A Certified Reduced Basis Approach to PDE-constrained Parameter Optimization with Quadratic Cost Functionals

Elizabeth Qian*, Martin Grepl†, Karen Veroy‡, and Karen Willcox§

January 15, 2016

Abstract

Parameter optimization problems constrained by partial differential equations (PDEs) appear in many science and engineering applications. Solving these optimization problems may require a prohibitively large number of computationally expensive PDE solves, especially if there are many variable parameters. It is therefore advantageous to replace expensive high-dimensional PDE solvers (e.g. finite element) with lower-dimension surrogate models. In this paper, we use the reduced basis (RB) model reduction method in conjunction with a trust region optimization framework to accelerate PDE-constrained parameter optimization. New *a posteriori* error bounds on the RB cost and cost gradient for quadratic cost functionals are presented, and used to guarantee convergence to the optimum of the high-fidelity model. The proposed certified RB trust region approach thus requires only a minimal number of high-order solves, used to update the RB model if the approximation is no longer sufficiently accurate. We consider problems governed by elliptic PDEs and present numerical results for a thermal fin model problem with six parameters.

1 Introduction

PDE-constrained parameter optimization problems form a broad class of problems with applications across engineering and science disciplines. In an engineering design setting, PDE-constrained parameter optimization is used to determine the optimal parameters of a system of interest, such as an airplane wing or a thermal fin. Because typical optimization algorithms require many simulations of the system dynamics, using classical discretization techniques (e.g. finite element) to solve these problems may be prohibitively expensive. One way to address this issue is to make use of surrogate models of reduced dimension.

Numerous model reduction approaches exist and have been successfully applied to PDE-constrained optimization problems. In this paper, we build on the reduced basis (RB) method, a projection-based model reduction technique first developed in the late 1970s that supports rigorous *a posteriori* error estimation (see [1] for a review). Like other projection-based methods, the RB method is traditionally divided into a computationally expensive offline phase, during which the reduced basis is built to be globally accurate over the entire parameter domain, and an online phase, in which the model may be efficiently evaluated at any parameter within the admissible domain, making it particularly suited to many-query and real-time contexts. However, although optimization is a many-query context, the traditional online-offline decomposition is not particularly efficient for optimization, because typical gradient-based optimization algorithms require only local accuracy of the model along the optimization trajectory. A significant portion of the computational expenditure in ensuring global accuracy of the model is thus wasted.

The trust region optimization framework provides us with a natural way to break with the traditional online-offline divide and progressively build the RB model along the optimization trajectory. This work builds on several contributions on the use of surrogate models in trust region optimization. Alexandrov et al. show convergence of the trust region approach when using models that satisfy the first-order condition [5]. Others have explored

*Department of Aeronautics and Astronautics, MIT, Cambridge, MA 02139 (elizqian@mit.edu)

†Institute for Geometry and Applied Mathematics, RWTH Aachen University, 52062 Aachen, Germany (grepl@igpm.rwth-aachen.de)

‡Aachen Institute for Advanced Study in Computational Engineering Science (AICES), RWTH Aachen University, 52062 Aachen, Germany (veroy@aices.rwth-aachen.de)

§Department of Aeronautics and Astronautics, MIT, Cambridge, MA 02139 (kwillcox@mit.edu)

the use of specific model reduction methods, including POD [6, 7], Krylov interpolatory methods [8], and, in the stochastic context, sparse grids [9]. In particular, we make use of a result from Yue and Meerbergen which guarantees convergence of a trust region framework using surrogate models for which error bounds exist [8].

In this paper, we use reduced basis models and the associated *a posteriori* error bounds to demonstrate convergence of a trust region reduced basis framework for parametrized linear elliptic PDEs with quadratic cost functionals. Section 2 presents the problem formulation and introduces our notation. Section 3 gives an overview of the reduced basis method and presents the reduced basis error bounds needed for the trust region algorithm. In Section 4, we introduce the trust region framework, discuss the convergence theory, and present the combined trust region reduced basis optimization algorithm. Section 5 demonstrates the algorithm on a thermal fin model problem and performance of the algorithm is compared to that of a standard quasi-Newton method.

2 Problem formulation

Let Ω be a physical domain in $\mathbb{R}^d$ with Lipschitz continuous boundary $\partial\Omega$. We define the Sobolev space $X^e \equiv H_0^1(\Omega) \subset H^1(\Omega)$, where $H^1(\Omega) = \left\{ v \mid v \in L^2(\Omega), \nabla v \in \left(L^2(\Omega)\right)^d \right\}$ and $H_0^1(\Omega) = \left\{ v \mid v \in H^1(\Omega), v|_{\partial\Omega} = 0 \right\}$. We associate with X^e the inner product $(w,v)_{X^e} \equiv \int_\Omega \nabla w \cdot \nabla v + \int_\Omega wv$, $\forall w, v \in X^e$, as well as the induced norm $\|\cdot\|_{X^e} = \sqrt{(\cdot,\cdot)_{X^e}}$. Additionally, we define $Y^e \equiv L^2(\Omega)$, with associated inner product $(w,v)_{Y^e} \equiv \int_\Omega wv$, $\forall w, v \in Y^e$ and induced norm $\|\cdot\|_{Y^e} = \sqrt{(\cdot,\cdot)_{Y^e}}$. The superscript $\cdot^e$ indicates that we are dealing with the "exact" continuous domain.

We now introduce the weak form of the μ-parametrized elliptic PDE,

$$a(u^e(\mu), v; \mu) = f(v), \quad \forall v \in X^e, \tag{1}$$

and quadratic cost functional of the PDE solution vector,

$$J^e(\mu) = d(u^e(\mu), u^e(\mu)) + \ell(u^e(\mu)), \tag{2}$$

where μ is a parameter vector in the parameter domain $\mathcal{D} \subset \mathbb{R}^P$, $a(\cdot,\cdot;\mu)$ and $d(\cdot,\cdot)$ are X^e-continuous symmetric bilinear forms, and $\ell(\cdot)$ and $f(\cdot)$ are Y^e-continuous linear forms. We note that this cost definition fits the standard output least squares formulation. The optimization problem is to find the parameter μ^e_{opt} which minimizes the cost $J^e(\mu)$, i.e.,

$$\mu^e_{opt} = \arg\min_{\mu \in \mathcal{D}} J^e(\mu). \tag{3}$$

We make several assumptions on the bilinear forms that appear in the above formulation. First, we assume continuity of $a(\cdot,\cdot;\mu)$ and $d(\cdot,\cdot)$, i.e.,

$$a(w,v;\mu) \le \gamma_a(\mu) \|w\|_X \|v\|_X \le \gamma_{a0} \|w\|_X \|v\|_X, \quad \forall w, v \in X, \forall \mu \in \mathcal{D} \tag{4}$$

and

$$d(w,v) \le \gamma_d \|w\|_X \|v\|_X, \quad \forall w, v \in X, \forall \mu \in \mathcal{D}, \tag{5}$$

where $\gamma_a(\mu)$ and γ_d are the continuity constants of $a(\cdot,\cdot;\mu)$ and $d(\cdot,\cdot)$, respectively. We also assume coercivity of $a(\cdot,\cdot;\mu)$, i.e.,

$$0 < \alpha_0 \le \alpha(\mu) \equiv \inf_{v \in X} \frac{a(v,v;\mu)}{\|v\|_X^2}, \quad \forall \mu \in \mathcal{D}, \tag{6}$$

where $\alpha(\mu)$ is the coercivity constant, and that $a(\cdot,\cdot;\mu)$ exhibits affine dependence on the parameter μ, i.e. we can write the form as

$$a(w,v;\mu) = \sum_{q=1}^{Q_a} \Theta_a^q(\mu) a^q(w,v) \quad \forall w, v \in X, \forall \mu \in \mathcal{D}, \tag{7}$$

where Q_a is an integer. For simplicity, we assume that the forms $f(\cdot)$, $d(\cdot,\cdot)$, and $\ell(\cdot)$ are parameter-independent, although extensions to affine dependence are readily admitted [1].

To make use of gradient-based optimization methods, we require not only the cost but also its derivatives. Derivatives of PDE outputs may be efficiently calculated using adjoint methods, which allow derivatives in all

directions to be computed by solving just two equations. We thus introduce the adjoint (dual) problem associated with our primal problem and cost [2], given by

$$a\left(v,\psi^{e}\left(\mu\right);\mu\right)=2d\left(u^{e}\left(\mu\right),v\right)+\ell\left(v\right),\quad\forall v\in X^{e}. \tag{8}$$

For ease of notation, we assume a single parameter μ, but the extension to multiple parameters is straightforward. We may thus calculate the derivative of the cost function with respect to the parameter μ via

$$\nabla_{\mu}J^{e}\left(\mu\right)=-a_{\mu}\left(u^{e}\left(\mu\right),\psi^{e}\left(\mu\right);\mu\right), \tag{9}$$

where the sensitivity bilinear form $a_{\mu}\left(\cdot,\cdot;\mu\right)$ is the derivative of $a\left(\cdot,\cdot;\mu\right)$ with respect to the parameter μ. We assume that $a_{\mu}\left(\cdot,\cdot;\mu\right)$ is continuous, i.e.,

$$a_{\mu}\left(w,v;\mu\right)\leq\gamma_{a_{\mu}}\left(\mu\right)\left\|w\right\|_{X}\left\|v\right\|_{X}\leq\gamma_{a_{\mu}0}\left\|w\right\|_{X}\left\|v\right\|_{X},\quad\forall w,v\in X,\forall\mu\in\mathcal{D}. \tag{10}$$

Affine dependence of $a\left(\cdot,\cdot;\mu\right)$ implies that the sensitivity bilinear form $a_{\mu}\left(\cdot,\cdot;\mu\right)$ also exhibits affine dependence on the parameter μ, i.e. it follows from (7) that

$$a_{\mu}\left(w,v;\mu\right)=\sum_{q=1}^{Q_{a}}\frac{\partial\Theta_{a}^{q}\left(\mu\right)}{\partial\mu}a^{q}\left(w,v\right)\quad\forall w,v\in X,\forall\mu\in\mathcal{D}. \tag{11}$$

We assume that we have access to $\alpha_{LB}\left(\mu\right)$, a lower bound on $\alpha\left(\mu\right)$, and to $\gamma_{a_{\mu}}^{UB}\left(\mu\right)$, an upper bound on $\gamma_{a_{\mu}}\left(\mu\right)$. These can be calculated via either the "min-theta" approach for affine coercive problems [1], or more generally, via the successive constraint method [3]. We also assume we have access to the constant γ_{d} which can be calculated by solving a generalized eigenvalue problem.

Finally, we introduce the $\mathcal{N}$-dimensional finite element (FE) approximation space $X\subset X^{e}$ and define $Y\equiv Y^{e}$. X and Y inherit their inner product and norm definitions from X^{e} and Y^{e}. For use in the reduced basis method, we will assume that $\mathcal{N}$ is sufficiently large that $u^{e}\left(\mu\right)$ and $u\left(\mu\right)$ are indistinguishable (and likewise for $J^{e}\left(\mu\right)$ and $J\left(\mu\right)$). The weak form of our PDE in the finite element space is given by

$$a\left(u\left(\mu\right),v;\mu\right)=f\left(v;\mu\right),\quad\forall v\in X, \tag{12}$$

with adjoint

$$a\left(v,\psi\left(\mu\right);\mu\right)=2d\left(u\left(\mu\right),v\right)+\ell\left(v\right),\quad\forall v\in X, \tag{13}$$

and cost and cost gradient

$$J\left(\mu\right)=d\left(u\left(\mu\right),u\left(\mu\right)\right)+\ell\left(u\left(\mu\right)\right) \tag{14}$$

and

$$\nabla_{\mu}J\left(\mu\right)=-a_{\mu}\left(u\left(\mu\right),\psi\left(\mu\right);\mu\right) \tag{15}$$

The optimization problem then becomes to find μ_{opt} satisfying

$$\mu_{opt}=\arg\min_{\mu\in\mathcal{D}}J\left(\mu\right). \tag{16}$$

3 Reduced basis approximation

The reduced basis (RB) approximation method is a projection-based model reduction method where the N-dimensional approximation space is given by the span of N 'snapshots', i.e. FE solutions taken at various chosen parameters. For the primal problem, this space is denoted X_{N}^{pr} and is defined to be

$$X_{N}^{pr}=\text{span}\left\{\zeta_{n}\equiv u\left(\mu_{n}^{pr}\right),1\leq n\leq N\right\}. \tag{17}$$

The RB approximation $u_{N}\left(\mu\right)\in X_{N}^{pr}$ to the exact solution $u\left(\mu\right)$ is obtained through a standard Galerkin projection:

$$a\left(u_{N}\left(\mu\right),v;\mu\right)=f\left(v;\mu\right),\quad\forall v\in X_{N}^{pr} \tag{18}$$

For the dual problem, the RB approximation space is denoted X_N^{du} and is defined to be

$$X_N^{du} = \text{span}\left\{\Upsilon_n \equiv \psi\left(\mu_n^{du}\right), 1 \leq n \leq N\right\}, \tag{19}$$

and the RB dual approximation $\psi_N(\mu) \in X_N^{du}$ is given by

$$a\left(v, \psi_N(\mu); \mu\right) = 2d\left(u_N(\mu), v\right) + \ell(v), \quad \forall v \in X_N^{du}. \tag{20}$$

This allows calculation of the RB cost and cost gradient via

$$J_N(\mu) = d\left(u_N(\mu), u_N(\mu)\right) + \ell\left(u_N(\mu)\right), \tag{21}$$

and

$$\nabla_\mu J_N(\mu) = -a_\mu\left(u_N(\mu), \psi_N(\mu); \mu\right). \tag{22}$$

3.1 *A posteriori* error estimation

In addition to allowing for efficient solution approximation, the RB method also allows efficient *a posteriori* calculation of upper bounds on the error in the RB solution. In this section, we summarize the standard *a posteriori* error estimation result for the primal problem, referring the reader to [1] for the detailed development. This result is extended to the dual problem, and primal-dual *a posteriori* bounds on the error in the cost and cost gradient are presented.

Error bounds in the X-norm We first consider error in the reduced basis approximation to the primal problem, given by

$$e^{pr}(\mu) = u(\mu) - u_N(\mu) \tag{23}$$

To arrive at a bound on $e^{pr}(\mu)$, we consider the residual of the primal weak form (18),

$$r^{pr}(v; \mu) = f(v; \mu) - a\left(u_N(\mu), v; \mu\right), \forall v \in X, \tag{24}$$

and introduce its Riesz representation,

$$\left(\hat{r}^{pr}(\mu), v\right)_X = r^{pr}(v; \mu), \quad \forall v \in X. \tag{25}$$

It follows from the Riesz representation theorem that

$$\left\|\hat{r}^{pr}(\mu)\right\|_X = \left\|r^{pr}(\cdot; \mu)\right\|_{X'} = \sup_{v \in X} \frac{r^{pr}(v; \mu)}{\|v\|_X}, \quad \forall v \in X. \tag{26}$$

We may then arrive at a bound on the X-norm of $e^{pr}(\mu)$ in terms of the dual norm of the residual and the coercivity lower bound [1],

$$\left\|e^{pr}(\mu)\right\|_X \leq \Delta_N^{pr}(\mu) = \frac{\left\|\hat{r}^{pr}(\mu)\right\|_X}{\alpha_{LB}(\mu)}. \tag{27}$$

Next, we consider the error in the dual problem, given by

$$e^{du}(\mu) = \psi(\mu) - \psi_N(\mu). \tag{28}$$

To arrive at an analogous bound for this error, we require the residual of the dual weak form (20),

$$r^{du}(v; \mu) = 2d\left(u_N(\mu), v; \mu\right) + \ell(v; \mu) - a\left(v, \psi_N(\mu); \mu\right), \quad \forall v \in X. \tag{29}$$

Again, we use the Riesz representation theorem to note that

$$\left\|\hat{r}^{du}(\mu)\right\|_X = \left\|r^{du}(\cdot; \mu)\right\|_{X'} = \sup_{v \in X} \frac{r^{du}(v; \mu)}{\|v\|_X}, \quad \forall v \in X, \tag{30}$$

where the Riesz representation is given by

$$\left(\hat{r}^{du}(\mu), v\right)_X = r^{du}(v; \mu), \quad \forall v \in X, \tag{31}$$

allowing us to bound the X-norm of $e^{du}(\mu)$ [4] by

$$\left\|e^{du}(\mu)\right\|_X \leq \Delta_N^{du}(\mu) = \frac{\left\|\hat{r}^{du}(\mu)\right\|_X + 2\gamma_d^{UB}(\mu)\,\Delta_N^{pr}(\mu)}{\alpha_{LB}(\mu)}. \tag{32}$$

Error bounds for the cost and cost gradient The error in the cost functional follows from (14) and (21), i.e.,

$$e^J(\mu) = J(\mu) - J_N(\mu) = d(u(\mu), u(\mu)) - d(u_N(\mu), u_N(\mu)) + \ell(u(\mu) - u_N(\mu)), \tag{33}$$

for which we may obtain the bound [4]

$$\left|e^J(\mu)\right| \le \Delta_N^J(\mu) \equiv \frac{\|\hat{r}^{pr}(\mu)\|_X \left\|\hat{r}^{du}(\mu)\right\|_X + \gamma_d^{UB}(\mu) \|\hat{r}^{pr}(\mu)\|_X \Delta_N^{pr}(\mu)}{\alpha_{LB}(\mu)} + \left|r^{pr}(\psi_N(\mu); \mu)\right|. \tag{34}$$

For the error in the cost gradient, we obtain from (15) and (22)

$$e^{\nabla_\mu J}(\mu) = \nabla_\mu J(\mu) - \nabla_\mu J_N(\mu) = a_\mu(u(\mu), \psi(\mu); \mu) - a_\mu(u_N(\mu), \psi_N(\mu); \mu) \tag{35}$$

for which we may obtain the bound [4]

$$e^{\nabla_\mu J}(\mu) \le \Delta_N^{\nabla_\mu J}(\mu) = \gamma_{a_\mu}^{UB}(\mu) \left(\Delta_N^{pr}(\mu) \Delta_N^{du}(\mu) + \Delta_N^{pr}(\mu) \|\psi_N(\mu)\|_X + \|u_N(\mu)\|_X \Delta_N^{du}(\mu)\right). \tag{36}$$

3.2 Computational procedure

Like many model reduction methods, the RB method is traditionally divided into a computationally expensive offline phase and a computationally efficient online phase. During the offline phase, a greedy algorithm is used to choose the parameters μ_n^{pr} and μ_n^{du} at which snapshots are taken to minimize the maximum error in the RB approximation over the entire admissible parameter domain [1]. Snapshots are computed, and $\mathcal{N}$-dimensional matrices and vectors are used to precompute N-dimensional matrices and vectors for use during the online phase. This allows all computation during the online phase of the RB algorithm to involve only N-dimensional quantities, allowing efficient execution of many model evaluations. Details of this online-offline decomposition for the state approximation and error estimation are presented in [1]. The online-offline decompositions for the dual problem, sensitivity approximations, and associated error estimations are analogous [4], and are not given explicitly here.

4 Trust region framework

The canonical trust region optimization framework uses a model function $m^k(\mu)$ to approximate the objective function $J(\mu)$ at each iterate μ^k. The model function $m^k(\mu)$, which generally changes at each trust region iteration, is often a local quadratic Taylor expansion, although various other approximate models, including POD, have been considered in [5, 9, 6, 8]. Additionally, for each iterate, we specify a trust region radius δ^k, which allows us to define the kth trust region subproblem as

$$\min_s m^k(\mu^k + s) \qquad \text{s.t.} \|s\| \le \delta^k \tag{37}$$

To determine if the step s should be accepted, we compute the ratio $\rho^k = \frac{m(\mu^k) - m(\mu^{k+1})}{J(\mu^k) - J(\mu^{k+1})}$, a measure of how well the model predicts decrease in the true cost. The value of ρ^k is used to determine not only whether or not the optimization step is accepted, but also whether the trust region radius should increase, decrease, or stay the same for the next optimization subproblem. One criticism of this approach is that the computation of ρ^k requires evaluating the true objective function $J(\mu)$, which may be computationally expensive.

To address this criticism, we use the reduced basis cost $J_N^k(\mu)$ as our model function $m^k(\mu)$. The *a posteriori* error bounds which we have developed in Section 3 allow us to eliminate the evaluations of the true cost. Finally, we use a recent result from Yue and Meerbergen [8] to use the *a posteriori* error bounds to guarantee convergence of our trust region approach to the optimum of the high-fidelity model.

4.1 Convergence

Standard trust region convergence theory requires (i) that the model function m^k satisfy the first-order condition, i.e. the model function must match the true objective and gradient at the current iterate exactly, and (ii) that each iterate of the optimization meet a sufficient decrease condition. In [8], Yue and Meerbergen relax these requirements to consider the general setting of an unconstrained trust region optimization algorithm which makes use of surrogate models with the following properties:

1. a bound on the error in the model function exists over the entire parameter space,

2. at any point within the parameter domain, we may reduce the approximation error to within any given tolerance $\epsilon > 0$, and

3. the model function must be smooth with finite gradient everywhere.

Given the above conditions, Yue and Meerbergen replace the first-order condition with the following relaxed first-order condition (adapted to our notation from Sections 2 and 3):

$$\begin{array}{lcccc} \text{A.} & \left|J_N^k\left(\mu^k\right) - J\left(\mu^k\right)\right| \leq \Delta_N^{J,k}\left(\mu^k\right) & \textit{and} & \left\|\nabla_\mu J_N^k\left(\mu^k\right) - \nabla_\mu J\left(\mu^k\right)\right\| \leq \Delta_N^{\nabla J,k}\left(\mu^k\right) \\ \text{B.} & \frac{\Delta_N^{J,k}\left(\mu^k\right)}{J_N^k\left(\mu^k\right)} \leq \tau_J & \textit{and} & \frac{\Delta_N^{\nabla J,k}\left(\mu^k\right)}{\left\|\nabla_\mu J_N^k\left(\mu^k\right)\right\|} \leq \tau_{\nabla_\mu J} \end{array} \tag{38}$$

for any given $\tau_J > 0$ and $\tau_{\nabla J} > 0$. There are two parts to the relaxed first-order condition. Part A requires that error bounds exist for both the cost function and its gradient. Part B requires that we be able to infinitely refine the reduced model; i.e. for any given τ_J and $\tau_{\nabla J}$, we can build a reduced basis model for which the relative error bounds are lower than these tolerances. The sufficient decrease condition is similarly replaced, with what Yue and Meerbergen term an "error-aware sufficient decrease condition":

$$J_N^{k+1}\left(\mu^{k+1}\right) \leq J_N^k\left(\mu_{AGC}^k\right) \tag{39}$$

where μ_{AGC} is known as the "approximate generalized Cauchy point", defined to be a point that achieves sufficient decrease in the reduced basis model in a descent direction.

If the relaxed first-order condition is satisfied, and all iterates satisfy the error-aware sufficient decrease condition, Yue and Meerbergen show convergence of the trust region algorithm to the optimum of the high-fidelity model under mild assumptions [8] which are satisfied in our setting. We note that the *a posteriori* error bounds of reduced basis models satisfy part A of the relaxed first-order condition, and that these bounds may be reduced to zero by simply adding the finite element solutions at the current iterate $u\left(\mu^k\right)$ and $\psi\left(\mu^k\right)$ to the primal and dual bases, respectively, satisfying part B. Thus, to ensure convergence of the trust region reduced basis algorithm, we need only ensure that all iterates satisfy the error-aware sufficient decrease condition by rejecting steps which violate this condition.

4.2 Trust region reduced basis algorithm

The optimization subproblem for the trust region reduced basis (TRRB) algorithm is defined as follows:

$$\min_{\mu^{k+1}} J_N^k\left(\mu^{k+1}\right) \qquad \text{s.t.} \left|\frac{\Delta_N^{J,k}\left(\mu^{k+1}\right)}{J_N^k\left(\mu^{k+1}\right)}\right| \leq \epsilon_L \tag{40}$$

Any line search method may be used for the solution of the optimization subproblem; in Section 5 we use a BFGS method. The error bound on the cost functional, $\Delta_N^{J,k}(\mu)$, is used to implicitly define the trust region, requiring that iterates of the line search remain within regions for which the error bound is less than ϵ_L times the cost. If the line search steps outside of this region, we use backtracking to bring the line search back to a region where $\Delta_N^{J,k}(\mu)$ is sufficiently low.

For each subproblem solve, we have two possible stop criteria: either (i) the line search method locates a stationary point within the trust region, or (ii) the line search gets too close to the boundary of the current trust region, i.e.,

$$\left\|\nabla J_N^k(\mu)\right\| \leq \tau_{sub} \quad \text{or} \quad \beta\epsilon_L \leq \frac{\Delta_N^{J,k}(\mu)}{J_N^k(\mu)} \leq \epsilon_L \tag{41}$$

for some small $\tau_{sub} \geq 0$ and for some $\beta \in (0,1)$, generally close to 1. The latter criterion prevents the algorithm from expending too much effort optimizing close to the trust region boundary, because we heuristically assume that further attainable gains in this region are minimal. Overall convergence is reached when the norm of the true gradient is less than a tolerance $\tau \geq \tau_{sub}$, i.e. when

$$\left\|\nabla J\left(\mu^k\right)\right\| \leq \left\|\nabla J_N^k\left(\mu^k\right)\right\| + \Delta_N^{\nabla_\mu J,k}\left(\mu^k\right) \leq \tau. \tag{42}$$

The reduced model we employ is a progressively-built reduced basis model, where the primal and dual bases used to calculate $J_N^0(\mu)$ and $\nabla_\mu J^0(\mu)$ consist solely of $u\left(\mu^0\right)$ and $\psi\left(\mu^0\right)$, respectively. In the course of the

optimization, the iterates $u\left(\mu^k\right)$ and $\psi\left(\mu^k\right)$ are added to the RB model if the relative error bound gets close to exceeding the defined trust region tolerance (i.e. if the previous subproblem optimization terminates because the line search is close to the trust region boundary). In building and adding to our reduced basis this way, we automatically satisfy the relaxed first-order condition (38), because the reduced model is able to exactly represent the FE solution at each trust region iterate in our elliptic context.

Below, we summarize the algorithm steps:

1. **Initialization.** Let $k = 0$, and choose $\tau \geq \tau_{sub} \geq 0$, $\tau_{\nabla J} \in (0, 1)$, and $\beta \in (0, 1)$. Additionally, choose μ^0, the initial optimization point, $\kappa_{tr} < 1$, a decrease factor for the trust region size, and ϵ_L, the initial trust region error boundary. Generate the initial reduced basis model by computing truth solutions $u\left(\mu^0\right)$ and $\psi\left(\mu^0\right)$.

2. **Optimization sub-problem.** Minimize $J_N^k(\mu)$ with the stopping criteria

$$\left\|\nabla J_N^k(\mu)\right\| \leq \tau_{sub} \quad \text{or} \quad \beta\epsilon_L \leq \frac{\Delta_N^{J,k}(\mu)}{J_N^k(\mu)} \leq \epsilon_L \tag{43}$$

3. **Determine step acceptance.**

   (a) First, we note that a sufficient condition for (39) is

$$J_N^k\left(\mu^{k+1}\right) + \Delta_N^{J,k}\left(\mu^{k+1}\right) + \Delta_N^{J,k+1}\left(\mu^{k+1}\right) \leq J_N^k\left(\mu_{AGC}^k\right). \tag{44}$$

   Note that we do not have access to $\Delta_N^{J,k+1}\left(\mu^{k+1}\right)$. However, it is sufficient to instead check

$$J_N^k\left(\mu^{k+1}\right) + \Delta_N^{J,k}\left(\mu^{k+1}\right) < J_N^k\left(\mu_{AGC}^k\right). \tag{45}$$

   because we may add $u\left(\mu^{k+1}\right)$ and $\psi\left(\mu^{k+1}\right)$ to the RB model before the next subproblem solve to ensure $\Delta_N^{J,k+1}\left(\mu^{k+1}\right) = 0$, thus also satisfying the sufficient condition. This is cheap to check, and if it holds we may accept μ^{k+1}, add $u\left(\mu^{k+1}\right)$ and $\psi\left(\mu^{k+1}\right)$ to the reduced basis model, and go to Step 4.

   (b) Otherwise, we note that a necessary condition for (39) is

$$J_N^k\left(\mu^{k+1}\right) - \Delta_N^{J,k}\left(\mu^{k+1}\right) - \Delta_N^{J,k+1}\left(\mu^{k+1}\right) \leq J_N^k\left(\mu_{AGC}^k\right). \tag{46}$$

   So we check

$$J_N^k\left(\mu^{k+1}\right) - \Delta_N^{J,k}\left(\mu^{k+1}\right) \leq J_N^k\left(\mu_{AGC}^k\right). \tag{47}$$

   If this condition fails, satisfying (46) may require a large error bound in the next model, leading to inaccurate approximations, so we reject the iterate μ^{k+1}, shrink the trust region (set $\epsilon_L = \kappa_{tr}\epsilon_L$), and re-solve the optimization sub-problem (go to Step 2).

   (c) Otherwise, if (47) holds, we add $u\left(\mu^{k+1}\right)$ and $\psi\left(\mu^{k+1}\right)$ to the reduced basis model and check (39). If it holds, then we accept μ^{k+1} and go to Step 4.

   (d) Otherwise, we reject μ^{k+1}, set $\epsilon_L = \kappa_{tr}\epsilon_L$, and re-solve the optimization subproblem (go to Step 2).

4. **Convergence check.** If $\left\|\nabla J_N^{k+1}\left(\mu^{k+1}\right)\right\| + \Delta_N^{\nabla_\mu J,k+1}\left(\mu^{k+1}\right) \leq \tau$, return μ^{k+1} and stop.

5 Numerical tests

5.1 Model problem

To demonstrate our algorithm, we consider a two-dimensional thermal fin with a fixed geometry (Fig. 1) consisting of a central post and four horizontal subfins [10]. The fin conducts heat away from a uniform flux source at the root of the fin, Γ_{root}, through the post and subfins to the surrounding air. The fin is characterized by a six-dimensional parameter vector $\mu = (k_0, k_1, k_2, k_3, k_4, \text{Bi})^T$ containing the heat conductivities of different regions of the fin and the Biot number, a nondimensional heat transfer coefficient relating the convective heat transfer coefficient to the conductivity of the fin.

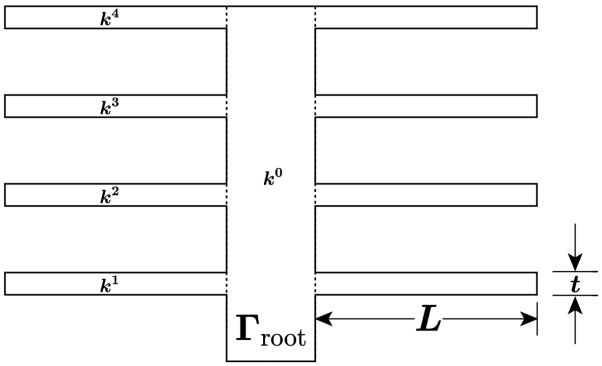

Figure 1: Thermal fin geometry

The steady-state temperature distribution within the fin, $u(\mu)$, is governed by the elliptic partial differential equation

$$-k_i \nabla^2 u_i = 0 \quad \text{in } \Omega_i, \quad i = 0, \ldots, 4 \ , \tag{48}$$

where ∇^2 is the Laplacian operator, u_i indicates the restriction of u to Ω_i, and Ω_i refers to the region of the fin with conductivity k_i, each with boundary $\partial\Omega_i$. To ensure continuity of temperature and heat flux at the subfin-post interfaces $\Gamma^i_{int} \equiv \partial\Omega_0 \cap \partial\Omega_i, i = 1, \ldots, 4$, we enforce the interface boundary conditions

$$u_0 \quad = \quad u_i$$

and

$$-\left(\nabla u_0 \cdot \hat{\mathbf{n}}_i\right) = -k_i \left(\nabla u_i \cdot \hat{\mathbf{n}}_i\right) \tag{49}$$

on each interface Γ^i_{int}, where $\hat{\mathbf{n}}_i$ is the outward normal on $\partial\Omega_i$. On the external fin boundaries we introduce a Neumann flux boundary condition at the root to model the heat source,

$$-\left(\nabla u_0 \cdot \hat{\mathbf{n}}_0\right) = -1 \text{ on } \Gamma_{root}, \tag{50}$$

and a Robin boundary condition at all other external boundaries to model convective heat losses,

$$-k_i \left(\nabla u_i \cdot \hat{\mathbf{n}}_i\right) = (\mathrm{Bi})\, u_i \text{ on } \Gamma^i_{ext}, \tag{51}$$

for $i = 0, \ldots, 4$, where Γ^i_{ext} is the boundary of Ω_i exposed to the surrounding air. We are interested in the average temperature at the root of the fin, $T_{root}(\mu)$, which can be expressed as $T_{root}(\mu) = \mathcal{L}(u(\mu))$, where

$$\mathcal{L}(v) = \int_{\Gamma_{root}} v. \tag{52}$$

For our optimization, we consider a thermal fin whose parameters are fixed but unknown, but for which we know the output T^{ref}_{root}. To infer the unknown parameters, we define a quadratic output $s(\mu)$

$$s(\mu) = \frac{1}{2}\left(\frac{T_{root}(\mu) - T^{ref}_{root}}{T^{ref}_{root}}\right)^2 \equiv \frac{\left\|\mathcal{L}(u(\mu)) - T^{ref}_{root}\right\|^2_D}{2\left(T^{ref}_{root}\right)^2}, \tag{53}$$

where D is a suitable Hilbert space of observation, and note that we can define $d(u,v) \equiv \frac{1}{2}(\mathcal{L}u, \mathcal{L}v)_D$ and $\ell(v) \equiv -\left(\mathcal{L}v, T^{ref}_{root}\right)_D$ to write the output as

$$s(\mu) = d(u(\mu), u(\mu)) + \ell(u(\mu)) + const. \tag{54}$$

Because we are interested in minimizing this function, we can ignore the constant term and minimize the following cost, which matches the formulation presented in Section 2, i.e.,

$$J(\mu) = d(u(\mu), u(\mu)) + \ell(u(\mu)). \tag{55}$$

5.2 Algorithm performance

To measure the performance of the algorithm, twenty different cost functions $J_m(\mu), m = 1, \ldots, 20$ are defined by generating a set of twenty randomly sampled points $\tilde{\mu}_m$ from within the six-dimensional parameter domain. The PDE is solved for each of these parameter values to obtain $T_{root}^{ref,m} = T_{root}(\tilde{\mu}_m)$. Finally, we define a regularization term for each cost function, given by

$$h(\mu) = \sum_{i=1}^{6} \left(\frac{\mu - \tilde{\mu}_i^m}{\tilde{\mu}_i^m} \right)^2 \tag{56}$$

such that the cost function to be optimized is

$$J(\mu) = d(u(\mu), u(\mu)) + \ell(u(\mu)) + h(\mu) \tag{57}$$

The optimization problem is then solved on a coarse, medium, and fine grid for each cost function, using both our trust region reduced basis approach and a BFGS algorithm which uses only full FE solves to obtain the necessary states and sensitivities. Statistics for these numerical trials are tabulated in Table 2 for a coarse (1333-node), medium (4970-node), and fine (17899-node) FEM grid (used for both the full BFGS solver as well as generating the basis functions for the reduced model). Algorithm parameters used for the TRRB tests are tabulated in Table 1. The parameters of the BFGS algorithm used for the full solver are identical to those used for the TRRB subproblem solver.

Parameter	Symbol	Value for numerical tests
"close" to TR boundary threshold	β	0.95
trust region boundary	ϵ_L	0.1
RB gradient error tolerance	$\tau_{\nabla J}$	0.1
subproblem convergence tolerance	τ_{sub}	1e-8
overall convergence tolerance	τ	1e-4

Table 1: Trust region reduced basis algorithm parameters used in numerical tests

Overall, the combined trust region reduced basis optimization approach makes significant gains (more than five-fold) in reducing the number of full FE solves required to reach the optimum. Whether or not this translates into run time speedups depends on the expense of the FE solve; for the coarsest grid tested, the standard finite element quasi-Newton approach outperforms the trust region reduced basis approach. On the medium grid, however, the trust region reduced basis approach begins to outperform the FE quasi-Newton approach, and these gains are even more pronounced for the finest grid tested. Additionally, we note that since the number of full and reduced solves needed by the two algorithms to converge to an optimum is roughly independent of the size of the grid, the trust region reduced basis approach is most appropriate to extremely large FE simulations, such as those that might be encountered in real-world problems (even our finest grid is relatively small).

Grid size	Coarse		Medium		Fine	
	TRRB	BFGS	TRRB	BFGS	TRRB	BFGS
Mean Time (s)	0.16	0.11	0.67	1.56	7.35	18.2
Min. Time (s)	0.08	0.05	0.42	0.48	4.37	6.69
Max. Time (s)	0.38	0.19	1.23	2.51	12.1	30.2
Mean # FE Solves	4.5	24.6	4.5	24.6	4.5	24.7
Min. # FE Solves	3	8	3	8	3	8
Max. # FE Solves	7	40	8	39	7	39
Mean # RB Solves	179.8	–	183.1	–	179.2	–
Min. # RB Solves	87	–	84	–	84	–
Max. # RB Solves	307	–	326	–	307	–

Table 2: Numerical performance of reduced basis trust region algorithm compared to full BFGS solver for 20 different cost functions on thermal fin problem

6 Conclusions

The proposed approach uses reduced basis methods in conjunction with a trust region optimization framework to make gains in the computational efficiency of parameter optimizations constrained by linear elliptic PDEs. In this approach, reduced basis models are leveraged to improve upon existing optimization approaches in several ways. First, reduced basis models are used as the model function within the trust region optimization, reducing the time for each optimization function evaluation. Second, the traditional trust region framework heuristically determines whether to accept or reject an optimization iterate by computing the ratio of actual cost function decrease to cost function decrease predicted by the model function, an approach which has been criticized because of the necessity of solving the full system simply to determine step acceptance. In the proposed trust region reduced basis framework, the *a posteriori* error bounds associated with the reduced basis models are used to choose when to accept and reject optimization iterates, allowing step acceptance to be decided without recourse to the full model. Third, the error bounds are used to intelligently determine when to update the reduced model. Fourth and finally, while previous work has shown heuristic convergence of similar frameworks making use of alternate model reduction methods, the existence of error bounds for RBM allows rigorous proof of convergence of the algorithm to a stationary point of the full model.

Acknowledgements. EQ acknowledges support from the US Fulbright Student Program, the National Science Foundation Graduate Research Fellowship, and the Hertz Foundation Graduate Fellowship for this work. The MIT authors also acknowledge the support of the US Department of Energy, Office of Advanced Scientific Computing Research (ASCR), Applied Mathematics Program, awards DE-FG02-08ER2585 and DE-SC0009297, as part of the DiaMonD Multifaceted Mathematics Integrated Capability Center.

References

[1] G. Rozza, D.B.P. Huynh, and A.T. Patera. Reduced Basis Approximation and a Posteriori Error Estimation for Affinely Parametrized Elliptic Coercive Partial Differential Equations. *Archives of Computational Methods in Engineering.* 15(3):229-275, 2008.

[2] F. Tröltzsch, *Optimal Control of Partial Differential Equations: Theory, Methods, and Applications*, volume 112, Graduate Studies in Mathematics. American Mathematical Society. 2010.

[3] D.B.P. Huynh, G. Rozza, and A.T. Patera. A successive constraint linear optimization method for lower bounds of parametric coercivity and inf-sup stability constants. *C.R. Math.* 345(8):473-478, 2007.

[4] E. Qian, M. Grepl, K. Veroy, and K. Willcox. A Certified Trust Region Reduced Basis Approach for PDE-constrained Optimization. In preparation.

[5] N. Alexandrov, J.E. Dennis, Jr., R.M. Lewis, and V. Torczon. A trust region framework for managing the use of approximation models in optimization. ICASE Report 97-50, Institute for Computer Applications in Science and Engineering, NASA Langley Research Center, Hampton, Virginia,1997.

[6] E. Arian, M. Fahl, and E.W. Sachs, "Trust Region Proper Orthogonal Decomposition for Flow Control." ICASE Report 2000-25, Institute for Computer Applications in Science and Engineering, NASA Langley Research Center, Hampton, Virginia, 2000.

[7] M.J. Zahr, and C. Farhat. Progressive construction of a parameteric reduced-order model for PDE-constrained optimization. *Int. J. Numer. Meth. Engng.* 102:1111-1135, 2015.

[8] Y. Yao and K. Meerbergen. Accelerating Optimization of Parametric Linear Systems by Model Order Reduction. *SIAM J. Optim.* 23(2):1344-1370, 2013.

[9] D.P. Kouri, M. Heinkenschloss, D. Ridzal, and B.G. Van Bloemen Waanders. Inexact Objective Function Evaluations in a Trust Region Algorithm for PDE-Constrained Optimization Under Uncertainty. *SIAM J. Sci. Comput.* 36(6):A3011-A3029, 2014.

[10] L. Machiels, Y. Maday, A. T. Patera, and D. V. Rovas, A blackbox reduced-basis output bound method for shape optimization, in Proceedings of the 12th International Conference on Domain Decompostion Methods, Chiba, Japan, 1999, T. Chan, T. Kako, H. Kawarada, and O. Pironneau, eds., DDM.org, 2001, pp. 429–436.