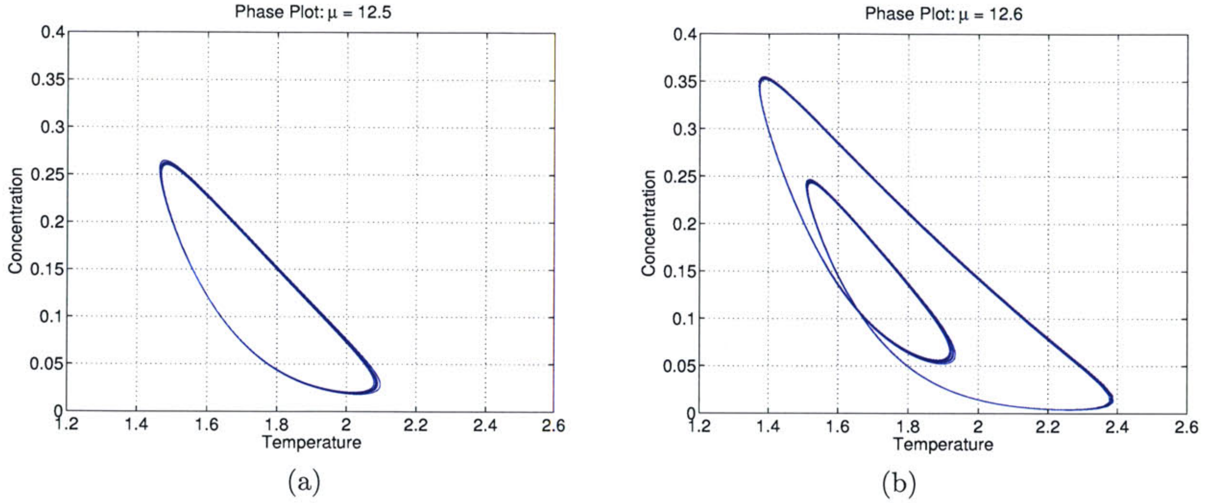


Figure 6-8: AP III: Outputs $s_1(\mu, t^k)$ and $s_2(\mu, t^k)$ in phase plane for (a) $\mu = 12.5$ and (b) $\mu = 12.6$.

and

$$W^c_{N_c} = \text{span}\{\zeta^c_m \equiv c(\mu^c_n, t^{k^c_n}),\ 1 \le n \le N_c\}, \quad 1 \le N_c \le N_{c,\max}, \tag{6.50}$$

where $T(\mu^T_n, t^{k^T_n})$ and $c(\mu^c_n, t^{k^c_n})$ are the solutions of (6.43) and (6.44) at time $t = t^{k^T_n}$ for $\mu = \mu^T_n$ and $t = t^{k^c_n}$ for $\mu = \mu^c_n$, respectively. We also define the nested collateral reduced-basis space

$$\begin{aligned} W^g_M &= \text{span}\{\xi_m \equiv g(c(\tilde{\mu}^g_m), T(\tilde{\mu}^g_m); \mu^g_m),\ 1 \le m \le M\} \\ &= \text{span}\{q_1, \ldots, q_M\}, \qquad 1 \le M \le M_{\max}, \end{aligned} \tag{6.51}$$

and nested set of interpolation points $T_M = \{t_1, \ldots, t_M\}$, $1 \le M \le M_{\max}$.

Our reduced-basis approximation is then: given $\mu \in \mathcal{D}$, $T_{N,M}(\mu, t^k) \in W^T_{N_T}$ and $c_{N,M}(\mu, t^k) \in W^c_{N_c}$

$$\begin{aligned} & m(T_{N,M}(\mu, t^k), v_T) + \Delta t\ a(T_{N,M}(\mu, t^k), v_T) \\ & \qquad -\Delta t\ \beta\ \Phi^2 \int_\Omega g_M^{c_{N,M}, T_{N,M}}(x; \mu, t^k)\ v_T = m(T_{N,M}(\mu, t^{k-1}), v_T), \quad \forall\, v_T \in W^T_{N_T}, \end{aligned} \tag{6.52}$$

$$\begin{aligned} & m(c_{N,M}(\mu, t^k), v_c) + \Delta t \text{ Le } a(c_{N,M}(\mu, t^k), v_c) \\ & \qquad +\Delta t\ \Phi^2 \int_\Omega g_M^{c_{N,M}, T_{N,M}}(x; \mu, t^k)\ v_c = m(c_{N,M}(\mu, t^{k-1}), v_c), \quad \forall\, v_c \in W^c_{N_c}, \end{aligned} \tag{6.53}$$

with inital conditions determined from $m(T(\mu, t^0), v_T) = m(T_0, v_T)$, $\forall v_T \in W^T_{N_T}$, and $m(c(\mu, t^0), v_c) = m(c_0, v_c)$, $\forall\, v_c \in W^c_{N_c}$; here $g_M^{c_{N,M}, T_{N,M}}(x; \mu, t^k)$ is the approximation to $g(c(\mu, t^k), T(\mu, t^k); \mu)$ given by

$$g_M^{c_{N,M}, T_{N,M}}(x; \mu, t^k) = \sum_{m=1}^{M} \varphi_{Mm}(\mu, t^k) q_m(x) \tag{6.54}$$

where the coefficients $\varphi_{Mm}(\mu, t^k)$ are determined from

$$\sum_{j=1}^{M} B_{ij}^M \varphi_{Mj}(\mu, t^k) = g(c(t_i; \mu, t^k), T(t_i; \mu, t^k); \mu), \quad 1 \leq i \leq M \tag{6.55}$$

and $B_{ij}^M = q_j(t_i)$, $1 \leq i, j \leq M$, $1 \leq M \leq M_{\max}$. Finally, we evaluate the outputs from

$$s_{N,M}^1(\mu, t^k) = \ell(T_{N,M}(\mu, t^k)), \quad \forall\, k \in \mathbb{K}, \tag{6.56}$$

and

$$s_{N,M}^2(\mu, t^k) = \ell(c_{N,M}(\mu, t^k)), \quad \forall\, k \in \mathbb{K}. \tag{6.57}$$

Offline-Online Procedure

The offline-online procedure follows directly from our previous discussion in Section 6.3.2. We first express

$$\begin{aligned} T_{N,M}(\mu, t^k) &= \sum_{n=1}^{N_T} \zeta_n^T \, T_{N,Mn}(\mu, t^k), & (6.58)\\ c_{N,M}(\mu, t^k) &= \sum_{n=1}^{N_c} \zeta_n^c \, c_{N,Mn}(\mu, t^k), & (6.59) \end{aligned}$$

and choose as test functions $v_T = \zeta_n^T$, $1 \leq n \leq N_T$ in (6.52), and $v_c = \zeta_n^c$, $1 \leq n \leq N_c$ in (6.53).

It then follows that $\underline{T}_{N,M}(\mu, t^k) = [T_{N,M1}(\mu, t^k) \dots T_{N,MN_T}(\mu, t^k)] \in \mathbb{R}^{N_T}$ and $\underline{c}_{N,M}(\mu, t^k) = [c_{N,M1}(\mu, t^k) \dots c_{N,MN_c}(\mu, t^k)], \in \mathbb{R}^{N_c}$, $\forall\, k \in \mathbb{K}$, satisfy

$$\begin{aligned} (M_N^T + \Delta t\ A_N^T)\ \underline{T}_{N,M}(\mu, t^k) - \Delta t\ \beta\ \Phi^2\ C^{N,M\,T}\ \varphi_M(\mu, t^k) &= M_N^T\ \underline{T}_{N,M}(\mu, t^{k-1}), & (6.60)\\ (M_N^c + \Delta t\ \mathrm{Le}\ A_N^c)\ \underline{c}_{N,M}(\mu, t^k) - \Delta t\ \Phi^2\ C^{N,M\,c}\ \varphi_M(\mu, t^k) &= M_N^c\ \underline{c}_{N,M}(\mu, t^{k-1}), & (6.61) \end{aligned}$$

with initial condition $M^T \underline{T}_{N,M}(\mu, t^0) = \underline{T}_{N,M}^0$, where $T_{N,Mi}^0 = m(T_0, \zeta_i^T)$, $1 \leq i \leq N_T$ and $M^c \underline{c}_{N,M}(\mu, t^0) = \underline{c}_{N,M}^0$, where $c_{N,Mi}^0 = m(c_0, \zeta_i^T)$, $1 \leq i \leq N_c$. Here, $\varphi_M(\mu, t^k) \in \mathbb{R}^M$ is determined from (6.55); $M_N^T \in \mathbb{R}^{N_T \times N_T}$, $A_N^T \in \mathbb{R}^{N_T \times N_T}$, $C^{N,M\,T} \in \mathbb{R}^{N_T \times M}$, $M_N^c \in \mathbb{R}^{N_c \times N_c}$, $A_N^c \in \mathbb{R}^{N_c \times N_c}$, and $C^{N,M\,c} \in \mathbb{R}^{N_c \times M}$ are *parameter-independent* matrices with entries $M_{Ni,j}^T = m(\zeta_i^T, \zeta_j^T)$, $1 \leq i, j \leq N_T$, $A_{Ni,j}^T = a(\zeta_i^T, \zeta_j^T)$, $1 \leq i, j \leq N_T$, $C_{i,j}^{N,M\,T} = \int_\Omega \zeta_i^T q_j$, $1 \leq i \leq N_T$, $1 \leq j \leq M$, $M_{Ni,j}^c = m(\zeta_i^c, \zeta_j^c)$, $1 \leq i, j \leq N_c$, $A_{Ni,j}^c = a(\zeta_i^c, \zeta_j^c)$, $1 \leq i, j \leq N_c$, $C_{i,j}^{N,M\,T} = \int_\Omega \zeta_i^c q_j$, $1 \leq i \leq N_c$, $1 \leq j \leq M$, respectively.

We can now substitute $\varphi_M(\mu, t^k) \in \mathbb{R}^M$ from (6.55) into (6.60) and (6.61) to obtain the coupled system of nonlinear algebraic equations

$$\begin{aligned}&(M_N^T + \Delta t\ A_N^T)\ \underline{T}_{N,M}(\mu, t^k) - \Delta t\ \beta\ \Phi^2 D^{N,M\,T}\\ &\qquad g\left(Z^{N,M\,c}\ \underline{c}_{N,M}(\underline{t}_M; \mu, t^k), Z^{N,M\,T}\ \underline{T}_{N,M}(\underline{t}_M; \mu, t^k); \mu\right) = M_N^T\ \underline{T}_{N,M}(\mu, t^{k-1}) \quad (6.62)\\ &(M_N^c + \Delta t \text{ Le } A_N^c)\ \underline{c}_{N,M}(\mu, t^k) - \Delta t\ \Phi^2 D^{N,M\,c}\\ &\qquad g\left(Z^{N,M\,c}\ \underline{c}_{N,M}(\underline{t}_M; \mu, t^k), Z^{N,M\,T}\ \underline{T}_{N,M}(\underline{t}_M; \mu, t^k); \mu\right) = M_N^c\ \underline{c}_{N,M}(\mu, t^{k-1}) \quad (6.63)\end{aligned}$$

with initial condition $M^T \underline{T}_{N,M}(\mu, t^0) = \underline{T}^0_{N,M}$ and $M^c \underline{c}_{N,M}(\mu, t^0) = \underline{c}^0_{N,M}$, which has to be solved using (say) Newton's Method for all $k \in \mathbb{K}$. Here, $D^{N,M\,T} = C^{N,M\,T}(B^M)^{-1} \in \mathbb{R}^{N_T \times M}$ and $D^{N,M\,c} = C^{N,M\,c}(B^M)^{-1} \in \mathbb{R}^{N_c \times M}$; and $Z^{N,M\ T} \in \mathbb{R}^{M \times N_T}$ and $Z^{N,M\,c} \in \mathbb{R}^{M \times N_c}$ are *parameter-independent* matrices with entries $Z_{i,j}^{N,M\,T} = \zeta_j^T(t_i)$, $1 \le j \le N_T, 1 \le i \le M$ and $Z_{i,j}^{N,M\,c} = \zeta_j^c(t_i)$, $1 \le j \le N_c, 1 \le i \le M$, and $\underline{t}_M = [t_1 \dots t_M] \in \mathbb{R}^M$ is the set of interpolation points.

Finally, we evaluate the output estimates from

$$s^1_{N,M}(\mu, t^k) = L_N^T\ \underline{T}_{N,M}(\mu, t^k), \quad \forall\, k \in \mathbb{K}, \tag{6.64}$$

and

$$s^2_{N,M}(\mu, t^k) = L_N^c\ \underline{c}_{N,M}(\mu, t^k), \quad \forall\, k \in \mathbb{K}, \tag{6.65}$$

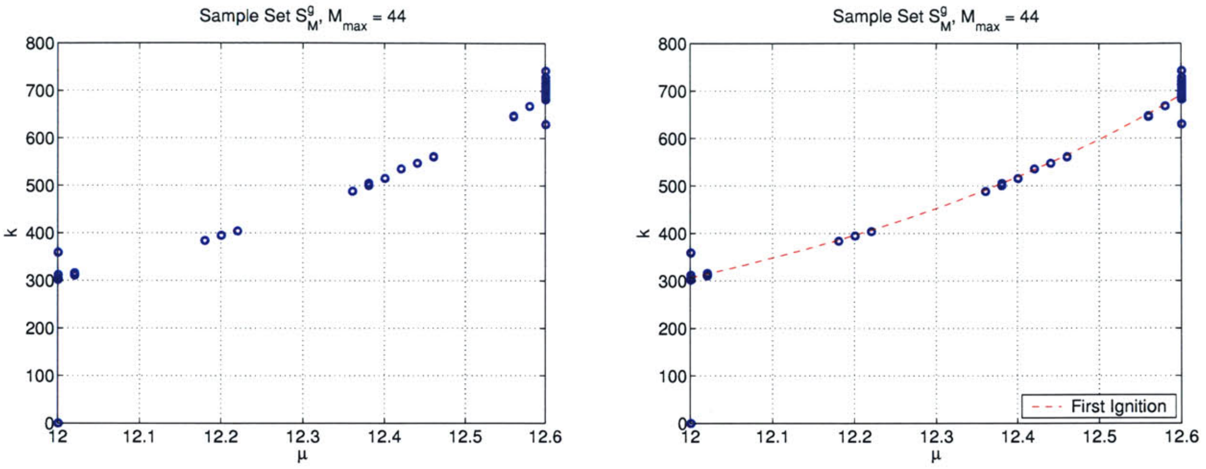
where $L_N^T \in \mathbb{R}^{N_T}$ and $L_N^c \in \mathbb{R}^{N_c}$ are the output vectors with entries $L^T_{Ni} = \ell(\zeta_i^T)$, $1 \le i \le N_T$ and $L^c_{Ni} = \ell(\zeta_i^c)$, $1 \le i \le N_c$, respectively.

The online-offline decomposition is now clear. In the offline stage — performed only once — we first construct the nested approximation space W_M^g and sets of interpolation points T_M, $1 \le M \le M_{\max}$; we then solve (6.43) and (6.44) for the ζ_j^T, $1 \le j \le N_T$ and ζ_j^c, $\le j \le N_c$, respectively, and store the parameter independent quantities M_N^T, A_N^T, $D^{M,N\ T}$, M_N^c, A_N^c, $D^{M,N\ c}$, $Z^{M,N\,T}$, $Z^{M,N\,c}$, and B^M. In the online stage — given a new parameter value μ — we solve (6.62) and (6.63) for $\underline{T}_{N,M}(\mu, t^k)$ and $\underline{c}_{N,M}(\mu, t^k)$ and evaluate $s^{1,2}_{N,M}(\mu, t^k)$ from (6.64) and (6.65). The operation count in the online stage is $O(\bar{\kappa} K (M N^2 + N^3))$, where $\bar{\kappa}$ is the average number of Newton steps per timestep and $N = N_T + N_c$; The operation count is thus *independent* of $\mathcal{N}$.

6.6.3 Numerical Results

We first consider the approximation (6.54) for the nonlinear function, $g(c(\mu, t^k), T(\mu, t^k); \mu)$, defined in (6.48). We sample on a regular grid Ξ^g of size 31 and set $(\mu_1^g, t^{k_1^g}) = (12, 1)$. Note that we need to precalculate and store the "truth" solutions for all timesteps and parameter points in Ξ^g. We next pursue the empirical interpolation method of Section 2.4 (using the $L^\infty(\Omega)$-norm surrogate) to construct S_M^g, W_M^g, T_M, and B^M, $1 \le M \le M_{\max}$, for $M_{\max} = 44$. The resulting sample set S_M^g is plotted in Figure 6-9(a). We observe that the samples are largely located along a curved line in parameter-time space. This curve represents the time of the first ignition, t_{ignition}, as a function of the parameter μ. To confirm this, we plot the ignition curve, i.e., the timestep where $s^1(\mu, t^k)$ reaches its maximum for each parameter value, in Figure 6-9(b).

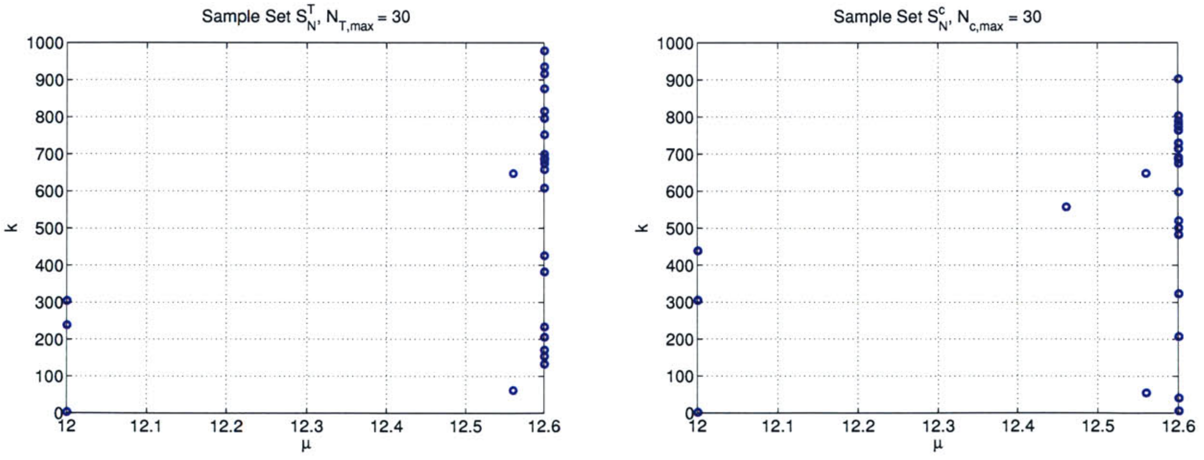
We next generate the sample sets $S_{N_T}^T$ and $S_{N_c}^c$ using our adaptive sampling procedure from Section 4.5 using the true errors in the energy norm, $|||e_T(\mu, t^k)|||$ and $|||e_c(\mu, t^k)|||$, instead of the error bound. Here, $e_T(\mu, t^k)(\mu) = T(\mu, t^k) - T_{N,M}(\mu, t^k)$, $e_c(\mu, t^k) = c(\mu, t^k) - c_{N,M}(\mu, t^k)$ and the



(a) (b)

Figure 6-9: AP III: Sample set S^g_M.

energy norm is defined as $|||v(\mu,t^k)|||^2 \equiv m(v(\mu,t^k),v(\mu,t^k)) + \sum_{k'=1}^{k} \Delta t\, a(v(\mu,t^{k'}),v(\mu,t^{k'}))$. We note that, although we use the true errors, our sampling procedure fails for this problem. The reason is that the reduced-basis approximation cannot capture the point ignition, t_{ignition}, without sufficiently resolving the transient behavior before the ignition occurs: the rate of change in the error (and norm) will be largest at the t_{ignition} — the adaptive procedure thus repeatedly suggests to pick the sample point corresponding to the point of ignition, t_{ignition}, instead of first choosing samples leading up to t_{ignition}. We thus employ the backup procedure described in Section 4.5.3. We plot the samples sets $S^T_{N_T}$ and $S^c_{N_c}$ in parameter-time space in Figure 6-10(a) and (b), respectively. We note that the samples sets $S^c_{N_c}$ and $S^T_{N_T}$ differ in size as well as in the specific samples chosen.

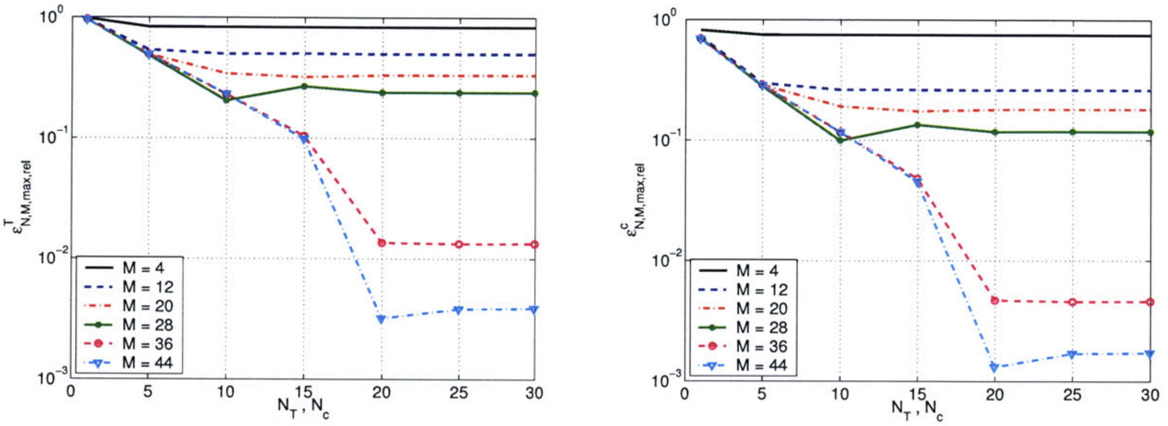


(a) (b)

Figure 6-10: AP III: Sample set (a) $S^T_{N_T}$ and (b) $S^c_{N_c}$.

We next present convergence results for the error in the energy norm. In Figure 6-11(a) and (b) we plot, as a function of N_T, N_c, and M, the maximum relative errors $\epsilon^{y_T}_{N,M,\mathrm{max,rel}}$ and $\epsilon^{y_c}_{N,M,\mathrm{max,rel}}$, respectively (see Section 6.5 for the definition of these quantities). We observe the typical behavior of the error convergence: the M-asymptotes level off at a lower and lower error as M increases.

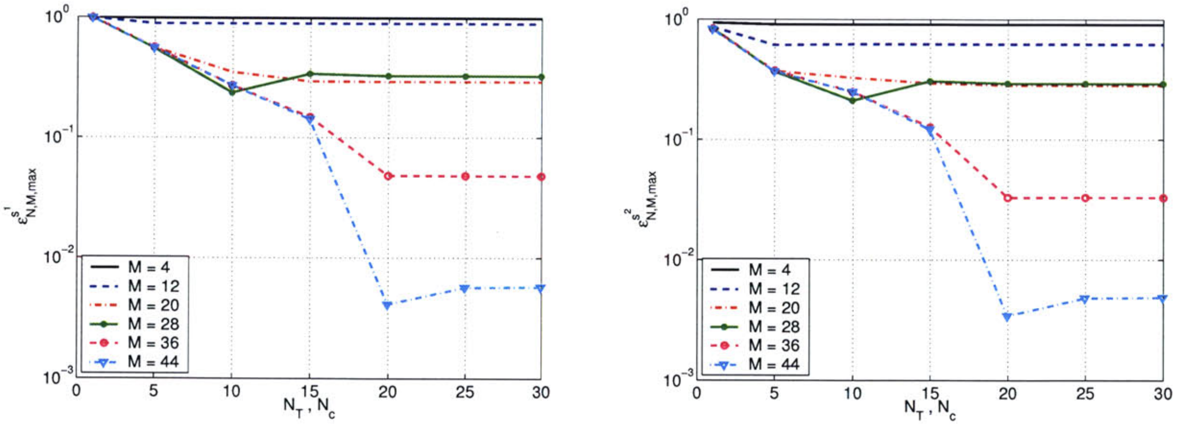
Finally, we present convergence results for the error in the outputs. In Figure 6-12(a) and (b) we plot, as a function of N_T, N_c, and M, the maximum relative output errors $\epsilon^{s^1}_{N,M,\mathrm{max,rel}}$ and $\epsilon^{s^2}_{N,M,\mathrm{max,rel}}$, respectively (see Section 6.5 for the definition of these quantities). The output error shows the same behavior as the error in the energy norm. To obtain a maximum relative error in both outputs of less than 1 percent, we require approximately $M = 44$, $N_T = 20$, and $N_c = 22$.



(a)

(b)

Figure 6-11: AP III: Convergence results for energy norm error.

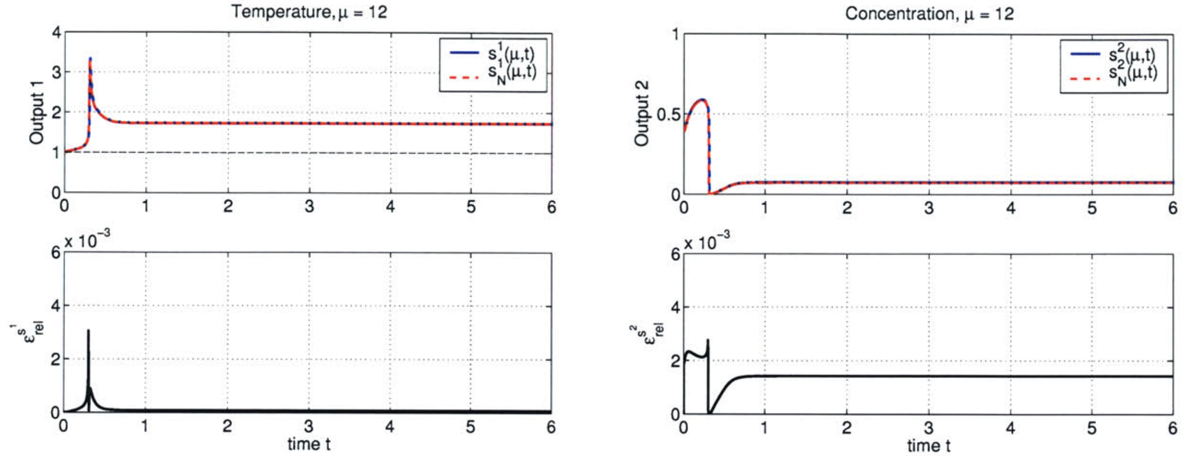


(a)

(b)

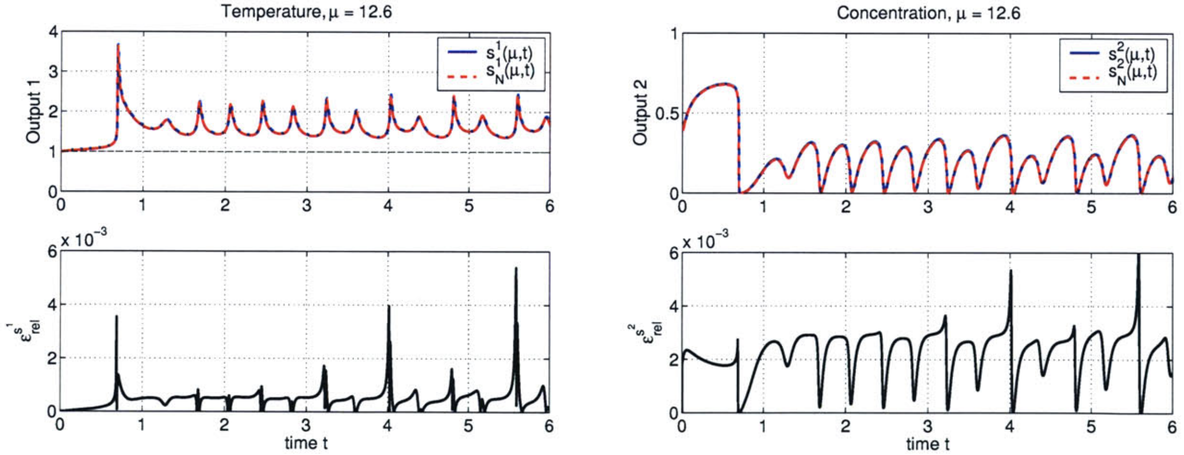
Figure 6-12: AP III: Convergence results for output error.

Finally, we present in Figures 6-13 and 6-14 the outputs and output estimates, $s^{1,2}(\mu, t^k)$ and $s^{1,2}_{N,M}(\mu, t^k)$, and the relative output errors, $\epsilon^{s^1}_{\mathrm{rel}}(\mu, t^k) = |s^1(\mu, t^k) - s^1_{N,M}(\mu, t^k)|/s^1_{\max}(\mu)$ and $\epsilon^{s^2}_{\mathrm{rel}}(\mu, t^k) = |s^2(\mu, t^k) - s^2_{N,M}(\mu, t^k)|/s^2_{\max}(\mu)$, as a function of (discrete) time for $\mu = 12.0$ and $\mu = 12.6$, respectively; here, $s^1_{\max}(\mu) = \max_{t^k \in \mathbb{I}} s^1(\mu, t^k)$ and $s^2_{\max}(\mu) = \max_{t^k \in \mathbb{I}} s^2(\mu, t^k)$



(a) (b)

Figure 6-13: AP III: Output $s(\mu, t^k)$ and output estimate $s_N(\mu, t^k)$, output error, and energy norm error as a function of time for $\mu = 12.0$.



(a) (b)

Figure 6-14: AP III: Output $s(\mu, t^k)$ and output estimate $s_N(\mu, t^k)$, output error, and energy norm error as a function of time for $\mu = 12.6$.

Chapter 7

Application to Real-Time Parameter Estimation and Inverse Problems

7.1 Introduction

In this chapter we employ the reduced-basis method and associated *a posteriori* error estimation for the efficient (real-time) solution of parameter estimation and inverse problems. To this end, we revisit several of the application problems discussed previously in this thesis — the fast and reliable evaluation of the input-output relationship will be the basis for the efficient and robust solution of the estimation problem.

We will start by formally introducing the notion of the "inverse" problem in our context and shortly review some standard approaches to solving these problems. We then incorporate the reduced-basis approximation in the inverse problem solution and discuss a method that can quantify the uncertainty due to measurement and modeling errors. The second part of this chapter is devoted to numerical tests based on the applications and problems introduced in earlier chapters.

7.2 Inverse Problems

7.2.1 Forward vs. Inverse Problem

Our main focus in the last chapters was the development of efficient and reliable numerical methods to evaluate input-output relationships governed by parametrized partial differential equations. More specifically, we were concerned with the following problem: given an input parameter μ in the admissible parameter set $\mathcal{D}$, we evaluate the output of interest, $s(\mu, t^k) = \ell(y(\mu, t^k))$, $\forall\, k \in \mathbb{K}$, where the state variable, $y(\mu, t^k)$, is the solution of a parametrized partial differential equation. Our methods, an all our efforts so far, are geared towards solving the (so called) "forward" problem — that is, the evaluation of the input-output map $\mu \longrightarrow s(\mu, t^k), \forall\, k \in \mathbb{K}$. Let us now assume we are interested in going the other way around, i.e., given the output $s(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, we ask: what is the parameter value μ that resulted in this output? This problem is commonly referred to as the "inverse" problem and its solution requires, in general, the repeated solution of the forward problem; this is where our methods come into play.

Inverse problems have received a lot of attention in engineering and science because of their practical importance in many applications, ranging from geophysics [115, 128], to ecology [14], image processing [27], heat transfer [19, 2, 80], continuum mechanics [12],physiology [11], medicine (e.g., hyperthermia treatment) [89, 31], and nondestructive evaluation [44]. The objective of the inverse problem is to determine unknown system parameters from observations (or measurements) of the state variables or outputs of the system.

We note that most of the literature on inverse problems can be divided into "theoretical" or "numerical" work. The former is concerned with developing concepts of uniqueness of solutions in parameter estimation problems and determining stability, i.e., whether the "identified" parameters depend continuously on the problem data. The latter, on the other hand, is concerned with developing efficient algorithms to solve inverse problems. Although we review some of the theoretical background, our main focus here is the latter.

7.2.2 Formulation

We henceforth assume that there exists a "true" parameter, μ^*, the corresponding "true" state, $y^*(t^k) \equiv y(\mu^*, t^k)$, and the resulting "true" output, $s^*(t^k) \equiv \ell(y^*(t^k))$, $\forall k \in \mathbb{K}$. In general, however, measurements of the complete state $y^*(t^k)$, $\forall\, k \in \mathbb{K}$ may not be possible; instead, only the output $s^*(t^k)$ — representing the observable part of $y^*(t^k)$ — is available for measurements. Furthermore, in actual practice the measurements themselves may be (i) corrupted by noise — we thus have no access to $s^*(t^k)$ itself but rather to a (noisy) measurement $z(t^k)$; and (ii) unavailable at all times t^k, $\forall\, k \in \mathbb{K}$, but only on a coarser timescale t^k, $\forall k \in \overline{\mathbb{K}}_{\rm exp}$, where (say) $\overline{\mathbb{K}}_{\rm exp} \equiv \{10, 20, \ldots, K\}$. We also note that the output $s^*(t^k)$, and hence $z(t^k)$, also depends on the control input $u(t^k)$. However, we shall assume here that $u(t^k)$ is known and may even be set by the operator, e.g., such as the heat input in the delamination example.

We can now formulate the inverse problem more precisely: Determine $\mu \in \mathcal{D}$ such that

$$s(\mu, t^k) = z(t^k), \quad \forall\, k \in \overline{\mathbb{K}}_{\rm exp}. \tag{7.1}$$

This problem, however, is generally "ill-posed," i.e., it is possible that no solution exists (which may often be the case in problems where experimental data is used), or that one has multiple solutions. Furthermore, even if a unique solution exists it might not depend continuously on the measurements. Given these complications, the question arises as to one can actually expect to identify μ^* from (7.1). The problem of *parameter identifiability* [53] is, in general, described as the injectivity of the input-output map $\mu \longrightarrow s(\mu, t^k)$, $\forall k \in \overline{\mathbb{K}}_{\rm exp}$. Many different notions of identifiability exist in the literature [14]; the following definition is taken from [14] (IV.3. Definition 3.1):

Definition 6. *(a) The parameter μ is identifiable at μ^* with respect to $\mathcal{D}$ if for any $\mu \in \mathcal{D}$, $s(\mu, t^k) = s(\mu^*, t^k)$, $\forall\, k \in \overline{\mathbb{K}}_{\rm exp}$ implies $\mu = \mu^*$.*
(b) The parameter μ is called identifiable with respect to $\mathcal{D}$ if it is identifiable at μ^* with respect to $\mathcal{D}$ for every $\mu^* \in \mathcal{D}$.

We note that the identifiability of μ does not only depend on the problem itself (i.e., the governing equation), but also on the outputs and even the number of observations taken for each output. Furthermore, a problem may be parameter identifiable considering the exact (or analytic) solution, but loses this property when considering the discretized problem; in some cases even the discretization technique can be decisive [14].

Solving the estimation problem (7.1) is thus not a trivial task. In applied problems additional complications arise from modeling errors and, as mentioned previously, measurement errors. From the latter it follows that even for the true parameter value μ^* we have $s(\mu^*, t^k) \neq z(t^k)$, whereas from the former it follows that the state $y(\mu^*, t^k)$ does not exactly replicate the true physical state. Because of all these complications, parameter estimation problems are usually considered as optimization problems.

7.2.3 Solution Methods

The first step in stating the optimization problem is the choice of a cost functional or error criterion $J(\mu, z)$; two options are usually considered: the *equation error criterion* and the *output error criterion*. The equation error criterion requires knowledge of the entire state $y(\mu, t^k)$ which is usually not readily available in actual practice. The output error criterion, also called output least squares (OLS) formulation, does not bear this disadvantage since only the output measurements are required. In our context, the OLS formulation can be stated as

$$J(\mu, z) = \frac{1}{2} \sum_{k \in \overline{\mathbb{K}}_{\text{exp}}} \|s(\mu, t^k) - z(t^k)\|^2, \tag{7.2}$$

where $\|\cdot\|$ denotes the usual Euclidean norm. Here, $z(t^k)$ denotes the (noisy) measurement and $s(\mu, t^k)$ is the output. We then obtain the parameter estimate μ^* by minimizing (7.2) over $\mu \in \mathcal{D}$ subject to the governing partial differential equations being satisfied. For example, assuming the governing PDEs are given by (4.3) and (4.4), we obtain

$$\begin{aligned} \mu^* = \arg\min_{\mu \in \mathcal{D}} J(\mu, z), \\ \text{s.t.} \quad (4.3), \ (4.4) \,. \end{aligned} \tag{7.3}$$

The OLS formulation has a wider applicability in practice and is thus more often used. However, it has two major disadvantages: first, it requires (repeated) solution of the governing equation; and second, the cost function is often very flat and not convex in μ. Reduced-order models, such as the one presented in this thesis, can be gainfully employed for the efficient solution of the governing equations. To obtain a well-posed problem and avoid the second difficulty, regularization techniques such as Tikhonov regularization [118], are often used. The regularized cost functional is given by

$$J_R(\mu, z) = \frac{1}{2} \sum_{k \in \overline{\mathbb{K}}_{\text{exp}}} \|s(\mu, t^k) - z(t^k)\|^2 + \frac{1}{2}\, \delta_R \, R(\mu) \tag{7.4}$$

where $R(\mu)$ is the regularization term and $\delta_R > 0$ is the regularization parameter. A common choice for the regularization term $R(\mu)$ is $R(\mu) = \|\mu - \hat{\mu}\|^2$, where $\hat{\mu}$ is some *a priori* estimate of the true parameter μ^*. By regularizing the problem, however, we introduce some assumptions *a priori* which affect the solution; the solution of the regularized problem is biased towards the *a priori* information included in the cost functional and in general different from the solution of the original problem. Furthermore, the uncertainty present in the original problem statement is not quantified and thus valuable information might be lost.

Many algorithms exist to solve the optimization problem (7.3). We mention especially the Levenberg-Marquardt algorithm, a quasi-Gauss-Newton method specialized for minimizing least squares problems, and the Broyden-Goldfarb-Fletcher-Shanno (BFGS) method, a widely used quasi-Newton technique. Other options are the Conjugate-Gradient method or even genetic algorithms. We note that if finite differences are used to approximate the gradient information, the Newton and Gauss-Newton methods become computationally increasingly expensive with the number of parameters that have to be estimated.

7.3 Integration of the Reduced-Basis Framework

In this section we discuss a solution method for inverse problems with which we can explicitly quantify the uncertainty in the problem formulation due to noise and modeling errors in the form of a "possibility region," $\mathcal{R}(\mu^*)$ [75]. The basic idea is: rather than aiming to find *one* regularized solution we strive to identify (almost) all parameter values that satisfy the constraints of the problem, e.g., the given measurements and governing equation. To this end, we first presume the existence of a region $\mathcal{Z}^*(\epsilon_{\text{exp}}, t^k)$ such that $z(t^k) \in \mathcal{Z}^*(\epsilon_{\text{exp}}, t^k)$, $\forall\, k \in \overline{\mathbb{K}}_{\text{exp}}$. We (plausibly) assume that the measurements $z(t^k)$ lie in a band around $s^*(t^k)$, bounded by the experimental error ϵ_{exp}; we thus define

$$\mathcal{Z}^*(\epsilon_{\text{exp}}, t^k) \equiv [s^*(t^k) - \epsilon_{\text{exp}}, s^*(t^k) + \epsilon_{\text{exp}}], \quad \forall\, k \in \overline{\mathbb{K}}_{\text{exp}}\,. \tag{7.5}$$

It is important to note that we assume here that ϵ_{exp} — and thus $\mathcal{Z}^*(\epsilon_{\text{exp}}, t^k)$ — are *known.* In actual practice, this may not always be the case and we thus have to quantify the measurement errors first before solving the inverse problem [73]. In some cases, however, it may be possible to deduce the measurement error from the experimental setup, e.g., from the limited thermal resolution of an IR camera used for temperature measurements.

We next note that, in general, there will exist multiple parameters μ such that the associated outputs $s(\mu, t^k)$ lie within $\mathcal{Z}^*(\epsilon_{\text{exp}}, t^k)$ for all $k \in \overline{K}$. We denote the parameter set containing these parameters by $\mathcal{P}(\mu^*)$, given by

$$\mathcal{P}(\mu^*) \equiv \{\mu \in \mathcal{D} \,|\, s(\mu, t^k) \in \mathcal{Z}^*(\epsilon_{\text{exp}}, t^k),\ \forall\, k \in \overline{\mathbb{K}}_{\text{exp}}\}. \tag{7.6}$$

When the output is induced by partial differential equations, however, evaluating $s(\mu, t^k)$ for a given μ is expensive. We thus incorporate the reduced-basis approximation into the solution process and "replace" the "truth" approximation output $s(\mu, t^k)$ by the output estimate $s_N(\mu, t^k)$. Thanks to our rigorous *a posteriori* error estimation procedures, we know that $s(\mu, t^k)$ satisfies $s_N^-(\mu, t^k) \leq s(\mu, t^k) \leq s_N^+(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, where the upper and lower output bounds are defined as $s_N^{\pm}(\mu, t^k) \equiv s_N(\mu, t^k) \pm \Delta_N^s(\mu, t^k)$, $\forall\, k \in \mathbb{K}$. We take this modeling uncertainty into account and define the possibility region, $\mathcal{R}^* \equiv \mathcal{R}(\mu^*)$, by

$$\mathcal{R}^* \equiv \mathcal{R}(\mu^*) \equiv \{\mu \in \mathcal{D} \,|\, [s_N^-(\mu, t^k), s_N^+(\mu, t^k)] \cap \mathcal{Z}^*(\epsilon_{\text{exp}}, t^k) \neq \emptyset,\ \forall\, k \in \overline{\mathbb{K}}_{\text{exp}}\}. \tag{7.7}$$

We note that $\mathcal{P}(\mu^*) \subset \mathcal{R}(\mu^*)$ and, since $\mu^* \in \mathcal{P}(\mu^*)$, it follows that $\mu^* \in \mathcal{R}(\mu^*)$. Furthermore, as our uncertainty due to modeling errors decreases, $\Delta_N^s(\mu, t^k) \to 0$, we obtain $\mathcal{R}(\mu^*) \to \mathcal{P}(\mu^*)$.

In our solution we account for the experimental (or measurement) errors, ϵ_{exp}, and (reduced-order) modeling error, $\Delta_N^s(\mu, t^k)$. Both error components have the same effect on the size of the possibility region: larger errors in the measurement and model lead to an increase in $\mathcal{R}(\mu^*)$, whereas the size of $\mathcal{R}(\mu^*)$ decreases if better measurements and/or a better model are available. This fact should also guide our choice of N for the reduced-basis approximation, i.e., if possible, we should avoid spoiling accurate measurements by an inaccurate model with large output bounds, $\Delta_N^s(\mu, t^k)$.

The reason is that measurement errors are, in general, harder to improve than the accuracy of our model. We also note that our solution *does not* incorporate modeling errors committed in designing our "truth" approximation. If the "truth" approximation does not replicate the physical system well, we cannot (and should not) expect the reduced-basis approximation to do so. This should always be remembered when interpreting results in actual practice.

We shortly return to our previous discussion of parameter identifiability. In actual practice, the constraints specifying the solution will not only be satisfied by the unique μ^*, but by a set of parameter points contained in the possibility region, $\mathcal{R}(\mu^*)$. However, it follows from Definition 6 that, absent measurement and modeling errors ($\epsilon_{\text{exp}} = \Delta_N^s(\mu, t^k) = 0$), the possibility "region" for an identifiable problem is just the unique parameter point μ^*, i.e., $\mathcal{R}(\mu^*) = \mu^*$. Although this condition will hardly ever be met in practice, we can numerically test and confirm this behavior. We simply "decrease" the measurement error gradually and plot the possibility region for each error level. We will use this as a regular test when discussing numerical results in Section 7.4-7.6.

Finally, we note that we can easily extend these ideas to treat problems with multiple outputs. We then have, of course, a separate measurement corresponding to each output — each additional measured output results in a set of additional conditions that have to be satisfied. We simply require that the condition stated in (7.7) has to be satisfied simultaneously *for all* outputs. Thus, increasing the number of outputs considered in the parameter estimation procedure will, in general, decrease the uncertainty of the parameter estimate. We will also observe this behavior in the examples to follow.

7.3.1 Construction of the Uncertainty Region

We now turn to the construction of the possibility region $\mathcal{R}(\mu^*)$. More precisely, we are interested in finding a set of boundary points, $\mu_j^{\mathcal{R}^*}$, of $\mathcal{R}(\mu^*)$ which we will then use in Section 7.3.2 to find a closed form description for $\mathcal{R}(\mu^*)$. To this end, we first find a parameter point $\mu \in \mathcal{R}(\mu^*)$ — referred to as the initial center μ_{IC}. To obtain μ_{IC} we solve

$$\mu_{\text{IC}} = \arg\min_{\mu \in \mathcal{D}} \frac{1}{2} \sum_{k \in \overline{\mathbb{K}}_{\text{exp}}} \|s_N(\mu, t^k) - z(t^k)\|^2, \tag{7.8}$$

subject to the governing equation (say) (4.17) being satisfied. We evaluate the output estimate, $s_N(\mu, t^k)$, from (4.20) or, if we employ the simple bound, from (4.110). In the numerical examples in Sections 7.4-7.6 we employ a Finite-Difference Levenberg-Marquardt scheme [64] to solve (7.8) and use the geometric center of the admissible parameter domain $\mathcal{D}$ as the initial guess. We recall that solving the optimization problem with the truth approximation output $s(\mu, t^k)$ would be computationally very expensive. In actual practice, we do not need solve for the true minimizer of (7.8); instead we stop the optimization procedure as soon as one iterate lies within $\mathcal{R}(\mu^*)$.

Given μ_{IC}, we perform a search along d_β directions determined by the angles β_j, $1 \le j \le d_\beta$, to find the boundary point μ_j^{R}, $1 \le j \le d_\beta$: we start with (say) $\beta_1 = 0$ and conduct a binary chop along this direction to find the boundary point $\mu_1^{\mathcal{R}^*}$, i.e., we first choose two parameters μ_{in} and μ_{out} located inside and outside $\mathcal{R}(\mu^*)$, respectively. We next calculate $\mu_{\text{mean}} = \frac{1}{2}(\mu_{\text{in}} + \mu_{\text{out}})$ and check whether μ_{mean} lies within $\mathcal{R}(\mu^*)$. If μ_{mean} is inside $\mathcal{R}(\mu^*)$ we set $\mu_{\text{in}} = \mu_{\text{mean}}$ and $\mu_{\text{out}} = \mu_{\text{out}}$, otherwise we set $\mu_{\text{in}} = \mu_{\text{in}}$ and $\mu_{\text{out}} = \mu_{\text{mean}}$. We repeat this process until $\|\mu_{\text{out}} - \mu_{\text{in}}\|$ is smaller than a desired tolerance $\Delta\mu^{\mathcal{R}^*}$. We next increment β_1 by $\Delta\beta = 360/d_\beta$ degrees to obtain $\beta_2 = \beta_1 + \Delta\beta$ and conduct a binary chop along the new direction to find $\mu_2^{\mathcal{R}^*}$. We repeat this process for all directions β_j, $1 \leq j \leq d_\beta$.

There are two important issues concerning the construction of $\mathcal{R}(\mu^*)$. First, the algorithm relies on the fact that $\mathcal{R}(\mu^*)$ is star-shaped with respect to μ_{IC}. If we find, or suspect, that this is not the case, we should restart the algorithm with a new initial center in $\mathcal{R}(\mu^*)$ to obtain the remaining part of the boundary. Second, $\mathcal{R}(\mu^*)$ might not be connected. In this case we have to map the boundary using different initial centers for each region separately.

Finally, we remark that the algorithm is not truly exhaustive, i.e., it is possible that we "miss out" on parameters which satisfy (7.7). However, we certainly decrease the uncertainty as compared to a single regularized solution.

7.3.2 Smallest Enclosing Ellipsoids

Given the set of boundary points $\mu_j^{\mathcal{R}^*}$, $1 \leq j \leq d_\beta$, we are interested in finding a closed-form description of the possibility region $\mathcal{R}(\mu^*)$. In general, the closed-form description is favorable because of (*i*) visualization, storage, or post-processing requirements, e.g., a possible subsequent design or optimization over all $\mu \in \mathcal{R}(\mu^*)$ would be much simplified given such description; and (*ii*) constructing $\mathcal{R}(\mu^*)$ becomes increasingly expensive with the number of parameter dimensions — we would thus like to characterize the possibility region with only a limited number of boundary points. One possible approach to obtain this closed-form description is to find the ellipsoid with minimum area (or volume) which contains the given set of (boundary) points. This problem, usually referred to as "smallest enclosing ellipsoids," has been widely studied [39, 87, 88, 107, 126].

An ellipsoid in $\mathbb{R}^d$ is defined as the set of points $x \in \mathbb{R}^d$ satisfying

$$(x - x_c)^T Q \, (x - x_c) = 1, \tag{7.9}$$

where $x_c \in \mathbb{R}^d$ is the center of the ellipsoid and $Q \in \mathbb{R}^{d \times d}$ is a positive definite matrix. Furthermore, the ellipsoid body is the set of points $x \in \mathbb{R}^d$ such that $(x - x_c)^T Q \, (x - x_c) \leq 1$. Given a set of points x_i, $1 \leq i \leq d$, the problem of finding the minimum enclosing ellipsoid can then be written as the convex program [72]

$$\begin{aligned} &\min \; -\log\det(Q) \\ &\text{s.t.} \begin{cases} (x_i - x_c)^T Q \, (x_i - x_c) \leq 1, \quad 1 \leq i \leq d, \\ Q \text{ positive definite,} \end{cases} \end{aligned} \tag{7.10}$$

which can be solved using Welzl's algorithm [126].

Here, we use a suboptimal but much simpler approach which can be cast as a linear program. We first note that (7.8) can also be written as

$$x^T P x + p^T x + c = 0, \tag{7.11}$$

where $x_c = -\frac{1}{2} P^{-1} p$ and $Q = P/(x_c^T P x_c - c)$; here $P \in \mathbb{R}^{d \times d}$, $p \in \mathbb{R}^d$, and $c \in \mathbb{R}$. Given the set of boundary points $\mu_j^{\mathcal{R}^*}$, $1 \leq j \leq d_\beta$, we then define the linear program

$$\begin{aligned} &\min\ \delta \\ &\text{s.t.} \begin{cases} {\mu_j^{\mathcal{R}^*}}^T P \mu_j^{\mathcal{R}^*} + p^T \mu_j^{\mathcal{R}^*} + c \leq 0, & 1 \leq j \leq d_\beta, \\ {\mu_j^{\mathcal{R}^*}}^T P \mu_j^{\mathcal{R}^*} + p^T \mu_j^{\mathcal{R}^*} + c \geq \delta, & 1 \leq j \leq d_\beta, \\ P_{11} = 1,\ P \text{ symmetric.} \end{cases} \end{aligned} \tag{7.12}$$

Here, we are essentially minimizing the largest diameter of the ellipsoid to the given set of boundary points subject to the constraint that all boundary points lie in the ellipse. We note, however, that the solution of this problem is not guaranteed to result in an ellipse, i.e., we may obtain a hyperbola or parabola.

We will now apply the above ideas to a few of the problems discussed earlier in this thesis. The general procedure for testing our methods is as follows: we first select the "true" parameter value $\mu^* \in \mathcal{D}$; given μ^*, we solve the "truth" finite element approximation and evaluate the output $s^*(t^k) = s(\mu^*, t^k)$; finally, given $s^*(t^k)$ we then construct $z(t^k)$ and $\mathcal{Z}^*(\epsilon_{\text{exp}}, t^k)$ in (7.5) by adding the measurement error, ϵ_{exp}, to the data. Throughout this chapter, we measure ϵ_{exp} in percent of the maximum output $s^*_{\max} \equiv \max_{k \in K} |s(\mu^*, t^k)|$, i.e., "$\epsilon_{\text{exp}} = 1\%$" is equivalent to $\epsilon_{\text{exp}} = 0.01 s^*_{\max}$. We then attempt to estimate μ^* given the inputs $z(t^k)$ and $\mathcal{Z}^*(\epsilon_{\text{exp}}, t^k)$ for the inverse procedure. We will not use actual physical experiments here and we thus do not have to be concerned about modeling errors in the "truth" approximation. All timing results presented are obtained on an Intel 750 MHz Pentium III processor running MATLAB 6.5.

7.4 Numerical Exercise 3: Banks and Kunisch

We first consider the one-dimensional convection-diffusion problem discussed in Section 4.8.4. The specific transport system (4.165)-(4.166) models the movement of fluids and transport of substances within the brain. Knowledge of this process is important in understanding the transport of large protein molecules in brain interstitial fluid (ISF) and hence fundamental to understanding cerebral function. The physiological investigations are focused on the issue whether the flow in brain tissue is governed by simple diffusion or diffusion plus convection. Given the measured outputs $s_{N,j}$, $1 \leq j \leq 3$ — in actual practice obtained from animal tests — the goal is thus to determine the diffusivity q_1 and the velocity q_2.

We recall that the parameter and admissible parameter domain are $\mu \equiv (\mu_1, \mu_2) \equiv (q_1, q_2) \in \mathcal{D} \equiv [0.1, 1] \times [0.5, 5] \subset \mathbb{R}^{P=2}$; and the timestep and number of total timesteps are given by $\Delta t = 0.01$ and $K = 100$, respectively. Based on the numerical results and our previous discussion in Section 4.8.4 we employ the primal formulation with the simple bound defined in Proposition 9; we thus have no need for the dual problems corresponding to the three outputs.

7.4.1 Parameter Estimation

In [13, 14], the authors consider the model problem (4.165)-(4.166) as a test case for their parameter estimation techniques. They discretize the original equations using quasi-modal techniques and cubic spline based schemes. The true parameter, $\mu^* = (0.3, 1.75)$, is estimated by solving the OLS formulation with a Finite-Difference Levenberg-Marquardt Algorithm.

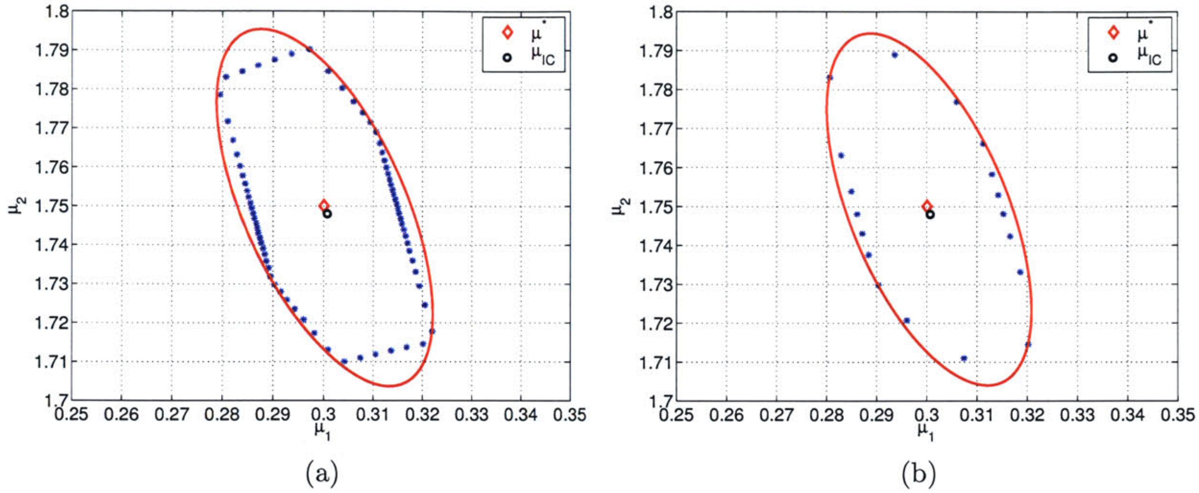
In general, modal techniques are only applicable to simple problems where the true modes do not depend on the parameter. Since the (to-be-estimated) parameter itself is unknown, it may not be possible to generate an approximation space that can represent the true modes sufficiently well. The use of (quasi-)modal techniques in parameter estimation problems is therefore very restricted in practice. Cubic spline based schemes, on the other hand, do not bear this disadvantage and can be applied to more general problems. Although the basis may also depend on the parameter for cubic splines, the necessary coefficient matrices can be precomputed and used in an offline-online fashion [14] for the constant parameter case. If the parameters are spatially varying, more elaborate techniques, e.g., series expansions, can be used to avoid the increasing computational cost due to repeated evaluation of the coefficient matrices.

Here, we consider the results summarized in [13, 14] for the case with three outputs and observed at only one point in time. The authors report that the parameter estimation scheme using the quasi-modal technique fails to converge for this case while the cubic spline method performs well: for $N_S = 8$ (the dimension of the approximation space) the parameter estimate is $\hat{\mu} = (0.3001, 1.7486)$. The authors also report that cubic splines in general perform better for transport systems and yield better results for a given amount of data. To see how the reduced-basis approximation performs compared to the cubic spline method, we solve (7.8) with $\epsilon_{\text{exp}} = 0$ and we choose $N = 8$ for the dimension of the reduced-basis space; as the stopping criterion we require that two consecutive iterates satisfy $\|\mu^k - \mu^{k+1}\|/\|\mu^k\| \leq 1\,\text{E}-10$. We obtain, after 7 iterations and 0.5 sec., the estimate $\mu_{\text{IC}} = (0.2999, 1.7501)$. We observe that the reduced-basis approximation performs equally well in estimating the unknown parameter. However, the solution to (7.8) alone does not quantify the uncertainty in the system due to noise or modeling errors and is therefore only the first step in our robust estimation procedure.

Before analyzing different aspects of this procedure, we present a typical solution to the inverse problem. We assume that $\mu^* = (0.3, 1.75)$, $\epsilon_{\text{exp}} = 1\%$, and $\overline{\mathbb{K}}_{\text{exp}} = \{10\}$ (i.e., the output is measured only at one (discrete) timestep $t^k = 10\Delta t$). We choose $N = 12$ for our reduced-basis approximation and solve (7.8) for the initial center μ_{IC}. After 3 iterations and 0.28 sec. we obtain $\mu_{\text{IC}} = (0.3007, 1.7480)$ (note that we stop the optimization as soon as one iterate lies within $\mathcal{R}(\mu^*)$). We next evaluate the boundary points $\mu_j^{\mathcal{R}^*}$, $1 \leq j \leq 72$, with $\Delta\beta = 5°$ $(d_\beta = 72)$ and $\Delta\mu^{\mathcal{R}^*} = 1\,\text{E}-4$: we require 827 forward solutions obtained in a total of 16.9 sec. Finally, given the boundary points $\mu_j^{\mathcal{R}^*}$, $1 \leq j \leq 72$, we solve (7.12) for the enclosing ellipse containing $\mathcal{R}$.

In Figure 7-1(a) we plot the "true" parameter value, μ^*, the initial center, μ_{IC}, the boundary points $\mu_j^{\mathcal{R}^*}$ and the enclosing ellipse. We observe that the initial center is close to the true parameter value. However, only the possibility region $\mathcal{R}(\mu^*)$ renders a clear picture of the uncertainty in the problem. We also observe that the enclosing ellipse gives a very good (and tight) description of the actual possibility region marked by the blue boundary points — this is certainly problem specific and may not always be the case. Finally, we notice that we do not need such a small increment $\Delta\beta$ — and hence so many boundary points $\mu_j^{\mathcal{R}^*}$ — to find the enclosing ellipse in actual practice. In general, a larger $\Delta\beta$ suffices to capture almost all (in the probabilistic sense) parameter values $\mu \in \mathcal{R}(\mu^*)$. To this end, we plot the corresponding result for $\Delta\beta = 20°$ in Figure 7-1(b): the enclosing ellipse is only slightly different, but the number of forward solutions and time to construct $\mathcal{R}(\mu^*)$ dropped by approximately a factor of 4 to 212 and 4.26 sec., respectively.

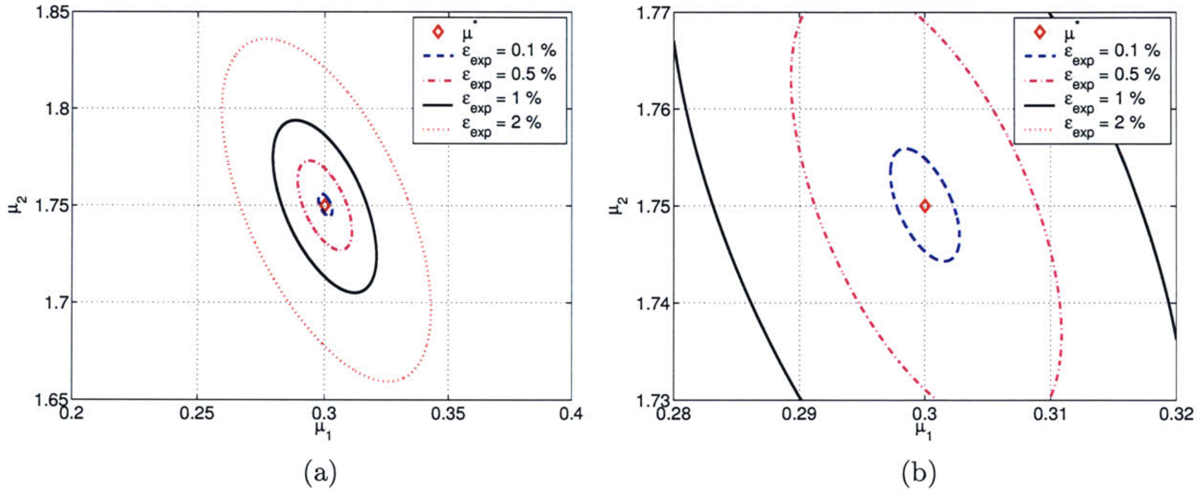
We next consider the sensitivity of the possibility region with respect to the measurement error ϵ_{exp}. We assume that the true parameter is $\mu^* = (0.3, 1.75)$ and the measurements are taken at $\overline{\mathbb{K}}_{\text{exp}} = \{10\}$. We choose $N = 12$ for our reduced-basis model and set $\Delta\beta = 10°$ to construct the



(a) (b)

Figure 7-1: NE 3: True parameter value μ^*, initial center $\mu_{\rm IC}$, and possibility region $\mathcal{R}(\mu^*)$ for (a) $\Delta\beta = 5°$ and (b) $\Delta\beta = 20°$.

possibility region. We solve the estimation problem and plot the enclosing ellipses for $\epsilon_{\rm exp} = 0.1\%$, 0.5%, 1%, and 2%, in Figure 7-2(a) and a zoom in Figure 7-2(b). As expected, the possibility region $\mathcal{R}(\mu^*)$ shrinks with a decreasing measurement error.

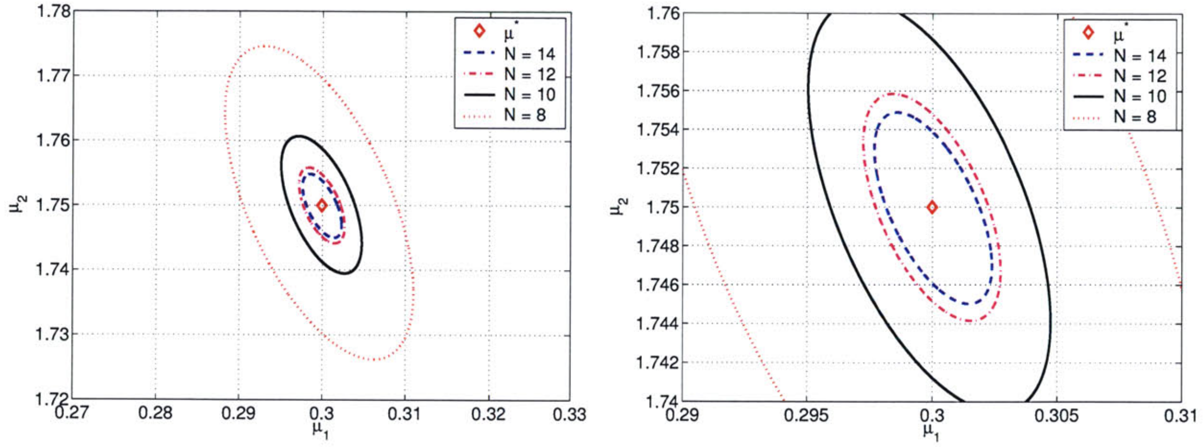


(a) (b)

Figure 7-2: NE 3: Possibility regions $\mathcal{R}(\mu^*)$ as a function of measurement error.

We obtain a similar result for the sensitivity with respect to modeling errors due to our reduced-basis approximation. We assume that the true parameter is $\mu^* = (0.3, 1.75)$, the measurements are taken at $\overline{\mathbb{K}}_{\rm exp} = \{10\}$, and the measurement error is now fixed at $\epsilon_{\rm exp} = 0.1\%$. We also set $\Delta\beta = 10°$ to construct the possibility region. In Section 4.8.4 we observed that the accuracy of the reduced-basis approximation — and hence the modeling errors introduced — strongly depend on the dimension of the reduced-basis space. To show this effect on the solution of the parameter estimation problem, we solve the inverse problem using four reduced-basis approximations with

N = 6, 10, 12, and 14. The enclosing ellipses are plotted in Figure 7-3(a) and (b): as expected, a larger dimension of the reduced-basis space and thus smaller modeling errors result in a smaller possibility region $\mathcal{R}(\mu^*)$.



(a)

(b)

Figure 7-3: NE 3: Possibility regions $\mathcal{R}(\mu^*)$ as a function of N.

The current example shows the behavior anticipated for an identifiable problem: the possibility region shrinks with decreasing measurement and modeling errors and eventually reduces to the single parameter point μ^*. We confirm the last conjecture by setting the measurement error to zero, $\epsilon_{\text{exp}} = 0\%$, and using the "best" reduced-basis approximation with $N = 16$ resulting in the smallest modeling errors. We solve the estimation problem for $\mu^* = (0.3, 1.75)$ with $\Delta\beta = 5°$ and $\Delta\mu^{\mathcal{R}(\mu^*)} = 1\,\text{E}-8$ and plot the boundary points of the possibility region in Figure 7-4 (note the scaling of the axis). The initial center μ_{IC} differs from the true parameter point μ^* only in the fifth decimal place and the maximum deviation for any point within $\mathcal{R}(\mu^*)$ from μ_{IC} is only 0.025%.

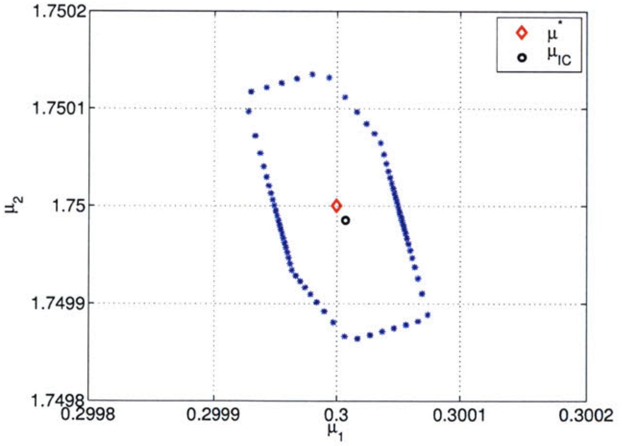


Figure 7-4: NE 3: Possibility regions $\mathcal{R}(\mu^*)$ for $\epsilon_{\text{exp}} = 0\%$ and $N = 16$.

We next analyze the behavior of the parameter estimation procedure as more measurements in time become available. We shall assume that the measurements are taken at $\overline{\mathbb{K}}_{\text{exp}} = \mathbb{K}_i$, $1 \leq i \leq 4$, where $\mathbb{K}_1 \equiv \{10\}$, $\mathbb{K}_2 \equiv \{10, 20, 30\}$, $\mathbb{K}_3 \equiv \{10, 20, \ldots, 50\}$, and $\mathbb{K}_4 \equiv \{10, 20, \ldots, 100\}$, i.e., we assume that we have access to 1, 3, 5, or 10 measurements in time. We shall consider the true parameter $\mu^* = (0.3, 1.75)$ and the measurement error $\epsilon_{\text{exp}} = 0.1$; we choose the dimension of the reduce-basis space to be $N = 12$ and set $\Delta\beta = 5°$. We solve the estimation problem and plot the possibility regions in Figure 7-5. We note that each additional measurement in time acts as an additional constraint in constructing the possibility region. The regions thus shrink with the number of measurements and are strictly contained within each other.

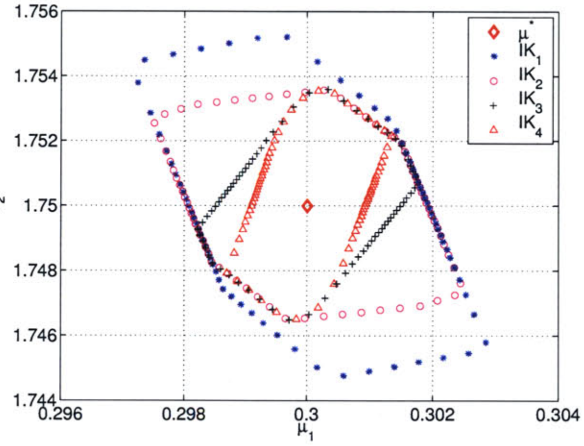


Figure 7-5: NE 3: Possibility regions $\mathcal{R}(\mu^*)$ for $\mathbb{K}_1 \equiv \{10\}$, $\mathbb{K}_2 \equiv \{10, 20, 30\}$, $\mathbb{K}_3 \equiv \{10, 20, \ldots, 50\}$, and $\mathbb{K}_4 \equiv \{10, 20, \ldots, 100\}$.

Finally, we show that size of the possibility region, $\mathcal{R}(\mu^*)$, does not only vary with the size of measurement and modeling errors, but may also depend on where the true parameter, μ^*, lies within the admissible parameter set $\mathcal{D}$. We consider four different values for μ^*, each lying close to one of the "corners" of $\mathcal{D}$. We solve the parameter estimation problem (with $\epsilon_{\text{exp}} = 0.1$, $\overline{\mathbb{K}}_{\text{exp}} = \{10\}$, $N = 12$, and $\Delta\beta = 10°$) and present the enclosing ellipses in Figure 7-6 — note that the scaling of the axis is the same in all plots. We observe that the uncertainty increases for larger values of the diffusivity μ_1 but is fairly insensitive with respect to the velocity μ_2.

We recall that the approach taken in [13, 14] results in only one (possibly regularized) solution to the parameter estimation problem. Our approach presented here, on the other hand, renders a more complete picture of the problem specific features: we can explicitly quantify the uncertainty in the solution due to (*i*) modeling and measurement errors; (*ii*) the number of observations in time; and (*iii*) the actual value of the to-be-estimated parameter. This knowledge may also help to design experiments to obtain actual data. As the number of parameters increases, the applicability of (quasi-)modal techniques would be even more restricted, whereas the detriment to cubic spline based schemes (the increased dimension of the approximation space) would probably be small. The efficiency of our approach — including the construction of $\mathcal{R}^*$ — would certainly deteriorate because the computational cost to determine the possibility region increases exponentially with the number of parameters.

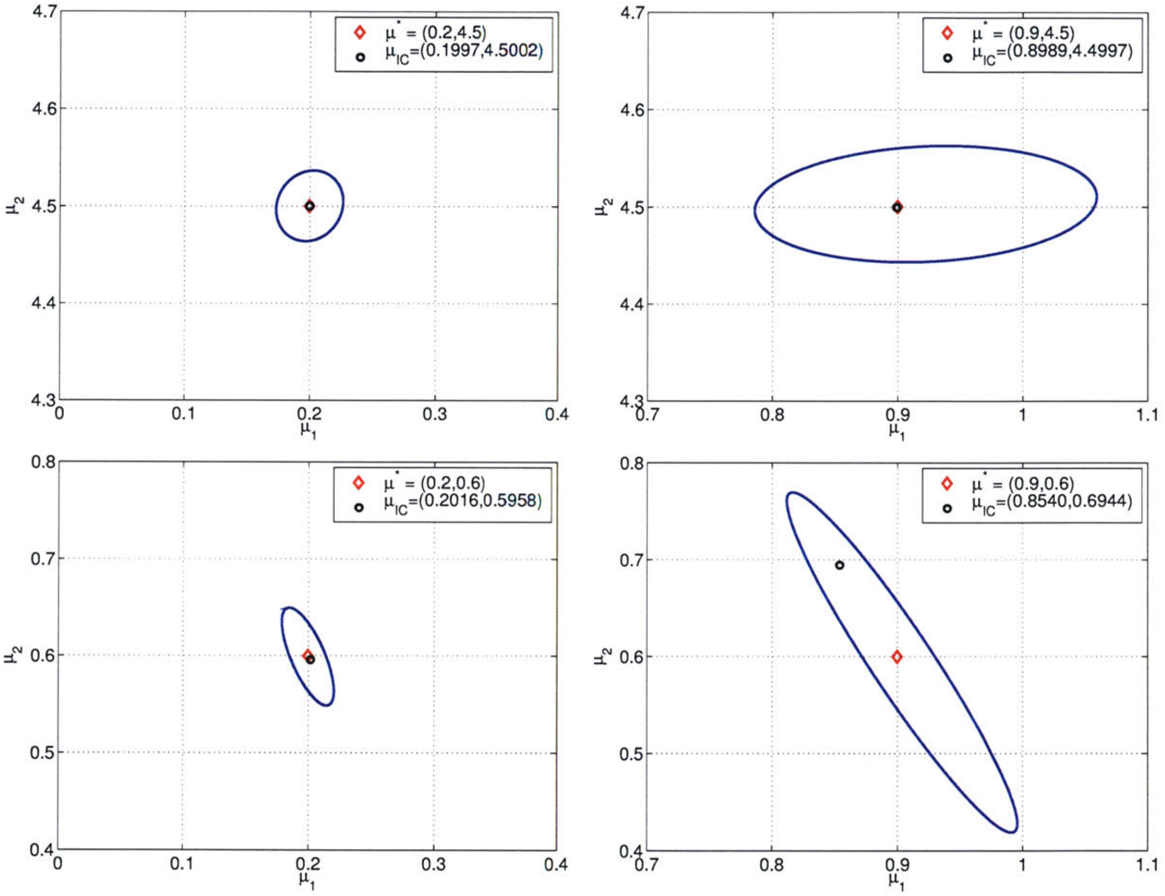


Figure 7-6: NE 3: Possibility regions $\mathcal{R}(\mu^*)$ for different parameter values μ^*.

7.5 AP I: Nondestructive Evaluation of Delamination

We now turn to the nondestructive evaluation problem introduced in Section 1.1.1. The surface of the tested structure is exposed to a heat source for $t \in [0, 0.5]$ and an IR imaging system is used to monitor the surface temperature for $t \in \bar{I} = [0, 10]$. The measurement noise is due to the limited thermal resolution of the IR imaging system. We shall assume here that the experimental error varies in the range $0.3 - 5.0\%$ [114]. Our goal is to infer the delamination length, w_{del}, and the conductivity ratio, $\varkappa$, from the measured surface temperatures $s_1(\mu, t^k)$ and $s_2(\mu, t^k)$.

We recall that the parameter and admissible parameter range are given by $\mu \equiv (\mu_1, \mu_2) \equiv (w_{\text{del}}/2, \varkappa) \in \mathcal{D} \equiv [1, 10] \times [0.4, 1.8]$ (note that we consider the half-width of the delamination as the parameter), the timestep is $\Delta t = 5\,\text{E}{-2}$, the time interval of interest is $\bar{I} = [0, 10]$, and the number of timesteps is $K = 200$. The heat input is thus applied for the first ten timesteps: $u(t^k) = q(t^k) = 1$ for $1 \leq k \leq 10$ and $u(t^k) = q(t^k) = 0$ for $k \geq 11$. We employ the reduced-basis approximation generated in Section 4.7 to solve the parameter estimation problem; based on our discussion of the convergence and computational efficiency results, we use the primal-dual formulation here.

7.5.1 Estimation of Delamination Length

To begin, we present a sample solution of the parameter estimation procedure: we shall assume that the true parameter $\mu^* = (4, 1.2)$ has to be estimated, that the measurement error is $\epsilon_{\text{exp}} = 0.5\%$, and that we are privy to measurements taken at $\mathbb{K}_{\text{exp}} = \{10, 20, \ldots, 200\}$. We next choose the dimension of the primal and dual reduced-basis approximations to be $N = N_{\text{pr}} = N_{\text{du},1} = N_{\text{du},2} = 50$. Given the noisy measurements, we first solve (7.8) for μ_{IC}; we obtain $\mu_{\text{IC}} = (3.991, 1.198)$ in 6.16 sec. after only 3 iterations. We next construct the boundary of the possibility region with the tolerance $\Delta\mu^{\mathcal{R}^*} = 1\,\text{E}{-4}$ and solve (7.12) for the enclosing ellipse. The result is shown in Figure 7-7. Since the scalings of the two parameters μ_1 and μ_2 are very different (reflected in a high eccentricity of the enclosing ellipse) we do not use a constant increment $\Delta\beta$ for the search angle β; instead we set $\beta_j = \frac{1}{2}(1 - \cos(j\pi/15))180°$ for $0 \leq j \leq 15$, and $\beta_j = \frac{1}{2}(3 - \cos(j\pi/15))180°$ for $16 \leq j \leq 28$; we thus have $d_\beta = 28$ search directions. Solving for the boundary points requires a total of 345 forward solutions (note that each evaluation involves the solution of the primal and the two dual problems) and takes a total of 211.6 sec.

We next consider the sensitivity of the parameter estimation with respect to measurement and modeling errors. We present in Figure 7-8(a), as a function of ϵ_{exp}, the enclosing ellipses for $\mu^* = (4, 1.2)$; here $N = N_{\text{pr}} = N_{\text{du},1} = N_{\text{du},2} = 50$ and $\mathbb{K}_{\text{exp}} = \{10, 20, \ldots, 200\}$. We observe that the uncertainty in the parameter estimate shrinks with a decreasing measurement error. We note that a 2% error in the temperature measurements results in a maximum uncertainty of approximately 5% in the estimated delamination width. In Figure 7-8(b), we present, as a function of $N = N_{\text{pr}} = N_{\text{du},1} = N_{\text{du},2}$, the possibility regions for $\mu^* = (4, 1.2)$ and a constant measurement error $\epsilon_{\text{exp}} = 0.3\%$. We plot the actual boundary points, $\mu_j^{\mathcal{R}^*}$, here because the influence of N is more visible. It is interesting to note that smaller modeling errors, i.e., larger N, only increases the accuracy in the μ_1 direction — the accuracy in the μ_2 direction is limited by the measurement error (a smaller ϵ_{exp} results in a decrease of $\mathcal{R}(\mu^*)$ also on the μ_2 direction). However, our primary interest lies in estimating the delamination half-width μ_1 — we only estimate μ_2 as a means to quantify the (maximum) uncertainty in μ_1. We also note that the "cut" edges, e.g., for the $N = 60$ region, are the result of the limited resolution of the search angles β_j.

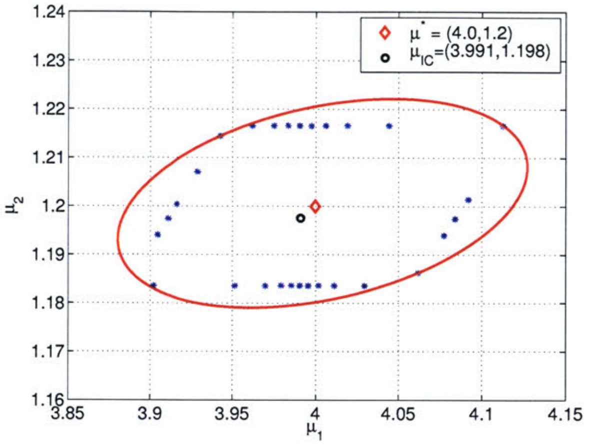
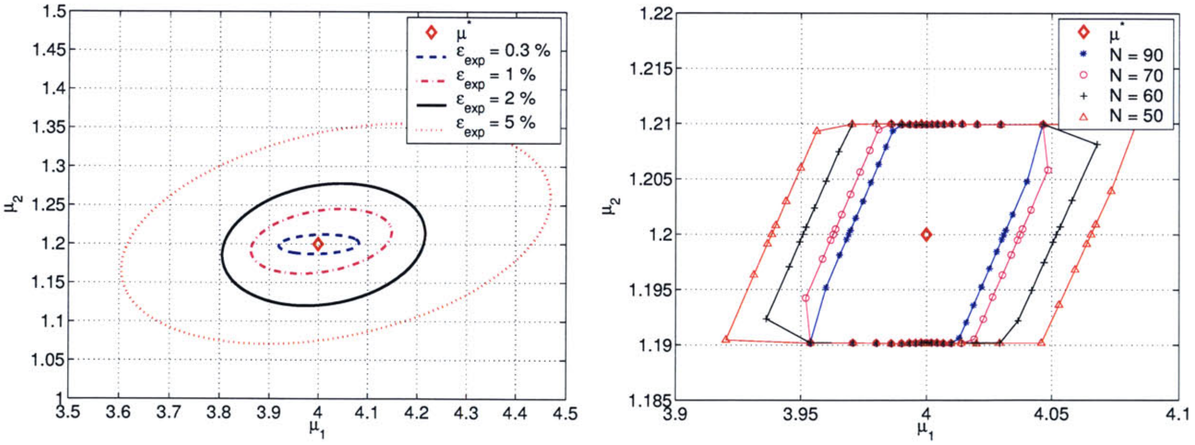


Figure 7-7: AP I: Possibility region $\mathcal{R}(\mu^*)$ for $\mu^*=(4,1.2)$.



(a)

(b)

Figure 7-8: AP I: Possibility region $\mathcal{R}(\mu^*)$ for $\mu^*=(4,1.2)$ as a function of (a) ϵ_{exp} and (b) N.

ϵ_{exp}	$\Delta\mu_1$	$\Delta\mu_2$
0.3%	0.16262	0.025187
1.0%	0.28626	0.083625
2.0%	0.41102	0.15794
5.0%	0.86509	0.28518

Table 7.1: AP I: Uncertainty in the parameter estimate.

In actual practice, we may only be interested in quantifying the absolute uncertainty of the parameter estimate instead of the enclosing ellipse — especially as the number of parameters increases and visualizing the results becomes harder. To this end, we introduce the "bounding box," i.e., the smallest box which contains the minimum enclosing ellipse. We present in Figures 7-9 the ellipses and corresponding bounding boxes for the results shown in Figure 7-8(a). The length and height of each box now correspond to the uncertainty of the parameter estimate, denoted by $\Delta\mu_1$ and $\Delta\mu_2$, for the given measurement error. We tabulate the results for the four different measurement errors in Table 7.1. We observe that the uncertainty increases almost linearly with the measurement error.

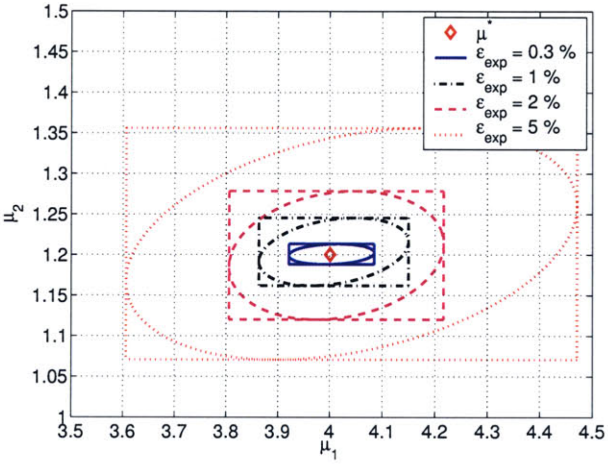


Figure 7-9: AP I: Possibility region $\mathcal{R}(\mu^*)$ and bounding box for $\mu^* = (4, 1.2)$ as a function of ϵ_{exp}.

We next consider the sensitivity with respect to which measurement is used for the solution of the inverse problem. We shall assume that the true parameter is $\mu^* = (8, 1.6)$, the measurement error is $\epsilon_{\text{exp}} = 0.5\%$, and measurements are taken at $\overline{\mathbb{K}} = \{10, 20, \ldots, K_f\}$; we also choose $N = N_{\text{pr}} = N_{\text{du},1} = N_{\text{du},2} = 50$. We next define $\mathbb{S}$ to be the set of measurements used in the parameter estimation procedure, i.e., $\mathbb{S} = \{1\}$ means that only the first measurement employed. We present in Figure 7-10 the two enclosing ellipses for $\mathbb{S} = \{1\}$ and $\mathbb{S} = \{2\}$. We note that using only the second measurement (on top of the undamaged region) results in a very poor estimate of the delamination width. Employing only the first measurement, on the other hand, results in a very good estimate of the delamination width. If we combine both estimates we effectively obtain the shaded region as our possibility region: $\mathbb{S} = \{1\}$ limits the uncertainty in μ_1 whereas $\mathbb{S} = \{2\}$ limits the uncertainty in μ_2. We observe that choosing the "correct" location for the sensors can considerably decrease the uncertainty; our analysis can thus help in guiding the placement of sensors in actual practice.

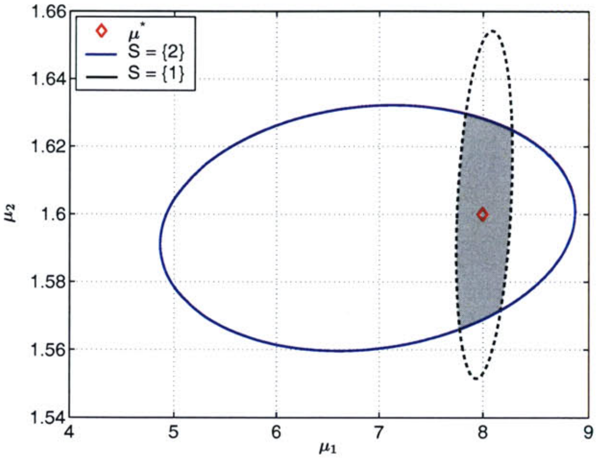


Figure 7-10: AP I: Possibility region $\mathcal{R}(\mu^*)$ for $\mu^* = (8, 1.6)$ for $\mathbb{S} = \{1\}$ and $\mathbb{S} = \{2\}$.

We now investigate the sensitivity of the parameter estimation with respect to the number of measurements in time. To this end, we shall assume that $\mu^* = (4, 1.2)$, $\epsilon_{\text{exp}} = 0.3\%$, and $N = N_{\text{pr}} = N_{\text{du},1} = N_{\text{du},2} = 50$. We solve the inverse problem for $\overline{\mathbb{K}}_{\text{exp}} = \mathbb{K}_i$, $1 \leq i \leq 3$, where $\mathbb{K}_1 \equiv \{5, 10, \ldots, 200\}$, $\mathbb{K}_2 \equiv \{10, 20, \ldots, 200\}$, and $\mathbb{K}_3 \equiv \{20, 40, \ldots, 200\}$. We plot the possibility regions for these three cases in Figures 7-11. We observe that the possibility regions corresponding to $\mathbb{K}_1$ and $\mathbb{K}_2$ coincide; it is thus sufficient to consider measurements taken at only every 10th timestep. However, the uncertainty does increase if measurements are taken only for all $k \in \mathbb{K}_3$. Since the size of $\overline{\mathbb{K}}_{\text{exp}}$ strongly affects the computational efficiency — recall the $O(K(K+1)N_{\text{pr}}N_{\text{du}})$ complexity to evaluate the output estimate — the choice of $\overline{\mathbb{K}}_{\text{exp}}$ should reflect the trade-off between the acceptable uncertainty and computational cost.

We also point out that the length of the observation interval is very important in obtaining a good estimate. We shall assume that the true parameter is $\mu^* = (7.5, 1.2)$ and the measurement error is $\epsilon_{\text{exp}} = 0.5\%$; we also choose $N = N_{\text{pr}} = N_{\text{du},1} = N_{\text{du},2} = 50$. We first recall from Figure 4-10 that the difference in the temperature measurements for different parameter values becomes visible only after a certain transient. Furthermore, the time for the transient increases with increasing values of μ_1 and μ_2. We plot the enclosing ellipses for $\overline{\mathbb{K}} = \{10, 20, \ldots, K_f\}$ with $K_f = 100$ and $K_f = 200$ in Figure 7-12. We observe that the longer observation time considerably decreases the uncertainty in the estimate.

Finally, we consider the dependence of the inverse solution on the actual delamination width and the conductivity ratio. We assume that $\epsilon_{\text{exp}} = 0.3\%$ and $\overline{\mathbb{K}}_{\text{exp}} \equiv \{10, 20, \ldots, 200\}$ and set $N = N_{\text{pr}} = N_{\text{du},1} = N_{\text{du},2} = 50$. We next solve the parameter estimation procedure for the four "true" parameters $\mu^* = (3, 1.6)$, $(3, 0.6)$, $(8, 1.6)$, and $(8, 0.6)$. The enclosing ellipses are shown in Figure 7-13 (note that the scaling of the axis is the same in all plots). We observe that the uncertainty increases with increasing delamination width, w_{del}. Unfortunately, an accurate estimate of w_{del} is especially important if w_{del} is large since the safety of the structure may be influenced. We also just showed that additional temperature measurements in time do not help to decrease this uncertainty — however, additional temperature measurements on the surface, i.e., between the

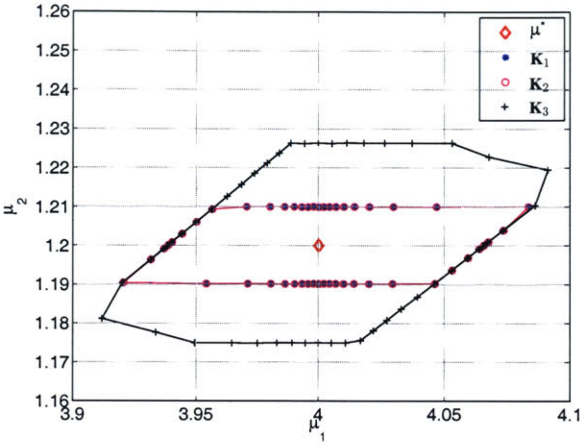


Figure 7-11: AP I: Possibility region $\mathcal{R}(\mu^*)$ for $\mu^* = (4, 1.2)$ and $\mathbb{K}_1 \equiv \{5, 10, \ldots, 200\}$, $\mathbb{K}_2 \equiv \{10, 20, \ldots, 200\}$, $\mathbb{K}_3 \equiv \{20, 40, \ldots, 200\}$.

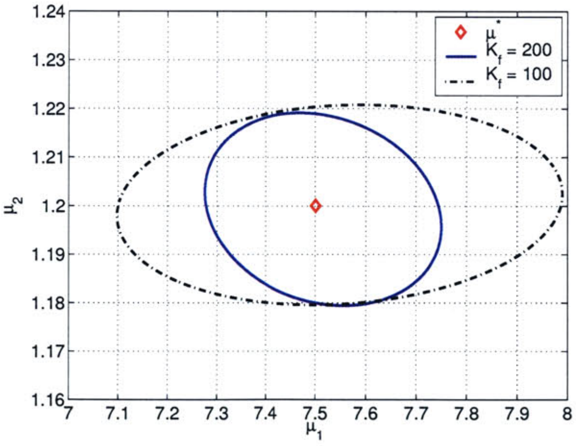


Figure 7-12: AP I: Possibility region $\mathcal{R}(\mu^*)$ for $\mu^* = (7.5, 1.2)$ and $\mathbb{K} \equiv \{10, 20, \ldots, K_f\}$.

locations where s_1 and s_2 are measured, would most likely lead to an improved estimate.

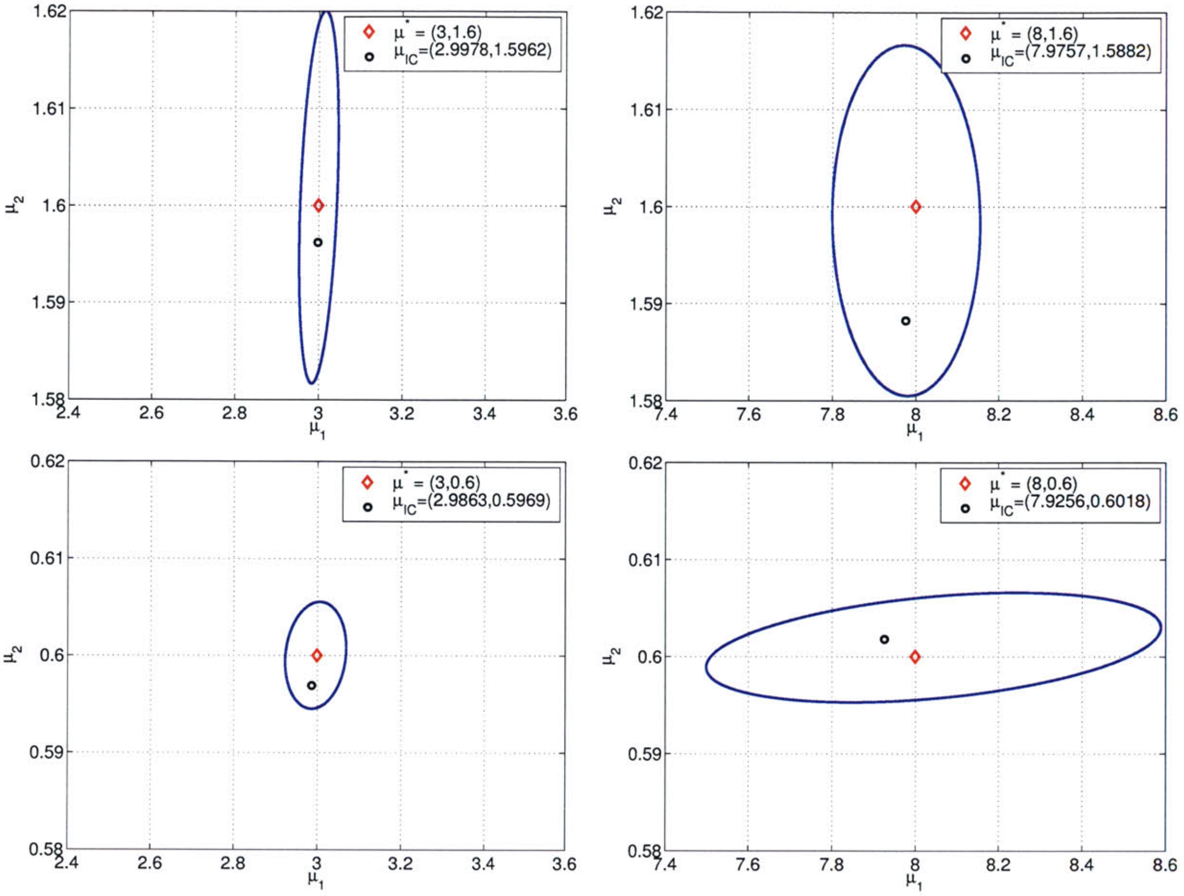


Figure 7-13: AP I: Possibility regions $\mathcal{R}(\mu^*)$ for different "true" parameter values μ^*.

7.6 AP II: Dispersion of Pollutants

As the last application in this chapter we consider the pollutant dispersion problem introduced in Section 1.1.1. Our goal is the following: Given the concentration measurements (or outputs), $s_i(\mu, t^k)$, $1 \leq i \leq 4$, at the first four sensor locations, we need to determine the source location (and possibly the diffusivity) of a pollutant dispersing in a fixed flow field $\mathbf{U}$. The description of the problem is detailed in Section 5.6, where we also discussed the reduced-basis approximation.

We recall the input parameter, μ, given by $\mu \equiv \{\mu_1, \mu_2, \mu_3\} \equiv \{\kappa, x_1^{\mathrm{PS}}, x_2^{\mathrm{PS}}\} \in \mathcal{D} \equiv [0.05, 0.5] \times [2.9, 3.1] \times [0.3, 0.5] \subset \mathbb{R}^{P=3}$, and the spatial domain, Ω^{PS}, defined as $\Omega^{\mathrm{PS}} \equiv [2.9, 3.1] \times [0.3, 0.5]$. The time interval of interest is $\bar{I} = [0, 1]$ and the timestep is $\Delta t = 5\,\mathrm{E}{-3}$; we thus have $K = 200$. For notational convenience, we also define $\mathbb{S}$ as the set of outputs used in the parameter estimation procedure, i.e., $\mathbb{S} = \{1, 2, 3, 4\}$ means that measurements at all four outputs are used in the solution of the inverse problem.

7.6.1 Estimation of Source Location

To begin, we shall assume that the diffusivity is known, $\mu_1 = 0.1$, and only the location of the source term, (μ_2, μ_3), has to be found. We present a sample solution of the inverse problem in Figure 7-14. Here, the true parameter is $\mu^*_{2,3} = (2.93, 0.47)$ and the measurement error is $\epsilon_{\rm exp} = 1.0\%$. The measurements are taken at 10 points in time, $\overline{\mathbb{K}}_{\rm exp} = \{10, 20, \ldots, 100\}$, and we assume that we have access to all four outputs, $\mathbb{S} = \{1, 2, 3, 4\}$. We choose $N = 140$ and $M = 40$ for the dimension of the reduced-basis and nonaffine function approximation space, respectively. We first solve (7.8) for $\mu_{\rm IC} = (\mu_{2,\rm IC}, \mu_{3,\rm IC})$; we obtain $\mu_{\rm IC} = (2.927, 0.469)$ after 4 iterations and a total of 8.34 sec. We next construct the possibility region $\mathcal{R}(\mu^*)$: we choose the tolerance $\Delta\mu^{\mathcal{R}^*} = 1\,\mathrm{E}{-}4$ for the binary chop and increment the search angle by $\Delta\beta = 20°$; we thus have a total of $d_\beta = 18$ search directions (we use these values for $\Delta\mu^{\mathcal{R}^*}$ and $\Delta\beta$ throughout this section). Solving for the $\mu_j^{\mathcal{R}^*}$, $1 \le j \le 18$, requires 175 forward solutions and takes a total of 158.9 sec. Given the boundary points, $\mu_j^{\mathcal{R}^*}$, we solve (7.12) for the enclosing ellipse. We note that the ellipse is nicely centered around the unknown parameter value $\mu^*_{2,3}$.

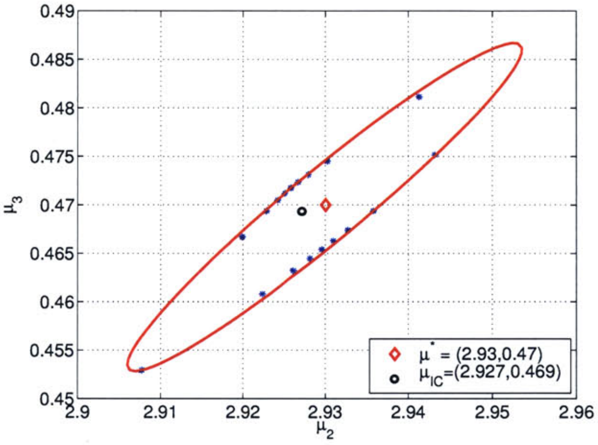
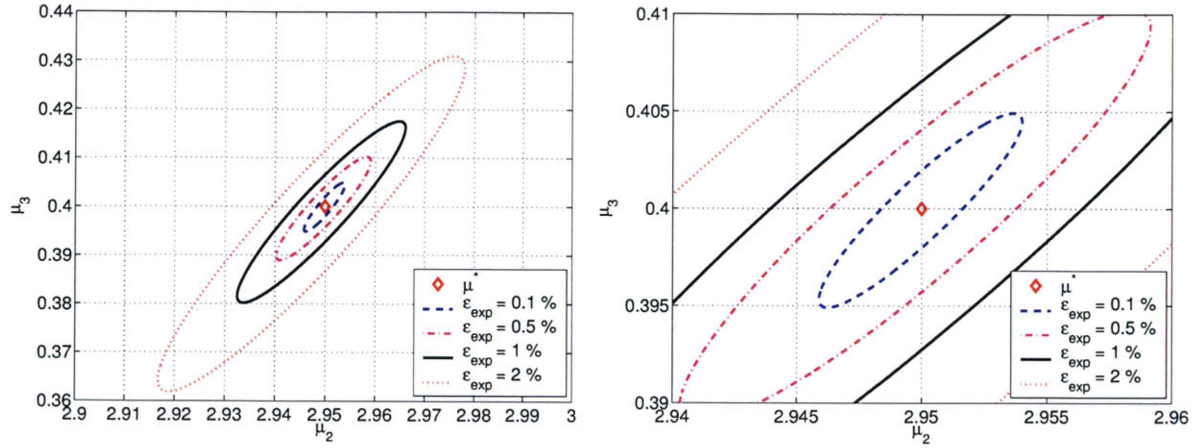


Figure 7-14: AP II: Possibility region $\mathcal{R}(\mu^*)$ for $\mu^* = (2.93, 0.47)$.

We next consider the sensitivity with respect to measurement and modeling errors. We shall assume that the source is located at $\mu^*_{2,3} = (2.95, 0.4)$, the measurements are taken at $\overline{\mathbb{K}}_{\rm exp} = \{10, 20, \ldots, 100\}$, and all four sensors provide data, $\mathbb{S} = \{1, 2, 3, 4\}$. We plot in Figures 7-15 and 7-16 the enclosing ellipses as a function of $\epsilon_{\rm exp}$ (for $N = 140$, $M = 40$ fixed) and as a function of (N, M) (for $\epsilon_{\rm exp} = 0.5\%$ fixed), respectively. We observe, as in the previous examples, that smaller measurement and modeling errors result in a smaller possibility region and thus smaller uncertainty in the parameter estimate. We also note from 7-16 that there is a big difference between the $(N, M) = (100, 40)$ and $(100, 40)$ regions, whereas this difference is less pronounced for larger values of (N, M). We should therefore always ensure that the reduced-basis approximation satisfies a certain maximum acceptable error tolerance.

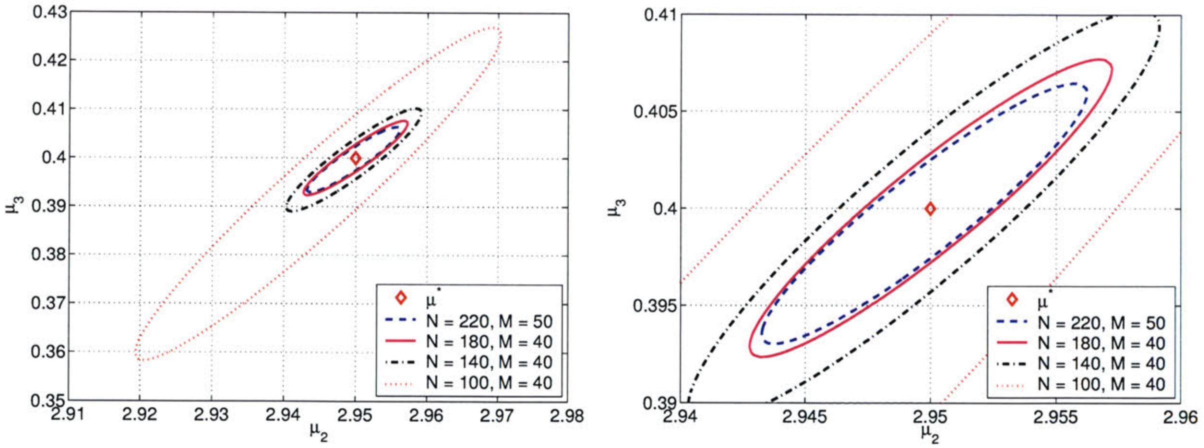
We now turn to the question which sensor measurements are most crucial for obtaining an accurate parameter estimate. We shall assume that the source is located at $\mu^*_{2,3} = (2.95, 0.4)$, that the measurements are taken at $\overline{\mathbb{K}}_{\rm exp} = \{10, 20, \ldots, 100\}$, and that the measurement error is $\epsilon_{\rm exp} = 0.1\%$. We choose $N = 140$ and $M = 40$ for the reduced-basis approximation. We recall



(a)

(b)

Figure 7-15: AP II: Possibility region $\mathcal{R}(\mu^*)$ for $\mu^* = (2.95, 0.4)$ as a function of ϵ_{exp}.



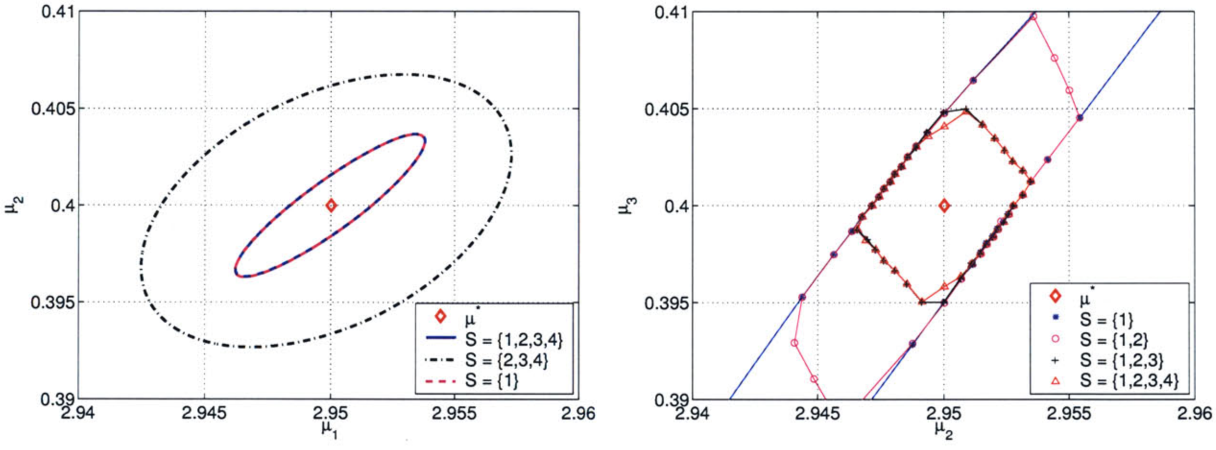
(a)

(b)

Figure 7-16: AP II: Possibility region $\mathcal{R}(\mu^*)$ for $\mu^* = (2.95, 0.4)$ as a function of N and M.

from Figures 5-5 and 5-6 that the dispersion of the pollutant strongly depends on the diffusivity μ_1 — we thus consider two different values for the diffusivity, $\mu_1 = 0.05$ and $\mu_1 = 0.5$. We first present in Figure 7-17(a) the enclosing ellipses for $\mu_1 = 0.05$ for different combinations $\mathbb{S}$. We immediately notice that the first sensor measurement alone is sufficient in estimating $\mu^*_{2,3}$; including the remaining sensor measurements in the inverse solution does not yield a smaller possibility region $\mathcal{R}(\mu^*)$. However, if only the measurement of the first sensor fails, $\mathcal{R}(\mu^*)$ increases considerably. This result could be expected since the flow is clearly convection-dominated for $\mu_1 = 0.05$ and the first sensor lies directly in the track of the pollution cloud. For $\mu_1 = 0.5$ the picture is different: we plot the boundary points, $\mu_j^{\mathcal{R}^*}$, in Figure 7-17(a) for different combinations $\mathbb{S}$.[1] We observe that the measurements from sensor 1 and 3 are now equally important. Furthermore, even sensor 4 contributes very slightly to decreasing the uncertainty. Since the flow is now diffusion dominated, the vicinity of a sensor to the pollution source plays a much more important role than the exact flow field $\mathbf{U}$. We thus note that a find grid of sensors is important for obtaining sharp estimates for all possible source locations and diffusivities.

Concerning the sensitivity of the parameter estimate with respect to additional measurements in time (for a fixed $\mathbb{S}$), we observed that (*i*) taking more observations over the same time period, e.g., $\overline{\mathbb{K}}_{\text{exp}} = \{5, 10, 15, \ldots, 100\}$, does not lead to an improved parameter estimate, and (*ii*) taking observations over a longer time period, i.e., $\overline{\mathbb{K}}_{\text{exp}} = \{10, 20, \ldots, 200\}$, resulted in only a very slight decrease of the possibility region.



(a)

(b)

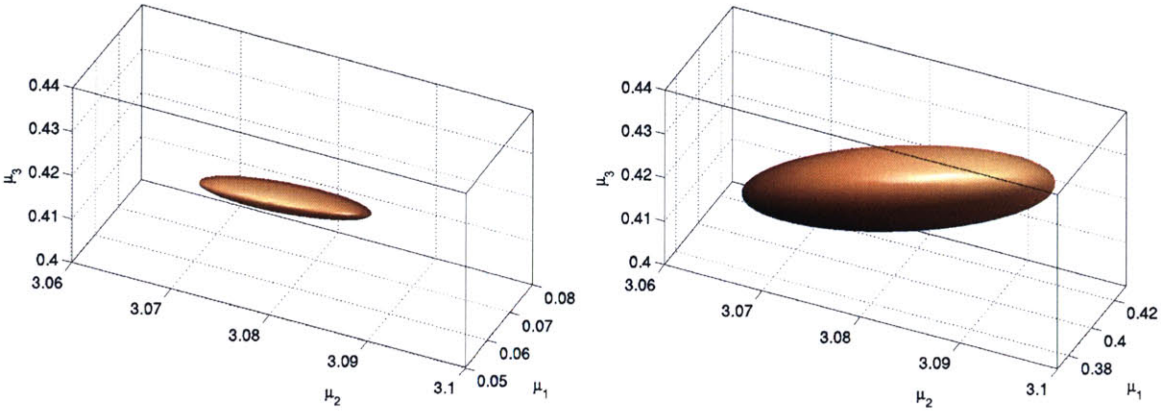
Figure 7-17: AP II: Possibility region $\mathcal{R}(\mu^*)$ for $\mu^*_{2,3} = (2.95, 0.4)$ with (a) $\mu_1 = 0.05$ and (b) $\mu_1 = 0.5$.

We now turn to the problem of simultaneously estimating the diffusivity, μ_1, as well as the source location, (μ_2, μ_3). We shall assume that measurements are taken at $\overline{\mathbb{K}}_{\text{exp}} = \{10, 20, \ldots, 100\}$ and that all four sensors provide data, $\mathbb{S} = \{1, 2, 3, 4\}$. We choose $N = 140$ and $M = 40$ for our reduced-basis approximation.

We first consider a fixed measurement error, $\epsilon_{\text{exp}} = 1.0\%$, and the two "true" parameters $\mu^*_{\text{a}} = (0.06, 3.08, 0.42)$ and $\mu^*_{\text{b}} = (0.4, 3.08, 0.42)$. Note that the location of the source is the same in

[1] We plot the boundary points, $\mu_j^{\mathcal{R}^*}$, here because the slight but important differences are not visible when only considering the enclosing ellipses.

both cases and only the diffusivities are different. Given μ_{a}^*, we generate the noisy measurements and solve (7.8) for the initial center μ_{IC}; we obtain $\mu_{\mathrm{IC}} = (0.0619, 3.078, 0.419)$ after 4 iterations and 11.5 sec. To construct the (now three-dimensional) possibility region, we search along 18 directions in three-dimensional parameter space; we require a total of 169 forward solutions obtained in 148.9 sec. Finally, we solve (7.12) for the enclosing ellipsoid, which is plotted in Figure 7-19(a). We proceed similarly for the true parameter μ_{b}^*: the initial center is $\mu_{\mathrm{IC}} = (0.399, 3.080, 0.420)$ and the corresponding ellipsoid is shown in Figure 7-19(b). We observe that the higher diffusivity results in a considerably larger possibility region $\mathcal{R}(\mu^*)$ (the scaling of the μ_2 and μ_3 axis is the same in both plots). The smaller uncertainty in estimating μ_{a}^* is due to the fact that the features of the concentration field over time are far more distinct for low diffusivities — thus allowing for a sharper estimate.

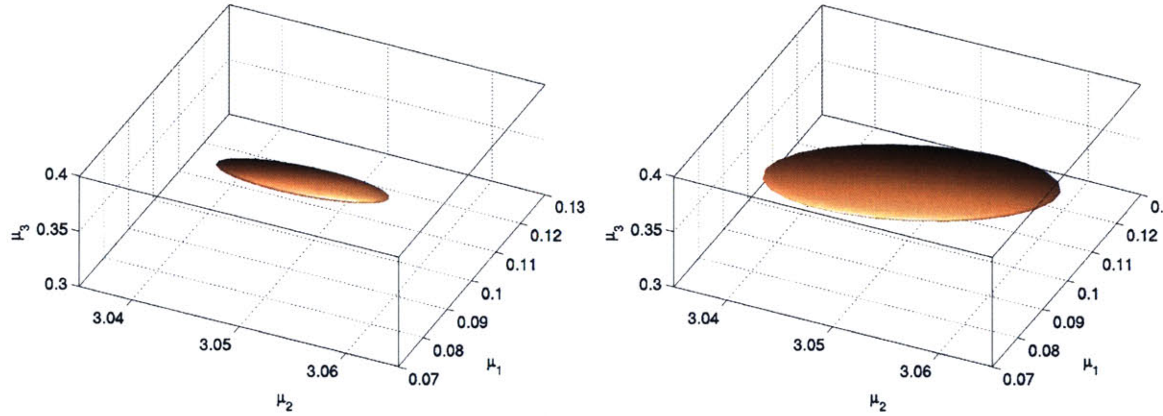


(a) (b)

Figure 7-18: AP II: Possibility region $\mathcal{R}(\mu^*)$ for (a) $\mu_{\mathrm{a}}^* = (0.06, 3.08, 0.42)$ and (b) $\mu_{\mathrm{b}}^* = (0.4, 3.08, 0.42)$.

We shortly return to the sensitivity with respect to the measurement error. Given the true parameter $\mu^* = (0.1, 3.05, 0.35)$, we solve the inverse problem and plot the enclosing ellipsoid for $\epsilon_{\mathrm{exp}} = 0.5\%$ and $\epsilon_{\mathrm{exp}} = 1.0\%$ in Figure 7-19(a) and (b), respectively. As noticed before, the uncertainty in determining the actual source location (μ_2^*, μ_3^*) increases of course. However, the influence of the error on the accuracy of the diffusivity estimate, μ_1, is considerably larger. The maximum possible deviation from the true diffusivity is close to 30%.

Finally, we recall that the allowable range of the source location, Ω^{PS}, is fairly small compared to the whole domain Ω. The reason we cannot handle a source whose location varies significantly over the domain is certainly related to (*i*) the small diffusivities, i.e., convection plays a very important role in the solution; and (*ii*) the very complex flow field, i.e., only a slight difference in the source location can result in very different dispersion patterns. Thus, in the present example the limiting factor is the dimension of the reduced-basis space, N, and not the dimension M of the nonaffine function approximation for the source term. If we were to increase the size of Ω^{PS}, the increase in M will probably be tolerable, whereas the increase in N may be prohibitive. However, there may be other cases, e.g., different flow fields and/or higher diffusivities, where the size of M becomes important.



(a) (b)

Figure 7-19: AP II: Possibility region $\mathcal{R}(\mu^*)$ for $\mu^* = (0.1, 3.05, 0.35)$ with (a) $\epsilon_{\text{exp}} = 0.5\%$ and (b) $\epsilon_{\text{exp}} = 1.0\%$.

Chapter 8

Application to Optimal Control

8.1 Introduction

In Chapter 7 we introduced a robust parameter estimation technique for systems governed by parabolic partial differential equations and successfully applied the proposed method to several numerical examples. The efficient and reliable solution of the partial-differential-equation-induced input-output relationship — afforded through the reduced-basis method and associated *a posteriori* error estimation introduced in earlier chapter — was the basis for our approach. Besides the characterization context there are several other classes of applications which require repeated and often real-time evaluation of input-output relationships. One of these applications, and also the one we consider in this chapter, is optimal control [24].

To begin, we shortly review the optimal control formulation and introduce the reduced-basis approximation into the solution of the problem. As a specific example we then consider the startup control of a welding process. We in fact combine the optimal control formulation with the tools developed in Chapter 7, thus pursuing an integrated estimation-control framework: first, we estimate the unknown system parameters; and second, given the parameter estimates, we solve for the optimal control input to obtain the desired system behavior.

8.2 Optimal Control Problem

Optimal control problems arise in many engineering applications when a known desired behavior is to be imposed on a dynamical system. The problem is usually cast in an optimization setting: a cost functional containing the control and state (or output) histories is set up to quantify the performance and controller trade-off. The optimal control input is then found from minimizing this cost function subject to the governing equations, and possibly constraints on the control and output, being satisfied. A variety of methods have been successfully applied to solve optimal control problems [24]: Riccati equations, shooting methods, control parametrization, sequential quadratic programming, and also general gradient methods such as steepest descent, Newton-Raphson, or conjugate gradient techniques. However, while Riccati equations are limited to only small state dimensions, all of the other techniques usually involve an iterative optimization process.

The solution of optimal control problems thus requires repeated and often real-time evaluation of input-output relationships. If the dynamics are described by partial differential equations, the cost quickly becomes prohibitively large [34, 45, 57, 78], and hence reduced-order models are often employed. Applications of reduced-order models in optimal control range from fluid flow [47, 48, 49, 95], to hyperthermia treatment [65, 66] to thermal processing of semiconductors [82] and canned foods [10].

8.2.1 Formulation

We consider here two formulations for the optimal control problem. We start with the standard Linear-Quadratic-Regulator (LQR) problem. Defining $\underline{u} \equiv [u(t^1)\ u(t^2)\ \ldots\ u(t^K)]^T \in \mathbb{R}^K$, the (discrete-time) quadratic cost functional to be minimized is given by

$$\begin{aligned} J(\underline{u};\mu) = \frac{1}{2}\left(s(\mu,t^K) - s_{\mathrm{d}}(t^K)\right)^T W_T \left(s(\mu,t^K) - s_{\mathrm{d}}(t^K)\right) \\ + \frac{\Delta t}{2}\sum_{k=1}^{K}\left\{\left(s(\mu,t^k) - s_{\mathrm{d}}(t^k)\right)^T W_R \left(s(\mu,t^k) - s_{\mathrm{d}}(t^k)\right) + u(t^k)\, W_U\, u(t^k)\right\}, \end{aligned} \tag{8.1}$$

where $s(\mu,t^K)$ and $s_{\mathrm{d}}(t^K)$ are the actual and desired outputs, $u(t^k)$, $\forall\, k \in \mathbb{K}$ is the control input, and W_T, W_R, and W_U, are symmetric positive (semi-)definite weighting matrices influencing the trade-off between tracking performance and controller cost. The first term penalizes the deviation of the output from the desired output $s_{\mathrm{d}}(t^K)$ at the final time; the second term penalizes the deviation of the output from the desired trajectory $s_{\mathrm{d}}(t^k)$ during the time interval of interest; and the last term reflects the cost of the control action. Note that we use the notation $J(\underline{u};\mu)$ to explicitly signify the dependence of the cost function on the parameter μ through the output $s(\mu,t^k)$.

The optimal control input, $u^*(t^k)$, $\forall\, k \in \mathbb{K}$, is then found by minimizing $J(\underline{u};\mu)$ subject to the initial conditions and the governing equations, (say) (4.3) and (4.4); constraints on the control input itself, such as non-negativity requirements, may also be present. We can thus state the problem as: Given a $\mu \in \mathcal{D}$, $u^*(t^k)$, $\forall\, k \in \mathbb{K}$, is the solution of

$$\min_{\underline{u}\in\mathbb{R}^K}\ J(\underline{u};\mu) \tag{8.2}$$

$$\text{s.t.} \begin{cases} (4.3),\ (4.4)\ \text{(say)} \\ u_{\mathrm{LB}} \le u(t^k) \le u_{\mathrm{UB}}, \quad \forall\, k \in \mathbb{K}. \end{cases}$$

Here, u_{LB} and $\le u_{\mathrm{UB}}$ are the lower and upper bound on the control input, respectively. The LQR formulation results in a set of linear stationarity conditions and is therefore the most common set-up for optimal control problems.

We now consider a different formulation of the optimal control problem that leads to a linear programming problem. Again, our goal is to track a desired output history, $s_{\mathrm{d}}(t^k)$. To this end, we minimize the maximum deviation from the desired output trajectory, i.e., $\max_{k\in K}|s(\mu,t^k) - s_{\mathrm{d}}(\mu,t^k)|$, subject to the governing equations, and possibly constraints on the control input. This problem can be written as: Given a $\mu \in \mathcal{D}$, $u^*(t^k)$, $\forall\, k \in \mathbb{K}$, is the solution of

$$\min_{\substack{\gamma \in \mathbb{R} \\ \underline{u} \in \mathbb{R}^K}} \quad \gamma \tag{8.3}$$

$$\text{s.t.} \begin{cases} |s(\mu, t^k) - s_{\rm d}(\mu, t^k)| \le \gamma, & \forall\, k \in \mathbb{K} \\ (4.3),\ (4.4)\ (\text{say}) \\ u_{\rm LB} \le u(t^k) \le u_{\rm UB}, & \forall\, k \in \mathbb{K}. \end{cases}$$

We recover the linear programming formulation by simply replacing $|s(\mu, t^k) - s_{\rm d}(\mu, t^k)| \le \gamma$ with the two constraints $s(\mu, t^k) - s_{\rm d}(\mu, t^k) \le \gamma$ and $-s(\mu, t^k) + s_{\rm d}(\mu, t^k) \le \gamma$. Note that we do not penalize the total control energy spent in this formulation. However, we may additionally include a term of the form $\sum_{k=1}^{K} |u(t^k)|$ into the cost function, which can again be reformulated as a linear programming problem [21].

So far we have assumed that the input parameter μ is (exactly) known before we attempt to solve the optimal control problem. Unfortunately, this may not always be the case. In Chapter 7 we introduced a robust parameter estimation procedure which accounts for measurement and modeling errors in the form of a possibility region $\mathcal{R}(\mu)$ — $\mathcal{R}(\mu)$ contains all (in the probabilistic sense) parameters which satisfy the constraints of the problem. In such a case we *do not* know μ exactly, but only that $\mu \in \mathcal{R}(\mu)$. We may explicitly introduce this uncertainty into the optimal control problem and pursue the min-max formulation $\underline{u}^*(t) = \arg\min_{\underline{u} \in \mathbb{R}^K} \max_{\mu \in \mathcal{R}(\mu)} J(\underline{u}; \mu)$ subject to the constraints stated in (8.2). We can also follow a similar approach for the linear programming problem (8.3). There may, of course, be an issue concerning computational feasibility: we saw in the last chapter that we can solve (7.8) for the initial center, $\mu_{\rm IC}$, very fast. However, constructing the possibility region $\mathcal{R}(\mu)$ requires more effort. Depending on the problem at hand, solving (8.2) using the estimated parameter value may be sufficient and the only choice concerning computational efficiency. If the robust solution is crucial and the computation of $\mathcal{R}(\mu)$ is fast enough, on the other hand, following the robust approach is preferable.

Finally, we note that a design exercise may result in a similar problem formulation. The parameters are now design variables that can be chosen by the user, e.g., the location of controllers or sensors, and our goal is to minimize $J(\underline{u}; \mu)$ over all $\mu \in \mathcal{D}$, where $\mathcal{D}$ is the admissible design space.

8.2.2 Optimization Procedure

We state the optimal control problems (8.2) and (8.3) in the last section in terms of the truth approximation $y(\mu, t^k)$ and $s(\mu, t^k)$. Although solution techniques tailored to partial-differential-equation-constrained optimal control problems have been developed [78], they are computationally expensive and thus real-time performance is hard to achieve. We therefore employ the reduced-basis method in the statement and solution of the problem and replace the truth approximation $y(\mu, t^k)$ (and $s(\mu, t^k)$) by their reduced-basis approximations $y_N(\mu, t^k)$ (and $s_N(\mu, t^k)$) in (8.1), (8.2) and (8.3); e.g., we have

$$\begin{aligned} J_N(\underline{u};\mu) = \frac{1}{2}\left(s_N(\mu,t^K) - s_{\mathrm{d}}(t^K)\right)^T W_T \left(s_N(\mu,t^K) - s_{\mathrm{d}}(t^K)\right) \\ + \frac{\Delta t}{2}\sum_{k=1}^{K}\left\{\left(s_N(\mu,t^k) - s_{\mathrm{d}}(t^k)\right)^T W_R \left(s_N(\mu,t^k) - s_{\mathrm{d}}(t^k)\right) + u(t^k)\, W_U\, u(t^k)\right\}. \end{aligned} \quad (8.4)$$

Since we do not aim at developing new techniques for solving optimal control problems in this thesis, we pursue a straightforward "impulse approach" for their solution similar to the one in [17]. We already remarked in Section 4.2.3 that the solution of any LTI system can be written as the convolution of the impulse response with the control input. This property is also the basis and justification for constructing the reduced-basis approximation with an impulse approach: it is sufficient that the reduced-basis subspace approximates well the (parameter-dependent) impulse response to obtain good approximation properties for all possible control input histories. As in (4.14), the output of any LTI system can be written as

$$s(\mu,t^k) = \sum_{j=1}^{k} g^s(\mu,t^{k-j+1})\, u(t^j), \quad \forall\, k \in \mathbb{K}, \quad (8.5)$$

where $g^s(\mu,t^k)$ is the output for a unit impulse control input $u(t^k) = \delta_{1k}$, $\forall\, k \in \mathbb{K}$. Similarly, we obtain $s_N(\mu,t^k) = \sum_{j=1}^{k} g_N^s(\mu,t^{k-j+1})\, u(t^j)$, $\forall\, k \in \mathbb{K}$, where $g_N^s(\mu,t^k)$ is the solution of (4.17) and (4.110) for $u(t^k) = \delta_{1k}$, $\forall\, k \in \mathbb{K}$. Defining $\underline{s}_N(\mu) = [s_N(\mu,t^1)\ s_N(\mu,t^2)\dots s_N(\mu,t^K)]^T$ and $\underline{u} = [u(t^1)\ u(t^2)\dots u(t^K)]$ we can write the input-output relationship as

$$\underline{s}_N(\mu) = \underline{G}_N(\mu)\ \underline{u}, \quad (8.6)$$

where the matrix $\underline{G}_N(\mu) \in \mathrm{I\!R}^{K\times K}$ is lower-triangular and contains the impulse response $g_N^s(\mu,t^k)$, $\forall k \in \mathbb{K}$. Note that $\underline{G}(\mu)$ depends on μ and thus has to be evaluated online for every new parameter value μ.

It directly follows from (8.6) and (8.4) that $J_N(u;\mu)$ can be written as

$$J_N(\underline{u};\mu) = \frac{1}{2}\underline{u}^T\, \underline{H}(\mu)\, \underline{u} + b^T(\mu)\, \underline{u} + c(\mu). \quad (8.7)$$

Here, $\underline{H}(\mu) \in \mathrm{I\!R}^{K\times K}$ is a parameter-dependent matrix containing the impulse response, $b(\mu) \in \mathrm{I\!R}^K$ is a parameter-dependent vector containing the impulse response and information regarding the desired trajectory, and $c(\mu)$ is a scalar containing only information regarding the desired trajectory. We then solve the optimization problem

$$\begin{aligned} \min_{\underline{u}\in\mathrm{I\!R}^K}\ & \frac{1}{2}\underline{u}^T\, \underline{H}(\mu)\, \underline{u} + b^T(\mu)\, \underline{u} + c(\mu) \\ \text{s.t.}\quad & u_{\mathrm{LB}} \le u(t^k) \le u_{\mathrm{UB}}, \quad \forall\, k \in \mathbb{K}, \end{aligned} \quad (8.8)$$

for the optimal control input $u^*(t^k)$, $\forall\, k \in \mathbb{K}$.

Similarly, invoking (8.6) we can write the linear programming problem as

$$\min_{\substack{\gamma \in \mathbb{R} \\ \underline{u} \in \mathbb{R}^K}} \quad \gamma \tag{8.9}$$

$$\text{s.t.} \begin{cases} \underline{G}\, \underline{u} - \underline{s}_{\mathrm{d}} \le \gamma\, e \\ -\underline{G}\, \underline{u} + \underline{s}_{\mathrm{d}} \le \gamma\, e \\ u_{\mathrm{LB}} \le u(t^k) \le u_{\mathrm{UB}}, \quad \forall\, k \in \mathbb{K}. \end{cases}$$

where $\underline{s}_{\mathrm{d}} = [s_{\mathrm{d}}(t^1)\ s_{\mathrm{d}}(t^2) \ldots s_{\mathrm{d}}(t^K)]^T \in \mathbb{R}^K$ and $e = [1\ 1 \ldots 1]^T \in \mathbb{R}^K$. We note, however, that using the impulse approach (8.6) to recast the optimal control problem is only efficient for the linear programming problem (8.9); in the LQR case the computational cost to form $\underline{H}(\mu)$ is $O(K^3/3)$ and is thus only efficient for K very small.

We shortly remark on the possibility to extend the previous discussion and consider Model Predictive Control (MPC), also referred to as receding horizon control or moving horizon control [3, 67, 70, 92, 96]. In MPC, the optimal control problem defined over a very long (infinite) time period, is essentially split into a series of short term (finite) horizon optimal control problems. The optimal control input or feedback law is *not* computed *once offline* for the entire time period, instead a series finite horizon optimal control problems is solved *repeatedly online* for the consecutive time periods using the current state of the plant as the initial condition. In contrast to an open-loop optimal control law, MPC can therefore react to perturbations in the system parameters, in effect "closing the loop" of the control implementation. However, for MPC to be applicable, the plant dynamics have to be sufficiently "slow" compared to the time required to solve the optimal control problem so as to permit the implementation. Reduced-order models thus lend themselves ideally to Model Predictive Control.

8.3 AP IV: Control of Welding Quality

We now turn to the welding application introduced in Section 1.1.1. The equation governing the temperature distribution in the joint-section is the unsteady convection-diffusion equation (1.19) with initial condition (1.20). The heat input from the welding torch is modelled as a Gaussian distribution centered at the torch position $x^{\mathrm{T}} \equiv (3.5, 1)$, given by

$$q_{\mathrm{w}}(x;\mu) = \frac{\eta_{\mathrm{w}}}{2\pi\sigma_{\mathrm{w}}^2}\, e^{-((x_1 - x_1^{\mathrm{T}})^2 + (x_2 - x_2^{\mathrm{T}})^2)/(2\sigma_{\mathrm{w}}^2)}, \tag{8.10}$$

where η_{w} is the efficiency and σ_{w} is the distribution parameter. We use the notation $q_{\mathrm{w}}(\cdot;\mu)$ to signify the dependence on the parameter $\mu \equiv (\mu_1, \mu_2) \equiv (\eta_{\mathrm{w}}, \sigma_{\mathrm{w}}) \in \mathcal{D} \subset \mathbb{R}^{P=2}$, where $\mathcal{D} = [0.1, 0.4] \times [0.15, 0.65]$ [112].

We shall make the following two assumptions. First, we assume that we are interested in achieving a fixed desired weld pool depth $d_{\mathrm{w,d}} = 0.5$. In [111, 112], the temperature distribution is searched for the isotherm corresponding to the melting temperature to deduce the pool depth d_{w}. However, searching the isotherms requires knowledge of the truth approximation state $y(\mu, t^k)$, $\forall k \in \mathbb{K}$ and the computational cost thus scales with the dimension $\mathcal{N}$ of the truth approximation. The basic premise for the computational efficiency of the reduced-basis method is the $\mathcal{N}$-independent computational complexity in the online stage — we should therefore avoid the isotherm search. To this end, we introduce a "fictitious" output, s_3, at the desired weld pool depth, $d_{\mathrm{w,d}}$, measuring the average temperature over a small domain. We then simply require this temperature output to be equal to the melting temperature, i.e., $s_3(\mu, t^k) = s_3^* \equiv 1,\ \forall\, k \in \mathbb{K}$. We then guarantee that the melting isotherm reaches the desired depth. In Figure 8-1 we show a sketch of the joint-section with the torch position and the three outputs, the two measurements s_1 and s_2 on the bottom of the workpiece and the additional output s_3. Note that we place the output s_3 downwind of the welding torch because of the convective term. We also note that we may consider several different desired weld pool depths by simply introducing a "fictitious" output at each desired depth level.

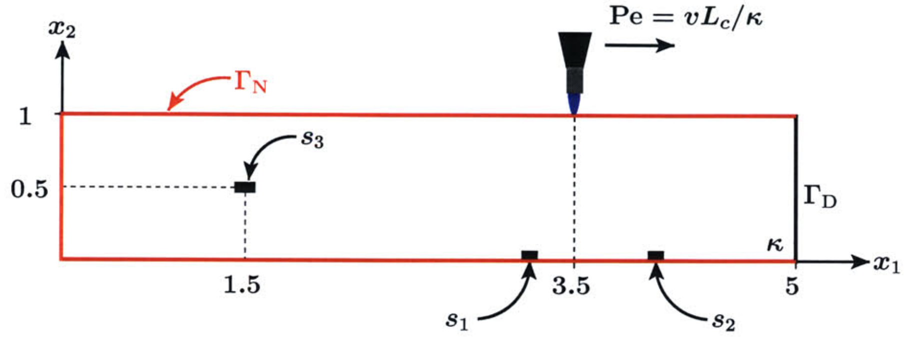


Figure 8-1: AP IV: Control of welding quality.

Our second assumption is related to the control input. In actual practice, the welding process can be controlled through the velocity of the torch, Pe, and the total heat input, $u(t)$. Here, we consider only the total heat input, $u(t)$, as the single control input and assume that the velocity is fixed, Pe $= 6$ [1]. We restrict our attention to this case because controlling the velocity would result in a nonlinear — or more precisely, bilinear — control law and our impulse approach to construct the reduced-basis approximation would not be valid.

The domain Ω, a typical point in which is (x_1, x_2), is given by $\Omega \equiv [0,5] \times [0,1]$. We shall assume that the temperature is equal to ambient temperature on Γ_{D}, and that the remaining boundaries, Γ_{N}, are insulated. The time-discrete weak form of the governing equation (1.19) for the temperature $T(\mu, t^k) \in Y$ is (B.2) (we use Crank-Nicolson for the time integration) with initial condition $T(\mu, t^0) = 0$, where $Y \subset Y^{\mathrm{e}} \equiv \{v \,|\, v \in H^1(\Omega), v = 0|_{\Gamma_{\mathrm{D}}}\}$ is a linear finite element truth approximation subspace of dimension $\mathcal{N} = 3720$. The bilinear and linear forms are given by $m(w,v) \equiv \int_\Omega w\, v$, $a^{\mathrm{CD}}(w,v) \equiv \int_\Omega \nabla w \cdot \nabla v + \int_\Omega v\, (\mathbf{U} \cdot \nabla w) + \frac{1}{2} \int_\Omega v\, w\ (\nabla \cdot \mathbf{U})$, and $b(v; q_{\mathrm{w}}(x;\mu)) \equiv \int_\Omega q_{\mathrm{w}}(x;\mu)\, v$, where $\mathbf{U} = [\mathrm{Pe}\ 0]^T$ and $q_{\mathrm{w}}(x;\mu)$ is defined in (8.10). We note that a and m do not depend on the parameter, and that the parameter dependence of $b(v; g(x;\mu))$ is *nonaffine* — we hence require the theory developed in Chapter 5. We also define the inner products $(w,v)_X \equiv m(w,v)$ and $(w,v)_Y \equiv \int_\Omega \nabla w \cdot \nabla v$; we may thus choose $\hat{\alpha}_a = 1$. The outputs $s_q(\mu, t^k)$, $1 \le q \le 3$, are given by $s_q(\mu, t^k) - |\Omega_{s_q}|^{-1} \int_{\Omega_{s_q}} v$, $1 \le q \le 3$, where $\Omega_{s_1} = [3.16, 3.29] \times [0, 0.07]$, $\Omega_{s_2} = [4, 4.12] \times [0, 0.07]$, and $\Omega_{s_3} = [1.42, 1.58] \times [0.47, 0.53]$. We shall consider the time interval $\bar{I} = [0,2]$ [2] and a timestep $\Delta t = 2\,\mathrm{E}{-}2$; we thus have $K = 100$.

[1] We could, of course, consider Pe as a parameter and construct a reduced-basis approximation for a certain velocity range. The desired velocity can then be chosen at the beginning of the welding process.

[2] Note that the time is also non-dimensionalized here. The time interval $\bar{I} = [0,2]$ corresponds to a "real" time interval from 0 to 10 sec.

We plot in Figure 8-2 the temperature distribution $T(\mu, t^k)$ and corresponding isotherms for $\mu = (0.3, 0.4)$ at three different discrete timesteps for $u(t^k) = 10$, $\forall k \in \mathbb{K}$. Because of the convective term the isotherms are shifted to the left from the torch position. The shape of the isotherms also justifies our choice for the location of the output s_3.

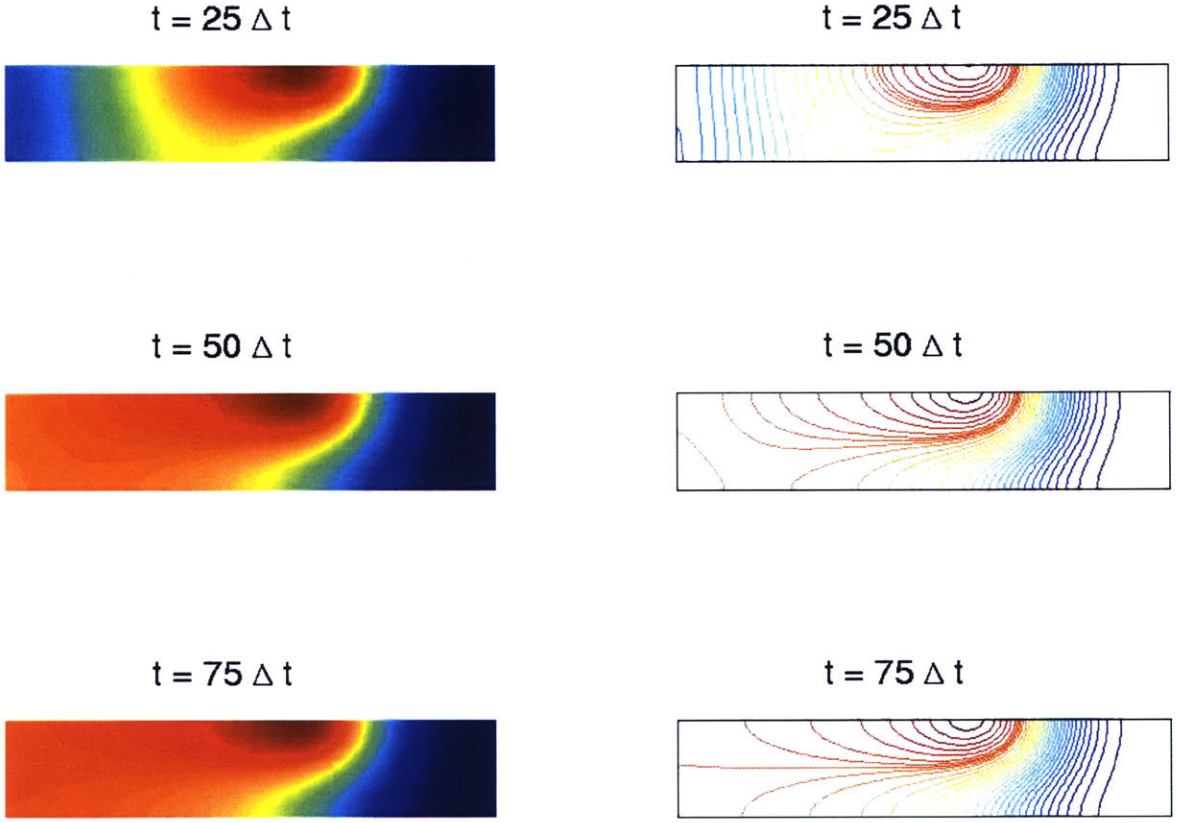


Figure 8-2: AP IV: Temperature distribution $T(\mu, t^k)$ and isotherms for $\mu = (0.3, 0.4)$ at $t = t^{25}$, t^{50}, and t^{75}.

8.3.1 Reduced-Basis Approximation

We first consider the approximation to the nonaffine function $q_{\rm w}(x; \mu)$ defined in (8.10). We note that the nonaffine parameter dependence of $q_{\rm w}(x; \mu)$ is only due to μ_2. We could thus include only μ_2 in the definition of $q_{\rm w}$ and consider μ_1 separately. Nevertheless, here we chose to define $q_{\rm w}$ to be a function of both parameters.[3] We choose for Ξ^g a deterministic grid of 41×41 parameter points over $\mathcal{D}$ and we choose $(\mu_1^g, \mu_2^g) = (0.4, 0.15)$. Next, we pursue the empirical interpolation method of Section 2.4 (using the $L^\infty(\Omega)$-norm) to construct S_M^g, W_M^g, T_M, and B^M, $1 \leq M \leq M_{\rm max}$, for $M_{\rm max} = 17$.

We next generate the sample set S_N^y and associated reduced basis space W_N^y according to the adaptive sampling procedure described in Section 4.5 with $M = M_{\rm max}$ for the nonaffine function

[3] Note that we could even treat μ_1 implicitly by including it in the control input $u(t)$.

approximation. We construct the reduced-basis approximation here using the impulse approach — since the system is LTI, it is sufficient that the reduced-basis subspace approximates well the parameter-dependent impulse response. This fact is crucial for applying our method to optimal control problems, because the optimal control input is not known in advance. We initialize the procedure with $\mu_1^y = (0.1, 0.4)$ and $t^{k_1^y} = 1\Delta t$ and set the desired error tolerance (for the relative error in the energy norm) to $\epsilon_{\text{tol,min}} = 1\,\text{E}-4$. We sample on a parameter test sample $\Xi_{\text{F}} \in (\mathcal{D})^{400}$ of size 400 (a regular 20×20 grid); we require $N_{\max} = 54$ basis functions to obtain the desired accuracy.

We note that we do not pursue the primal-dual formulation here. The numerical results for the output errors and output bounds are obtained using the simple bound defined in Proposition 16.

To begin, we present convergence results for the nonaffine function approximation. We present in Table 8.1 $\varepsilon^*_{M,\max}$, $\overline{\rho}_M$, Λ_M, $\overline{\eta}_M$, and $\varkappa_M$ as a function of M (see Section 2.4.3 for the definitions of these quantities; here Ξ^g_{Test} is a test sample of size 225). We observe that the maximum error $\varepsilon^*_{M,\max}$ converges very rapidly with M; that the Lebesgue constant provides a reasonably sharp measure of the interpolation-induced error; that the error estimator effectivity is reasonably close to unity (recall that $\hat{\varepsilon}_M(\mu) \le \varepsilon_M(\mu)$, $1 \le M \le M_{\max}$). The reason that $M_{\max}$ is fairly small here is because $q_{\text{w}}(x;\mu)$ is only nonaffine in μ_2 as mentioned previously.

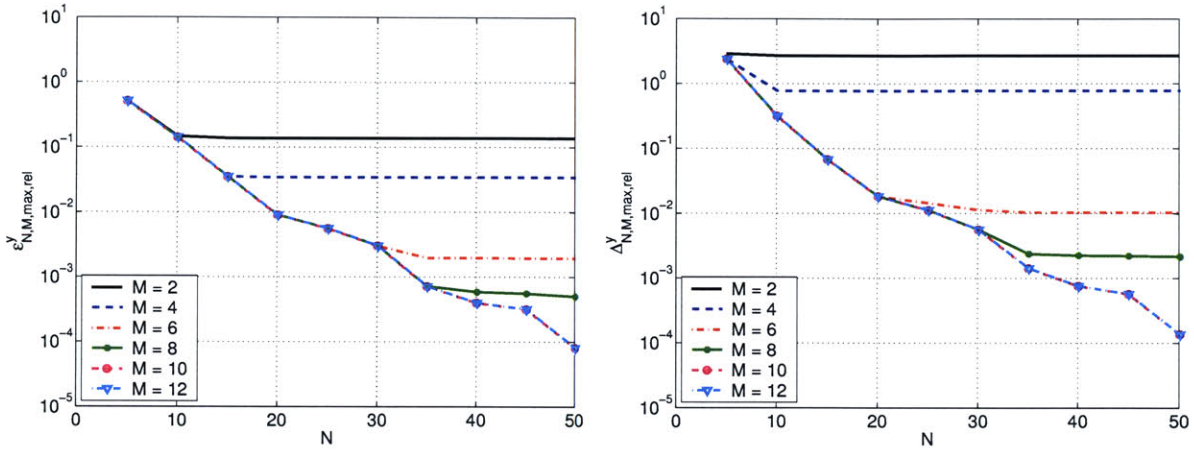
M	$\varepsilon^*_{M,\max}$	$\overline{\rho}_M$	Λ_M	$\overline{\eta}_M$	$\varkappa_M$
2	8.34 E-02	0.57	1.12	0.86	1.22
4	1.32 E-02	0.48	2.61	0.79	2.41
6	5.29 E-04	0.51	3.57	0.26	3.27
8	1.10 E-04	0.56	2.86	0.73	4.92
10	6.24 E-06	0.51	5.63	0.87	6.00
12	3.53 E-07	0.36	3.62	0.58	7.17
14	3.81 E-08	0.32	5.98	0.50	8.36
16	2.72 E-08	0.22	6.92	0.32	8.57

Table 8.1: AP IV: $\varepsilon^*_{M,\max}$, $\overline{\rho}_M$, Λ_M, $\overline{\eta}_M$, and $\varkappa_M$ as a function of M.

We now turn to the convergence results and error bounds for the reduced-basis approximation. In Figure 8-3(a) and (b) we plot, as a function of N and M, the maximum relative error in the energy norm $\epsilon^y_{N,M,\max,\text{rel}}$ and the maximum relative error bound $\Delta^y_{N,M,\max,\text{rel}}$; here, $\epsilon^y_{N,M,\max,\text{rel}}$ is the maximum over Ξ_{Test} of $|||e(\mu, t^K)|||/|||y(\mu_y, t^K)|||$ and $\Delta^y_{N,M,\max,\text{rel}}$ is the maximum over Ξ_{Test} of $\Delta^y_{N,M}(\mu, t^K)/|||y(\mu_y, t^K)|||$, where $\Xi_{\text{Test}} \in (\mathcal{D})^{225}$ is a an input sample of size 225 (a 15×15 random grid), and $\mu_y \equiv \arg\max_{\mu \in \Xi_{\text{Test}}} |||y(\mu, t^K)|||$. We observe the typical behavior in the error and error bound convergence curves. Also, the separation points of the asymptotes are different for $\epsilon^y_{N,M,\max,\text{rel}}$ and $\Delta^y_{N,M,\max,\text{rel}}$; to obtain the best possible error bounds we should base our choice on the $\Delta^y_{N,M,\max,\text{rel}}$ curves.

In Table 8.2 we present, as a function of N and M, $\epsilon^y_{N,M,\max,\text{rel}}$, $\Delta^y_{N,M,\max,\text{rel}}$, and the average effectivity $\overline{\eta}^y$, where $\overline{\eta}^y$ is the average over $\Xi_{\text{Test}} \times \mathbb{I}$ of $\Delta^y_{N,M}(\mu, t^k)/|||y(\mu, t^k) - y_N(\mu, t^k)|||$. Here, we select the (N, M) combinations from Figure 8-3 which roughly correspond to the separation points. We note that the effectivities for the bound of the energy norm error are very good.

We next present in Tables 8.3, 8.4, and 8.5 the maximum relative output error $\epsilon^s_{N,M,\max,\text{rel}}$,



(a)

(b)

Figure 8-3: AP IV: (a) Maximum relative error in the energy norm and (b) error bound.

N	M	$\epsilon^y_{N,M,\max,\mathrm{rel}}$	$\Delta^y_{N,M,\max,\mathrm{rel}}$	$\overline{\eta}^y$
10	4	1.42 E − 01	7.82 E − 01	6.68
20	6	9.05 E − 03	1.83 E − 02	2.00
30	8	3.02 E − 03	5.60 E − 03	1.96
40	10	3.97 E − 04	7.55 E − 04	1.85
50	10	8.01 E − 05	1.35 E − 04	1.67

Table 8.2: AP IV: Convergence rate and effectivities as a function of N and M.

the maximum relative output bound $\Delta^s_{N,M,\max,\text{rel}}$, and the average effectivity $\overline{\eta}^s_{N,M}$ as a function of N and M for outputs 1, 2, and 3, respectively. Here, $\epsilon^s_{N,M,\max,\text{rel}}$ is the maximum over Ξ_{Test} of $|s(\mu, t_\eta(\mu)) - s_N(\mu, t_\eta(\mu))|/s_{\max}$, $\Delta^s_{N,M,\max,\text{rel}}$ is the maximum over Ξ_{Test} of $\Delta^s_{N,M}(\mu, t^K)/|s_{\max}|$ and $\overline{\eta}^s$ is the average over Ξ_{Test} of $\Delta^s_{N,M}(\mu, t_\eta(\mu))/|s(\mu, t_\eta(\mu)) - s_{N,M}(\mu, t_\eta(\mu))|$, where $t_\eta(\mu) \equiv \arg\max_{t^k \in \mathbb{I}} |s(\mu, t^k) - s_N(\mu, t^k)|$ and $s_{\max} \equiv \max_{t^k \in \mathbb{I}} \max_{\mu \in \Xi_{\text{Test}}} |s(\mu, t^k)|$. The error in all three outputs converges approximately at the same rate. The magnitude of the output effectivities is similar to the previous numerical examples and is still acceptable for the simple output bound. We require $N = 45 - 50$ and $M = 10$ to obtain an accuracy in the output bounds of approximately 1%.

N	M	$\epsilon^s_{N,M,\max,\text{rel}}$	$\Delta^s_{N,M,\max,\text{rel}}$	$\overline{\eta}^s_{N,M}$
10	4	2.82 E−01	3.15 E+01	102
20	6	1.72 E−02	7.39 E−01	58.3
30	8	2.63 E−03	2.26 E−01	58.4
40	10	2.57 E−04	3.04 E−02	104
50	10	1.07 E−04	5.44 E−03	59.3

Table 8.3: AP IV: Maximum relative output error, output bound, and effectivities for output 1.

N	M	$\epsilon^s_{N,M,\max,\text{rel}}$	$\Delta^s_{N,M,\max,\text{rel}}$	$\overline{\eta}^s_{N,M}$
10	4	4.00 E−01	6.53 E+01	126
20	6	2.68 E−02	1.53 E+00	41.3
30	8	7.80 E−03	4.68 E−01	56.0
40	10	5.24 E−04	6.31 E−02	39.5
50	10	2.23 E−04	1.13 E−02	27.9

Table 8.4: AP IV: Maximum relative output error, output bound, and effectivities for output 2.

N	M	$\epsilon^s_{N,M,\max,\text{rel}}$	$\Delta^s_{N,M,\max,\text{rel}}$	$\overline{\eta}^s_{N,M}$
10	4	1.01 E−01	2.12 E+01	165
20	6	6.18 E−03	4.96 E−01	51.0
30	8	1.37 E−03	1.52 E−01	103
40	10	3.48 E−04	2.04 E−02	130
50	10	4.88 E−05	3.65 E−03	79.9

Table 8.5: AP IV: Maximum relative output error, output bound, and effectivities for output 3.

Finally, in Table 8.6 we present, as a function of N and M, the online computational times to calculate $s_{N,M}(\mu, t^k)$ and $\Delta^s_{N,M}(\mu, t^k)$, $\forall\, k \in \mathbb{K}$. The values are normalized with respect to the computational time for the direct calculation of the truth approximation output $s(\mu, t^k) = \ell(y(\mu, t^k))$, $\forall\, k \in \mathbb{K}$. The computational savings for $N = 50$ and $M = 10$ are approximately a factor of 280. Although we do not employ the primal-dual formulation here, the computational savings are considerable because of the very fast convergence — and hence fairly small N — of our reduced-basis approximation.

N	M	$s_{N,M}(\mu,t^k),\ \forall\, k \in \mathbb{K}$	$\Delta^s_{N,M}(\mu,t^k),\ \forall\, k \in \mathbb{K}$	$s(\mu,t^k),\ \forall\, k \in \mathbb{K}$
10	4	3.81 E-04	1.44 E-03	1
20	6	5.72 E-04	1.56 E-03	1
30	8	7.70 E-04	1.72 E-03	1
40	10	1.06 E-03	2.04 E-03	1
50	10	1.43 E-03	2.16 E-03	1

Table 8.6: AP IV: Online computational times (normalized with respect to the time to solve for $s(\mu, t^k),\ \forall\, k \in \mathbb{K}$).

8.3.2 Estimation of Weld Pool Depth

Our goal is the in-process control of the weld pool depth d_{w} to the desired value $d_{\mathrm{w,d}}$. However, the pool depth depends on the unknown parameter μ through the temperature distribution in the joint-section. We thus have to estimate the parameter μ first before we can proceed and control d_{w}. To this end, we split the estimation-control process into the following series of problems solved in consecutive time intervals:

1. For $t \in \mathbb{I}_1 = \{t^0, \ldots, t^{k_1}\}$: The welding process is started with a nominal control input $u_{\mathrm{n}}(t^k)$ and temperature measurements, $z_1(t^k)$ and $z_2(t^k)$, are taken at several discrete points in time at the two measurement points at the bottom of the plate.

2. For $t \in \mathbb{I}_2 = \{t^{k_1}, \ldots, t^{k_2}\}$: Given the measured temperatures, we solve the inverse problem for the parameter estimate μ_{IC} (and the possibility region $\mathcal{R}(\mu^*)$).

3. For $t \in \mathbb{I}_3 = \{t^{k_2}, \ldots, t^{k_3}\}$: Given the parameter estimate μ_{IC}, we solve the optimal control problem for the time interval $\mathbb{I}_4 = \{t^{k_3}, \ldots, t^{k_4}\}$ with the estimated output $s_3(\mu_{\mathrm{IC}}, t_3)$ as initial condition.

4. For $t \in \mathbb{I}_4 = \{t^{k_3}, \ldots, t^{k_4}\}$: We apply the optimal control input $u^*(t^k)$ to the welding process.

We note that the finite element "truth" approximation with a specific true parameter value μ^* serves as our "real-world" welding process. We obtain the temperature measurements $z_1(t^k)$ and $z_2(t^k)$ in step 1 by adding noise to the truth approximation outputs $s_1(\mu^*, t^k)$ and $s_2(\mu^*, t^k)$. In step 4 we apply the optimal control input $u^*(t^k)$ to the truth approximation and check whether we obtain the desired temperature, $s_3^* = 1$, in the output $s_3(\mu^*, t^k)$.

We start with step 1 and 2: we shall assume that the true parameter value is given by $\mu^* = (0.34, 0.46)$ and that temperature is measured for $\mathbb{I}_1 = \{t^0, \ldots, t^{10}\}$ at the discrete timesteps $\overline{\mathbb{K}} = \{2, 4, 6, 8, 10\}$ (this corresponds to a sampling frequency of approximately 5 Hz). Based on our discussion in the last section we choose $N = 50$ and $M = 10$ for the reduced-basis approximation.

We present a sample solution of the parameter estimation procedure in Figure 8-4. We assume that $\epsilon_{\mathrm{exp}} = 1.0\%$ and solve (7.8) for μ_{IC}: we obtain $\mu_{\mathrm{IC}} = (0.339, 0.462)$ after 3 iterations in 1.29 sec [4]. We need 233 forward solutions and a total of 35.0 sec. to generate the boundary points $\mu^{\mathcal{R}^*}$ (here, $\Delta\beta = 20°$ and $\Delta\mu^{\mathcal{R}} = 1\,\mathrm{E}{-}5$). Finally, we solve (7.12) for the enclosing ellipse.

[4] All timing results presented are obtained on an Intel 750 MHz Pentium III processor running MATLAB 6.5.

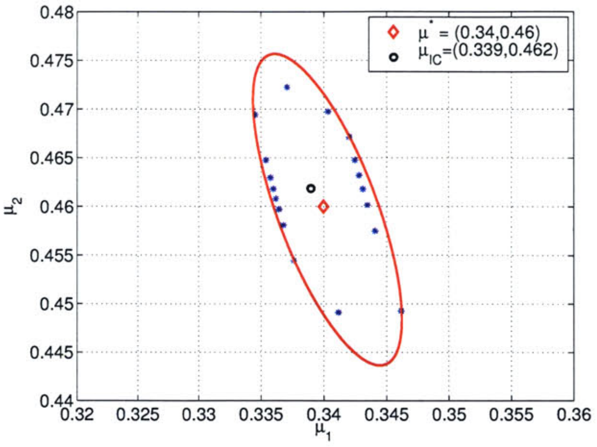


Figure 8-4: AP IV: Possibility region $\mathcal{R}(\mu)$ for $\mu^* = (0.34, 0.46)$.

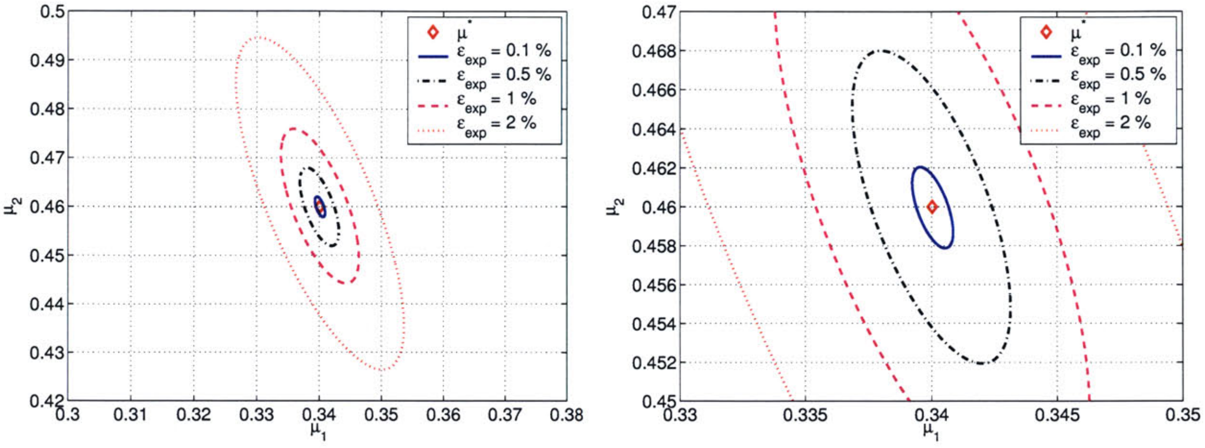
We also investigate the sensitivity of the parameter estimate with respect to measurement and modeling errors. We present in Figures 8-5 and 8-6 the enclosing ellipses as a function of ϵ_{exp} (for $N = 50$ and $M = 10$) and as a function of N and M (for $\epsilon_{\text{exp}} = 0.5\%$), respectively. Again, smaller measurement and modeling errors result in a smaller possibility region $\mathcal{R}(\mu^*)$. We observe that we could decrease N to 40 without incurring a severe loss in the accuracy of the parameter estimate.

In actual practice, we should not wait until all measurements are available before we start the parameter estimation process. In fact, the time intervals $\mathbb{I}_1$ and $\mathbb{I}_2$ may overlap: since solving (7.8) is an iterative process, we can continuously add measurements into the solution process as they become available. We test this approach by starting the iterative solution of (7.8) with only the first two measurements, $\overline{\mathbb{K}} = \{2, 4\}$. After each iteration of the Levenberg-Marquardt Algorithm we include the next measurement into the solution process. We obtain $\mu_{\text{IC}} = (0.339, 0.463)$ after only 3 iterations and 1.25 sec. We thus reduced the total time to obtain the parameter estimate μ_{IC} by 0.6 sec., i.e., the time to take 3 measurements at a sampling rate of 5 Hz.

8.3.3 Control of Weld Pool Depth

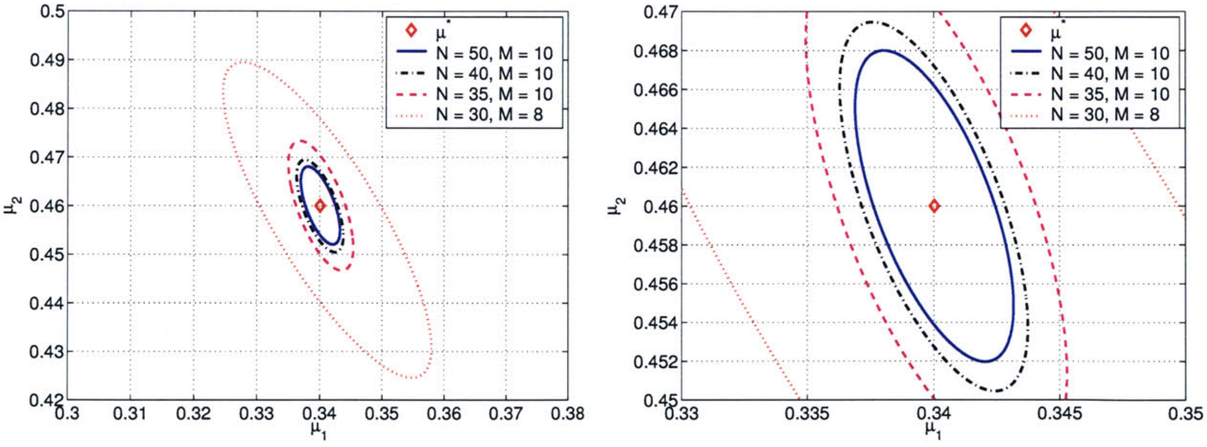
Given the solution of the parameter estimation procedure, we can now turn to the optimal control problem. Since the construction of the possibility region $\mathcal{R}(\mu^*)$ takes a considerable amount of time, we do not consider the min-max problem discussed at the end of Section 8.2.1 for the in-process control. Instead, we simply use μ_{IC} as our estimate for the true parameter μ^* and solve the optimal control problem given μ_{IC}. We only present results here for the LQR problem (8.8) with $W_U = 1$, $W_T = 1\,\text{E}+4$, and $W_R = 1\,\text{E}+6$; we solve (8.8) using the `quadprog` routine from MATLAB.

We shall assume that the nominal input is $u_{\text{n}}(t^k) = 30$, $1 \leq k \leq k_3$, and that the control input is bounded from below and above by $u_{\text{LB}} = 25 \leq u(t^k) \leq u_{\text{UB}} = 50$, $\forall\, k \in \mathbb{K}$, i.e., there exist a minimum and maximum limit for the heat input from the welding torch. Note that we *solve* the optimal control problem during the time interval $\mathbb{I}_3$ and *apply* the optimal control law during the time interval $\mathbb{I}_4$. Since the state of the system at the beginning of $\mathbb{I}_4$ (at the discrete time t^{k_3}) serves as the initial condition for the optimal control problem, we need to predict the time frame necessary for solving the optimal control problem. To begin, we simply choose $k_3 = 30$ and confirm



(a) (b)

Figure 8-5: AP IV: Possibility region $\mathcal{R}(\mu^*)$ for $\mu^* = (0.34, 0.46)$ as a function of ϵ_{exp}.



(a) (b)

Figure 8-6: AP IV: Possibility region $\mathcal{R}(\mu^*)$ for $\mu^* = (0.34, 0.46)$ as a function of N and M.

afterwards if our choice is justified. We also assume that the control horizon is $t^{k_4} = t^{100}$; the period actively controlled is thus $\mathbb{I}_4 = \{t^{30}, \ldots, t^{100}\}$ (corresponding to the time period from 3 to 10 sec.).

To begin, we consider the case with a 1% error in the temperature measurements for the parameter estimation. At the end of the last section we obtained the estimate, $\mu_{\rm IC} = (0.339, 0.463)$, for the true parameter value, $\mu^* = (0.34, 0.46)$, after 1.25 sec. Since we started the parameter estimation procedure after the second measurement was taken, this corresponds to $\mu_{\rm IC}$ being available 1.65 sec. after the welding process is started. Also, the discrete timestep $k_3 = 30$ corresponds to a real time of 3 sec. — we thus need to solve the optimal control problem within 1.35 sec.

Given $s_3^* = 1$ and $\mu_{\rm IC} = (0.339, 0.463)$, we solve the LQR optimal control problem: in Figure 8-7 we present the optimal control input $u^*(t^k)$, the output $s_3(\mu^*, t^k)$ obtained from applying $u^*(t^k)$ to the truth finite element approximation, and the deviation from the desired temperature $|s_3^* - s_3(\mu^*, t^k)|$. Here, we have assumed that the sampling frequency for the control input is $f_{\rm s} = 10$ Hz. Note that the control input is set to the nominal value $u_{\rm n}(t^k)$ for the first three seconds — during this time the parameter estimation and optimal control problem are solved. We observe that approximately 2 seconds after the controller starts the error in the output remains at less than 1% for the remainder of the controlled time interval. We also note that the control input turns off one second before the final time: this is an "artifact" of the thermal inertia of the system as well as the location of the "fictitious" output s_3 downwind of the heat source.

The time to solve for the optimal control input is approximately 3 seconds — with our current implementation we cannot reach the required solution time of 1.35 sec. One option to satisfy the time constraint is, of course, to consider a shorter time interval. Another option, and the one we pursue here, is to presume a smaller sampling frequency, $f_{\rm s}$, for updating the control input. This results in a smaller number of unknowns in the optimization and thus a faster solution time. In Figure 8-8(a) and (b) we present the optimal control input $u^*(t^k)$, the output $s_3(\mu^*, t^k)$, and the output error $|s_3^* - s_3(\mu^*, t^k)|$ for $f_{\rm s} = 5$ Hz and $f_{\rm s} = 2$ Hz, respectively. We note that the error level is very similar despite the lower sampling frequencies. The solution time, however, decreases to approximately 1.8 sec. for $f_{\rm s} = 5$ Hz and to 1.0 sec. for $f_{\rm s} = 2$ Hz. We can thus achieve the required solution time for $f_{\rm s} = 2$ Hz without a serious detriment to the tracking performance.

We next consider two more test cases with increasingly larger measurement errors $\epsilon_{\rm exp}$ during the parameter estimation procedure. In general, a larger measurement error results in a poorer parameter estimate and we thus expect the steady-state error of the controlled output to be larger. We first solve the parameter estimation problem with $\epsilon_{\rm exp} = 2\%$ and $\epsilon_{\rm exp} = 5\%$ and obtain $\mu_{\rm IC} = (0.339, 0.466)$ and $\mu_{\rm IC} = (0.334, 0.473)$, respectively. The solution time is approximately 1.25 sec. in both cases. Given the parameter estimates, we solve the optimal control problem for $f_{\rm s} = 5$ Hz and $f_{\rm s} = 2$ Hz. The results are presented in Figures 8-9 and 8-10. We observe that difference in the output error is very small for $\epsilon_{\rm exp} = 1\%$ and $\epsilon_{\rm exp} = 2\%$ (note that the parameter estimates are very similar). The output error for $\epsilon_{\rm exp} = 5\%$, on the other hand, is now above one percent due to the larger error in the parameter estimate $\mu_{\rm IC}$.

Finally, we note that the approach presented here can be considered as the first step in a model predictive control framework. Once the optimal control problem is solved and implemented, we repeat the estimation-control process: we first take new measurements and update the parameter estimate accordingly; given the updated parameter estimate, we solve the optimal control problem for the next upcoming time interval. Proceeding in this fashion, we can obtain a sampling period for the parameter estimation and control update of approximately 3 seconds. The controller can thus

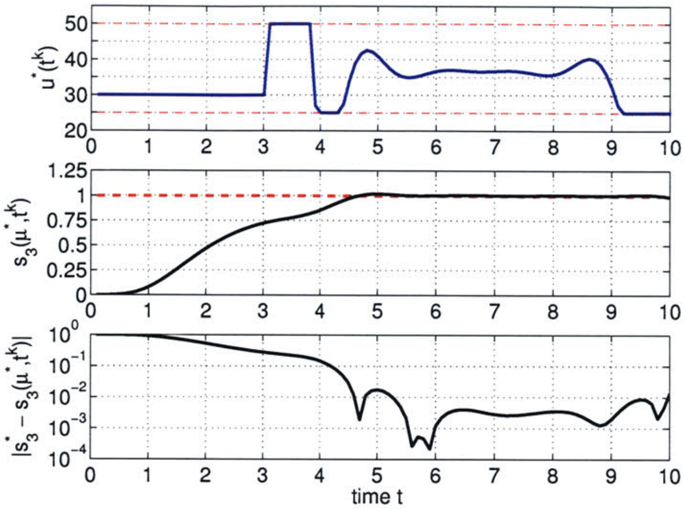


Figure 8-7: AP IV: Optimal control input $u^*(t^k)$, output $s_3(\mu^*, t^k)$, and output error $|d_{\mathrm{w,d}} - s_3(\mu^*, t^k)|$, for $\mu_{\mathrm{IC}} = (0.339, 0.463)$, $\epsilon_{\exp} = 1\%$ and $f_{\mathrm{s}} = 10$ Hz.

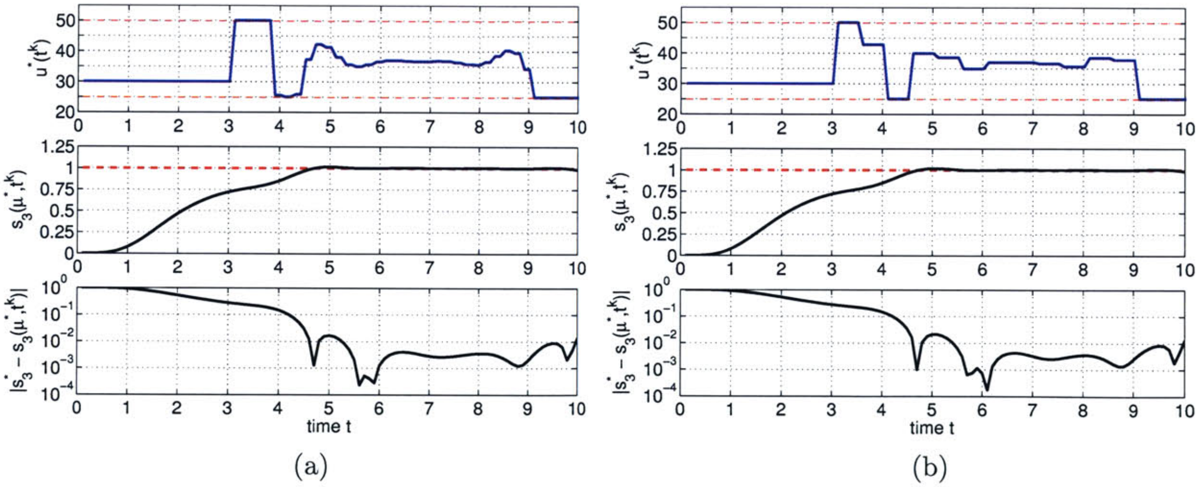
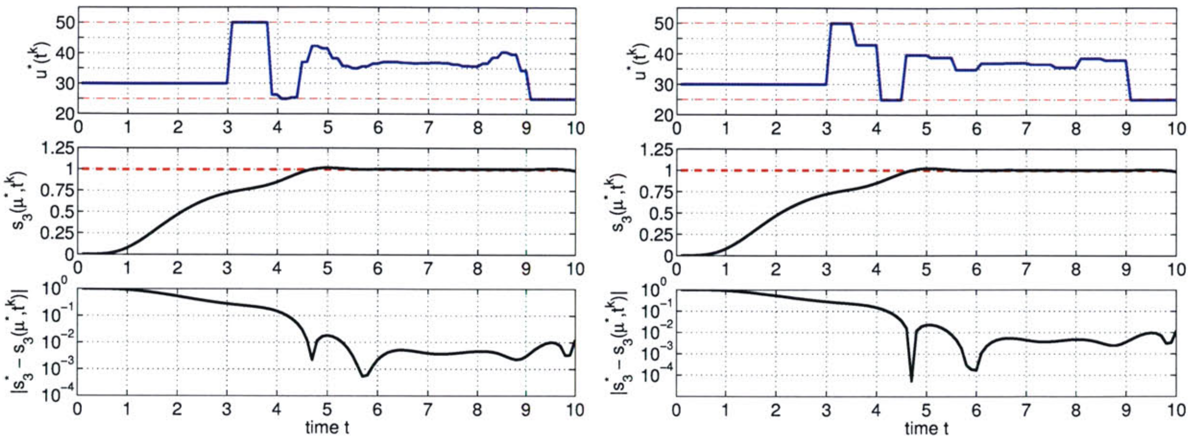
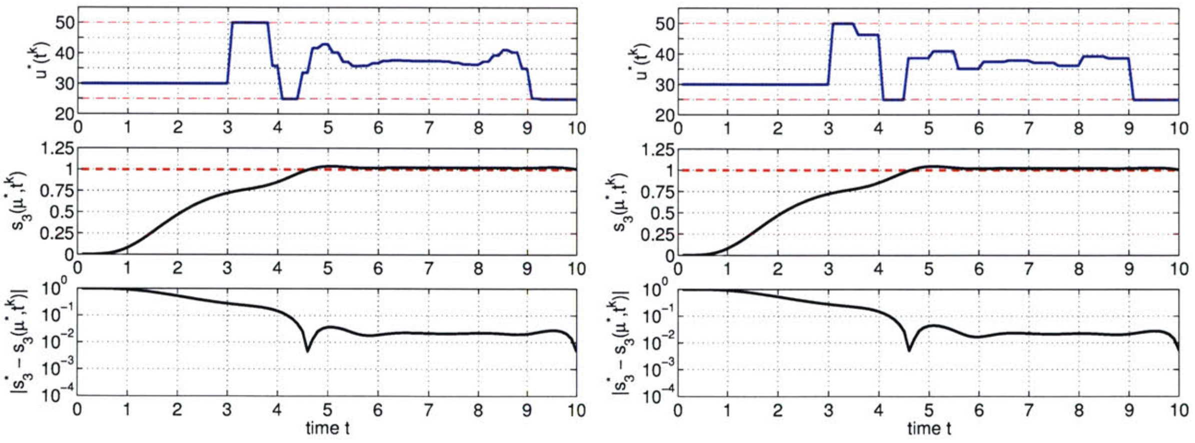


Figure 8-8: AP IV: Optimal control input $u^*(t^k)$, output $s_3(\mu^*, t^k)$, and output error $|d_{\mathrm{w,d}} - s_3(\mu^*, t^k)|$, for $\mu_{\mathrm{IC}} = (0.339, 0.463)$, $\epsilon_{\exp} = 1\%$ and (a) $f_{\mathrm{s}} = 5$ Hz, (b) $f_{\mathrm{s}} = 2$ Hz .



(a) (b)

Figure 8-9: AP IV: Optimal control input $u^*(t^k)$, output $s_3(\mu^*, t^k)$, and output error $|d_{\mathrm{w,d}} - s_3(\mu^*, t^k)|$, for $\mu_{\mathrm{IC}} = (0.339, 0.466)$, $\epsilon_{\mathrm{exp}} = 2\%$ and (a) $f_{\mathrm{s}} = 5$ Hz, (b) $f_{\mathrm{s}} = 2$ Hz .



(a) (b)

Figure 8-10: AP IV: Optimal control input $u^*(t^k)$, output $s_3(\mu^*, t^k)$, and output error $|d_{\mathrm{w,d}} - s_3(\mu^*, t^k)|$, for $\mu_{\mathrm{IC}} = (0.334, 0.473)$, $\epsilon_{\mathrm{exp}} = 5\%$ and (a) $f_{\mathrm{s}} = 5$ Hz, (b) $f_{\mathrm{s}} = 2$ Hz .

react to disturbances and changes in the system parameters. However, as mentioned previously, the variation of the system parameters has to be slow compared to the time to solve the inverse and optimal control problems.

Chapter 9

Concluding Remarks

9.1 Summary

The main goal of this thesis is the development of reduced-basis methods for problems governed by parametrized parabolic partial differential equations. The essential components are (*i*) rapidly convergent reduced-basis approximations — Galerkin projection onto a space W_N spanned by solutions of the governing partial differential equation at N selected points in parameter-time 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*) offline-online 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.

Many model-order reduction techniques for time-dependent linear and nonlinear systems are proposed in the literature. However, almost all of these techniques consider time the *only* variable and do not accommodate parametric variation nor *a posteriori* error estimation. In the reduced-basis method, we simply treat time as an additional, albeit special, parameter. Instead of generating a reduced-order model for a particular time-varying system, we create a model valid for general parametric families of systems. Our results show that we obtain good approximation properties for all parameters in the admissible domain.

We also improve and extend on earlier work on reduced-basis methods for parabolic problems in several directions: we rigorously treat (*i*) temporal forcing/control inputs that are *not known a priori*, and (*ii*) outputs that are also (scalar) functions of time. We develop a new *a posteriori* error estimation procedure that provides rigorous bounds for the error in the energy norm and in the output at all (discrete) timesteps. This generalization allows us to treat a wider class of applications and pursue a more rational way of constructing the parameter-time sample set — our "greedy" adaptive procedure optimally selects the samples and thus helps avoid ill-conditioning of the reduced-order model. The procedure performs very well in practice and sometimes results in non-obvious parameter samples that would be hard to predict. Furthermore, we propose an impulse approach to construct the basis for LTI systems — especially important for optimal control applications.

We propose adjoint procedures in the context of problems with affine parameter dependence. The primal-dual formulation results in a square effect of the output estimate and output bound convergence. Despite the additional cost due to the solution of the dual problem, the overall computational efficiency for certain problems may increase considerably as compared to a "primal-only" approach. However, the primal-dual formulation is less advantageous when considering problems with either many outputs or a large number of timesteps — one such example being the pollution problem discussed several times.

In Chapter 4 the assumption of affine parameter dependence is critical for computational efficiency. Unfortunately, it is also rather limiting: the pollution problem (with a varying source location) and the welding problem both display a nonaffine parameter dependence. We treat these problems in Chapter 5, where we introduce a collateral reduced-basis approximation for the nonaffine terms and employ an empirical interpolation method to calculate the coefficients for the nonaffine approximation. We also introduce *a posteriori* error bounds and offline-online computational decompositions which retain the online *$\mathcal{N}$-independence*; although our error bounds are rigorous only under certain conditions on the nonaffine function approximation, we observe in the numerical results that our methods perform well even if these conditions are not met. Nevertheless, we have to carefully choose the approximation order N and M of the reduced-basis and nonaffine function approximation, respectively.

We also consider certain classes of nonlinear parabolic problems in Chapter 6. The treatment of the nonlinear case is similar to the nonaffine case in that we now employ the empirical interpolation method to approximate the *nonlinear* term. Here, it is difficult to generate an explicit affine approximation of the nonlinear term since the field variable is not known in advance. The greedy adaptive sampling procedure ensures good approximation properties, but is very (maybe prohibitively) expensive in the nonlinear case. We also propose *a posteriori* error estimation procedures and offline-online decompositions which are valid even in the presence of highly nonlinear terms. As a specific example in the class of reaction-diffusion equations we consider the self-ignition of a coal stockpile. Although we do not have error bounds available for this problem, our results show that the reduced-basis method performs very well in approximating and capturing the highly nonlinear behavior.

Finally, we integrate the reduced-basis method into two representative applications requiring repeated and rapid evaluation of input-output relationships: robust parameter estimation in Chapter 7 and optimal control in Chapter 8. The examples presented in these chapters, and the numerical tests performed throughout this thesis, demonstrate the applicability, effectivity, and efficiency of our proposed method. We rigorously and efficiently quantify the uncertainty due to measurement and reduced-basis modeling errors in inverse problems — we can thus pursue a *real-time* and *robust* parameter estimation procedure which would have been intractable with conventional finite element methods. Furthermore, we can consider in-process (optimal) control of certain engineering problems, which is of special interest in the model predictive control framework.

9.2 Future Work

We conclude by giving some suggestions for future work. In Chapter 4 we presented several numerical results for applying a primal-dual formulation in the reduced-basis context. We observed that we can obtain a specific desired accuracy for the output estimate and output bound for different combinations of N_{pr} and N_{du}, the dimensions of the primal and dual spaces. The choice of N_{pr} and

N_{du} is, of course, also important for the computational efficiency of the method. An interesting question is thus to find the "best" relative choice of N_{pr} and N_{du} which minimizes the computational cost for a given desired approximation accuracy. The solution will certainly be different for every specific problem at hand, but it would be good if a general guideline is found.

On a related issue, we observed that the primal-dual formulation greatly improves the convergence rate of the output estimate and output bound. However, the output effectivities reported showed that the output bounds are not necessarily sharper as compared to the primal-only approach. Following the approach in [121], we presented preliminary results that lead to output effectivities of $O(1)$ by adding the residual correction term to the output bound instead of improving the output estimate. Nevertheless, the required dimensions of the primal and dual reduced-basis approximations to obtain a specific accuracy did not decrease. A hybrid approach, on the other hand, may result in a smaller N_{pr} and N_{du} and improve the computational cost.

In this thesis, we presented adjoint techniques only for affine problems. It would be good to extend the methods in Chapters 4 and 5 to consider primal-dual formulations for nonaffine problems. The nonlinear case is far more complicated and certainly requires additional effort (see [121] for the application of adjoint techniques to reduced-basis approximations of the steady Navier-Stokes equation).

The extension of the methods presented here to more general nonlinear parabolic problems is another interesting topic of research. One possibility are problems involving nonlinearity in the Laplacian, e.g., anisotropic or nonlinear diffusion problems widely used in image processing. The most interesting choice with the widest applicability is to consider the unsteady Navier-Stokes equation. Combining the theory developed in [121] and in this thesis might be a first step in this direction. However, even the theory for nonlinear problems presented here should be improved upon, especially the extremely high computational cost for the greedy procedure.

Although we applied our method to several real-world problems, the work in this thesis is still mostly theoretic. Implementing the reduced-basis method in actual practice, e.g., in the "real-world" control or parameter estimation framework, would prove the real potential of the proposed method. We also note that in this thesis we only considered problems with up to three parameters, whereas in real-world problems the number of parameters might be considerably higher. However, considering problems with $O(10)$ parameters may vastly reduce the computational savings, since even for our few-parameter problems N reached up to 200. The question of how the reduced-basis approximation scales with the dimension of the parameter space is thus also very interesting to investigate.

Appendix A

Offline-Online Computational Procedure

A.1 Reduced-Basis Approximation

We summarize here the reduced-basis approximations and necessary quantities for the dual problem and the output estimate (for the primal problem, see Section 4.3.3).

For the dual problem we define $\underline{\Psi}_N(\mu, t^k) = [\Psi_{N\,1}(\mu, t^k) \quad \Psi_{N\,2}(\mu, t^k) \;\ldots\; \Psi_{N\,N_{\mathrm{du}}}(\mu, t^k)]^T$ and obtain from (4.18) that

$$\Big(M_N^{\mathrm{du}}(\mu) + \Delta t \; A_N^{\mathrm{du}}(\mu)\Big) \; \underline{\Psi}_N(\mu, t^k) = M_N^{\mathrm{du}}(\mu) \; \underline{\Psi}_N(\mu, t^{k+1}), \quad \forall\, k \in \mathbb{K}, \tag{A.1}$$

where

$$M_N^{\mathrm{du}}(\mu) = \sum_{q=1}^{Q_m} \Theta_m^q(\mu) \; M_N^{\mathrm{du}\,q}, \qquad A_N^{\mathrm{du}}(\mu) = \sum_{q=1}^{Q_a} \Theta_a^q(\mu) \; A_N^{\mathrm{du}\,q}, \tag{A.2}$$

with entries

$$\begin{array}{rcll} M_{N\,i,j}^{\mathrm{du}\,q} & = & m^q(\zeta_i^{\mathrm{du}}, \zeta_j^{\mathrm{du}}), & 1 \le i,j \le N_{\mathrm{du,max}},\; 1 \le q \le Q_m; \\ A_{N\,i,j}^{\mathrm{du}\,q} & = & a^q(\zeta_i^{\mathrm{du}}, \zeta_j^{\mathrm{du}}), & 1 \le i,j \le N_{\mathrm{du,max}},\; 1 \le q \le Q_a; \text{ and} \\ L_{N\,i}^{\mathrm{du}} & = & \ell(\zeta_i^{\mathrm{du}}), & 1 \le i \le N_{\mathrm{du,max}}. \end{array} \tag{A.3}$$

Note that $\underline{\Psi}_N(\mu, t^{K+1})$ is calculated from $M_N^{\mathrm{du}}(\mu) \; \underline{\Psi}_N(\mu, t^{K+1}) = L_N^{\mathrm{du}}$.

Finally, we evaluate the output estimate, $\forall\, k \in \mathbb{K}$, from

$$\begin{aligned} s_N(\mu, t^k) = {} & L_N^{\mathrm{pr}\,T} \, \underline{y}_N(\mu, t^k) + \Delta t \; \sum_{k'=1}^{k} \underline{\Psi}_N^T(\mu, t^{K-k+k'}) \\ & \times \Big\{ B_N^{\mathrm{du}}(\mu) \, u(t^{k'}) - A_N^{\mathrm{pr,du}}(\mu) \, \underline{y}_N(\mu, t^{k'}) - \frac{1}{\Delta t} M_N^{\mathrm{pr,du}}(\mu) \; \Big(\underline{y}_N(\mu, t^{k'}) - \underline{y}_N(\mu, t^{k'-1})\Big) \Big\} \end{aligned} \tag{A.4}$$

where

$$\begin{aligned} M_N^{\mathrm{pr,du}}(\mu) &= \sum_{q=1}^{Q_m} \Theta_m^q(\mu)\ M_N^{\mathrm{pr,du}\,q}, \\ A_N^{\mathrm{pr,du}}(\mu) &= \sum_{q=1}^{Q_a} \Theta_a^q(\mu)\ A_N^{\mathrm{pr,du}\,q}, \\ B_N^{\mathrm{du}}(\mu) &= \sum_{q=1}^{Q_b} \Theta_b^q(\mu)\ B_N^{\mathrm{du}\,q}, \end{aligned} \tag{A.5}$$

with entries

$$\begin{aligned} M_{N\,i,j}^{\mathrm{pr,du}\,q} &= m^q(\zeta_i^{\mathrm{du}}, \zeta_j^{\mathrm{pr}}), \quad 1 \le i \le N_{\mathrm{du,max}},\ 1 \le j \le N_{\mathrm{pr,max}},\ 1 \le q \le Q_m; \\ A_{N\,i,j}^{\mathrm{pr,du}\,q} &= a^q(\zeta_i^{\mathrm{du}}, \zeta_j^{\mathrm{pr}}), \quad 1 \le i \le N_{\mathrm{du,max}},\ 1 \le j \le N_{\mathrm{pr,max}},\ 1 \le q \le Q_a; \\ B_{N\,i}^{\mathrm{du}\,q} &= b^q(\zeta_i^{\mathrm{du}}), \quad 1 \le i \le N_{\mathrm{du,max}},\ 1 \le q \le Q_b; \\ L_{N\,i}^{\mathrm{pr}} &= \ell(\zeta_i^{\mathrm{pr}}), \quad 1 \le i \le N_{\mathrm{pr,max}}. \end{aligned} \tag{A.6}$$

The offline-online procedure is described in Section 4.3.3.

A.2 *A Posteriori* Error Estimation

In this section we discuss the calculation of the primal and dual error bound. For the primal error bound, we first note from standard duality arguments that

$$\begin{aligned} \varepsilon_{N_{\mathrm{pr}}}^{\mathrm{pr}}(\mu, t^k) &\equiv \sup_{v \in Y} \frac{R^{\mathrm{pr}}(v; \mu, t^k)}{\|v\|_Y} && \text{(A.7)} \\ &= \|\hat{e}^{\mathrm{pr}}(\mu, t^k)\|_Y, && \text{(A.8)} \end{aligned}$$

where $\hat{e}^{\mathrm{pr}}(\mu, t^k) \in Y$ is given by

$$(\hat{e}^{\mathrm{pr}}(\mu, t^k), v)_Y = R^{\mathrm{pr}}(v; \mu, t^k), \qquad \forall\, v \in Y; \tag{A.9}$$

(A.9) is effectively a Poisson problem for each $t^k \in \mathbb{I}$.

From (4.21) and the affine assumptions (4.9)-(4.11) it thus follows that $\hat{e}^{\mathrm{pr}}(\mu, t^k)$ satisfies

$$\begin{aligned} (\hat{e}^{\mathrm{pr}}(\mu, t^k), v)_Y = \sum_{q=1}^{Q_b} \Theta_b^q(\mu)\, b^q(v)\, u(t^k) - \sum_{n=1}^{N_{\mathrm{pr}}} \Bigg\{ \sum_{q=1}^{Q_a} \Theta_a^q(\mu)\, y_{Nn}(\mu, t^k)\, a^q(\zeta_n^{\mathrm{pr}}, v) \\ + \sum_{q=1}^{Q_m} \frac{1}{\Delta t} \Theta_m^q(\mu) \left(y_{Nn}(\mu, t^k) - y_{Nn}(\mu, t^{k-1}) \right) m^q(\zeta_n^{\mathrm{pr}}, v) \Bigg\},\ \forall\, v \in Y. \end{aligned} \tag{A.10}$$

It is clear from linear superposition that we can express $\hat{e}(\mu, t^k)$ as

$$\hat{e}^{\mathrm{pr}}(\mu) = \sum_{q=1}^{Q_b} \Theta_b^q(\mu)\, u(t^k)\, \mathcal{B}_q^{\mathrm{pr}} - \sum_{n=1}^{N_{\mathrm{pr}}} \left\{ \sum_{q=1}^{Q_a} \Theta_a^q(\mu)\, y_{Nn}(\mu, t^k)\, \mathcal{A}_{q,n}^{\mathrm{pr}} \right.$$
$$\left. + \sum_{q=1}^{Q_m} \frac{1}{\Delta t} \Theta_m^q(\mu) \left(y_{Nn}(\mu, t^k) - y_{Nn}(\mu, t^{k-1}) \right) \mathcal{M}_{q,n}^{\mathrm{pr}} \right\}, \quad \text{(A.11)}$$

where we calculate $\mathcal{B}_q^{\mathrm{pr}} \in Y$, $\mathcal{A}_{q,n}^{\mathrm{pr}} \in Y$, and $\mathcal{M}_{q,n}^{\mathrm{pr}} \in Y$ from

$$\begin{aligned}
(\mathcal{B}_q^{\mathrm{pr}}, v)_Y &= b^q(v), && \forall\, v \in Y,\ 1 \le q \le Q_b, \\
(\mathcal{A}_{q,n}^{\mathrm{pr}}, v)_Y &= a^q(\zeta_n^{\mathrm{pr}}, v), && \forall\, v \in Y,\ 1 \le n \le N_{\mathrm{pr,max}},\ 1 \le q \le Q_a, \\
(\mathcal{M}_{q,n}^{\mathrm{pr}}, v)_Y &= m^q(\zeta_n^{\mathrm{pr}}, v), && \forall\, v \in Y,\ 1 \le n \le N_{\mathrm{pr,max}},\ 1 \le q \le Q_m,
\end{aligned} \quad \text{(A.12)}$$

respectively; note $\mathcal{B}$, $\mathcal{A}$, and $\mathcal{M}$ are parameter independent.

From (A.8) and (A.11) it follows that

$$\begin{aligned}
\varepsilon_{N_{\mathrm{pr}}}^{\mathrm{pr}}(\mu, t^k)^2 \;=\;& \sum_{q,q'=1}^{Q_b} \Theta_b^q(\mu)\, \Theta_b^{q'}(\mu)\, u(t^k)\, u(t^k)\, \Lambda_{qq'}^{\mathrm{pr},bb} \\
&+ \sum_{q=1}^{Q_b} \sum_{n=1}^{N_{\mathrm{pr}}} \Theta_b^q(\mu)\, u(t^k) \Bigg(\sum_{q'=1}^{Q_a} \Theta_a^{q'}(\mu)\, y_{Nn}(\mu, t^k)\, \Lambda_{qq'n}^{\mathrm{pr},ab} \\
&\qquad + \sum_{q'=1}^{Q_m} \Theta_m^{q'}(\mu) \left(y_{Nn}(\mu, t^k) - y_{Nn}(\mu, t^{k-1}) \right) \Lambda_{qq'n}^{\mathrm{pr},mb} \Bigg) \\
&+ \sum_{n,n'=1}^{N_{\mathrm{pr}}} \Bigg\{ \sum_{q,q'=1}^{Q_a} \Theta_a^q(\mu)\, \Theta_a^{q'}(\mu)\, y_{Nn}(\mu, t^k)\, y_{Nn'}(\mu, t^k)\, \Lambda_{qnq'n'}^{\mathrm{pr},aa} \\
&\qquad + \sum_{q,q'=1}^{Q_m} \Theta_m^q(\mu)\, \Theta_m^{q'}(\mu) \left(y_{Nn}(\mu, t^k) - y_{Nn}(\mu, t^{k-1}) \right) \\
&\qquad\qquad \times \left(y_{Nn'}(\mu, t^k) - y_{Nn'}(\mu, t^{k-1}) \right) \Lambda_{qnq'n'}^{\mathrm{pr},mm} \\
&\qquad + \sum_{q=1}^{Q_a} \sum_{q'=1}^{Q_m} \Theta_a^q(\mu)\, \Theta_m^{q'}(\mu)\, y_{Nn}(\mu, t^k) \\
&\qquad\qquad \times \left(y_{Nn'}(\mu, t^k) - y_{Nn'}(\mu, t^{k-1}) \right) \Lambda_{qnq'n'}^{\mathrm{pr},am} \Bigg\},
\end{aligned} \quad \text{(A.13)}$$

where the parameter-independent quantities Λ^{pr} are defined as

$$\begin{aligned}
\Lambda^{\mathrm{pr},bb}_{qq'} &= (\mathcal{B}^{\mathrm{pr}}_q, \mathcal{B}^{\mathrm{pr}}_{q'})_Y, && 1 \le q, q' \le Q_b; \\
\Lambda^{\mathrm{pr},ab}_{qq'n} &= -2\, (\mathcal{B}^{\mathrm{pr}}_q, \mathcal{A}^{\mathrm{pr}}_{q',n})_Y, && 1 \le q \le Q_b,\ 1 \le q' \le Q_a,\ 1 \le n \le N_{\mathrm{pr,max}}; \\
\Lambda^{\mathrm{pr},mb}_{qq'n} &= -\tfrac{2}{\Delta t}\, (\mathcal{B}^{\mathrm{pr}}_q, \mathcal{M}^{\mathrm{pr}}_{q',n})_Y, && 1 \le q \le Q_b,\ 1 \le q' \le Q_m,\ 1 \le n \le N_{\mathrm{pr,max}}; \\
\Lambda^{\mathrm{pr},aa}_{qnq'n'} &= (\mathcal{A}^{\mathrm{pr}}_{q,n}, \mathcal{A}^{\mathrm{pr}}_{q',n'})_Y, && 1 \le q, q' \le Q_a,\ 1 \le n, n' \le N_{\mathrm{pr,max}}; \\
\Lambda^{\mathrm{pr},am}_{qnq'n'} &= \tfrac{2}{\Delta t}\, (\mathcal{A}^{\mathrm{pr}}_{q,n}, \mathcal{M}^{\mathrm{pr}}_{q',n'})_Y, && 1 \le q \le Q_a,\ 1 \le q' \le Q_m,\ 1 \le n, n' \le N_{\mathrm{pr,max}}; \\
\Lambda^{\mathrm{pr},mm}_{qnq'n'} &= \tfrac{1}{\Delta t^2} (\mathcal{M}^{\mathrm{pr}}_{q,n}, \mathcal{M}^{\mathrm{pr}}_{q',n'})_Y, && 1 \le q, q' \le Q_m,\ 1 \le n, n' \le N_{\mathrm{pr,max}}.
\end{aligned} \tag{A.14}$$

The computational procedure for the dual error bound follows arguments similar to the primal error bound presented in (A.7)-(A.11). Thus, we first solve for $\mathcal{A}^{\mathrm{du}}_{q,n} \in Y$, and $\mathcal{M}^{\mathrm{du}}_{q,n} \in Y$ from

$$\begin{aligned}
(\mathcal{A}^{\mathrm{du}}_{q,n}, v)_Y &= a^q(v, \zeta^{\mathrm{du}}_n), && \forall\, v \in Y,\ 1 \le n \le N_{\mathrm{du,max}},\ 1 \le q \le Q_a, \\
(\mathcal{M}^{\mathrm{du}}_{q,n}, v)_Y &= m^q(v, \zeta^{\mathrm{du}}_n), && \forall\, v \in Y,\ 1 \le n \le N_{\mathrm{du,max}},\ 1 \le q \le Q_m,
\end{aligned} \tag{A.15}$$

respectively, and then evaluate the dual norm from

$$\begin{aligned}
\varepsilon^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu, t^k)^2 \;=\; \sum_{n,n'=1}^{N_{\mathrm{du}}} \Bigg\{ & \sum_{q,q'=1}^{Q_a} \Theta^q_a(\mu)\, \Theta^{q'}_a(\mu)\, \Psi_{Nn}(\mu, t^k)\, \Psi_{Nn'}(\mu, t^k)\, \Lambda^{\mathrm{du},aa}_{qnq'n'} \\
&+ \sum_{q,q'=1}^{Q_m} \Theta^q_m(\mu)\, \Theta^{q'}_m(\mu)\, \Big(\Psi_{Nn}(\mu, t^k) - \Psi_{Nn}(\mu, t^{k+1})\Big) \\
&\qquad \times \Big(\Psi_{Nn'}(\mu, t^k) - \Psi_{Nn'}(\mu, t^{k+1})\Big)\, \Lambda^{\mathrm{du},mm}_{qnq'n'} \\
&+ \sum_{q=1}^{Q_a} \sum_{q'=1}^{Q_m} \Theta^q_a(\mu)\, \Theta^{q'}_m(\mu)\, \Psi_{Nn}(\mu, t^k) \\
&\qquad \times \Big(\Psi_{Nn'}(\mu, t^k) - \Psi_{Nn'}(\mu, t^{k+1})\Big)\, \Lambda^{\mathrm{du},am}_{qnq'n'} \Bigg\},
\end{aligned} \tag{A.16}$$

where the parameter-independent quantities Λ^{du} are defined as

$$\begin{aligned}
\Lambda^{\mathrm{du},aa}_{qnq'n'} &= (\mathcal{A}^{\mathrm{du}}_{q,n}, \mathcal{A}^{\mathrm{du}}_{q',n'})_Y, && 1 \le q, q' \le Q_a,\ 1 \le n, n' \le N_{\mathrm{du,max}}; \\
\Lambda^{\mathrm{du},am}_{qnq'n'} &= \tfrac{2}{\Delta t}\, (\mathcal{A}^{\mathrm{du}}_{q,n}, \mathcal{M}^{\mathrm{du}}_{q',n'})_Y, && 1 \le q \le Q_a,\ 1 \le q' \le Q_m,\ 1 \le n, n' \le N_{\mathrm{du,max}}; \\
\Lambda^{\mathrm{du},mm}_{qnq'n'} &= \tfrac{1}{\Delta t^2} (\mathcal{M}^{\mathrm{du}}_{q,n}, \mathcal{M}^{\mathrm{du}}_{q',n'})_Y, && 1 \le q, q' \le Q_m,\ 1 \le n, n' \le N_{\mathrm{du,max}}.
\end{aligned} \tag{A.17}$$

Finally, for the contribution due to the error of the dual problem at the final time we first solve for $\mathcal{L}^{\Psi_f} \in Y$ and $\mathcal{M}^{\Psi_f}_{q,n} \in Y$ from

$$\begin{aligned}
(\mathcal{L}^{\Psi_f}, v)_X &= \ell(v), && \forall\, v \in Y, \\
(\mathcal{M}^{\Psi_f}_{q,n}, v)_X &= m^q(v, \zeta^{\mathrm{du}}_n), && \forall\, v \in Y,\ 1 \le n \le N_{\mathrm{du,max}},\ 1 \le q \le Q_m,
\end{aligned} \tag{A.18}$$

respectively; we then evaluate the dual norm from

$$\varepsilon^{\Psi_f}_{N_{\rm du}}(\mu)^2 = \Lambda^{\Psi_f,\ell\ell} + \sum_{n=1}^{N_{\rm du}} \sum_{q=1}^{Q_m} \Theta^q_m(\mu)\, \Psi_{Nn}(\mu, t^{K+1}) \times \left\{ \Lambda^{\Psi_f,\ell m}_{qn} + \sum_{n'=1}^{N_{\rm du}} \sum_{q'=1}^{Q_m} \Theta^{q'}_m(\mu)\, \Psi_{Nn'}(\mu, t^{K+1}) \Lambda^{\Psi_f,mm}_{qnq'n'} \right\}, \qquad \text{(A.19)}$$

where the parameter-independent quantities Λ^{Ψ_f} are defined as

$$\begin{aligned} \Lambda^{\Psi_f,\ell\ell} &= (\mathcal{L}^{\Psi_f}, \mathcal{L}^{\Psi})_X; \\ \Lambda^{\Psi_f,\ell m}_{qn} &= -2\,(\mathcal{M}^{\Psi_f}_{q,n}, \mathcal{L}^{\Psi_f})_X, \quad 1 \le q \le Q_m, \ 1 \le n \le N_{\rm du,max}; \\ \Lambda^{\Psi_f,mm}_{qnq'n'} &= (\mathcal{M}^{\Psi_f}_{q,n}, \mathcal{M}^{\Psi_f}_{q',n'})_X, \quad 1 \le q, q' \le Q_m, \ 1 \le n, n' \le N_{\rm du,max}. \end{aligned} \qquad \text{(A.20)}$$

The offline-online decomposition is now clear. In the offline stage we first compute the quantities $\mathcal{B}^{\rm pr}$, $\mathcal{L}^{\Psi_f}$, $\mathcal{A}^{\rm pr,du}$, and $\mathcal{M}^{\rm pr,du,\Psi_f}$ from (A.12), (A.15), and (A.18) and then evaluate the $\Lambda^{\rm pr,du,\Psi_f}$ from (A.14), (A.17), and (A.20); this requires (to leading order) $O((N_{\rm pr,max} + N_{\rm du,max})(Q_a + Q_m))$ expensive "truth" finite element solutions, and $O((N^2_{\rm pr,max} + N^2_{\rm du,max})(Q^2_a + Q_a Q_m + Q^2_m))$ $\mathcal{N}$-inner products. In the online stage, given a new parameter value μ and associated reduced-basis solutions $\underline{y}_N(\mu, t^k)$ and $\underline{\Psi}_N(\mu, t^k)$, we perform the sums (A.13), (A.16), and (A.19) and evaluate the error bound from

$$\Delta^s(\mu, t^k) = \left(\frac{\Delta t}{\hat{\alpha}_a(\mu)} \sum_{k'=1}^{k} \varepsilon^{\rm pr}_{N_{\rm pr}}(\mu, t^{k'})^2 \right)^{\frac{1}{2}} \left(\frac{\Delta t}{\hat{\alpha}_a(\mu)} \sum_{k'=K-k+1}^{K} \varepsilon^{\rm du}_{N_{\rm du}}(\mu, t^{k'})^2 + \frac{\varepsilon^{\Psi_f}_{N_{\rm du}}(\mu)^2}{\hat{\alpha}_m(\mu)} \right)^{\frac{1}{2}}, \quad \forall\, k \in \mathbb{K}; \qquad \text{(A.21)}$$

it directly follows that the online operation count for $\Delta^s(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, is $O(K(N^2_{\rm pr} + N^2_{\rm du})(Q^2_a + Q_a Q_m + Q^2_m))$. Thus, all requisite online calculations are *independent* of the dimension of the underlying "truth" finite element space, $\mathcal{N}$.

Appendix B

Time-Discretization: Crank-Nicolson

B.1 Abstract Formulation

We derived here the results for the Crank-Nicolson time integration scheme corresponding to the results presented in Chapter 4 for the Euler-Backward scheme. We consider the time interval $I \equiv]0, t_f]$ ($\bar{I} \equiv [0, t_f]$) and divide $\bar{I}$ into K subintervals of equal length $\Delta t = \frac{t_f}{K}$ and define $t^k \equiv k\Delta t$, $0 \leq k \leq K \equiv \frac{t_f}{\Delta t}$, and $\mathbb{I} \equiv \{t^0, \ldots, t^k\}$; for notational convenience, we also introduce $\mathbb{K} \equiv \{1, \ldots, K\}$. Again, our results must be stable as $\Delta t \to 0$, $K \to \infty$. We also recall our reference finite element approximation space $Y \subset Y^{\mathrm{e}}\,(\subset X^{\mathrm{e}})$ of very large dimension $\mathcal{N}$.

B.1.1 Primal Problem

We can now directly state the reference (or "truth") finite element approximation: Given a parameter $\mu \in \mathcal{D}$, we evaluate the (single) output $s(\mu, t^k) \in \boldsymbol{R}$ from

$$s(\mu, t^k) = \ell(y(\mu, t^k)), \quad \forall\, k \in \mathbb{K}, \tag{B.1}$$

where $y(\mu, t^k) \in Y$, $\forall\, k \in \mathbb{K}$ satisfies

$$\begin{aligned} m(y(\mu, t^k), v; \mu) &+ \frac{\Delta t}{2}\; a(y(\mu, t^k) + y(\mu, t^{k-1}), v; \mu) \\ &= \; m(y(\mu, t^{k-1}), v; \mu) + \frac{\Delta t}{2}\; b(v; \mu)\; (u(t^k) + u(t^{k-1})), \quad \forall\, v \in Y, \end{aligned} \tag{B.2}$$

with initial condition $y(\mu, t^0) = 0$.

B.1.2 Dual Problem

We also introduce a dual problem which evolves backward in time. Invoking the LTI property we can express the adjoint for the output at time t^L, $1 \leq L \leq K$, as $\psi_L(\mu, t^k) = \Psi(\mu, t^{K-L+k})$, $1 \leq k \leq L$, where $\Psi(\mu, t^k) \in Y$ satisfies

$$m(v, \Psi(\mu, t^k); \mu) + \frac{\Delta t}{2}\; a(v, \Psi(\mu, t^k) + \Psi(\mu, t^{k+1}); \mu) \;=\; m(v, \Psi(\mu, t^{k+1}); \mu), \qquad \forall\, v \in Y,\ \forall\, k \in \mathbb{K}, \quad \text{(B.3)}$$

with final condition

$$m(v, \Psi(\mu, t^{K+1}); \mu) \equiv \ell(v), \qquad \forall\, v \in Y. \qquad \text{(B.4)}$$

Again, to obtain $\psi_L(\mu, t^k)$, $1 \leq k \leq L$, $\forall L \in \mathbb{K}$, we solve *once* for $\Psi(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, and then appropriately shift the result — we do not need to solve K separate dual problems.

B.2 Reduced-Basis Approximation

We now introduce the nested sample sets $S^{\rm pr}_{N_{\rm pr}} = \{\tilde{\mu}^{\rm pr}_1 \in \tilde{\mathcal{D}}, \ldots, \tilde{\mu}^{\rm pr}_{N_{\rm pr}} \in \tilde{\mathcal{D}}\}$, $1 \leq N_{\rm pr} \leq N_{\rm pr,max}$, and $S^{\rm du}_{N_{\rm du}} = \{\tilde{\mu}^{\rm du}_1 \in \tilde{\mathcal{D}}, \ldots, \tilde{\mu}^{\rm du}_{N_{\rm du}} \in \tilde{\mathcal{D}}\}$, $1 \leq N_{\rm du} \leq N_{\rm du,max}$, where $\tilde{\mu} \equiv (\mu, t^k)$ and $\tilde{\mathcal{D}} \equiv \mathcal{D} \times \mathbb{I}$; note that the samples must reside in the *parameter-time* space, $\tilde{\mathcal{D}}$. Here, $N_{\rm pr}$ and $N_{\rm du}$ are the dimensions of the reduced basis space for the primal and dual variables, respectively; in general, $S^{\rm pr}_{N_{\rm pr}} \neq S^{\rm du}_{N_{\rm du}}$ and in fact $N_{\rm pr} \neq N_{\rm du}$. We then define the associated nested Lagrangian [85] reduced-basis spaces

$$W^{\rm pr}_{N_{\rm pr}} = \text{span}\{\zeta^{\rm pr}_n \equiv y(\tilde{\mu}^{\rm pr}_n),\ 1 \leq n \leq N_{\rm pr}\}, \quad 1 \leq N_{\rm pr} \leq N_{\rm pr,max}, \qquad \text{(B.5)}$$

and

$$W^{\rm du}_{N_{\rm du}} = \text{span}\{\zeta^{\rm du}_n \equiv \Psi(\tilde{\mu}^{\rm du}_n),\ 1 \leq n \leq N_{\rm du}\}, \quad 1 \leq N_{\rm du} \leq N_{\rm du,max}, \qquad \text{(B.6)}$$

where $y(\tilde{\mu}^{\rm pr}_n)$ is the solution of (B.2) at time $t = t^{k^{\rm pr}_n}$ for $\mu = \mu^{\rm pr}_n$ and $\Psi(\tilde{\mu}^{\rm du}_n)$ is the solution of (B.3) at time $t = t^{k^{\rm du}_n}$ for $\mu = \mu^{\rm du}_n$.

B.2.1 Formulation

Our reduced-basis approximation $y_N(\mu, t^k)$ to $y(\mu, t^k)$ is then obtained by a standard Galerkin projection: given $\mu \in \mathcal{D}$, $y_N(\mu, t^k) \in W^{\rm pr}_{N_{\rm pr}}$, $\forall\, k \in \mathbb{K}$ satisfies

$$\begin{aligned} m(y_N(\mu, t^k), v; \mu) &+ \frac{\Delta t}{2}\; a(y_N(\mu, t^k) + y_N(\mu, t^{k-1}), v; \mu) \\ &= m(y_N(\mu, t^{k-1}), v; \mu) + \frac{\Delta t}{2}\; b(v; \mu)\; (u(t^k) + u(t^{k-1})), \quad \forall\, v \in W^{\rm pr}_{N_{\rm pr}}, \quad \text{(B.7)} \end{aligned}$$

with initial condition $y_N(\mu, t^0) = 0$. Similarly, we obtain the reduced-basis approximation $\Psi_N(\mu, t^k) \in W^{\rm du}_{N_{\rm du}}$ to $\Psi(\mu, t^k)$ as the solution of

$$\begin{aligned} m(v, \Psi_N(\mu, t^k); \mu) &+ \Delta t\; a(v, \Psi_N(\mu, t^k) + \Psi_N(\mu, t^{k+1}); \mu) \\ &= m(v, \Psi_N(\mu, t^{k+1}); \mu), \quad \forall\, v \in W^{\rm du}_{N_{\rm du}},\ \forall\, k \in \mathbb{K}, \quad \text{(B.8)} \end{aligned}$$

with final condition

$$m(v, \Psi_N(\mu, t^{K+1}); \mu) \equiv \ell(v), \quad \forall\, v \in W^{\rm du}_{N_{\rm du}}. \qquad \text{(B.9)}$$

Finally, we evaluate the output estimate, $s_N(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, from

$$s_N(\mu, t^k) \equiv \ell(y_N(\mu, t^k)) + \sum_{k'=1}^{k} R^{\mathrm{pr}}(\tfrac{1}{2}(\Psi_N(\mu, t^{K-k+k'}) + \Psi_N(\mu, t^{K-k+k'+1})); \mu, t^{k'})\ \Delta t, \qquad \text{(B.10)}$$

where

$$\begin{aligned} R^{\mathrm{pr}}(v; \mu, t^k) \equiv \frac{1}{2}\, b(v; \mu)\, (u(t^k) + u(t^{k-1})) - \frac{1}{2}\, a(y_N(\mu, t^k) + y_N(\mu, t^{k-1}), v; \mu)& \\ - \frac{1}{\Delta t} m(y_N(\mu, t^k) - y_N(\mu, t^{k-1}), v; \mu), \quad \forall\, v \in Y,\ \forall\, k \in \mathbb{K},& \qquad \text{(B.11)} \end{aligned}$$

is the primal residual. Note that here $N \equiv (N_{\mathrm{pr}}, N_{\mathrm{du}})$.

B.2.2 *A Priori* Convergence Theory

We now consider the rate at which $y_N(\mu, t^k)$ converges to $y(\mu, t^k)$. We first note from (B.2) and (B.7) that $e^{\mathrm{pr}}(\mu, t^k) = y(\mu, t^k) - y_N(\mu, t^k)$ satisfies

$$\begin{aligned} m(e^{\mathrm{pr}}(\mu, t^k), v; \mu) + \frac{\Delta t}{2}\ a(e^{\mathrm{pr}}(\mu, t^k) + e^{\mathrm{pr}}(\mu, t^{k-1}), v; \mu) = m(e^{\mathrm{pr}}(\mu, t^{k-1}), v; \mu),& \\ \forall\, v \in W^{\mathrm{pr}}_{N_{\mathrm{pr}}},\ \forall\, k \in \mathbb{K},& \qquad \text{(B.12)} \end{aligned}$$

with initial condition $e^{\mathrm{pr}}(\mu, t^0) = y(\mu, t^0) - y_N(\mu, t^0) = 0$ since $y(\mu, t^0) = y_N(\mu, t^0) = 0$ by assumption. We next let $w_N(t^k) \in W^{\mathrm{pr}}_{N_{\mathrm{pr}}}$ be the projection of $y(\mu, t^k)$ with respect to the "m" scalar product and choose $v \equiv w_N(t^k) - y_N(\mu, t^k) + w_N(t^{k-1}) - y_N(\mu, t^{k-1}) = e^{\mathrm{pr}}(\mu, t^k) + e^{\mathrm{pr}}(\mu, t^{k-1}) - (y(\mu, t^k) - w_N(t^k)) - (y(\mu, t^{k-1}) - w_N(t^{k-1}))$ in (B.12). We then obtain

$$\begin{aligned} &m(e^{\mathrm{pr}}(\mu, t^k) - e^{\mathrm{pr}}(\mu, t^{k-1}), e^{\mathrm{pr}}(\mu, t^k) + e^{\mathrm{pr}}(\mu, t^{k-1}); \mu) \\ &\quad + \frac{\Delta t}{2}\ a(e^{\mathrm{pr}}(\mu, t^k) + e^{\mathrm{pr}}(\mu, t^{k-1}), e^{\mathrm{pr}}(\mu, t^k) + e^{\mathrm{pr}}(\mu, t^{k-1}); \mu) \\ &\qquad = m(e^{\mathrm{pr}}(\mu, t^k) - e^{\mathrm{pr}}(\mu, t^{k-1}), y(\mu, t^k) - w_N(t^k) + y(\mu, t^{k-1}) - w_N(t^{k-1}); \mu) \\ &\qquad\quad + \frac{\Delta t}{2}\ a(e^{\mathrm{pr}}(\mu, t^k) + e^{\mathrm{pr}}(\mu, t^{k-1}), y(\mu, t^k) - w_N(t^k) + y(\mu, t^{k-1}) - w_N(t^{k-1}); \mu), \end{aligned}$$

which can be written as

$$\begin{aligned} &m(e^{\mathrm{pr}}(\mu, t^k), e^{\mathrm{pr}}(\mu, t^k); \mu) - m(e^{\mathrm{pr}}(\mu, t^{k-1}), e^{\mathrm{pr}}(\mu, t^{k-1}); \mu) \\ &\quad + \frac{\Delta t}{2}\ a(e^{\mathrm{pr}}(\mu, t^k) + e^{\mathrm{pr}}(\mu, t^{k-1}), e^{\mathrm{pr}}(\mu, t^k) + e^{\mathrm{pr}}(\mu, t^{k-1}); \mu) \\ &\quad = m(y(\mu, t^k) - w_N(t^k) - (y(\mu, t^{k-1}) - w_N(t^{k-1})), y(\mu, t^k) - w_N(t^k) + y(\mu, t^{k-1}) - w_N(t^{k-1}); \mu) \\ &\qquad + \frac{\Delta t}{2}\ a(e^{\mathrm{pr}}(\mu, t^k) + e^{\mathrm{pr}}(\mu, t^{k-1}), y(\mu, t^k) - w_N(t^k) + y(\mu, t^{k-1}) - w_N(t^{k-1}); \mu), \qquad \text{(B.13)} \end{aligned}$$

since $m(z, y(\mu, t^k) - w_N(t^k)) = 0, \ \forall\, z \in W^{\rm pr}_{N_{\rm pr}}$. We next note that

$$\begin{aligned}
&\frac{\Delta t}{2}\; a(e^{\rm pr}(\mu,t^k) + e^{\rm pr}(\mu,t^{k-1}), y(\mu,t^k) - w_N(t^k) + y(\mu,t^{k-1}) - w_N(t^{k-1}); \mu) \\
&\qquad = \; \frac{\Delta t}{4}\; a(e^{\rm pr}(\mu,t^k) + e^{\rm pr}(\mu,t^{k-1}), e^{\rm pr}(\mu,t^k) + e^{\rm pr}(\mu,t^{k-1}); \mu) - \frac{\Delta t}{2}\; a(v,v;\mu) \\
&\qquad\quad +\frac{\Delta t}{4}\; a(y(\mu,t^k) - w_N(t^k) + y(\mu,t^{k-1}) - w_N(t^{k-1}), \\
&\qquad\qquad\qquad\qquad y(\mu,t^k) - w_N(t^k) + y(\mu,t^{k-1}) - w_N(t^{k-1}); \mu) \\
&\qquad \le \; \frac{\Delta t}{4}\; a(e^{\rm pr}(\mu,t^k) + e^{\rm pr}(\mu,t^{k-1}), e^{\rm pr}(\mu,t^k) + e^{\rm pr}(\mu,t^{k-1}); \mu) \\
&\qquad\quad +\frac{\Delta t}{4}\; a(y(\mu,t^k) - w_N(t^k) + y(\mu,t^{k-1}) - w_N(t^{k-1}), \\
&\qquad\qquad\qquad\qquad y(\mu,t^k) - w_N(t^k) + y(\mu,t^{k-1}) - w_N(t^{k-1}); \mu).
\end{aligned} \tag{B.14}$$

From (B.13) and (B.14) it follows that

$$\begin{aligned}
&m(e^{\rm pr}(\mu,t^k), e^{\rm pr}(\mu,t^k); \mu) - m(e^{\rm pr}(\mu,t^{k-1}), e^{\rm pr}(\mu,t^{k-1}); \mu) \\
&\quad +\frac{\Delta t}{4}\; a(e^{\rm pr}(\mu,t^k) + e^{\rm pr}(\mu,t^{k-1}), e^{\rm pr}(\mu,t^k) + e^{\rm pr}(\mu,t^{k-1}); \mu) \\
&\qquad = \; m(y(\mu,t^k) - w_N(t^k), y(\mu,t^k) - w_N(t^k); \mu) \\
&\qquad\quad -m(y(\mu,t^{k-1}) - w_N(t^{k-1}), y(\mu,t^{k-1}) - w_N(t^{k-1}); \mu) \\
&\qquad\quad +\frac{\Delta t}{4}\; a(y(\mu,t^k) - w_N(t^k) + y(\mu,t^{k-1}) - w_N(t^{k-1}), \\
&\qquad\qquad\qquad\qquad y(\mu,t^k) - w_N(t^k) + y(\mu,t^{k-1}) - w_N(t^{k-1}); \mu).
\end{aligned} \tag{B.15}$$

Summing from $k' = 1$ to k and invoking the coercivity and continuity of a and m we thus obtain

$$\begin{aligned}
&\alpha_m(\mu)\ \|e^{\rm pr}(\mu,t^k)\|_X^2 + \alpha_a(\mu)\ \Delta t \sum_{k'=1}^{k} \|\tfrac{1}{2}(e^{\rm pr}(\mu,t^{k'}) + e^{\rm pr}(\mu,t^{k'-1}))\|_Y^2 \\
&\qquad \le \; \sum_{k'=1}^{k} \inf_{w_N(t^{k'}) \in W^{\rm pr}_{N_{\rm pr}}} \gamma_a(\mu)\ \Delta t\ \|\tfrac{1}{2}(y(\mu,t^{k'}) - w_N(t^{k'}) + y(\mu,t^{k'-1}) - w_N(t^{k'-1}))\|_Y^2 \\
&\qquad\quad + \inf_{w_N(t^k) \in W^{\rm pr}_{N_{\rm pr}}} \{\gamma_m(\mu)\ \|y(\mu,t^k) - w_N(t^k)\|_X^2 \\
&\qquad\qquad\qquad + \gamma_a(\mu)\ \Delta t\ \|\tfrac{1}{2}(y(\mu,t^k) - w_N(t^k) + y(\mu,t^{k-1}) - w_N(t^{k-1}))\|_Y^2\}.
\end{aligned}$$

B.2.3 Offline-Online Computational Procedure

The offline-online computation procedure is by construction very similar to the Euler-Backward time integration scheme in Section 4.3.3. We therefore only state the matrix form of (B.7) and (B.8) and summarize the operation count.

We first express $y_N(\mu, t^k)$ and $\Psi_N(\mu, t^k)$ as

$$y_N(\mu, t^k) = \sum_{n=1}^{N_{\rm pr}} y_{Nn}(\mu, t^k)\ \zeta_n^{\rm pr}, \tag{B.16}$$

and

$$\Psi_N(\mu, t^k) = \sum_{n=1}^{N_{\rm du}} \Psi_{Nn}(\mu, t^k)\ \zeta_n^{\rm du}, \tag{B.17}$$

respectively. We then choose as test functions $v = \zeta_n^{\rm pr}$, $1 \le n \le N_{\rm pr}$, for the primal problem (B.7) and $v = \zeta_n^{\rm du}$, $1 \le n \le N_{\rm du}$, for the dual problem (B.8).

It then follows from (B.7) that $\underline{y}_N(\mu, t^k) = [y_{N\,1}(\mu, t^k)\ \ y_{N\,2}(\mu, t^k)\ \ \ldots\ y_{N\,N_{\rm pr}}(\mu, t^k)]^T \in \mathbb{R}^{N_{\rm pr}}$ satisfies

$$\begin{aligned}&(2\,M_N^{\rm pr}(\mu) + \Delta t\, A_N^{\rm pr}(\mu))\ \underline{y}_N(\mu, t^k)\\ &\qquad = (2\,M_N^{\rm pr}(\mu) - \Delta t\, A_N^{\rm pr}(\mu))\ \underline{y}_N(\mu, t^{k-1}) + \Delta t\ B_N^{\rm pr}(\mu)\ (u(t^k) + u(t^{k-1})), \quad \forall\, k \in \mathbb{K}, \end{aligned}\tag{B.18}$$

with initial condition $y_{N\,n}(\mu, t^0) = 0$, $1 \le n \le N_{\rm pr}$. For the dual problem we define $\underline{\Psi}_N(\mu, t^k) = [\Psi_{N\,1}(\mu, t^k)\ \ \Psi_{N\,2}(\mu, t^k)\ \ \ldots\ \Psi_{N\,N_{\rm du}}(\mu, t^k)]^T$ and obtain from (B.8) that

$$\Big(2\,M_N^{\rm du}(\mu) + \Delta t\, A_N^{\rm du}(\mu)\Big)\ \underline{\Psi}_N(\mu, t^k) = \Big(2\,M_N^{\rm du}(\mu) - \Delta t\, A_N^{\rm du}(\mu)\Big)\ \underline{\Psi}_N(\mu, t^{k+1}), \quad \forall\, k \in \mathbb{K}, \tag{B.19}$$

Note that $\underline{\Psi}_N(\mu, t^{K+1})$ is calculated from $M_N^{\rm du}(\mu)\ \underline{\Psi}_N(\mu, t^{K+1}) = L_N^{\rm du}$.

Finally, we evaluate the output estimate, $\forall\, k \in \mathbb{K}$, from

$$\begin{aligned} s_N(\mu, t^k) = L_N^{{\rm pr}\,T}\, \underline{y}_N(\mu, t^k) + \Delta t\ &\sum_{k'=1}^{k} \frac{1}{2}\, (\underline{\Psi}_N^T(\mu, t^{K-k+k'}) + \underline{\Psi}_N^T(\mu, t^{K-k+k'+1}))\\ \times \Bigg\{ \frac{1}{2}\, B_N^{\rm du}(\mu)\ &\Big(u(t^{k'}) + u(t'^{-1})\Big) - \frac{1}{2}\, A_N^{\rm pr,du}(\mu)\ \Big(\underline{y}_N(\mu, t^{k'}) + \underline{y}_N(\mu, t^{k'-1})\Big)\\ &- \frac{1}{\Delta t} M_N^{\rm pr,du}(\mu)\ \Big(\underline{y}_N(\mu, t^{k'}) - \underline{y}_N(\mu, t^{k'-1})\Big)\Bigg\}. \end{aligned}\tag{B.20}$$

Note that the quantities $M_N^{\rm pr,du}$, $A_N^{\rm pr,du}$ $B_N^{\rm pr,du}$ $L_N^{\rm pr,du}$ are already defined in A.5.

In the online stage — performed many times, for each new parameter value μ — we first assemble the reduced-basis matrices (4.55), (A.2), and (A.5); this requires $O((N_{\rm pr}^2 + N_{\rm du}^2 + N_{\rm pr}N_{\rm du})(Q_a + Q_m))$ operations. We then solve the primal and dual problem for $\underline{y}_N(\mu, t^k)$ and $\underline{\Psi}_N(\mu, t^k)$, respectively; since the reduced-basis matrices are in general full, the operation count (based on LU factorization and our LTI assumption) is $O(N_{\rm pr}^3 + N_{\rm du}^3 + K(N_{\rm pr}^2 + N_{\rm du}^2))$. Finally, given $\underline{y}_N(\mu, t^k)$ and $\underline{\Psi}_N(\mu, t^k)$ we evaluate the output estimate $s_N(\mu, t^k)$ from (B.20) at a cost of $O(2kN_{\rm pr}N_{\rm du})$; note that the calculation of all outputs $s_N(\mu, t^k)$, $\forall\, k \in \mathbb{K}$, requires $O(K(K+1)N_{\rm pr}N_{\rm du})$ operations.

B.3 *A Posteriori* Error Estimation

B.3.1 Preliminaries

To begin, we recall the definition of $\hat{\alpha}_a(\mu) : \mathcal{D} \to \boldsymbol{R}_+$ in (4.57) and $\hat{\alpha}_m(\mu) : \mathcal{D} \to \boldsymbol{R}_+$ in (4.58) as lower bounds for the coercivity constants $\alpha_a(\mu)$ and $\alpha_m(\mu)$, respectively. As in Section 4.4.1 we define the dual norm of the primal residual

$$\varepsilon^{\text{pr}}_{N_{\text{pr}}}(\mu, t^k) \equiv \sup_{v \in Y} \frac{R^{\text{pr}}(v; \mu, t^k)}{\|v\|_Y}, \quad \forall\, k \in \mathbb{K}, \tag{B.21}$$

and the dual norm of the dual residual

$$\varepsilon^{\text{du}}_{N_{\text{du}}}(\mu, t^k) \equiv \sup_{v \in Y} \frac{R^{\text{du}}(v; \mu, t^k)}{\|v\|_Y}, \quad \forall\, k \in \mathbb{K}, \tag{B.22}$$

where

$$\begin{aligned} R^{\text{du}}(v; \mu, t^k) \equiv & -\frac{1}{2} a(v, \Psi_N(\mu, t^k) + \Psi_N(\mu, t^{k+1}); \mu) \\ & - \frac{1}{\Delta t} m(v, \Psi_N(\mu, t^k) - \Psi_N(\mu, t^{k+1}); \mu), \quad \forall\, v \in Y,\ \forall\, k \in \mathbb{K}, \end{aligned} \tag{B.23}$$

is the dual residual.

We now present and prove the bounding properties for the errors in the primal variable, the dual variable, and the output estimate for the Crank-Nicolson time integration scheme. Throughout this section we assume that the "truth" solutions $y(\mu, t^k)$ and $\Psi(\mu, t^k)$ satisfy (B.2) and (B.3), respectively, and the corresponding reduced-basis approximations $y_N(\mu, t^k)$ and $\Psi_N(\mu, t^k)$ satisfy (B.7) and (B.8), respectively.

B.3.2 Error Bound Formulation

Primal Variable

We obtain the following result for the error in the primal variable.

Proposition 19. *Let $e^{\text{pr}}(\mu, t^k) \equiv y(\mu, t^k) - y_N(\mu, t^k)$ be the error in the primal variable and define the "spatio-temporal" energy norm*

$$\begin{aligned} |||v(\mu, t^k)|||^{\text{pr}}_{\text{CN}} \equiv \Bigg(& m(v(\mu, t^k), v(\mu, t^k); \mu) \\ & + \sum_{k'=1}^{k} a(\tfrac{1}{2}(v(\mu, t^{k'}) + v(\mu, t^{k'-1})), \tfrac{1}{2}(v(\mu, t^{k'}) + v(\mu, t^{k'-1})); \mu)\ \Delta t \Bigg)^{\frac{1}{2}}, \quad \forall\, v \in Y. \end{aligned} \tag{B.24}$$

The error in the primal variable is then bounded by

$$|||e^{\text{pr}}(\mu, t^k)|||^{\text{pr}}_{\text{CN}} \le \Delta^{\text{pr}}_{N_{\text{pr}}}(\mu, t^k), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{B.25}$$

where the error bound $\Delta^{\rm pr}_{N_{\rm pr}}(\mu, t^k)$ *is defined as*

$$\Delta^{\rm pr}_{N_{\rm pr}}(\mu, t^k) \equiv \left(\frac{\Delta t}{\hat{\alpha}_a(\mu)} \sum_{k'=1}^{k} \varepsilon^{\rm pr}_{N_{\rm pr}}(\mu, t^{k'})^2 \right)^{\frac{1}{2}} , \tag{B.26}$$

and $\varepsilon^{\rm pr}_{N_{\rm pr}}(\mu, t^k)$ *is the dual norm of the primal residual defined in (B.21).*

Proof. We immediately derive from (B.2) and (B.11) that $e^{\rm pr}(\mu, t^k) = y(\mu, t^k) - y_N(\mu, t^k)$ satisfies

$$\begin{aligned} m(e^{\rm pr}(\mu, t^k), v; \mu) + \frac{\Delta t}{2}\, a(e^{\rm pr}(\mu, t^k) + e^{\rm pr}(\mu, t^{k-1}), v; \mu)& \\ = m(e^{\rm pr}(\mu, t^{k-1}), v; \mu) + \Delta t\, R^{\rm pr}(v; \mu, t^k),& \quad \forall\, v \in Y,\ \forall\, k \in \mathbb{K}, \end{aligned} \tag{B.27}$$

where $e^{\rm pr}(\mu, t^0) = 0$ since $y(\mu, t^0) = y_N(\mu, t^0) = 0$ by assumption. We now choose $v = e^{\rm pr}(\mu, t^k) + e^{\rm pr}(\mu, t^{k-1})$, and invoke (B.21) to obtain

$$\begin{aligned} m(e^{\rm pr}(\mu, t^k), e^{\rm pr}(\mu, t^k); \mu) - m(e^{\rm pr}(\mu, t^{k-1}), e^{\rm pr}(\mu, t^{k-1}); \mu)& \\ + \frac{\Delta t}{2}\, a(e^{\rm pr}(\mu, t^k) + e^{\rm pr}(\mu, t^{k-1}), e^{\rm pr}(\mu, t^k) + e^{\rm pr}(\mu, t^{k-1}); \mu)& \\ \le \Delta t\, \varepsilon^{\rm pr}_{N_{\rm pr}}(\mu, t^k)\ \|e^{\rm pr}(\mu, t^k) + e^{\rm pr}(\mu, t^{k-1})\|_Y,& \quad \forall\, k \in \mathbb{K}. \end{aligned} \tag{B.28}$$

We now apply (4.69) with $c = \varepsilon^{\rm pr}_{N_{\rm pr}}(\mu, t^k)$, $d = \frac{1}{2}\|e^{\rm pr}(\mu, t^k) + e^{\rm pr}(\mu, t^{k-1})\|_Y$, and $\rho = \hat{\alpha}_a(\mu)^{\frac{1}{2}}$ to get

$$\begin{aligned} \varepsilon^{\rm pr}_{N_{\rm pr}}(\mu, t^k)\ \|e^{\rm pr}(\mu, t^k) + e^{\rm pr}(\mu, t^{k-1})\|_Y& \\ \le \frac{1}{\hat{\alpha}_a(\mu)}\, \varepsilon^{\rm pr}_{N_{\rm pr}}(\mu, t^k)^2 + \frac{\hat{\alpha}_a(\mu)}{4}\, \|e^{\rm pr}(\mu, t^k) + e^{\rm pr}(\mu, t^{k-1})\|^2_Y.& \end{aligned} \tag{B.29}$$

Combining (B.28) and (B.29), and invoking (4.7) and (4.57), we obtain

$$\begin{aligned} m(e^{\rm pr}(\mu, t^k), e^{\rm pr}(\mu, t^k); \mu) - m(e^{\rm pr}(\mu, t^{k-1}), e^{\rm pr}(\mu, t^{k-1}); \mu)& \\ + \Delta t\, a(\tfrac{1}{2}(e^{\rm pr}(\mu, t^k) + e^{\rm pr}(\mu, t^{k-1})), \tfrac{1}{2}(e^{\rm pr}(\mu, t^k) + e^{\rm pr}(\mu, t^{k-1})); \mu)& \\ \le \frac{\Delta t}{\hat{\alpha}_a(\mu)}\, \varepsilon^{\rm pr}_{N_{\rm pr}}(\mu, t^k)^2,& \quad \forall\, k \in \mathbb{K}, \end{aligned} \tag{B.30}$$

We now perform the sum from $k' = 1$ to k and recall that $e^{\rm pr}(\mu, t^0) = 0$, leading to

$$\begin{aligned} m(e^{\rm pr}(\mu, t^k), e^{\rm pr}(\mu, t^k); \mu)& \\ + \sum_{k'=1}^{k} \Delta t\, a(\tfrac{1}{2}(e^{\rm pr}(\mu, t^{k'}) + e^{\rm pr}(\mu, t^{k'-1})), \tfrac{1}{2}(e^{\rm pr}(\mu, t^{k'}) + e^{\rm pr}(\mu, t^{k'-1})); \mu)& \\ \le \frac{\Delta t}{\hat{\alpha}_a(\mu)} \sum_{k'=1}^{k} \varepsilon^{\rm pr}_{N_{\rm pr}}(\mu, t^{k'})^2,& \quad \forall\, k \in \mathbb{K}, \end{aligned} \tag{B.31}$$

which is the result stated in Proposition 19. □

Dual Variable

The treatment of the final condition of the dual problem for the Crank-Nicolson integration scheme is the same as for the Euler-Backward scheme. We recall from Lemma 7 and (4.74) that

$$m(e^{\mathrm{du}}(\mu,t^{K+1}),e^{\mathrm{du}}(\mu,t^{K+1});\mu) \le \hat{\alpha}_m(\mu)\,\Delta^{\Psi_f}_{N_{\mathrm{du}}}(\mu)^2; \tag{B.32}$$

where $\Delta^{\Psi_f}_{N_{\mathrm{du}}}(\mu)$ is defined in 4.76.

The bounding property for the dual problem is given in the following proposition.

Proposition 20. *Let $e^{\mathrm{du}}(\mu,t^k) \equiv \Psi(\mu,t^k) - \Psi_N(\mu,t^k)$ be the error in the dual variable and define*

$$|||v(\mu,t^k)|||^{\mathrm{du}}_{\mathrm{CN}} \equiv \Big(m(v(\mu,t^k),v(\mu,t^k);\mu) \\ + \sum_{k'=k}^{K} a(\tfrac{1}{2}(v(\mu,t^{k'})+v(\mu,t^{k+1})),\tfrac{1}{2}(v(\mu,t^{k'})+v(\mu,t^{k+1}));\mu)\ \Delta t\Big)^{\frac{1}{2}}. \tag{B.33}$$

The error in the dual variable is then bounded by

$$|||e^{\mathrm{du}}(\mu,t^k)|||^{\mathrm{du}}_{\mathrm{CN}} \le \Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k), \quad \forall\,\mu\in\mathcal{D},\ \forall\,k\in\mathbb{K}, \tag{B.34}$$

where the error bound $\Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k)$ is defined as

$$\Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k) \equiv \left(\frac{\Delta t}{\hat{\alpha}_a(\mu)}\sum_{k'=k}^{K}\varepsilon^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^{k'})^2 + \hat{\alpha}_m(\mu)\,\Delta^{\Psi_f}_{N_{\mathrm{du}}}(\mu)^2\right)^{\frac{1}{2}}, \tag{B.35}$$

and $\varepsilon^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k)$ is the dual norm of the dual residual defined in (B.22).

Proof. We immediately derive from (B.3) and (B.23) that $e^{\mathrm{du}}(\mu,t^k) = \Psi(\mu,t^k) - \Psi_N(\mu,t^k)$ satisfies

$$m(v,e^{\mathrm{du}}(\mu,t^k);\mu) + \frac{\Delta t}{2}\,a(v,e^{\mathrm{du}}(\mu,t^k)+e^{\mathrm{du}}(\mu,t^{k+1});\mu) \\ = m(v,e^{\mathrm{du}}(\mu,t^{k+1});\mu) + \Delta t\ R^{\mathrm{du}}(v;\mu,t^k), \quad \forall\,v\in Y,\ \forall\,k\in\mathbb{K}, \tag{B.36}$$

with final condition $m(v,e^{\mathrm{du}}(\mu,t^{K+1});\mu) = R^{\Psi_f}(v;\mu)$, $\forall\ v\in Y$. Choosing $v = e^{\mathrm{du}}(\mu,t^k) + e^{\mathrm{du}}(\mu,t^{k+1})$, and invoking (B.22) we obtain

$$m(e^{\mathrm{du}}(\mu,t^k),e^{\mathrm{du}}(\mu,t^k);\mu) - m(e^{\mathrm{du}}(\mu,t^{k+1}),e^{\mathrm{du}}(\mu,t^{k+1});\mu) \\ + \frac{\Delta t}{2}\,a(e^{\mathrm{du}}(\mu,t^k)+e^{\mathrm{du}}(\mu,t^{k+1}),e^{\mathrm{du}}(\mu,t^k)+e^{\mathrm{du}}(\mu,t^{k+1});\mu) \\ \le \Delta t\ \varepsilon^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k)\ \|e^{\mathrm{du}}(\mu,t^k)+e^{\mathrm{du}}(\mu,t^{k+1})\|_Y, \quad \forall\,k\in\mathbb{K}. \tag{B.37}$$

We now apply (4.69) with $c = \varepsilon^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k)$, $d = \frac{1}{2}\|e^{\mathrm{du}}(\mu,t^k)+e^{\mathrm{du}}(\mu,t^{k+1})\|_Y$, and $\rho = \hat{\alpha}_a(\mu)^{\frac{1}{2}}$.

Invoking (4.7) and (4.57), we arrive at

$$\begin{aligned} m(e^{\mathrm{du}}(\mu,t^k), e^{\mathrm{du}}(\mu,t^k);\mu) - m(e^{\mathrm{du}}(\mu,t^{k+1}), e^{\mathrm{du}}(\mu,t^{k+1});\mu) \\ + \Delta t\; a(\tfrac{1}{2}(e^{\mathrm{du}}(\mu,t^k) + e^{\mathrm{du}}(\mu,t^{k+1})), \tfrac{1}{2}(e^{\mathrm{du}}(\mu,t^k) + e^{\mathrm{du}}(\mu,t^{k+1}));\mu) \\ \leq \frac{\Delta t}{\hat{\alpha}_a(\mu)}\, \varepsilon^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k)^2, \quad \forall\, k \in \mathbb{K}, \end{aligned} \tag{B.38}$$

We now perform the sum from $k' = k$ to K and invoke (B.32) to obtain

$$\begin{aligned} m(e^{\mathrm{du}}(\mu,t^k), e^{\mathrm{du}}(\mu,t^k);\mu) \\ + \sum_{k'=k}^{K} \Delta t\; a(\tfrac{1}{2}(e^{\mathrm{du}}(\mu,t^{k'}) + e^{\mathrm{du}}(\mu,t^{k'+1})), \tfrac{1}{2}(e^{\mathrm{du}}(\mu,t^{k'}) + e^{\mathrm{du}}(\mu,t^{k'+1}));\mu) \\ \leq \frac{\Delta t}{\hat{\alpha}_a(\mu)} \sum_{k'=k}^{K} \varepsilon^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^{k'})^2 + \hat{\alpha}_m(\mu)\, \Delta^{\Psi_f}_{N_{\mathrm{du}}}(\mu)^2, \quad \forall\, k \in \mathbb{K}, \end{aligned} \tag{B.39}$$

which is the result stated in Proposition 20. □

Output Bound

Finally, the error bound for the output estimate is given in the following proposition.

Proposition 21. *Let the output of interest, $s(\mu,t^k)$, and the reduced-basis output estimate, $s_N(\mu,t^k)$, be given by*

$$s(\mu,t^k) = \ell(y(\mu,t^k)), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{B.40}$$

and

$$\begin{aligned} s_N(\mu,t^k) = \ell(y_N(\mu,t^k)) + \sum_{k'=1}^{k} R^{\mathrm{pr}}(\tfrac{1}{2}(\Psi_N(\mu,t^{K-k+k'}) + \Psi_N(\mu,t^{K-k+k'+1}));\mu,t^{k'})\,\Delta t, \\ \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \end{aligned} \tag{B.41}$$

respectively. The error in the output of interest is then bounded by

$$|s(\mu,t^k) - s_N(\mu,t^k)| \leq \Delta^s(\mu,t^k), \quad \forall\, \mu \in \mathcal{D},\ \forall\, k \in \mathbb{K}, \tag{B.42}$$

where the output bound $\Delta^s(\mu,t^k)$ is defined as

$$\Delta^s(\mu,t^k) \equiv \Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^k)\ \Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^{K-k+1}), \tag{B.43}$$

and $\Delta^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^k)$ and $\Delta^{\mathrm{du}}_{N_{\mathrm{du}}}(\mu,t^k)$ are defined in Propositions 19 and 20, respectively.

Proof. To begin, we recall the definition of the dual problem for the output at time t^L, $L \in \mathbb{K}$, given by

$$m(v, \psi_L(\mu, t^k); \mu) + \frac{\Delta t}{2}\, a(v, \psi_L(\mu, t^k) + \psi_L(\mu, t^{k+1}); \mu) = m(v, \psi_L(\mu, t^{k+1}); \mu), \qquad \forall\, v \in Y,\ (K \geq) L \geq k \geq 1, \quad \text{(B.44)}$$

with final condition $m(v, \psi_L(\mu, t^{L+1}); \mu) \equiv \ell(v)$, $\forall\, v \in Y$. We now choose $v = \frac{1}{2}(e^{\text{pr}}(\mu, t^k) + e^{\text{pr}}(\mu, t^{k-1}))$ in (B.44) and sum from $k = 1$ to L, to obtain

$$\frac{1}{2}\sum_{k'=1}^{L} m(e^{\text{pr}}(\mu, t^{k'}) + e^{\text{pr}}(\mu, t^{k'-1}), \psi_L(\mu, t^{k'}) - \psi_L(\mu, t^{k'+1}); \mu) + \sum_{k'=1}^{L} \frac{\Delta t}{4}\, a(e^{\text{pr}}(\mu, t^{k'}) + e^{\text{pr}}(\mu, t^{k'-1}), \psi_L(\mu, t^{k'}) + \psi_L(\mu, t^{k'+1}); \mu) = 0. \quad \text{(B.45)}$$

which can be rewritten in the form

$$\frac{1}{2}\sum_{k'=1}^{L} m(e^{\text{pr}}(\mu, t^{k'}) - e^{\text{pr}}(\mu, t^{k'-1}), \psi_L(\mu, t^{k'}) + \psi_L(\mu, t^{k'+1}); \mu) + \sum_{k'=1}^{L} \frac{\Delta t}{4}\, a(e^{\text{pr}}(\mu, t^{k'}) + e^{\text{pr}}(\mu, t^{k'-1}), \psi_L(\mu, t^{k'}) + \psi_L(\mu, t^{k'+1}); \mu) = m(e^{\text{pr}}(\mu, t^L), \psi_L(\mu, t^{L+1}); \mu) \quad \text{(B.46)}$$

where we used the fact that $e^{\text{pr}}(\mu, t^0) = 0$. We now note from the final condition of the dual problem that $m(e^{\text{pr}}(\mu, t^L), \psi_L(\mu, t^{L+1}); \mu) = \ell(e^{\text{pr}}(\mu, t^L))$ to obtain

$$\ell(e^{\text{pr}}(\mu, t^L)) = \frac{1}{2}\sum_{k'=1}^{L} m(e^{\text{pr}}(\mu, t^{k'}) - e^{\text{pr}}(\mu, t^{k'-1}), \psi_L(\mu, t^{k'}) + \psi_L(\mu, t^{k'+1}); \mu) + \sum_{k'=1}^{L} \frac{\Delta t}{4}\, a(e^{\text{pr}}(\mu, t^{k'}) + e^{\text{pr}}(\mu, t^{k'-1}), \psi_L(\mu, t^{k'}) + \psi_L(\mu, t^{k'+1}); \mu). \quad \text{(B.47)}$$

We next choose $v = \frac{1}{2}(\psi_L(\mu, t^k) + \psi_L(\mu, t^{k+1}))$ in the error equation for the primal variable, (B.27), and sum from $k = 1$ to L, to find

$$\frac{1}{2}\sum_{k'=1}^{L} m(e^{\text{pr}}(\mu, t^{k'}) - e^{\text{pr}}(\mu, t^{k'-1}), \psi_L(\mu, t^{k'}) + \psi_L(\mu, t^{k'+1}); \mu) + \sum_{k'=1}^{L} \frac{\Delta t}{4}\, a(e^{\text{pr}}(\mu, t^{k'}) + e^{\text{pr}}(\mu, t^{k'-1}), \psi_L(\mu, t^{k'}) + \psi_L(\mu, t^{k'+1}); \mu) = \sum_{k'=1}^{L} R^{\text{pr}}(\tfrac{1}{2}(\psi_L(\mu, t^{k'}) + \psi_L(\mu, t^{k'+1})); \mu, t^{k'})\ \Delta t. \quad \text{(B.48)}$$

From (B.47) and (B.48) we thus obtain

$$\begin{aligned}
\ell(e^{\mathrm{pr}}(\mu,t^L)) &= \sum_{k'=1}^{L} R^{\mathrm{pr}}(\tfrac{1}{2}(\psi_L(\mu,t^{k'})+\psi_L(\mu,t^{k'+1}));\mu,t^{k'})\ \Delta t. &\text{(B.49)}\\
&= \sum_{k'=1}^{L} R^{\mathrm{pr}}(\tfrac{1}{2}(\Psi(\mu,t^{K-L+k'})+\Psi(\mu,t^{K-L+k'+1}));\mu,t^{k'})\ \Delta t. &\text{(B.50)}
\end{aligned}$$

From the definition of $s(\mu,t^k)$ and $s_N(\mu,t^k)$, and (B.50) we now obtain

$$\begin{aligned}
s(\mu,t^k)-s_N(\mu,t^k) &= \ell(e^{\mathrm{pr}}(\mu,t^k))\\
&\quad -\sum_{k'=1}^{k} R^{\mathrm{pr}}(\tfrac{1}{2}(\Psi_N(\mu,t^{K-k+k'})+\Psi_N(\mu,t^{K-k+k'+1}));\mu,t^{k'})\ \Delta t &\text{(B.51)}\\
&= \sum_{k'=1}^{k} R^{\mathrm{pr}}(\tfrac{1}{2}(e^{\mathrm{du}}(\mu,t^{K-k+k'})+e^{\mathrm{du}}(\mu,t^{K-k+k'+1}));\mu,t^{k'})\ \Delta t. &\text{(B.52)}
\end{aligned}$$

Invoking (B.21) and the Cauchy-Schwarz inequality we arrive at

$$\begin{aligned}
|s(\mu,t^k)-s_N(\mu,t^k)| &\le \sum_{k'=1}^{k} \varepsilon^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^{k'})\ \|\tfrac{1}{2}(e^{\mathrm{du}}(\mu,t^{K-k+k'})+e^{\mathrm{du}}(\mu,t^{K-k+k'+1}))\|_Y\ \Delta t &\text{(B.53)}\\
&\le \left(\sum_{k'=1}^{k} \varepsilon^{\mathrm{pr}}_{N_{\mathrm{pr}}}(\mu,t^{k'})^2\ \Delta t\right)^{\frac{1}{2}}\\
&\qquad \times\left(\sum_{k'=1}^{k} \|\tfrac{1}{2}(e^{\mathrm{du}}(\mu,t^{K-k+k'})+e^{\mathrm{du}}(\mu,t^{K-k+k'+1}))\|_Y^2\ \Delta t\right)^{\frac{1}{2}}. &\text{(B.54)}
\end{aligned}$$

Let us first bound the second term on the right hand side. From (4.7) and the fact that $\hat{\alpha}_a(\mu)\le$

$\alpha(\mu)$, $\forall\, \mu \in \mathcal{D}$, we obtain

$$
\begin{aligned}
\sum_{k'=1}^{k} \|\tfrac{1}{2}(e^{\rm du}(\mu,t^{K-k+k'}) + e^{\rm du}(\mu,t^{K-k+k'+1}))\|_Y^2 \; \Delta t & \\
\leq \; \frac{1}{\hat{\alpha}_a(\mu)} \sum_{k'=1}^{k} a(\tfrac{1}{2}(e^{\rm du}(\mu,t^{K-k+k'}) + e^{\rm du}(\mu,t^{K-k+k'+1})), & \\
\tfrac{1}{2}(e^{\rm du}(\mu,t^{K-k+k'}) + e^{\rm du}(\mu,t^{K-k+k'+1})); \mu) \; \Delta t & \qquad \text{(B.55)} \\
= \; \frac{1}{\hat{\alpha}_a(\mu)} \sum_{k'=K-k+1}^{K} a(\tfrac{1}{2}(e^{\rm du}(\mu,t^{k'}) + e^{\rm du}(\mu,t^{k'+1})), & \\
\tfrac{1}{2}(e^{\rm du}(\mu,t^{k'}) + e^{\rm du}(\mu,t^{k'+1})); \mu) \, \Delta t & \qquad \text{(B.56)} \\
\leq \; \frac{1}{\hat{\alpha}_a(\mu)} \Bigg(\sum_{k'=K-k+1}^{K} a(\tfrac{1}{2}(e^{\rm du}(\mu,t^{k'}) + e^{\rm du}(\mu,t^{k'+1})), & \\
\tfrac{1}{2}(e^{\rm du}(\mu,t^{k'}) + e^{\rm du}(\mu,t^{k'+1})); \mu) \; \Delta t & \\
+ \; m(e^{\rm du}(\mu,t^{K-k+1}), e^{\rm du}(\mu,t^{K-k+1}); \mu) \Bigg) & \qquad \text{(B.57)} \\
= \; \frac{1}{\hat{\alpha}_a(\mu)} \left(|||e^{\rm du}(\mu,t^{K-k+1})|||^{\rm du}_{\rm CN} \right)^2, & \qquad \text{(B.58)}
\end{aligned}
$$

where the second inequality follows from the coercivity of $m(\cdot,\cdot;\mu)$ and the last equality from the definition (B.33) of the $|||\cdot|||^{\rm du}_{\rm CN}$-norm. Finally, inserting (B.58) into (B.54) and invoking (B.25) and (B.34), we obtain

$$
|s(\mu,t^k) - s_N(\mu,t^k)| \leq \Delta^{\rm pr}_{N_{\rm pr}}(\mu,t^k) \; \Delta^{\rm du}_{N_{\rm du}}(\mu,t^{K-k+1}), \qquad \text{(B.59)}
$$

which is the result stated in Proposition 21. □

B.3.3 Offline-Online Computational Procedure

The offline-online computational procedure is very similar to the bound calculation for the Euler-Backward scheme described in detail in Section A.2. Since the derivation in Section A.2 can easily be adjusted to the Crank-Nicolson scheme we omit the detailed equations here and only summarize the operation count.

The computational cost in the offline stage is (to leading order) $O((N_{\rm pr,max}+N_{\rm du,max})(Q_a+Q_m))$ solutions of the underlying "truth" finite element approximation and $O((N^2_{\rm pr,max} + N^2_{\rm du,max})(Q_a^2 + Q_aQ_m + Q_m^2))$ $\mathcal{N}$-inner products; the storage requirement is $O((N^2_{\rm pr,max} + N^2_{\rm du,max})(Q_a^2 + Q_aQ_m + Q_m^2))$. In the online stage — given a new parameter value μ and associated reduced-basis solutions $\underline{y}_N(\mu,t^k)$ and $\underline{\Psi}_N(\mu,t^k)$, $\forall\, k \in \mathbb{K}$ — the computational cost to evaluate $\Delta^s(\mu,t^k)$, $\forall\, k \in \mathbb{K}$, is $O(K(N^2_{\rm pr} + N^2_{\rm du})(Q_a^2 + Q_aQ_m + Q_m^2))$. Thus, all online calculations needed are *independent* of $\mathcal{N}$.

Appendix C

AP I: Delamination (Reference Domain)

C.1 Reference Domain

The fixed reference domain $\Omega \equiv [0, 30] \times [0, 11]$ for $\mu_{1,\mathrm{ref}} = 5$ (the geometry only depends on μ_1 and not on μ_2) of the delamination problem is shown in Figure C-1. We first divide Ω into 10 subdomains, Ω^i, $1 \leq i \leq 10$. The delamination is indicated by the magenta horizontal line, Γ_{del}, between the domains Ω^1, Ω^2 and Ω^5, Ω^7, respectively. We note that we only have to consider geometric variations, i.e., a stretch in the x_1-direction, in regions Ω^2, Ω^3, Ω^7, and Ω^8; the remaining regions do not vary with the delamination width w_{del}. We also note that the "fictitious" regions Ω_6 and Ω_{10} are only introduced for easier reference — the two outputs are defined as the average temperatures over these regions. The affine mapping does not require a distinction between the regions Ω_9 and Ω_{10} and the regions Ω_5 and Ω_6, respectively. We assume homogeneous Neumann boundary conditions on Γ_{N} and homogeneous Dirichlet boundary conditions on Γ_{D}.

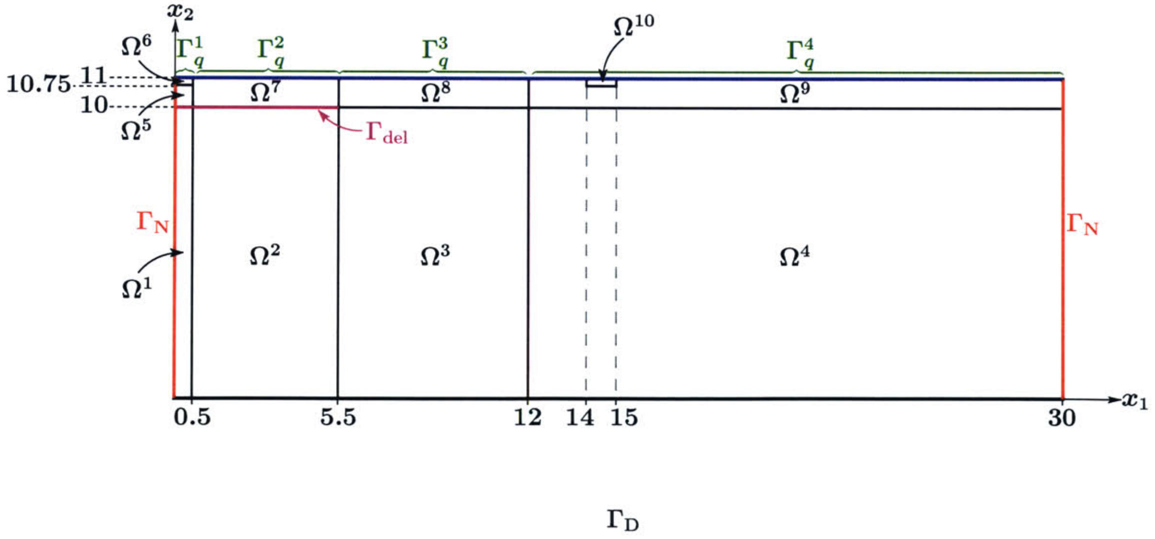


Figure C-1: AP I: Reference domain

C.2 Affine Decomposition

The affine decomposition resulting from the geometric mapping is given as follows (note that the output functional, ℓ, does not depend on μ):

Bilinear form $m(\cdot,\cdot;\mu)$: $Q_m = 3$

$$\begin{aligned}
\Theta^1_m(\mu) &= 1 & m^1(w,v) &= \int_{\Omega^1\cup\Omega^4\cup\Omega^5\cup\Omega^6\cup\Omega^9\cup\Omega^{10}} w\,v \\
\Theta^2_m(\mu) &= \frac{\mu_1-0.5}{5} & m^2(w,v) &= \int_{\Omega^2\cup\Omega^7} w\,v \\
\Theta^3_m(\mu) &= \frac{12-\mu_1}{6.5} & m^3(w,v) &= \int_{\Omega^3\cup\Omega^8} w\,v
\end{aligned} \tag{C.1}$$

Bilinear form $a(\cdot,\cdot;\mu)$: $Q_a = 10$

$$\begin{aligned}
\Theta^1_a(\mu) &= 1 & a^1(w,v) &= \int_{\Omega^1\cup\Omega^4} \nabla w\cdot\nabla v \\
\Theta^2_a(\mu) &= \mu_2 & a^2(w,v) &= \int_{\Omega^5\cup\Omega^6\cup\Omega^9\cup\Omega^{10}} \nabla w\cdot\nabla v \\
\Theta^3_a(\mu) &= \frac{5}{\mu_1-0.5} & a^3(w,v) &= \int_{\Omega^2} w_x\,v_x \\
\Theta^4_a(\mu) &= \frac{\mu_1-0.5}{5} & a^4(w,v) &= \int_{\Omega^2} w_y\,v_y \\
\Theta^5_a(\mu) &= \frac{6.5}{12-\mu_1} & a^5(w,v) &= \int_{\Omega^3} w_x\,v_x \\
\Theta^6_a(\mu) &= \frac{12-\mu_1}{6.5} & a^6(w,v) &= \int_{\Omega^3} w_y\,v_y \\
\Theta^7_a(\mu) &= \mu_2\frac{5}{\mu_1-0.5} & a^7(w,v) &= \int_{\Omega^7} w_x\,v_x \\
\Theta^8_a(\mu) &= \mu_2\frac{\mu_1-0.5}{5} & a^8(w,v) &= \int_{\Omega^7} w_y\,v_y \\
\Theta^9_a(\mu) &= \mu_2\frac{6.5}{12-\mu_1} & a^9(w,v) &= \int_{\Omega^8} w_x\,v_x \\
\Theta^{10}_a(\mu) &= \mu_2\frac{12-\mu_1}{6.5} & a^{10}(w,v) &= \int_{\Omega^8} w_y\,v_y
\end{aligned} \tag{C.2}$$

Linear form $b(\cdot;\mu)$: $Q_b = 3$

$$\begin{aligned}
\Theta^1_b(\mu) &= 1 & b^1(w,v) &= \int_{\Gamma^1_1\cup\Gamma^4_q} v\,d\Gamma \\
\Theta^2_b(\mu) &= \frac{\mu_1-0.5}{5} & b^2(w,v) &= \int_{\Gamma^2_q} v\,d\Gamma \\
\Theta^3_b(\mu) &= \frac{12-\mu_1}{6.5} & b^3(w,v) &= \int_{\Gamma^3_q} v\,d\Gamma
\end{aligned} \tag{C.3}$$

Bibliography

[1] M. Ainsworth and J. T. Oden. A posteriori *error estimation in finite element analysis.* Pure and applied mathematics. Wiley-Interscience, New York, 2000.

[2] O.M. Alifanov. *Inverse Heat Transfer Problems.* Springer, New York, 1994.

[3] F. Allgöwer and A. Zheng. *Nonliner Model Predictive Control.* Birkhäuser, 2000.

[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] R. Aris. *The Mathematical Thoery of Diffusion and Reaction in Permeable Catalysts, Vol.1.* Clarendon Press, 1975.

[6] R. Aris. *The Mathematical Thoery of Diffusion and Reaction in Permeable Catalysts, Vol.2.* Clarendon Press, 1975.

[7] J.A. Atwell and B.B. King. Proper orthogonal decomposition for reduced basis feedback controllers for parabolic equations. *Mathematical and Computer Modelling*, 33:1–19, 2001.

[8] Z. J. Bai. Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems. *Applied Numerical Mathematics*, 43(1-2):9–44, October 2002.

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

[10] E. Balsa-Canto, A.A. Alonso, and J.R. Banga. Reduced-order models for nonlinear distributed process systems and their application in dynamic optimization. *Industrial & Engineering Chemistry Research*, 43(13):3353–3363, June 2004.

[11] H.T. Banks. Parameter identification techniques for physiological control systems. In F. Hoppensteadt, editor, *Mathematical Aspects of Physiology*, volume 19 of *Lectures in Applied Mathematics*, pages 361–383. AMS, Providence, RI, 1981.

[12] H.T. Banks and J.M. Crowley. Parameter identification in continuum models. *Journal of Astronautical Science*, 33:85–94, 1985.

[13] H.T. Banks and P. Kareiva. Parameter estimation techniques for transport equations with application to population dispersal and tissue bulk flow models. *Journal of Mathematical Biology*, 17(3):253–273, 1983.

[14] H.T. Banks and K. Kunisch. *Estimation Techniques for Distributed Parameter Systems.* Systems & Control: Foundations & Applications. Birkhäuser, 1989.

[15] M. Barrault, N. C. Nguyen, Y. Maday, and A. T. Patera. An "empirical interpolation" method: Application to efficient reduced-basis discretization of partial differential equations. *C. R. Acad. Sci. Paris, Série I.*, 339:667–672, 2004.

[16] A. Barrett and G. Reddien. On the reduced basis method. *Z. Angew. Math. Mech.*, 75(7):543–549, 1995.

[17] R. Bartlett, P.T. Lin, A.G. Salinger, J.N. Shadid, and B. van Bloemen Wanders. Development of transport/inversion algorithms and capabilities for countermeasures to chem/bio/rad attacks in support of homeland security. In *Toward Real-Time & Online PDE-Constrained Optimization, Second Sandia Workshop on PDE-Constrained Optimization*, May 2004. Poster.

[18] J.V. Beck and K.J. Arnold. *Parameter Estimation.* Wiley, 1977.

[19] J.V. Beck, B. Blackwell, and C.R. St. Clair Jr. *Inverse Heat Conduction.* Wiley, New York, 1985.

[20] R. Becker and R. Rannacher. Weighted *a posteriori* error control in finite element methods. In *ENUMATH 95 Proc.* World Sci. Publ. Singapore, 1997.

[21] D. Bertsimas and J.N. Tsitsiklis. *Introduction to Linear Optimization.* Athena Scientific, 1997.

[22] N.F. Britton. *Reaction-Diffusion Equations and Their Applications to Biology.* Academic Press, 1986.

[23] K. Brooks, V. Balakotaiah, and D. Luss. Effect of natural convection on spontaneous combustion of coal stockpiles. *AIChE Journal*, 34(3):353–365, 1988.

[24] A.E. Bryson and Y.-C. Ho. *Applied Optimal Control.* Taylor and Francis, revised printing edition, 1975.

[25] T.T. Bui, M. Damodaran, and K. Wilcox. Proper orthogonal decomposition extensions for parametric applications in transonic aerodynamics (AIAA Paper 2003-4213). In *Proceedings of the 15th AIAA Computational Fluid Dynamics Conference*, June 2003.

[26] G. Caristi and E. Mitidieri, editors. *Reaction diffusion systems*, volume 194 of *Lecture notes in pure and applied mathematics.* Marcel Dekker, 1997.

[27] B. Chalmond. *Modeling and Inverse Problems in Image Analysis.* Springer, 2003.

[28] J. Chen and S-M. Kang. Model-order reduction of nonlinear mems devices through arclength-based karhunen-loéve decomposition. In *Proceeding of the IEEE international Symposium on Circuits and Systems*, volume 2, pages 457–460, 2001.

[29] Y. Chen and J. White. A quadratic method for nonlinear model order reduction. In *Proceeding of the international Conference on Modeling and Simulation of Microsystems*, pages 477–480, 2000.

[30] E.A. Christensen, M. Brøns, and J.N. Sørensen. Evaluation of proper orthogonal decomposition-based decomposition techniques applied to parameter-dependent nonturbulent flows. *SIAM J. Scientific Computing*, 21(4):1419–1434, 2000.

[31] S.T. Clegg and R.B. Roemer. Reconstruction of experimental hyperthermia temperature distributions: Application of state and parameter estimation. *ASME Journal of Biomechanical Engineering*, 115:380–388, 1993.

[32] G. Continillo, S. Crescitelli, and M. Giona, editors. *Nonlinear Dynamics and Control in Process Engineering - Recent Advances.* Springer, 2002.

[33] J. W. Demmel. *Applied numerical linear algebra.* SIAM, Philadelphia, 1997.

[34] W. Desch, F. Kappel, and K. Kunisch, editors. *Control and Estimation of Distributed Parameter Systems*, volume 126 of *International Series of Numerical Mathematics.* Birkhäuser, 1998.

[35] Earl H. Dowell and Kenneth C. Hall. Modeling of fluid structure interaction. *Annu. Rev. Fluid. Mech.*, 33:445–490, 2001.

[36] N.H. El-Farra and P.D. Christofides. Coordinating feedback and switching for control of spatially distributed processes. *Computers and Chemical Engineering*, 28:111–128, 2004.

[37] 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.

[38] Nancy Garcia. Airport tests bio/chem technologies. *Sandia Technology*, 5(2):23–24, 2003.

[39] B. Gärtner and S. Schönherr. Smallest enclosing ellipses – fast and exact. Technical Report B 97-03, Freie Univsersität Berlin, 1997.

[40] M. A. Grepl, N. C. Nguyen, K. Veroy, A. T. Patera, and G. R. Liu. Certified rapid solution of parametrized partial differential equations for real-time applications. In *Proceedings of the 2^{nd} Sandia Workshop of PDE-Constrained Optimization: Towards Real-Time and On-Line PDE-Constrained Optimization*, SIAM Computational Science and Engineering Book Series, 2004. Submitted.

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

[42] P.M. Gresho and R.L. Sani. *Incompressible Flow and the Finite Element Method: Advection-Diffusion and Isothermal Laminar Flow.* John Wiley & Sons, 1998.

[43] M. D. Gunzburger. *Finite Element Methods for Viscous Incompressible Flows: A Guide to Theory, Practice, and Algorithms.* Academic Press, Boston, 1989.

[44] G.F. Hawkins, E.C. Johnson, and J.P. Nokes. Detecting manufacturing flaws in composite retrofits. *SPIE*, 3587:97–104, 1999.

[45] K.-H. Hoffmann, G. Leugering, and F. Tröltzsch, editors. *Optimal Control of Partial Differential Equations*, volume 133 of *International Series of Numerical Mathematics.* Birkhäuser, 1998.

[46] L. Hörmander. *Linear Partial Differential Operators*, volume 1. Springer-Verlag, 1964.

[47] K. Ito and S. S. Ravindran. A reduced basis method for control problems governed by PDEs. In W. Desch, F. Kappel, and K. Kunisch, editors, *Control and Estimation of Distributed Parameter Systems*, pages 153–168. Birkhäuser, 1998.

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

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

[50] Pierre Jamet. Galerkin-type approximations which are discontinuous in time for parabolic equations in a variable domain. *SIAM Journal on Numerical Analysis*, 15(5):912–928, Oct. 1978.

[51] C. Johnson. *Numerical solution of partial differential equations by the finite element method.* Cambridge University Press, 1987.

[52] V.J. Keilis-Borok and T.B. Yanovskaya. Inverse problems of seismology (structural review). *Geophys. J. Royal Astr. Soc.*, 13:223–234, 1967.

[53] C. Kravaris and J.H. Seinfeld. Identification of parameters in distributed systems by regularization. *SIAM Journal on Control and Optimization*, 23:217–241, 1985.

[54] P. Krysl, S. Lall, and J.E. Marsden. Dimensional model reduction in non-linear finite element dynamics of solids and structures. *International Journal for Numerical Methods in Engineering*, 51:479–504, 2001.

[55] S. Lall, J. E. Marsden, and S. Glavaski. A subspace approach to balanced truncation for model reduction of nonlinear control systems. *Int. J. Robust Nonlinear Control*, 12:519–535, 2002.

[56] M. Lin Lee. Estimation of the error in the reduced basis method solution of differential algebraic equation systems. *SIAM Journal on Numerical Analysis*, 28(2):512–528, 1991.

[57] J.L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations.* Springer, 1971.

[58] J.L. Lions. *Some Aspects of the Optimal Control of Distributed Parameter Systems.* Regional Conference Series in Applied Mathematics. SIAM, 1972.

[59] J. Lumley and P. Blossey. Control of turbulence. *Annu. Rev. Fluid. Mech.*, 30:311–327, 1998.

[60] 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.

[61] 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(3):289–294, 2002.

[62] N.D. Malmuth, W.F. Hall, B.I. Davis, and C.D. Rosen. Transient thermal phenomena and weld geometry in gtaw. *Welding Journal*, 53(9):388s–400s, 1974.

[63] C. G. Markidakis and I. Babuska. On the stability of the discontinuous galerkin method for the heat equation. *SIAM Journal of Numerical Analysis*, 34(1):389–401, 1997.

[64] D.W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[65] M. Mattingly, E.A. Bailey, A.W. Dutton, R.B. Roemer, and S. Devasia. Reduced-order modeling for hyperthermia: An extended balanced-realization-based approach. *IEEE Transactions on Biomedical Engineering*, 45(9):1154–1162, September 1998.

[66] M. Mattingly, R.B. Roemer, and S. Devasia. Exact temperature tracking for hyperthermia: A model-based approach. *IEEE Transactions on Control Systems Technology*, 8(6):979–992, November 2000.

[67] D.Q. Mayne, J.B. Rawlings, C.V. Rao, and P.O.M. Scokaert. Constrained model predictive control: stability and optimality. *Automatica*, 36(6):789–814, 2000.

[68] M. Meyer and H. G. Matthies. Efficient model reduction in non-linear dynamics using the karhunen-loève expansion and dual-weighted-residual methods. *Computational Mechanics*, 31(1-2):179–191, May 2003.

[69] B.C. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, 26(1):17–32, 1981.

[70] M. Morari and J.H. Lee. Model predictive control: Past, present and future. *Comp. and Chem. Eng.*, 23(4/5):667–682, 1999.

[71] 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.

[72] Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*, volume 13 of *SIAM studies in applied mathematics*. Society of Industrial and Applied Mathematics, 1993.

[73] M. Neuilly. *Modelling and Estimation of Measurement Errors*. Lavoisier, 1999.

[74] N. C. Nguyen. *Reduced-Basis Approximation and* A Posteriori *Error Bounds for Nonaffine and Nonlinear Partial Differential Equations: Application to Inverse Analysis*. PhD thesis, Singapore-MIT Alliance, National University of Singapore., 2005. In progress.

[75] N. C. Nguyen, G. R. Liu, and A. T. Patera. Reduced-basis methods for inverse problems in partial differential equations. In *Proceedings Singapore-MIT Alliance Symposium*, January 2004.

[76] N. C. Nguyen, K. Veroy, and A. T. Patera. Certified real-time solution of parametrized partial differential equations. In S. Yip, editor, *Handbook of Materials Modeling*, pages 1523–1558. Springer, 2005.

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

[78] I. B. Oliveira. *A "HUM" Conjugate Gradient Algorithm for Constrained Nonlinear Optimal Control: Terminal and Regulator Problems.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, February 2002.

[79] I. B. Oliveira and A. T. Patera. Reduced-basis techniques for rapid reliable optimization of systems described by affinely parametrized coercive elliptic partial differential equations. *Optimization and Engineering*, 2004. Submitted.

[80] M.N. Özişik and H.R.B. Orlande. *Inverse Heat Transfer.* Taylor & Francis, New York, 2000.

[81] M. Paraschivoiu and A. T. Patera. A hierarchical duality approach to bounds for the outputs of partial differential equations. *Comp. Meth. Appl. Mech. Engrg.*, 158(3-4):389–407, June 1998.

[82] H.M. Park, T.Y. Yoon, and O.Y. Kim. Optimal control of rapid thermal processing systems by empirical reduction of modes. *Ind. Eng. Chem. Res.*, 38:3964–3975, 1999.

[83] J. S. Peterson. The reduced basis method for incompressible viscous flow calculations. *SIAM J. Sci. Stat. Comput.*, 10(4):777–786, July 1989.

[84] J.R. Phillips. Projection-based approaches for model reduction of weakly nonlinear systems, time-varying systems. In *IEEE Transactions On Computer-Aided Design of Integrated Circuit and Systems*, volume 22, pages 171–187, 2003.

[85] 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.

[86] T. A. Porsching and M. Lin Lee. The reduced basis method for initial value problems. *SIAM Journal on Numerical Analysis*, 24(6):1277–1287, 1987.

[87] M.J. Post. A minimum spanning ellipse algorithm. In *Proc. 22nd IEEE Symposium on Foundations of Computer Science*, October 1981.

[88] M.J. Post. Computing minimum spanning ellipses. Technical Report CS-82-16, Brown University, Department of Computer Science, 1982.

[89] J.K. Potocki and H.S. Tharp. Reduced-order modeling for hyperthermia control. *IEEE Transactions on Biomedical Engineering*, 39:1265–1273, 1992.

[90] F. Press. Earth models obtained by monte-carlo inversion. *Journal of Geophysics Research*, 73(16):5223–5234, 1968.

[91] 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.

[92] S.J. Qin and T.A. Badgwell. An overview of industrial model predictive control technology. In J.C. Kantor, C.E. Garcia, and B. Carnahan, editors, *Fifth International Conference on Chemical Process Control — CPC V*, pages 232–256. American Institute of Chemical Engineers, 1996.

[93] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical Mathematics*, volume 37 of *Texts in Applied Mathematics*. Springer, New York, 1991.

[94] A. Quarteroni and A. Valli. *Numerical Approximation of Partial Differential Equations*. Springer, 2nd edition, 1997.

[95] S. S. Ravindaran. A reduced-order approach for optimal control of fluids using proper orthogonal decomposition. *Int. J. Numer. Meth. Fluids*, 34:425–448, 2000.

[96] J.B. Rawlings. Tutorial overview of model predictive control. *IEEE Control Systems Magazine*, 20(3):38–52, 2000.

[97] M. Rewienski and J. White. A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices. In *IEEE Transactions On Computer-Aided Design of Integrated Circuit and Systems*, volume 22, pages 155–170, 2003.

[98] 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.

[99] D.H. Rothman. Nonlinear inversion, statistical mechanics, and residual statics estimation. *Geophysics*, 50:2797–2807, 1985.

[100] D.H. Rothman. Automatic estimation of large residual statics corrections. *Geophysics*, 51(332-346), 1986.

[101] D. V. Rovas, L. Machiels, and Y. Maday. Reduced-basis output bound methods for parabolic problems. *IMA Journal of Applied Mathematics*, 2004. Submitted.

[102] D.V. Rovas. *Reduced-Basis Output Bound Methods for Parametrized Partial Differential Equations*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, October 2002.

[103] A.G. Salinger, R. Aris, and J.J. Derby. Modeling the spontaneous ignition of coal stockpiles. *AIChE Journal*, 40(6):991–1004, 1994.

[104] J. M. A. Scherpen. Balancing for nonlinear systems. *Systems and Control Letters*, 21:143–153, 1993.

[105] D. Schmal, J.H. Duyzer, and J.W. van Heuven. A model for the spontaneous heating of coal. *Fuel*, 64:963–972, 1985.

[106] Dominik Schötzau and Christoph Schwab. Time discretization of parabolic problems by the hp-version of the discontinuous galerkin finite element method. *SIAM Journal of Numerical Analysis*, 38(3):837–875, 2000.

[107] B.W. Silverman and D.M. Titterington. Minimum covering ellipses. *SIAM J. Sci. Statist. Comput.*, 1:401–409, 1980.

[108] L. Sirovich. Turbulence and the dynamics of coherent structures, part 1: Coherent structures. *Quarterly of Applied Mathematics*, 45(3):561–571, October 1987.

[109] 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.

[110] Y. Solodukhov. *Reduced-Basis Methods Applied to Locally Non-Affine and Non-Linear Partial Differential Equations.* PhD thesis, Massachusetts Institute of Technology, 2005.

[111] J.-B. Song and D.E. Hardt. Estimation of weld bead depth for in-process control. In *Automation of Manufacturing Processes*, volume DSC-22 of *ASME Winter Annual Meeting*, pages 39–45, 1990.

[112] J.-B. Song and D.E. Hardt. Closed-loop control of weld pool depth using a thermally based depth estimator. *Welding Journal*, 72(10):S471–S478, 1993.

[113] J.-B. Song and D.E. Hardt. Dynamic modeling and adaptive control of the gas metal arc welding process. *Journal of Dynamic Systems, Measurement, and Control*, 116:405–413, 1994.

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

[115] N. Sun, N.-Z. Sun, M. Elimelech, and J.N. Ryan. Sensitivity analysis and parameter identifiability for colloid transport in geochemically heterogeneous porous media. *Water Resources Research*, 37(2):209–222, 2001.

[116] A. Tarantola. *Inverse problem theory and methods for model parameter estimation.* Siam, 2005.

[117] Vidar Thomee. *Galerkin Finite Element Methods for Parabolic Problems*, chapter The Discontinuous Galerkin Time Stepping Method, pages 181–208. Springer Series in Computational Mathematics. Springer, June 1997.

[118] A.N. Tikhonov and V.Y. Arsenin. *Solutions of Ill-Pose Problems.* Wiley, New York, 1977.

[119] San Francisco International Airport unveils chem/bio defense collaboration with Sandia. Nancy garcia. *SandiaLabNews*, 55(9), 2003.

[120] K. Veroy. *Reduced-Basis Methods Applied to Problems in Elasticity: Analysis and Applications.* PhD thesis, Massachusetts Institute of Technology, 2003.

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

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

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

[124] 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.

[125] D.S. Weile, E. Michielssen, and K. Gallivan. Reduced-order modeling of multiscreen frequency-selective surfaces using Krylov-based rational interpolation. *IEEE Transactions on Antennas and Propagation*, 49(5):801–813, May 2001.

[126] E. Welzl. Smallest enclosing disks (balls and ellipsoids). In H. Mauerer, editor, *New Results and New Trends in Computer Science*, volume 555 of *Lecture Notes in Computer Science*, pages 359–370. Springer-Verlag, 1991.

[127] K. Willcox and J. Peraire. Balanced model reduction via the proper orthogonal decomposition. In *15th AIAA Computational Fluid Dynamics Conference*. AIAA, June 2001.

[128] W.W.-G. Yeh. Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water Resource Research*, 22(2):95–108, 1986.