PREPRINT 453

igpm | RWTH AACHEN UNIVERSITY

# A certified trust region reduced basis approach to PDE-constrained optimization

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

Institut für Geometrie und Praktische Mathematik
Templergraben 55, 52062 Aachen, Germany

* Department of Aeronautics and Astronautics, MIT, Cambridge, MA (elizqian@mit.edu, kwillcox@mit.edu),
† Institute for Geometry and Applied Mathematics, RWTH Aachen University, Aachen, Germany (grepl@igpm.rwth-aachen.de)
‡ Aachen Institute for Advanced Study in Computational Engineering Science (AICES), RWTH Aachen University, Aachen, Germany (veroy@aices.rwth-aachen.de)

JUNE 2016

# A CERTIFIED TRUST REGION REDUCED BASIS APPROACH TO PDE-CONSTRAINED OPTIMIZATION

ELIZABETH QIAN*, MARTIN GREPL†, KAREN VEROY‡, AND KAREN WILLCOX*

**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 the dimension of the design space is large. It is therefore advantageous to replace expensive high-dimensional PDE solvers (e.g., finite element) with lower-dimension surrogate models. In this paper, the reduced basis (RB) model reduction method is used in conjunction with a trust region optimization framework to accelerate PDE-constrained parameter optimization. Novel *a posteriori* error bounds on the RB cost and cost gradient for quadratic cost functionals (e.g., least squares) are presented, and used to guarantee convergence to the optimum of the high-fidelity model. The proposed certified RB trust region approach uses high-fidelity solves to update the RB model only if the approximation is no longer sufficiently accurate, reducing the number of full-fidelity solves required. We consider problems governed by elliptic and parabolic PDEs and present numerical results for a thermal fin model problem in which we are able to reduce the number of full solves necessary for the optimization by up to 86%.

**Key words.** model reduction, optimization, trust region methods, partial differential equations, reduced basis methods, error bounds, parametrized systems

**AMS subject classifications.** 35J20, 35K10, 49K20, 65K10, 65M15, 90C06, 90C30

## 1. Introduction.

Optimization problems governed by partial differential equations (PDEs) appear in many settings across engineering and science disciplines, including engineering design optimization, optimal control problems, and inverse problems. Because typical optimization algorithms require numerous PDE evaluations, using classical discretization techniques (e.g., finite element) to solve these problems may be time-consuming, and in some cases, prohibitively expensive. One way to accelerate the solution of these problems is to replace expensive PDE evaluations with cheaper surrogate models. In this paper, we consider surrogate models based on projection-based reduced models.

The use of surrogate models in optimization has an extensive literature (see, e.g., the review in [11]). Our interest is in formulations that retain convergence guarantees even when approximate information is employed throughout the optimization solution process. Trust region methods are one class of approaches that have a rich history of convergence results; see, for example [9] for a detailed discussion of trust region methods. Traditionally, trust region methods replaced high-fidelity objective function evaluations with local linear or quadratic Taylor expansions. These local approximations automatically satisfy first-order consistency conditions (i.e., the approximate model's objective and gradient evaluations are locally exact), which in turn provides guarantees that the resulting optimization solution will satisfy the optimality conditions of the original high-fidelity system. The influence of inexact gradient information is considered in [7, 25], and of inexact gradient and function information in [6, 8, 9]. In [1], the authors consider a trust region framework with more general

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

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

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

approximation models of varying fidelity and show how adaptive corrections may be used to achieve the first-order consistency conditions required to achieve a provably convergent formulation for general approximation models.

In addition to providing a theoretical framework that yields a convergent surrogate-based optimization formulation, trust region methods also provide an iterative framework for adaptation of the surrogate to the optimization problem of interest. Generating globally accurate surrogate models is typically prohibitively expensive, particularly when the underlying system is governed by PDEs. Thus, approaches that tailor the surrogate model, in our case a projection-based reduced model, to the optimization problem are of particular interest. While a number of adaptation approaches have been proposed for projection-based reduced models (see, e.g., [10, 17, 20, 22]), the challenge in the optimization setting is that regions of interest are not known *a priori*. Iterative approaches that adapt the reduced model as the optimization progresses have been considered in [5, 21]. In this paper we similarly adapt the reduced model as the optimization progresses, while also constructing our adaptation so as to rigorously address the convergence of the resulting optimization formulation.

We use the reduced basis (RB) method, a projection based reduced-order modeling method, together with a trust region approach. The use of projection-based reduced models as surrogates in trust region optimization was first explored using proper orthogonal decomposition (POD) in [2]. In [2], the authors assume an upper bound on the inexactness of the function and gradient information resulting from the POD model and prove convergence of their algorithm using the results from [6, 25]. Unfortunately, verification of this upper bound in practice requires evaluation of the high-fidelity model. In [30], the authors assume the existence of error bounds for the surrogate function and gradient evaluations, and use this to prove convergence of the trust region method to a high-fidelity optimum relying only on surrogate evaluations. However, due to the absence of a rigorous and efficiently evaluable error bound for the Krylov-Pade interpolatory reduced model considered in [30], the authors introduce a heuristic bound and thus are able to demonstrate only heuristic convergence. Heuristic error indicators have also been applied to trust region optimization for POD models [31], and in a stochastic context, an approximation based on sparse grids [16].

The reduced basis method is a reduced-order modeling technique for parametrized PDEs which supports rigorous *a posteriori* error estimation (see [24] for a review). We propose a reduced basis trust region method for solving optimization problems constrained by elliptic and parabolic PDEs which avoids the costly offline phase of the traditional RB method and iteratively builds the reduced model along the optimization trajectory as the algorithm progresses. After introducing the problem statement in section 2, we present the following contributions:

1. In section 3, we present *a posteriori* error bounds for the optimization cost functional and its gradient. Our bounds are based on a primal-dual formulation and are rigorous and efficiently computable. The dual formulation permits us to efficiently evaluate the gradient of the cost functional and at the same time derive error bounds for the cost functional which are superlinearly convergent with respect to the primal and dual error bound.
2. The error bounds play a crucial role in the reduced basis trust region method introduced in section 4—they allow us to show convergence of the proposed approach to the (unknown) high-fidelity optimum. Furthermore, they allow us to efficiently control the accuracy of the reduced basis surrogate model during the optimization. We avoid the computationally expensive offline phase and build the reduced model adaptively along the optimization trajectory,

thus keeping the number of high-fidelity solves to a minimum.

In section 5, we present numerical results for parameter optimization problems constrained by elliptic and parabolic PDEs. We consider a thermal fin model problem with up to six variable parameters and compare the performance of our proposed reduced basis trust region approach to that of a traditional optimization using high-fidelity PDE evaluations. We also compare to a "classical" RB approach, where the reduced model is first generated during an offline stage and then used for the optimization in the online stage.

**2. Problem formulation.** In this section we introduce the PDE-constrained parameter optimization problem for both the elliptic and parabolic settings.

**2.1. Preliminaries.** Let $\Omega$ be a physical domain in $\mathbb{R}^d$ with Lipschitz continuous boundary $\partial\Omega$. We define the Hilbert space $X^e$ such that $H^1(\Omega) \supset X^e \supset H_0^1(\Omega)$ and $Y^e := L^2(\Omega)$, where $H^1(\Omega) = \left\{ v \mid v \in L^2(\Omega), \nabla v \in \left(L^2(\Omega)\right)^d \right\}$, $H_0^1(\Omega) = \left\{ v \mid v \in H^1(\Omega),\, v|_{\partial\Omega} = 0 \right\}$, and $L^2(\Omega)$ is the space of square-integrable functions over $\Omega$. We associate with $X^e$ and $Y^e$ the inner products $(w,v)_{X^e}$ and $(w,v)_{Y^e}$ as well as the induced norms $\|\cdot\|_{X^e} = \sqrt{(\cdot,\cdot)_{X^e}}$ and $\|\cdot\|_{Y^e} = \sqrt{(\cdot,\cdot)_{Y^e}}$, respectively; for example, $(w,v)_{X^e} := \int_\Omega \nabla w \cdot \nabla v + \int_\Omega wv$, $\forall w, v \in X^e$ and $(w,v)_{Y^e} := \int_\Omega wv$, $\forall w, v \in Y^e$. We denote the corresponding dual spaces by $X^{e\prime}$ and $Y^{e\prime}$. The superscript $\cdot^e$ indicates that we are dealing with the "exact" continuous domain. Finally, let $\mathcal{D} \subset \mathbb{R}^P$ be a $P$-dimensional parameter set in which our $P$-tuple parameter $\mu := (\mu_1, \ldots, \mu_P)$ resides.

We now define the conforming $\mathcal{N}$-dimensional finite element (FE) approximation space $X \subset X^e$ and define $Y := Y^e$, inheriting inner product and norm definitions from $X^e$ and $Y^e$, respectively. For the parabolic case, we directly consider a time-discrete framework associated to the time interval $I := ]0, t_f]$, where $\bar{I} := [0, t_f]$ is divided into $K$ uniform subintervals of length $\Delta t = \frac{t_f}{K}$. We introduce $\mathbb{K} := \{1, \ldots, K\}$ for notational convenience, and define $t^k := k\Delta t, \forall k \in \mathbb{K}$, and finally, $\mathbb{I} := \{t^0, \ldots, t^k\}$. We shall assume that $\mathcal{N}$ and $K$ are large enough – i.e. $X$ is sufficiently rich and the time-discretization sufficiently fine – such that the FE approximation guarantees a desired accuracy over the whole parameter domain $\mathcal{D}$.

We introduce the parameter-dependent bilinear form $a(\cdot,\cdot;\mu) : X \times X \to \mathbb{R}$ and its derivative in the $i$th component of $\mu$, $a_{\mu_i}(\cdot,\cdot;\mu) : X \times X \to \mathbb{R}, \forall i \in \{1, \ldots, P\}$. We also introduce the parameter-independent bilinear forms $m(\cdot,\cdot) : X \times X \to \mathbb{R}$ and $d(\cdot,\cdot) : X \times X \to \mathbb{R}$. We assume that all bilinear forms are continuous:

$$0 < \gamma_a(\mu) := \sup_{w \in X\setminus\{0\}} \sup_{v \in X\setminus\{0\}} \frac{a(w,v;\mu)}{\|w\|_X \|v\|_X} \leq \gamma_0^a < \infty, \quad \forall \mu \in \mathcal{D}, \tag{1}$$

$$0 < \gamma_{a_{\mu_i}}(\mu) := \sup_{w \in X\setminus\{0\}} \sup_{v \in X\setminus\{0\}} \frac{a_{\mu_i}(w,v;\mu)}{\|w\|_X \|v\|_X} \leq \gamma_0^{a_{\mu_i}} < \infty, \quad \forall \mu \in \mathcal{D}, \quad i = 1, .., P \tag{2}$$

$$0 < \gamma_m := \sup_{w \in X\setminus\{0\}} \sup_{v \in X\setminus\{0\}} \frac{m(w,v)}{\|w\|_Y \|v\|_Y} < \infty, \tag{3}$$

$$0 < \gamma_d := \sup_{w \in X\setminus\{0\}} \sup_{v \in X\setminus\{0\}} \frac{d(w,v)}{\|w\|_X \|v\|_X} < \infty, \tag{4}$$

and symmetric, i.e. $\forall\, w, v \in X, \forall\, \mu \in \mathcal{D}$, $a(v, w; \mu) = a(w, v; \mu)$, $a_{\mu_i}(v, w; \mu) = a_{\mu_i}(w, v; \mu)$, $m(v, w) = m(w, v)$, and $d(v, w) = d(w, v)$. Additionally, we assume that $a(\cdot, \cdot; \mu)$ and $m(\cdot, \cdot)$ are coercive:

$$0 < \alpha_0^a \leq \alpha(\mu) := \inf_{v \in X} \frac{a(v, v; \mu)}{\|v\|_X^2}, \quad \forall \mu \in \mathcal{D}, \tag{5}$$

$$0 < \alpha_0^m := \inf_{v \in X} \frac{m(v, v)}{\|v\|_Y^2}, \quad \forall \mu \in \mathcal{D}. \tag{6}$$

We next introduce two $X$-continuous linear functionals, the parameter-dependent $f(\cdot; \mu) : X \to \mathbb{R}$ and the parameter-independent $\ell(\cdot) : X \to \mathbb{R}$. Finally, we assume that all parameter-dependent linear and bilinear forms depend affinely on the parameter $\mu$ and can be expressed as

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

$$f(v; \mu) = \sum_{q=1}^{Q_f} \Theta_f^q(\mu) f^q(v), \quad \forall v \in X, \mu \in \mathcal{D}, \tag{8}$$

where $Q_a$ and $Q_f$ are some (preferably) small integers, the functions $\Theta_a^q(\mu), \Theta_f^q(\mu) : \mathcal{D} \to \mathbb{R}$ are continuous and depend on $\mu$, but the continuous bilinear and linear forms $a^q(\cdot, \cdot) : X \times X \to \mathbb{R}$ and $f^q : X \to \mathbb{R}$ do not depend on $\mu$. For simplicity, we assume that the forms $m(\cdot, \cdot)$, $d(\cdot, \cdot)$, and $\ell(\cdot)$ are parameter-independent, although extensions to affine parameter dependence are readily admitted [24].

For the development of the *a posteriori* error bounds we also require the following ingredients. We assume that we are given a positive lower bound $\alpha_{LB}(\mu) : \mathcal{D} \to \mathbb{R}_+$ for the coercivity constant $\alpha(\mu)$ defined in (5) such that

$$0 < \alpha_0^a \leq \alpha_{LB}(\mu) \leq \alpha(\mu), \quad \forall \mu \in \mathcal{D}. \tag{9}$$

Furthermore, we assume that we have an upper bound available for the continuity constants $\gamma_{a_{\mu_i}}(\mu)$ defined in (2) such that

$$\gamma_{a_{\mu_i}}^{UB}(\mu) \geq \gamma_{a_{\mu_i}}(\mu), \quad \forall \mu \in \mathcal{D}. \tag{10}$$

Various recipes exist to obtain such bounds [15, 24].

**2.2. Elliptic PDE-constrained optimization.** We consider the constrained minimization of the output least-squares cost functional

$$\min_{\mu \in \mathcal{D}} \ \|\mathcal{L}(u(\mu)) - g_{\text{ref}}\|_{\mathbb{D}}^2 + \lambda \mathcal{R}(\mu) \tag{11a}$$

$$\text{s.t. } u(\mu) \in X \text{ satisfies} \quad a(u(\mu), v; \mu) = f(v; \mu), \qquad \forall v \in X, \tag{11b}$$

where $\mathcal{L} : X \to \mathbb{R}$ is a linear (output) functional and $g_{\text{ref}}$ is a reference output, e.g. obtained from experimental measurements. Furthermore, $\mathbb{D}$ is a suitable Hilbert space of observations with inner product $(\cdot, \cdot)_{\mathbb{D}}$ and induced norm $\|\cdot\|_{\mathbb{D}} = \sqrt{(\cdot, \cdot)_{\mathbb{D}}}$, and $\mathcal{R} : \mathcal{D} \to \mathbb{R}$ with $\lambda \in \mathbb{R}_+$ together form a scaled regularization term.

Note that we may expand (11a) with

$$(\mathcal{L}(u) - g_{\text{ref}}, \mathcal{L}(u) - g_{\text{ref}})_{\mathbb{D}} = (\mathcal{L}(u), \mathcal{L}(u))_{\mathbb{D}} - 2(\mathcal{L}(u), g_{\text{ref}})_{\mathbb{D}} + (g_{\text{ref}}, g_{\text{ref}})_{\mathbb{D}}.$$

Thus, defining $d(w, v) := (\mathcal{L}(w), \mathcal{L}(v))_{\mathbb{D}}, \forall w, v \in X$ and $\ell(v) := -2(\mathcal{L}(v), g_{\text{ref}})_{\mathbb{D}}, \forall v \in X$ and dropping the constant term $(g_{\text{ref}}, g_{\text{ref}})_{\mathbb{D}}$, we obtain the following equivalent formulation for the optimization problem:

$$\min_{\mu \in \mathcal{D}} J(\mu) \quad \text{where} \quad J(\mu) := d(u(\mu), u(\mu)) + \ell(u(\mu)) + \lambda \mathcal{R}(\mu) \tag{12a}$$

$$\text{s.t. } u(\mu) \in X \text{ satisfies } a(u(\mu), v; \mu) = f(v; \mu), \qquad \forall v \in X. \tag{12b}$$

In the sequel, we will use this more general quadratic cost formulation in developing the theory of the method.

Gradient-based optimization methods require access to the cost derivatives, which may be efficiently calculated using adjoint methods. We thus introduce the FE adjoint (dual) problem associated with our primal problem and cost in (12) [26] as follows: Given $\mu \in \mathcal{D}$ and the associated solution $u(\mu)$ to (12b), find $p(\mu) \in X$ satisfying

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

The derivative of the cost function with respect to the $i$th parameter $\mu_i$ may then be calculated via

$$\frac{\partial J(\mu)}{\partial \mu_i} = f_{\mu_i}(p(\mu); \mu) - a_{\mu_i}(u(\mu), p(\mu); \mu) + \lambda \frac{\partial \mathcal{R}(\mu)}{\partial \mu_i}. \tag{14}$$

**2.3. Parabolic PDE-constrained optimization.** The parabolic optimization formulation is analogous to the elliptic case. We therefore directly consider the following (time-discrete) constrained minimization problem with quadratic cost:

$$\min_{\mu \in \mathcal{D}} J(\mu) \quad \text{where} \quad J(\mu) := \Delta t \sum_{k=1}^{K} \left[ d\big(u^k(\mu), u^k(\mu)\big) + \ell\big(u^k(\mu)\big) \right] + \lambda \mathcal{R}(\mu) \tag{15a}$$

$$\text{s.t. } u^k(\mu) \in X \text{ satisfies}$$

$$\frac{m\big(u^k(\mu) - u^{k-1}(\mu), v\big)}{\Delta t} + a\big(u^k(\mu), v; \mu\big) = f(v; \mu) y\big(t^k\big), \quad \forall v \in X, k \in \mathbb{K}, \tag{15b}$$

$$\text{with initial condition}$$

$$u^0(\mu) = 0, \tag{15c}$$

where $y\big(t^k\big)$ is a (known) time-dependent forcing input and we assume zero initial conditions for simplicity. Note that we consider an Euler-Backward discretization for the time integration; however, we can also readily treat higher-order schemes such as Crank-Nicolson. We next introduce the dual problem: Given $\mu \in \mathcal{D}$ and the associated solution $u^k(\mu),\ k \in \mathbb{K}$ to (15b), the adjoint field variable $p^k(\mu) \in X, K \geq k \geq 1$, satisfies

$$\frac{m\big(v, p^k(\mu) - p^{k+1}(\mu)\big)}{\Delta t} + a\big(v, p^k(\mu); \mu\big) = 2d\big(u^k(\mu), v\big) + \ell(v), \quad \forall v \in X, k \in \mathbb{K}, \tag{16}$$

with final condition $p^{K+1}(\mu) = 0$. Note that the adjoint field variable evolves backward in time. The derivative of the cost function with respect to the $i$th parameter $\mu_i$ may then be calculated via

$$\frac{\partial J(\mu)}{\partial \mu_i} = \Delta t \sum_{k=1}^{K} \left[ f_{\mu_i}\big(p^k(\mu)\big) - a_{\mu_i}\big(u^k(\mu), p^k(\mu); \mu\big) \right] + \lambda \frac{\partial \mathcal{R}(\mu)}{\partial \mu_i}. \tag{17}$$

**3. Reduced basis method.** The RB method is a projection-based model reduction method for parametrized PDEs [24]. Traditionally, it consists of an expensive, time-consuming offline phase, in which the reduced basis is built, and an inexpensive online phase, during which the pre-built RB may be exploited for rapid and certified simulations of the PDE at any parameter within the admissible parameter domain. In this section, we present primal-dual RB approximations and associated novel *a posteriori* error estimation procedures for the elliptic and parabolic PDE-constrained parameter optimization problems introduced in the last section. To this end, we employ the RB approximations as surrogate models in the optimization problems (12) and (15) and develop new rigorous and efficiently evaluable error bounds for the cost functional and its gradient. In this work, we leverage these new error bounds to break from the offline/online paradigm in the optimization, i.e., we build the RB approximation during the iterative optimization procedure on-the-fly. Our error bounds guide the RB updates and at the same time allow us to guarantee convergence of the surrogate optimization to the (unknown) optimal solution of the original (FE) optimization problem. We note, however, that the results presented here also apply to the traditional offline/online RB setting.

Subsections 3.1 and 3.2 present the RB approximation and error estimation results for the elliptic and parabolic case, respectively. Subsection 3.3 discusses the computational aspects of the RB approximation.

**3.1. Elliptic problems.** This section introduces the RB approximation and error estimation results for the elliptic optimization problem (12).

**3.1.1. Approximation.** Given $X$-orthogonal sets of primal and dual basis vectors $\zeta_n$ and $\psi_n$, $n = 1, \ldots, N$, we denote the $N$-dimensional primal and dual RB approximation spaces by $X_N^{\mathrm{pr}}$ and $X_N^{\mathrm{du}}$, defined as

$$X_N^{\mathrm{pr}} := \mathrm{span}\{\zeta_n, 1 \le n \le N\} = \mathrm{span}\{u(\mu_n^{\mathrm{pr}}), 1 \le n \le N\},$$
$$X_N^{\mathrm{du}} := \mathrm{span}\{\psi_n, 1 \le n \le N\} = \mathrm{span}\{p(\mu_n^{\mathrm{du}}), 1 \le n \le N\},$$

respectively. We will comment on how $\mu_n^{\mathrm{pr}}$ and $\mu_n^{\mathrm{du}}$ are chosen in sections 3.3 and 4.1. For simplicity, we shall assume in this paper that the dimensions of the primal and dual RB spaces are equivalent. The following derivation, however, directly extends to the case with different dimensions.

The RB approximation is then obtained via a Galerkin projection: Given $\mu \in \mathcal{D}$, the RB primal approximation $u_N(\mu) \in X_N^{\mathrm{pr}}$ satisfies

$$a(u_N(\mu), v; \mu) = f(v; \mu), \quad \forall v \in X_N^{\mathrm{pr}}, \tag{18}$$

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

$$a(v, p_N(\mu); \mu) = 2d(u_N(\mu), v) + \ell(v), \quad \forall v \in X_N^{\mathrm{du}}. \tag{19}$$

The RB cost functional and its derivative with respect to the $i$th parameter can the be computed from

$$J_N(\mu) = d(u_N(\mu), u_N(\mu)) + \ell(u_N(\mu)) + \lambda \mathcal{R}(\mu), \tag{20}$$

and

$$\frac{\partial J_N(\mu)}{\partial \mu_i} = f_{\mu_i}(p_N(\mu); \mu) - a_{\mu_i}(u_N(\mu), p_N(\mu); \mu) + \lambda \frac{\partial \mathcal{R}(\mu)}{\partial \mu_i}. \tag{21}$$

**3.1.2. *A posteriori* error estimation.** We turn to the *a posteriori* error bounds. We first require

DEFINITION 1. *The residuals of the primal and dual equations are defined by*

$$r^{\rm pr}(v;\mu) := f(v;\mu) - a(u_N(\mu), v;\mu), \qquad \forall v \in X,\ \forall \mu \in \mathcal{D}, \tag{22}$$

$$r^{\rm du}(v;\mu) := 2d(u_N(\mu), v) + \ell(v) - a(v, p_N(\mu);\mu), \qquad \forall v \in X,\ \forall \mu \in \mathcal{D}. \tag{23}$$

We also define the primal and dual errors as follows:

$$e_{pr}(\mu) := u(\mu) - u_N(\mu) \ \text{ and } \ e_{du}(\mu) := p(\mu) - p_N(\mu). \tag{24}$$

We can now prove

LEMMA 2. *Let $u(\mu)$ and $u_N(\mu)$ be the solutions to (12b) and (18), respectively. Furthermore, let $p(\mu)$ and $p_N(\mu)$ be the solutions to the associated dual equation (13) and (19). The error in the primal variable, $e_{pr}(\mu) = u(\mu) - u_N(\mu)$, is then bounded by*

$$\|e_{pr}(\mu)\|_X \le \Delta_N^{\rm pr}(\mu) := \frac{\|r^{\rm pr}(\cdot;\mu)\|_{X'}}{\alpha_{LB}(\mu)}, \quad \forall \mu \in \mathcal{D}, \tag{25}$$

*and the error in the dual variable, $e_{du}(\mu) = p(\mu) - p_N(\mu)$, by*

$$\|e_{du}(\mu)\|_X \le \Delta_N^{\rm du}(\mu) := \frac{\left\|r^{\rm du}(\cdot;\mu)\right\|_{X'} + 2\gamma_d \Delta_N^{\rm pr}(\mu)}{\alpha_{LB}(\mu)}, \quad \forall \mu \in \mathcal{D}. \tag{26}$$

*Proof.* The bound (25) is standard, see e.g. [24]. We follow an analogous procedure to show (26). We first note from (13) and (23) that the dual error satisfies

$$a(v, e_{du}(\mu);\mu) = r^{\rm du}(v;\mu) + 2d(e_{pr}(\mu), v). \tag{27}$$

Choosing $v = e_{du}(\mu)$ and invoking (5), (9), and (4) we obtain

$$\alpha_{LB}(\mu)\|e_{du}(\mu)\|_X^2 \le \left\|r^{\rm du}(\cdot;\mu)\right\|_{X'}\|e_{du}(\mu)\|_X + 2\gamma_d\|e_{pr}(\mu)\|_X\|e_{du}(\mu)\|_X.$$

The result (26) then follows from (25). ∎

We may now consider the error in the cost functional and its gradient.

THEOREM 3. *The error in the cost functional, $e^J(\mu) := J(\mu) - J_N(\mu)$, satisfies*

$$|e^J(\mu)| \le \Delta_N^J(\mu) := \left\|r^{\rm du}(\mu)\right\|_{X'}\Delta_N^{\rm pr}(\mu) + \gamma_d\,\Delta_N^{\rm pr}(\mu)^2 + |r^{\rm pr}(p_N(\mu);\mu)|, \ \ \forall \mu \in \mathcal{D}, \tag{28}$$

*where $\Delta_N^{\rm pr}(\mu)$ is the primal bound defined in Lemma 2.*

*Proof.* It follows from (12a) and (20) that

$$e^J(\mu) = d(u(\mu), u(\mu)) - d(u_N(\mu), u_N(\mu)) + \ell(e_{pr}(\mu)).$$

Adding and subtracting $r^{\rm pr}(p_N(\mu);\mu)$ on the right-hand side and recalling the primal error-residual relationship, $a(e_{pr}(\mu), v;\mu) = r^{\rm pr}(v;\mu)$, $\forall v \in X$, we obtain

$$\begin{aligned} e^J(\mu) = d(u(\mu), u(\mu)) - d(u_N(\mu), u_N(\mu)) + \ell(e_{pr}(\mu)) \\ + r^{\rm pr}(p_N(\mu);\mu) - a(e_{pr}(\mu), p_N(\mu);\mu). \end{aligned}$$

If we also add and subtract the term $2d(u_N(\mu), e_{pr}(\mu))$ on the right-hand side and invoke (23), it follows that

$$\begin{aligned} e^J(\mu) = r^{\mathrm{du}}(e_{pr}(\mu);\mu) + d(u(\mu),u(\mu)) - d(u_N(\mu),u_N(\mu)) \\ + r^{\mathrm{pr}}(p_N(\mu);\mu) - 2d(u_N(\mu),e_{pr}(\mu)). \end{aligned}$$

Expanding $d(u_N(\mu), e_{pr}(\mu)) = d(u_N(\mu), u(\mu) - u_N(\mu))$, we obtain

$$\begin{aligned} e^J(\mu) &= r^{\mathrm{du}}(e_{pr}(\mu);\mu) + r^{\mathrm{pr}}(p_N(\mu);\mu) + d(u(\mu),u(\mu)) \\ &\qquad -2d(u_N(\mu),u(\mu)) + d(u_N(\mu),u_N(\mu)) \\ &= r^{\mathrm{du}}(e_{pr}(\mu);\mu) + r^{\mathrm{pr}}(p_N(\mu);\mu) + d(e_{pr}(\mu),e_{pr}(\mu)), \end{aligned}$$

which yields

$$|e^J(\mu)| \le \left\|r^{\mathrm{du}}(\mu)\right\|_{X'} \|e_{pr}(\mu)\|_X + \gamma_d \|e_{pr}(\mu)\|_X^2 + |r^{\mathrm{pr}}(p_N(\mu);\mu)|,$$

where we used the continuity of the bilinear form $d$. The desired result directly follows from Lemma 2. ∎

Before presenting the result for the cost gradient we make several remarks. First, since our goal is to develop effective *a posteriori* error bounds for the cost functional as opposed to increasing the accuracy of the RB cost functional[1], we incorporate the residual correction term in the bound (28) instead of correcting the RB cost functional; see e.g. the discussion in [27]. Second, the dual problem plays two roles in our setting: it allows us to (*i*) efficiently compute the cost gradient from (21) without having to resort to sensitivity derivatives; and (*ii*) devise an *a posteriori* error bound for the cost functional which converges superlinearly with respect to the primal and dual bounds [23]. Finally, we note that certified reduced basis approximations for quadratic outputs have been previously considered in [14]. As opposed to the dual problem defined in (13) in this paper, the authors in [14] introduce a dual problem which is dependent on the RB solution $u_N(\mu)$, e.g., for $p(\mu) \in X$,

$$a(v, p(\mu);\mu) = d(u(\mu) + u_N(\mu), v) + \ell(v), \quad \forall v \in X$$

Although we would obtain a similar bound to (28) for the cost functional using this formulation, the dual variable $p(\mu)$ thus defined cannot be used to compute the cost gradient from (14). We now turn to the error bound for the cost gradient.

THEOREM 4. *The error in the cost gradient, $e^{\nabla J}(\mu) = \nabla_\mu J(\mu) - \nabla_\mu J_N(\mu)$, satisfies*

$$\left\|e^{\nabla J}(\mu)\right\| \le \Delta_N^{\nabla J}(\mu) := \left\|\Delta_N^{\nabla_\mu J}(\mu)\right\|, \tag{29}$$

*where $\|\cdot\|$ is the Euclidean norm and $\Delta_N^{\nabla_\mu J}(\mu)$ is a vector whose ith component is the bound on the error in the ith component of the gradient, given by*

$$\begin{aligned} \Delta_N^{\nabla_{\mu_i} J}(\mu) &= \|f_{\mu_i}(\cdot;\mu)\|_{X'} \Delta_N^{\mathrm{du}}(\mu) \\ &\quad + \gamma_{a_{\mu_i}}^{UB}(\mu)\big(\Delta_N^{\mathrm{pr}}(\mu)\Delta_N^{\mathrm{du}}(\mu) + \Delta_N^{\mathrm{pr}}(\mu)\|p_N(\mu)\|_X + \|u_N(\mu)\|_X \Delta_N^{\mathrm{du}}(\mu)\big), \end{aligned} \tag{30}$$

*where $\Delta_N^{\mathrm{pr}}(\mu)$ and $\Delta_N^{\mathrm{du}}(\mu)$ are the primal and dual error bounds defined in Lemma 2.*

[1] We will observe in section 5 that the RB cost functional as defined in (20) is sufficiently accurate for our purpose.

*Proof.* We consider the error in the derivative of the cost with respect to $\mu_i$, the $i$th element of the parameter vector $\mu$. It follows from (14) and (21) that

$$e^{\nabla_{\mu_i} J}(\mu) = f_{\mu_i}(e_{du}(\mu);\mu) - (a_{\mu_i}(u(\mu),p(\mu);\mu) - a_{\mu_i}(u_N(\mu),p_N(\mu);\mu)). \tag{31}$$

We next note that

$$\begin{aligned} a_{\mu_i}(u(\mu),p(\mu);\mu) - a_{\mu_i}(u_N(\mu),p_N(\mu);\mu) = \\ a_{\mu_i}(e_{pr}(\mu),e_{du}(\mu);\mu) + a_{\mu_i}(e_{pr}(\mu),p_N(\mu);\mu) + a_{\mu_i}(u_N(\mu),e_{du}(\mu);\mu). \end{aligned} \tag{32}$$

Plugging (32) into (31) and invoking (2) and (10) we obtain

$$\begin{aligned} e^{\nabla_{\mu_i} J}(\mu) \le \|f_{\mu_i}(\cdot;\mu)\|_{X'}\|e_{du}(\mu)\|_X + \gamma^{UB}_{a_{\mu_i}}(\mu)\|e_{pr}(\mu)\|_X\|e_{du}(\mu)\|_X \\ + \gamma^{UB}_{a_{\mu_i}}(\mu)\|e_{pr}(\mu)\|_X\|p_N(\mu)\|_X + \gamma^{UB}_{a_{\mu_i}}(\mu)\|u_N(\mu)\|_X\|e_{du}(\mu)\|_X \end{aligned} \tag{33}$$

The result (30) then follows from Lemma 2 and (29) is obtained by taking the norm of all components. ∎

We note that, as opposed to the *a posteriori* error bound for the cost functional, the bound for the gradient does not exhibit a superlinear convergence with respect to the primal and dual error. However, it is well-known in the literature [6, 8] that fairly large relative errors (of 50% or more) in the gradient are permissible in the trust-region framework without jeopardizing the overall convergence. Even our fairly crude bound for the cost gradient will thus be sufficient to guarantee convergence of the RB trust region approach; also see the discussion in subsection 4.1 and the numerical results in section 5.

**3.2. Parabolic problems.** This section introduces the RB approximation and error estimation results for the parabolic case.

**3.2.1. Approximation.** We first introduce the primal and dual reduced basis spaces

$$X_N^{\mathrm{pr}} = \mathrm{span}\{\zeta_n, 1 \le n \le N\}, \qquad X_N^{\mathrm{du}} = \mathrm{span}\{\psi_n, 1 \le n \le N\},$$

where the $\zeta_n$ (and the $\psi_n$), $n = 1,\ldots,N$, are mutually $X$-orthogonal basis functions. We comment on their construction in sections 3.3 and 4.1.

The primal and dual RB approximations are obtained from a Galerkin projection: Given $\mu \in \mathcal{D}$, the primal approximation $u_N^k(\mu) \in X_N^{\mathrm{pr}}$ to $u^k(\mu) \in X$ satisfies

$$\frac{m\big(u_N^k(\mu) - u_N^{k-1}(\mu), v\big)}{\Delta t} + a\big(u_N^k(\mu), v; \mu\big) = f(v;\mu)y\big(t^k\big), \quad \forall v \in X_N^{\mathrm{pr}}, \tag{34}$$

and the dual approximation $p_N^k(\mu) \in X_N^{\mathrm{du}}$ to $p^k(\mu) \in X$ is given by

$$\frac{m\big(v, p_N^k(\mu) - p_N^{k+1}(\mu)\big)}{\Delta t} + a\big(v, p_N^k(\mu);\mu\big) = 2d\big(u_N^k(\mu), v\big) + \ell(v), \quad \forall v \in X_N^{\mathrm{du}}. \tag{35}$$

We can then calculate the RB cost and its derivative with respect to the $i$th parameter via

$$J_N(\mu) := \Delta t \sum_{k=1}^{K} \big[d\big(u_N^k(\mu), u_N^k(\mu)\big) + \ell\big(u_N^k(\mu)\big)\big] + \lambda\mathcal{R}(\mu), \tag{36}$$

and

$$\frac{\partial J_N(\mu)}{\partial \mu_i} = \Delta t \sum_{k=1}^{K} \big[f_{\mu_i}\big(p_N^k(\mu)\big) - a_{\mu_i}\big(u_N^k(\mu), p_N^k(\mu);\mu\big)\big] + \lambda\frac{\partial\mathcal{R}(\mu)}{\partial\mu_i}. \tag{37}$$

**3.2.2. *A posteriori* error estimation.** The *a posteriori* error estimation procedure for the parabolic problem is analogous to that of the elliptic problem. In this section, we present the error bounds necessary for the trust region approach proposed in section 4, deferring proofs to Appendix A. We first introduce the residuals in

DEFINITION 5. *The residuals of the primal and dual equations are defined by*

$$r_{pr}^k(v;\mu) = f(v)y(t^k) - a\big(u_N^k(\mu), v;\mu\big) - \frac{1}{\Delta t} m\big(u_N^k(\mu) - u_N^{k-1}(\mu), v\big), \tag{38}$$

$$r_{du}^k(v;\mu) = 2d\big(u_N^k(\mu), v\big) + \ell(v) - a\big(v, p_N^k(\mu);\mu\big) - \frac{1}{\Delta t} m\big(v, p_N^k(\mu) - p_N^{k+1}(\mu)\big), \tag{39}$$

*for all $v \in X$ and all $\mu \in \mathcal{D}$.*

For the parabolic case, we also require the "spatio-temporal" energy norms for the primal and dual problem as follows.

DEFINITION 6. *The spatio-temporal energy norms for the primal and dual problem are given by*

$$|||v^k(\mu)|||_{pr} := \left[ m(v^k(\mu), v^k(\mu)) + \Delta t \sum_{k'=1}^{k} a(v^{k'}(\mu), v^{k'}(\mu);\mu) \right]^{\frac{1}{2}}, \quad \forall v \in X, \tag{40a}$$

$$|||v^k(\mu)|||_{du} := \left[ m(v^1(\mu), v^1(\mu)) + \Delta t \sum_{k'=k}^{K} a(v^{k'}(\mu), v^{k'}(\mu);\mu) \right]^{\frac{1}{2}}, \quad \forall v \in X. \tag{40b}$$

We may now prove the following results for the primal and dual RB errors.

LEMMA 7. *Let $u^k(\mu)$ and $u_N^k(\mu)$, $k \in \mathbb{K}$, be the solutions to (15b) and (34), respectively. Furthermore, let $p^k(\mu)$ and $p_N^k(\mu)$, $k \in \mathbb{K}$, be the solutions to the associated dual equation (16) and (35). Then, the following bounds for the error in the primal variable, $e_{pr}^k(\mu) = u^k(\mu) - u_N^k(\mu)$, and the dual variable, $e_{du}^k(\mu) = p^k(\mu) - p_N^k(\mu)$, hold for all $\mu \in \mathcal{D}$*

$$|||e_{pr}^K(\mu)|||_{pr} \le \Delta_{N,K}^{\mathrm{pr}}(\mu) := \left( \frac{\Delta t}{\alpha_{LB}(\mu)} \sum_{k=1}^{K} \big\| r_{pr}^k(\cdot;\mu) \big\|_{X'}^2 \right)^{\frac{1}{2}} \tag{41}$$

$$|||e_{du}^1|||_{du} \le \Delta_{N,1}^{\mathrm{du}}(\mu) := \left( 8\gamma_d^2 \left( \frac{\Delta_{N,K}^{\mathrm{pr}}(\mu)}{\alpha_{LB}(\mu)} \right)^2 + \frac{2\Delta t}{\alpha_{LB}(\mu)} \sum_{k=1}^{K} \big\| r_{du}^k(\cdot;\mu) \big\|_{X'}^2 \right)^{\frac{1}{2}}. \tag{42}$$

With Lemma 7 in hand, we may bound the parabolic cost and cost gradient as we did in Theorems 3 and 4 for the elliptic case.

THEOREM 8. *The error in the cost functional, $e^J(\mu) := J(\mu) - J_N(\mu)$, may be bounded by*

$$\begin{aligned} e^J(\mu) \le \Delta_N^J(\mu) := {} & \left( \Delta t \sum_{k=1}^{K} \big\| r_{du}^k(\cdot;\mu) \big\|_{X'}^2 \right)^{\frac{1}{2}} \frac{\Delta_{N,K}^{\mathrm{pr}}(\mu)}{\sqrt{\alpha_{LB}(\mu)}} \\ & + \frac{\gamma_d}{\alpha_{LB}(\mu)} (\Delta_{N,K}^{\mathrm{pr}}(\mu))^2 + \Delta t \left| \sum_{k=1}^{K} r_{pr}^k\big(p_N^k(\mu);\mu\big) \right|, \quad \forall \mu \in \mathcal{D}, \end{aligned} \tag{43}$$

*where $\Delta_{N,K}^{\mathrm{pr}}(\mu)$ is defined in Lemma 7.*

THEOREM 9. *The error in the cost gradient,* $e^{\nabla J}(\mu) = \nabla_\mu J(\mu) - \nabla_\mu J_N(\mu)$*, satisfies*

$$\left\|e^{\nabla J}(\mu)\right\| \leq \Delta_N^{\nabla J}(\mu) := \left\|\Delta_N^{\nabla_\mu J}(\mu)\right\|, \quad \forall \mu \in \mathcal{D}, \tag{44}$$

*where* $\Delta_N^{\nabla_\mu J}(\mu)$ *is a vector whose* $i$*th component is the bound on the error of the* $i$*th component of the gradient, given by*

$$\begin{aligned}\Delta_N^{\nabla_{\mu_i} J}(\mu) := &\left(\Delta t \sum_{k=1}^{K} \|f_{\mu_i}(\cdot;\mu)\|_{X'}^2\right)^{\frac{1}{2}} \frac{\Delta_{N,1}^{\mathrm{du}}(\mu)}{\sqrt{\alpha_{LB}(\mu)}} \\ &+ \frac{\gamma_{a_{\mu_i}}^{UB}(\mu)}{\alpha_{LB}(\mu)} \Delta_{N,K}^{\mathrm{pr}}(\mu)\Delta_{N,1}^{\mathrm{du}}(\mu) + \frac{\gamma_{a_{\mu_i}}^{UB}(\mu)}{\sqrt{\alpha_{LB}(\mu)}} \Delta_{N,K}^{\mathrm{pr}}(\mu) \left(\Delta t \sum_{k=1}^{K} \left\|p_N^k(\mu)\right\|_X^2\right)^{\frac{1}{2}} \\ &+ \frac{\gamma_{a_{\mu_i}}^{UB}(\mu)}{\sqrt{\alpha_{LB}(\mu)}} \Delta_{N,1}^{\mathrm{du}}(\mu) \left(\Delta t \sum_{k=1}^{K} \left\|u_N^k(\mu)\right\|_X^2\right)^{\frac{1}{2}}.\end{aligned} \tag{45}$$

*and* $\Delta_{N,K}^{\mathrm{pr}}(\mu)$ *and* $\Delta_{N,1}^{\mathrm{du}}(\mu)$ *are defined in Lemma 7.*

**3.3. Computational procedure.** Like other model reduction methods, the RB method is traditionally divided into a computationally expensive offline phase and a computationally efficient online phase. A detailed discussion of the necessary computations and computational cost can be found e.g. in [24], we thus only present a short summary and focus on the main ingredients and costs.

During the offline phase, the reduced basis for elliptic problems (resp. parabolic problems) is usually built incrementally using a greedy (resp. POD-greedy) algorithm [28]. The greedy algorithm chooses the parameters $\mu_n^{\mathrm{pr}}$ and $\mu_n^{\mathrm{du}}$ at which snapshots are taken by searching for the largest *a posteriori* error bound over a training parameter set. In the elliptic case, the snapshots $u(\mu_n^{\mathrm{pr}})$ and $p(\mu_n^{\mathrm{du}})$ are computed, orthonormalized, and added directly to the basis. In the parabolic case, the $X$-orthogonal projection of $u^k(\mu_n^{\mathrm{pr}})$ and $p^k(\mu_n^{\mathrm{du}})$, $k \in \mathbb{K}$, onto the current basis is computed, and the largest POD mode of the time history of the projection error is added to the basis. The costs of calculating the FE snapshots during the offline phase are thus $2N$ $\mathcal{N}$-dimensional $A(\mu)$-solves (one primal and one dual solve for each $N$) for the elliptic case, and $2NK$ $\mathcal{N}$-dimensional $A(\mu)$-solves in the parabolic case (the cost of time integration without LU-factorization for both the dual and the primal for each $N$). Here, $A(\mu)$ is the FE matrix corresponding to the bilinear form $a$.

Additionally, in order to facilitate efficient online error estimation, the offline phase requires $(Q_a+Q_f)\,\mathcal{N}$-dimensional solves of the $X$-inner product matrix (denoted $\mathbb{X}$) per vector added to the basis. Since the matrix $\mathbb{X}$ is parameter-independent, we may precompute its (sparse) LU-factorization once at the start of the optimization, allowing the necessary $\mathbb{X}$-solves to be efficiently executed offline.

As mentioned above, we also assume that we have access to $\alpha_{LB}(\mu)$, a lower bound on $\alpha(\mu)$, and to $\gamma_{a_\mu}^{UB}(\mu)$, an upper bound on $\gamma_{a_\mu}(\mu)$. These can be calculated via either the "min-theta" approach [24], or more generally, via the successive constraint method [15]. We also assume access to the continuity constant $\gamma_d$, which may be obtained via a generalized eigenvalue problem.

**4. Trust region framework.** The canonical trust region optimization framework solves a set of successive optimization subproblems, defined as

$$\min_s \mathcal{M}^k\big(\mu^k + s\big) \qquad \text{s.t.} \|s\| \leq \delta^k,$$

where $\mu^k$ is the current optimization iterate, $\mathcal{M}^k(\mu)$ is the model function used to approximate the true objective function $J(\mu)$, $\delta^k$ is the trust region radius, and we solve for $s$, the optimal step within the defined trust region; we refer to the book [9] for an extensive resource on trust region methods. The model function $\mathcal{M}^k(\mu)$ changes at each trust region iteration and is often a local quadratic Taylor expansion. Other surrogates, however, have also been considered in the literature [1, 2, 16, 18, 30].

To determine if the step $s$ should be accepted, the ratio $\rho^k = \frac{\mathcal{M}(\mu^k) - \mathcal{M}(\mu^{k+1})}{J(\mu^k) - J(\mu^{k+1})}$, a measure of how well the model predicts decrease in the true cost, is computed. The value of $\rho^k$ is used to determine not only whether or not the optimization step is accepted, but also whether and how to change trust region radius for the next optimization subproblem. One criticism of this approach in the POD or general surrogate model context is that the computation of $\rho^k$ requires evaluating the true objective function $J(\mu)$, which may be computationally expensive [2, 30].

In this work, the reduced basis cost $J_N^k(\mu)$ serves as the model function $\mathcal{M}^k(\mu)$. We note that the letter $k$ is used to indicate the $k$th trust region iterate as well as the $k$th timestep in the parabolic time integration; however, the meaning should be clear from context. The *a posteriori* error bounds developed in section 3 are used $(i)$ to minimize the number of true objective evaluations required, and $(ii)$ together with a recent result from Yue and Meerbergen [30], to guarantee convergence of the approach to the optimum of the high-fidelity model.

We again stress that we do not follow the traditional RB offline/online strategy here, i.e. generate the RB approximation in the offline stage and only revert to the online evaluation during the optimization. Instead, we generate the RB approximation *on the fly* during the optimization: we use the online evaluation to efficiently solve the trust region subproblems and update the reduced basis along the optimization trajectory only if the *a posteriori* error bounds indicate a need to do so. The offline and online stages thus intertwine and each reduced basis update requires an FE snapshot computation and update of the error bound computation as discussed in the last section.

**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. It has been shown, however, that trust region optimizations converge even if inexact model and gradient information is used [7, 13, 25]. In [30], Yue and Meerbergen relax the stringent first-order accuracy 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, [30] replaces the first-order condition with the following relaxed first-order condition (adapted to our notation from sections 2 and 3:

$$\text{(46a)}\quad \left|J_N^k(\mu^k) - J(\mu^k)\right| \le \Delta_N^{J,k}(\mu^k) \quad \text{and} \quad \left\|\nabla_\mu J_N^k(\mu^k) - \nabla_\mu J(\mu^k)\right\| \le \Delta_N^{\nabla J,k}(\mu^k)$$

$$\text{(46b)}\quad \frac{\Delta_N^{J,k}(\mu^k)}{J_N^k(\mu^k)} \le \tau_J \quad \text{and} \quad \frac{\Delta_N^{\nabla J,k}(\mu^k)}{\left\|\nabla_\mu J_N^k(\mu^k)\right\|} \le \tau_{\nabla J}$$

for any given $\tau_J > 0$ and $\tau_{\nabla J} > 0$. There are two parts to this condition: (46a) requires that error bounds exist for both the cost function and its gradient, while (46b) requires that the reduced basis model be able to meet arbitrarily small tolerances $\tau_J$ and $\tau_{\nabla J}$. The sufficient decrease condition is similarly replaced, with an "error-aware sufficient decrease condition" (EASDC):

$$J_N^{k+1}(\mu^{k+1}) \le J_N^k(\mu_{AGC}^k) \tag{47}$$

where $\mu_{AGC}$ is known as the "approximate generalized Cauchy point", a point that achieves sufficient decrease in the reduced basis model in a descent direction.

To ensure that all optimization iterates satisfy the EASDC, Yue and Meerbergen present a procedure designed to reject steps which violate this condition [30]. We summarize the procedure in [30] using our notation here and begin by noting that a sufficient condition for (47) is

$$J_N^k(\mu^{k+1}) + \Delta_N^{J,k}(\mu^{k+1}) + \Delta_N^{J,k+1}(\mu^{k+1}) \le J_N^k(\mu_{AGC}^k).$$

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

$$J_N^k(\mu^{k+1}) + \Delta_N^{J,k}(\mu^{k+1}) < J_N^k(\mu_{AGC}^k), \tag{48}$$

because we may update the RB model with basis functions taken at $\mu^{k+1}$ before the next subproblem solve to ensure that $\Delta_N^{J,k+1}(\mu^{k+1}) = 0$, thus satisfying the sufficient condition. We can check this cheaply, and if it holds we may accept the iterate $\mu^{k+1}$, updating the reduced basis model at $\mu^{k+1}$ as necessary.

Otherwise, we note that a necessary condition for (47) is

$$J_N^k(\mu^{k+1}) - \Delta_N^{J,k}(\mu^{k+1}) - \Delta_N^{J,k+1}(\mu^{k+1}) \le J_N^k(\mu_{AGC}^k), \tag{49}$$

so we check

$$J_N^k(\mu^{k+1}) - \Delta_N^{J,k}(\mu^{k+1}) \le J_N^k(\mu_{AGC}^k). \tag{50}$$

If this condition fails, satisfying (49) may require a large error bound in the next model, leading to inaccurate approximations, so we reject the iterate $\mu^{k+1}$, shrink the trust region (set $\epsilon_L = \kappa_{tr}\epsilon_L$ for some $\kappa_{tr} \in (0,1)$), and re-solve the optimization sub-problem. Otherwise, if (50) holds, we update the model at $\mu^{k+1}$ and check (47). If it holds, then we accept $\mu^{k+1}$. Otherwise, we reject $\mu^{k+1}$, shrink the trust region, and re-solve the optimization subproblem.

If the relaxed first-order condition is satisfied, and all iterates satisfy the EASDC, Yue and Meerbergen show convergence of the trust region algorithm to the optimum of the high-fidelity model under mild assumptions [30] which are satisfied in our setting.

**4.2. Trust region reduced basis algorithm.** The optimization subproblem for the trust region reduced basis algorithm is defined as follows:

$$\min_{\mu^{k+1}} J_N^k\left(\mu^{k+1}\right) \quad \text{s.t.} \quad \left|\frac{\Delta_N^{J,k}\left(\mu^{k+1}\right)}{J_N^k(\mu^{k+1})}\right| \leq \epsilon_L \tag{51}$$

where $\epsilon_L$ is the maximum allowable relative error in the cost. We note that the error bound on the cost functional, $\Delta_N^{J,k}(\mu)$, is used to implicitly define the trust region; if the subproblem solver steps outside of this region, we use backtracking to return to a region where $\Delta_N^{J,k}(\mu)$ is sufficiently low.

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

$$\text{(a)} \ \left\|\nabla J_N^k(\mu)\right\| \leq \tau_{sub} \quad \text{or} \quad \text{(b)} \ \beta\epsilon_L \leq \frac{\Delta_N^{J,k}(\mu)}{J_N^k(\mu)} \leq \epsilon_L \tag{52}$$

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 where the model becomes inaccurate. Overall convergence is reached when the norm of the true gradient is less than a tolerance $\tau \geq \tau_{sub}$, i.e.

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

The reduced model employed is an iteratively-built reduced basis model that is updated only when the subproblem optimization terminates on condition (52b), indicating that our RB model is not sufficiently accurate. In the elliptic case, updating the RB model entails adding $u\left(\mu^k\right)$ and $\psi\left(\mu^k\right)$ to the primal and dual bases. In doing so, we automatically satisfy (46b), since the reduced basis is able to exactly represent the FE solution $\mu^k$. In the parabolic case, we may add singular modes from the primal and dual solutions at the current iterate until (46b) is satisfied.

The algorithm steps are summarized in Algorithm 1.

**5. Numerical tests.** In this section, we introduce a thermal fin model optimization problem. The optimization is then solved using three different approaches:

1. an FE-only approach, consisting of an interior point optimizer [3, 4, 29] as implemented in MATLAB's `fmincon` routine, using only the high-dimensional FE model for its function and gradient evaluations,
2. a traditional RB approach, consisting of an offline phase, in which a global reduced basis is built, and an online phase, in which MATLAB's `fmincon` interior point implementation is used to solve the optimization using reduced basis function and gradient evaluations, and
3. the trust region reduced basis algorithm presented in subsection 4.2, employing the BFGS quasi-Newton method to solve each trust region subproblem using only reduced evaluations, and solving the full model as needed to progressively build the reduced basis along the optimization trajectory.

In subsection 5.2 we present results regarding the quality of the reduced basis approximation employed in the traditional RB approach. Subsection 5.3 compares performance of the three optimization approaches for a two-parameter and a six-parameter optimization. The algorithm parameters used for the optimization tests and construction of the reduced bases for the traditional offline/online approach are shown in Table 1.

**Algorithm 1** Trust region reduced basis optimization

1: Initialize. Let $k = 0$, and choose $\tau \geq \tau_{sub} \geq 0$, $\tau_{\nabla J} \in (0,1)$, and $\beta \in (0,1)$. Additionally, choose $\mu^0$, $\epsilon_L$, and $\kappa_{tr} < 1$, a trust region decrease factor. Initialize the initial reduced basis model at $\mu^0$.
2: Solve the optimization subproblem (51) with termination criteria (52).
3: **if** the sufficient condition (48) holds **then**
4: Accept and update the reduced model at $\mu^{k+1}$ and go to Algorithm 1.
5: **else if** the necessary condition (50) fails **then**
6: Reject $\mu^{k+1}$, set $\epsilon_L = \kappa_{tr}\epsilon_L$ and return to Algorithm 1.
7: **else**
8: Update the model at $\mu^{k+1}$.
9: **if** the EASDC (47) holds **then**
10: Accept $\mu^{k+1}$ and go to Algorithm 1.
11: **else**
12: Reject $\mu^{k+1}$, set $\epsilon_L = \kappa_{tr}\epsilon_L$ and return to Algorithm 1.
13: **end if**
14: **end if**
15: If $\left\|\nabla J_N^{k+1}(\mu^{k+1})\right\| + \Delta_N^{\nabla_\mu J,k+1}(\mu^{k+1}) \leq \tau$, return $\mu^{k+1}$ and stop. Otherwise, go to Algorithm 1.

| Parameter | Symbol | Value for numerical tests |
|---|---|---|
| "close" to TR boundary threshold | $\beta$ | 0.95 |
| trust region boundary | $\epsilon_L$ | 0.1 |
| RB gradient error tolerance | $\tau_{\nabla J}$ | 0.1 |
| subproblem convergence tolerance | $\tau_{sub}$ | 1e-8 |
| overall convergence tolerance | $\tau$ | 5e-4 |

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

**5.1. Thermal fin model problem.** We consider a two-dimensional thermal fin with a fixed geometry (Fig. 1) consisting of a central post and four horizontal subfins, with interior $\Omega$ and boundary $\Gamma$ [19]. The fin conducts heat away from a uniform heat flux source at the root of the fin, $\Gamma_{\text{root}}$, 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, $k_i \in [0.1, 10]$, of the subfins and the central post and the Biot number, $\text{Bi} \in [0.01, 1]$, a nondimensional heat transfer coefficient relating the convective heat transfer coefficient to the conductivity of the fin. We will consider a two-parameter and a six-parameter optimization. In the two-parameter optimization, we fix $k_0 = 1$ and constrain the subfin conductivities to vary together (i.e., $k_1 = k_2 = k_3 = k_4$). In the six-parameter optimization, all six components of $\mu$ may vary independently.

**5.1.1. Elliptic model problem.** The temperature distribution within the fin, $u(\mu)$, is governed by the steady heat equation with a unit Neumann flux boundary condition at the root of the fin to model a heat source. We enforce a Robin boundary condition on all other external boundaries to model convective heat losses, and we enforce continuity of both $u$ and its gradient at interfaces between the fin post and subfins. The output of interest is the average temperature of the fin root, $T_{\text{root}}(\mu) =$

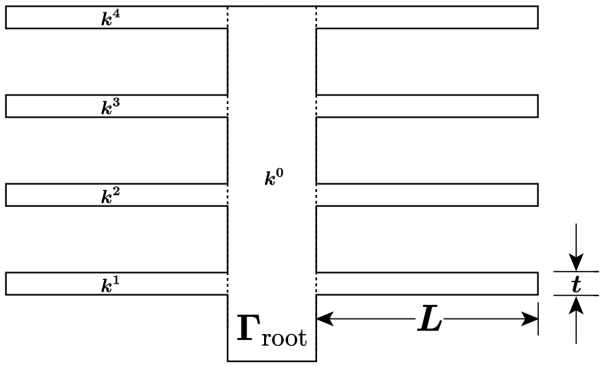


Fig. 1: Thermal fin geometry

$\mathcal{L}(u(\mu)) = \int_{\Gamma_{\text{root}}} u(\mu)$. For the high-fidelity discretization we consider a piecewise linear FE approximation space $X$ on a quasi-uniform unstructured mesh of dimension $\dim(X) = \mathcal{N} = 17899$.

For our optimization, we generate artificial experimental measurements $\hat{T}_{\text{root}}$ by considering a thermal fin whose parameters are fixed but unknown. We then aim to infer the unknown parameters by minimizing the output least-squares formulation

$$s(\mu) = \frac{1}{2}(T_{\text{root}}(\mu) - \hat{T}_{\text{root}})^2 = \frac{1}{2}\left\|\mathcal{L}(u(\mu)) - \hat{T}_{\text{root}}\right\|_{\mathbb{R}}^2. \tag{53}$$

To obtain a cost function of the form presented in subsection 2.2, we define $d(u,v) \equiv \frac{1}{2}(\mathcal{L}u, \mathcal{L}v)_{\mathbb{R}}$ and $\ell(v) \equiv -\left(\mathcal{L}v, \hat{T}_{\text{root}}\right)_{\mathbb{R}}$, drop the constant term $\frac{1}{2}(\hat{T}_{\text{root}}, \hat{T}_{\text{root}})_{\mathbb{R}}$, and introduce the regularization $\mathcal{R}(\mu) = \|\frac{\mu-\hat{\mu}}{\hat{\mu}}\|_{\mathbb{R}}^2$, where $\hat{\mu} \in \mathcal{D}$.

**5.1.2. Parabolic model problem.** We now consider the time-varying temperature distribution within the fin in the time interval $I = ]0, 10]$ governed by the time-dependent heat equation with a sinusoidal control input $y(t) = \cos(t)$ at the root of the fin. As in the elliptic problem, we enforce Robin boundary conditions at all other external boundaries and continuity of temperature and heat flux at all internal interfaces. In the parabolic problem, our output of interest is the average temperature of the entire fin at the current timestep, $T^k_{\text{avg}}(\mu) = \mathcal{L}(u^k(\mu)) = \int_\Omega u^k(\mu)$. Again, we generate artifical output data $\hat{T}^k_{\text{avg}}$ for all $k \in \mathbb{K}$ by considering a fin whose parameters are fixed but unknown. Thus, our output least-squares formulation is given by

$$s(\mu) = \Delta t \sum_{k=1}^{K} \frac{1}{2}(T^k_{\text{avg}}(\mu) - \hat{T}^k_{\text{avg}})^2 = \Delta t \sum_{k=1}^{K} \frac{1}{2}\left\|\mathcal{L}\big(u^k(\mu)\big) - \hat{T}^k_{\text{avg}}\right\|_{\mathbb{R}}^2. \tag{54}$$

Analogous to the elliptic case, we may obtain a least-squares cost functional of the form presented in subsection 2.3 by defining $d(u,v) \equiv \frac{1}{2}(\mathcal{L}u, \mathcal{L}v)_{\mathbb{R}}$ and $\ell(v) \equiv -\left(\mathcal{L}v, \hat{T}_{\text{avg}}\right)_{\mathbb{R}}$, dropping the constant term $\frac{1}{2}(\hat{T}^k_{\text{avg}}, \hat{T}_{\text{avg}})_{\mathbb{R}}$, and introducing the regularization $\mathcal{R}(\mu) = \|\frac{\mu-\hat{\mu}}{\hat{\mu}}\|_{\mathbb{R}}^2$, where $\hat{\mu} \in \mathcal{D}$.

**5.1.3. Problem data.** In subsection 5.3, we compare the performance of our trust region algorithm to that of the FE-only and RB-only `fmincon` interior point

approaches for the 2D and 6D elliptic and parabolic optimizations. Optimization trials in each case are run on ten different least-squares cost functionals, corresponding to ten randomly selected values for $\mu^*$ within the parameter domain. For each randomly selected $\mu^*$-value, we obtain $\hat{T}_{\text{root}}$ or $\hat{T}^k_{\text{avg}}$ from the high-fidelity FE model. The value $\mu^*$ is then used as the regularization function parameter $\hat{\mu}$. Table 2 specifies the data used to generate numerical results in subsequent sections.

| Parameter | Symbol | Elliptic | Parabolic |
|---|---|---|---|
| FE dimension | $\mathcal{N}$ | 17899 | |
| $d$-continuity constant | $\gamma_d$ | 0.7999 | 9.6970 |
| Number of time steps | $K$ | – | 100 |
| Regularization scaling factor | $\lambda$ | 1 | 0.01 |

Table 2: Problem data used for generation of numerical results. The $X$-inner product was defined as $(\cdot,\cdot)_X = a(\cdot,\cdot;\bar{\mu})$, for $\bar{\mu} = (1,1,1,1,1,0.1)^T$.

**5.2. Global reduced-basis approximation quality.** In order to compare the performance of the proposed trust region reduced basis approach to the performance of a traditional RB approach, we must first generate a reduced basis offline. To this end, we introduce a training set $\mathcal{D}_{\text{train}} \subset \mathcal{D}$ of size $n_{\text{train}}$, pick an initial parameter $\mu^{(1)}$ and the desired error tolerances $\tau$ and $\tau_{\nabla J}$. We then employ a slight variation of the greedy algorithm, which chooses a single parameter at which to add both the primal and dual solutions to their respective bases based on the *a posteriori* error bounds in the cost and cost gradient. The exact procedure employed is given in Algorithm 2. We note that the tolerance values $\tau$ and $\tau_{\nabla J}$ are the same values used for the optimization trials (Table 1). This ensures that error everywhere on the training grid is low enough to meet the convergence tolerances. However, we note that this is insufficient to guarantee low error over the entire parameter domain.

**Algorithm 2** Generate global reduced-basis

Choose $\mathcal{D}_{\text{train}} \subset \mathcal{D}$, $\tau > 0$, and $\tau_{\nabla J} > 0$
Initialize primal and dual reduced bases at $\mu^{(1)} \in \mathcal{D}_{\text{train}}$
**while** $\max_{\mu\in\mathcal{D}_{\text{train}}} \frac{\Delta_N^J(\mu)}{J_N(\mu)} > \tau$ or $\max_{\mu\in\mathcal{D}_{\text{train}}} \frac{\Delta_N^{\nabla_\mu J}(\mu)}{\|\nabla_\mu J_N(\mu)\|} > \tau_{\nabla J}$ **do**
  **if** $\max_{\mu\in\mathcal{D}_{\text{train}}} \frac{\Delta_N^J(\mu)}{J_N(\mu)} > \tau$ **then**
    $\mu^* \leftarrow \arg\max_{\mu\in\mathcal{D}_{\text{train}}} \frac{\Delta_N^J(\mu)}{J_N(\mu)}$
  **else**
    $\mu^* \leftarrow \arg\max_{\mu\in\mathcal{D}_{\text{train}}} \frac{\Delta_N^{\nabla_\mu J}(\mu)}{\|\nabla_\mu J_N(\mu)\|}$
  **end if**
  Update the reduced-basis at $\mu^*$
**end while**

We now present the standard convergence results for both the two- and the six-parameter case. Specifically, Tables 3-6 present, as a function of $N$, the maximum relative error bounds $\Delta^{\text{pr}}_{\text{rel,max}}$, $\Delta^{\text{du}}_{\text{rel,max}}$, $\Delta^{J}_{\text{rel,max}}$, $\Delta^{\nabla_\mu J}_{\text{rel,max}}$, as well as the average effectivities $\bar{\eta}_{\text{pr}}$, $\bar{\eta}_{\text{du}}$ $\bar{\eta}_J$, and $\bar{\eta}_{\nabla J}$, over a randomly generated test set $\Xi \subset \mathcal{D}$ of size

$n_{\text{train}} = 100$, i.e., for the elliptic case we have

$$\Delta^{\text{pr}}_{\text{rel,max}} = \max_{\mu\in\Xi} \frac{\Delta^{\text{pr}}_N(\mu)}{\|u(\mu)\|_X}, \quad \Delta^{\text{du}}_{\text{rel,max}} = \max_{\mu\in\Xi} \frac{\Delta^{\text{du}}_N(\mu)}{\|p(\mu)\|_X},$$
$$\Delta^{J}_{\text{rel,max}} = \max_{\mu\in\Xi} \frac{\Delta^{J}_N(\mu)}{J(\mu)}, \quad \Delta^{\nabla_\mu J}_{\text{rel,max}} = \max_{\mu\in\Xi} \frac{\Delta^{\nabla_\mu J}_N(\mu)}{\nabla_\mu J(\mu)},$$

and

$$\eta_{\text{pr}}(\mu) = \frac{\Delta^{\text{pr}}_N(\mu)}{\|e^{\text{pr}}(\mu)\|_X}, \quad \eta_{\text{du}}(\mu) = \frac{\Delta^{\text{du}}_N(\mu)}{\|e^{\text{du}}(\mu)\|_X},$$
$$\eta_J(\mu) = \frac{\Delta^{J}_N(\mu)}{|e^{J}(\mu)|}, \quad \eta_{\nabla_\mu J}(\mu) = \frac{\Delta^{\nabla_\mu J}_N(\mu)}{\|e^{\nabla_\mu J}(\mu)\|}.$$

We omit the definition of the corresponding quantities for the parabolic case since they are defined similarly.

We first observe that all of the bounds are generally sharper for the 2 parameter case than for the 6 parameter case. We also observe that the effectivities of the primal bounds are close to 1 for all cases considered thus indicating very sharp bounds. The dual effectivities are considerably larger, which is due to the primal error propagating to the dual problem and entering into the dual error bound formulation. We next note that the error bounds for the cost functional converge very fast and that we can thus achieve the required error tolerance for the trust region approach. Except for small $N$, the effectivities are approximately $\mathcal{O}(10-100)$ for the elliptic case and approximately $\mathcal{O}(100-1000)$ for the parabolic case, which still seems acceptable given the fast convergence of the reduced-basis approximation. Finally — and as anticipated — the bounds for the cost gradients perform worst of all. We do, however, achieve the required accuracy despite the large effectivities. We also recall our discussion at the end of subsection 3.1.2 that fairly large relative errors in the gradient are permissible in the trust region approach, and the result thus poses no impediment for our approach.

| $N$ | $\Delta^{\text{pr}}_{\text{rel,max}}$ | $\bar{\eta}_{\text{pr}}$ | $\Delta^{\text{du}}_{\text{rel,max}}$ | $\bar{\eta}_{\text{du}}$ | $\Delta^{J}_{\text{rel,max}}$ | $\bar{\eta}_J$ | $\Delta^{\nabla_\mu J}_{\text{rel,max}}$ | $\bar{\eta}_{\nabla_\mu J}$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 2.97 | 2.29 | 310 | 49.4 | 3.67 | 62.8 | 205 | 1162 |
| 4 | 2.36 | 2.22 | 80.5 | 47.1 | 0.55 | 54.9 | 37.1 | 439 |
| 6 | 1.91 | 2.76 | 62.2 | 44.9 | 3.1e-2 | 53.2 | 3.38 | 605 |
| 8 | 0.64 | 2.52 | 43.4 | 44.2 | 1.2e-2 | 39.9 | 1.76 | 1047 |
| 10 | 0.41 | 3.24 | 34.1 | 48.4 | 7.5e-3 | 40.3 | 1.66 | 1882 |

Table 3: Two-parameter elliptic thermal fin problem: convergence rate and effectivities of global reduced basis built using greedy algorithm (Algorithm 2)

| $N$ | $\Delta^{\text{pr}}_{\text{rel,max}}$ | $\bar{\eta}_{\text{pr}}$ | $\Delta^{\text{du}}_{\text{rel,max}}$ | $\bar{\eta}_{\text{du}}$ | $\Delta^{J}_{\text{rel,max}}$ | $\bar{\eta}_{J}$ | $\Delta^{\nabla_\mu J}_{\text{rel,max}}$ | $\bar{\eta}_{\nabla_\mu J}$ |
|---|---|---|---|---|---|---|---|---|
| 4 | 7.70 | 5.47 | 1.7e3 | 81.0 | 1.70 | 1542 | 422 | 2.4e3 |
| 8 | 5.83 | 5.39 | 8.1e2 | 85.8 | 0.84 | 250 | 323 | 2.4e3 |
| 16 | 1.70 | 6.25 | 1.2e2 | 144 | 2.0e-2 | 174 | 45.4 | 4.2e3 |
| 32 | 0.66 | 6.23 | 9.85 | 190 | 4.0e-4 | 97.9 | 0.68 | 6.6e4 |
| 48 | 0.16 | 6.88 | 2.00 | 190 | 1.8e-5 | 129 | 9.7e-2 | 4.9e5 |

Table 4: Six-parameter elliptic thermal fin problem: convergence rate and effectivities of global reduced basis built by greedy algorithm (Algorithm 2)

| $N$ | $\Delta^{\text{pr}}_{\text{rel,max}}$ | $\bar{\eta}_{\text{pr}}$ | $\Delta^{\text{du}}_{\text{rel,max}}$ | $\bar{\eta}_{\text{du}}$ | $\Delta^{J}_{\text{rel,max}}$ | $\bar{\eta}_{J}$ | $\Delta^{\nabla_\mu J}_{\text{rel,max}}$ | $\bar{\eta}_{\nabla_\mu J}$ |
|---|---|---|---|---|---|---|---|---|
| 5 | 0.54 | 2.47 | 420 | 1.2e2 | 43.0 | 3.4e2 | 2.3e4 | 8.9e3 |
| 15 | 0.27 | 2.16 | 83.1 | 2.6e2 | 0.34 | 1.5e3 | 4.1e2 | 6.0e3 |
| 30 | 2.8e-2 | 2.00 | 9.29 | 1.2e3 | 5.2e-3 | 6.4e2 | 11.5 | 1.6e4 |
| 45 | 4.8e-3 | 2.06 | 1.70 | 1.6e3 | 2.2e-4 | 1.4e2 | 2.96 | 1.1e4 |
| 60 | 1.7e-3 | 2.03 | 0.25 | 2.4e3 | 2.1e-5 | 2.4e2 | 1.14 | 4.0e4 |

Table 5: Two-parameter parabolic thermal fin problem: convergence rate and effectivities of global reduced basis built by greedy algorithm (Algorithm 2).

| $N$ | $\Delta^{\text{pr}}_{\text{rel,max}}$ | $\bar{\eta}_{\text{pr}}$ | $\Delta^{\text{du}}_{\text{rel,max}}$ | $\bar{\eta}_{\text{du}}$ | $\Delta^{J}_{\text{rel,max}}$ | $\bar{\eta}_{J}$ | $\Delta^{\nabla_\mu J}_{\text{rel,max}}$ | $\bar{\eta}_{\nabla_\mu J}$ |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.55 | 2.8 | 1.3e3 | 1.2e3 | 5.0 | 8.8e3 | 4.3e3 | 4.7e4 |
| 60 | 4.3e-2 | 3.0 | 29 | 1.6e3 | 3.1e-2 | 3.1e3 | 47 | 6.0e4 |
| 100 | 6.0e-3 | 3.0 | 6.4 | 2.4e3 | 8.6e-4 | 1.1e3 | 4.8 | 9.1e4 |
| 140 | 1.1e-3 | 3.0 | 1.1 | 3.3e3 | 2.1e-5 | 7.6e2 | 0.98 | 4.9e5 |
| 180 | 2.5e-4 | 3.2 | 0.17 | 2.1e3 | 7.8e-7 | 1.3e2 | 0.15 | 9.9e5 |

Table 6: Six-parameter parabolic thermal fin problem: convergence rate and effectivities of global reduced-basis built by greedy algorithm (Algorithm 2).

**5.3. Algorithm performance.** The optimization problem is solved using the FE-only interior point, traditional RB interior point, and trust region RB approaches. We consider ten random least-squares cost functions (as discussed in subsection 5.1.3) and solve the optimization for each cost function using the same set of ten random initial conditions, resulting in a total of 100 optimization trials. Algorithm parameters and problem data used are tabulated in Table 1 and Table 2. Performance results measured in terms of optimization run time and required number of FE evalutions for the elliptic case are presented in Figures 2 and 3, and in Figures 4 and 5 for the parabolic case. Note that in the parabolic case the number of FE evaluations stated in the figures corresponds to the number of full forward integrations in time.

Overall, the combined trust region reduced basis optimization approach consistently reduces the number of full FE evaluations required to locate an optimum relative to the two other optimization approaches tested. On average, compared to the FE-interior point method, the trust region RB approach requires 39% (30%) as many full solves in the 2D (6D) elliptic case. In the parabolic case, the trust region RB approach requires 30% as many full solves in the 2D case, and only 14% of the full

solves required in the 6D case. We note that in our numerical trials, Line 12 of Algorithm 1 is never reached — i.e., we never update the model (Line 8) at a potential iterate only to reject it, thus never 'wasting' a full solve. These results will vary for more general problems, but by defining the trust region systematically via the error estimates, we can influence how often this would occur.

By building the RB adaptively along the optimization trajectory, the trust region reduced basis method also reduces the number of full evaluations needed relative to the number of full solves needed in the offline phase of the traditional RB approach: on average, the trust region RB requires 37% and 12% of the full solves needed for the 2D and 6D elliptic cases, and 25% and 3% of the full solves required in the 2D and 6D parabolic cases. Finally, the trust region RB approach is able to reduce the number of evaluations of the reduced system relative to the offline phase of the traditional RB approach, by a mean factor of 3 (3000) in the 2D (6D) elliptic case, and by a mean factor of 6 (8600) in the 2D (6D) parabolic case.

We note that especially high-dimensional parameter optimization problems pose a considerable challenge for the traditional offline/online RB approach. Guaranteeing a certain desired accuracy over the whole parameter space requires a sufficiently fine test sample and thus results in a very expensive offline phase. Moreoever, most of this effort may be wasted since the (online) optimization requires a sufficient accuracy only along the optimization path. The trust region reduced basis optimization approach aims exactly at providing the required accuracy only along the optimization path, thus reducing the overall number (offline+online) of FE solves.

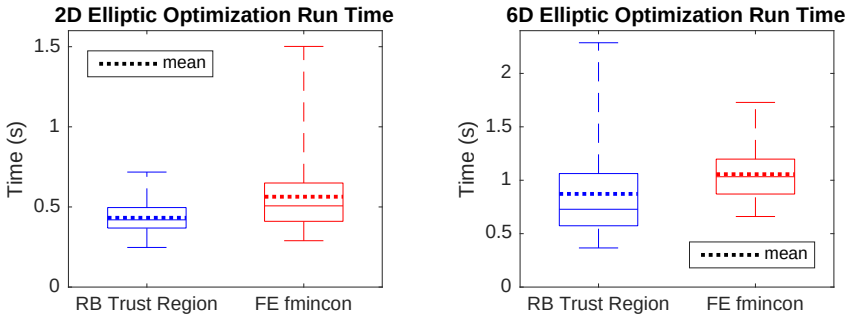


Fig. 2: Runtime comparison for optimizations constrained by elliptic PDEs. In contrast, the traditional offline-online reduced-basis approach for a 2D (6D) optimization runs in 0.04 (0.10) seconds online, but requires 1.6 (4800) seconds offline (on average).

Whether or not the achieved reduction in number of full FE solves translates into a run time speedup depends on the size and complexity of the problem. The trust region RB run time averages 83% of the FE-interior point run time in the elliptic 2D case, 82% in the elliptic 6D case, 72% in the parabolic 2D case, and 44% in the parabolic 6D case. In the elliptic case, these gains are small because the full elliptic problem can be solved relatively inexpensively. In the parabolic 2D case, the overhead involved in building the reduced-basis fills most of the time saved by the three-fold reduction in full solves. In contrast, in the parabolic 6D optimization, the seven-fold reduction in full solves makes the RB overhead a much smaller portion of the overall optimization time. The significant offline computational investment in the traditional

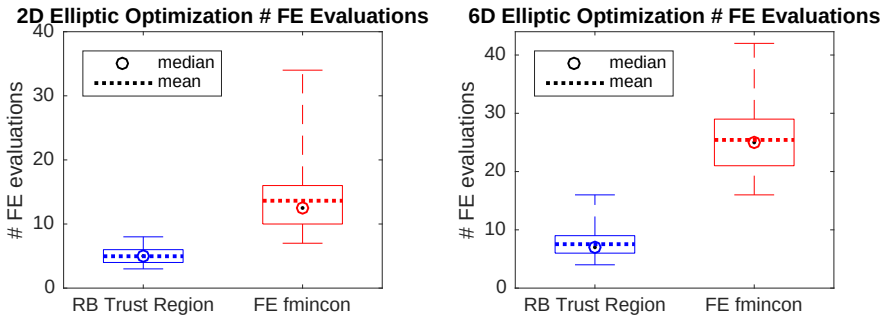


Fig. 3: Comparison of number of full model evaluations required for optimizations constrained by elliptic PDEs. The traditional offline-online reduced-basis approach requires 0 full evaluations online, and an average of 9 (48) full evaluations in the 2D (6D) case offline.

RB approach makes it slower than both our proposed approach and the FE-only approach for all cases tested.

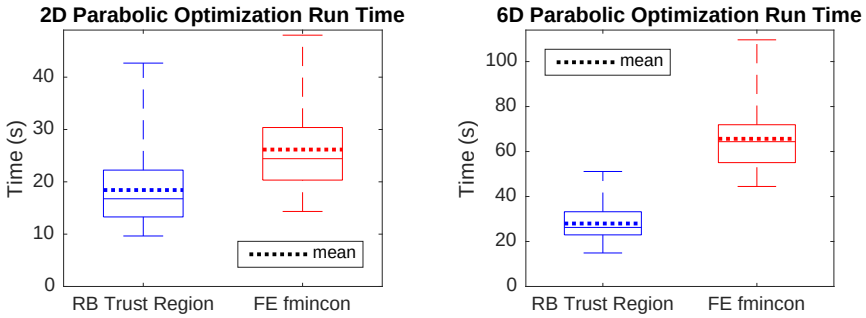


Fig. 4: Runtime comparison for optimizations constrained by parabolic PDEs. In contrast, the traditional offline-online reduced-basis approach for a 2D (6D) optimization runs in 0.23 (2.7) seconds online, but requires 125 seconds (60 hours) offline (on average).

We summarize the achieved gains in Table 7. Our results suggest that there may be potential for greater savings in optimizations of higher parameter dimension. Additionally, we note that the size of our FE discretization is fairly small, especially relative to the size of FE discretizations that might be encountered in real-world problems. Because the trust region approach is able to significantly reduce the number of FE solves relative to the FE-only approach, there is also potential for greater gains in problems with higher-dimensional discretizations.

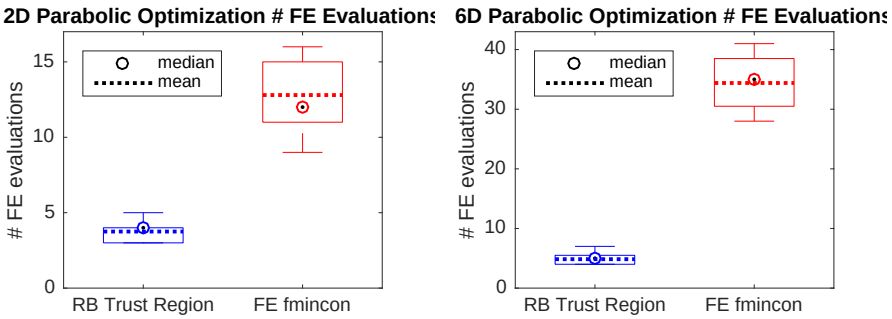


Fig. 5: Comparison of number of full model evaluations required for optimizations constrained by parabolic PDEs. The traditional offline-online reduced-basis approach requires 0 full evaluations online, and an average of 15 (151) full evaluations in the 2D (6D) case offline.

| | Trust region RB run time / FE-`fmincon` run time | | | Trust region RB # FE solves / FE-`fmincon` # FE solves | | |
|---|---|---|---|---|---|---|
| | min | mean | max | min | mean | max |
| 2D Elliptic | 0.20 | 0.83 | 1.48 | 0.09 | 0.39 | 0.75 |
| 6D Elliptic | 0.31 | 0.82 | 2.32 | 0.13 | 0.30 | 0.70 |
| 2D Parabolic | 0.38 | 0.72 | 1.32 | 0.20 | 0.30 | 0.44 |
| 6D Parabolic | 0.21 | 0.44 | 0.80 | 0.10 | 0.14 | 0.24 |

Table 7: Summary of gains made by proposed trust region reduced basis optimization approach relative to MATLAB `fmincon`. Time gains are seen in the majority of test cases, and a reduction in the number of required FE function evaluations is seen in all cases tested, with the largest reduction in the 6D parabolic optimization.

**6. Conclusions.** We have introduced a combined reduced basis trust region framework for PDE-constrained optimization of quadratic cost functionals, as well as novel *a posteriori* error bounds for the reduced basis cost approximation and its gradient. In this approach, reduced basis models are leveraged 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 reduced basis *a posteriori* error bounds are used to choose when to accept and reject trust region optimization iterates. Third, the error bounds are used to systematically determine when to update the reduced model. Fourth and finally, the existence of error bounds for RBM allows rigorous proof of convergence of the algorithm to a stationary point of the full model. We have implemented the proposed algorithm on a thermal fin model problem using least-squares cost functions with up to 6 variable parameters, and achieve reductions in the number of full evaluations needed relative to a high-fidelity interior point approach in all cases tested, with up to 3-fold gains in the elliptic case and 7-fold gains in the parabolic case.

**Acknowledgments.** EQ acknowledges support from the US Fulbright Student Program, the National Science Foundation Graduate Research Fellowship, and the Fannie and John Hertz Foundation 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. MG and KV acknowledge support by the Excellence Initiative of the German federal and state governments and the German Research Foundation through Grant GSC 111.

## Appendix A. Parabolic *a posteriori* error estimation.

### A.1. Proof of Lemma 7 .

*Proof.* We refer to [12] for the proof of (41). We thus only need to show (42). It follows from (16) and (39) that the error in the dual variable, $e^k_{\rm du} \in X$, satisfies

$$m\big(v, e^k_{\rm du} - e^{k+1}_{\rm du}\big) + \Delta t \cdot a\big(v, e^k_{\rm du}; \mu\big) = \Delta t \cdot 2d\big(e^k_{\rm pr}, v\big) + \Delta t \cdot r^k_{\rm du}(v; \mu). \tag{55}$$

Following the usual procedure, we choose $v = e^k_{\rm du}$ to obtain

$$m\big(e^k_{\rm du}, e^k_{\rm du} - e^{k+1}_{\rm du}\big) + \Delta t \cdot a\big(e^k_{\rm du}, e^k_{\rm du}; \mu\big) = 2\Delta t \cdot d\big(e^k_{\rm du}, e^k_{\rm pr}\big) + \Delta t \cdot r^k_{\rm du}\big(e^k_{\rm du}; \mu\big). \tag{56}$$

First, we invoke the Cauchy-Schwarz and Young's inequalities to get

$$m\big(e^k_{\rm du}, e^{k+1}_{\rm du}\big) \le \frac{1}{2} m\big(e^k_{\rm du}, e^k_{\rm du}\big) + \frac{1}{2} m\big(e^{k+1}_{\rm du}, e^{k+1}_{\rm du}\big). \tag{57}$$

Also, from Young's inequality it follows that

$$r^k_{\rm du}\big(e^k_{\rm du}(\mu); \mu\big) \le \frac{1}{\alpha_{LB}} \big\|r^k_{\rm du}(\cdot; \mu)\big\|^2_{X'} + \frac{\alpha_{LB}}{4} \big\|e^k_{\rm du}(\mu)\big\|^2_X, \tag{58}$$

and

$$d\big(e^k_{\rm du}, e^k_{\rm pr}\big) \le \frac{2\gamma_d^2}{\alpha_{LB}(\mu)} \big\|e^k_{\rm pr}\big\|^2_X + \frac{\alpha_{LB}}{8} \big\|e^k_{\rm du}\big\|^2_X; \tag{59}$$

where we also used the continuity of (4) of the bilinear form $d$. Substituting (57), (58), and (59) into (56) we obtain

$$\begin{aligned} &m\big(e^k_{\rm du}, e^k_{\rm du}\big) - m\big(e^{k+1}_{\rm du}, e^{k+1}_{\rm du}\big) + 2\Delta t \cdot a\big(e^k_{\rm du}, e^k_{\rm du}; \mu\big) \\ &\qquad \le \frac{2\Delta t}{\alpha_{LB}(\mu)} \big\|r^k_{\rm du}(\cdot; \mu)\big\|^2_{X'} + \frac{8\gamma_d^2 \Delta t}{\alpha_{LB}(\mu)} \big\|e^k_{\rm pr}\big\|^2_X + \alpha_{LB}(\mu)\Delta t \big\|e^k_{\rm du}\big\|^2_X \end{aligned} \tag{60}$$

Finally, summing over $k = 1..K$ it follow that

$$\begin{aligned} &m\big(e^1_{\rm du}, e^1_{\rm du}\big) + \Delta t \sum_{k=1}^K a(e^k_{\rm du}(\mu), e^k_{\rm du}(\mu); \mu) \le \\ &\qquad \sum_{k=1}^K \left( \frac{8\gamma_d^2 \Delta t}{\alpha_{LB}(\mu)} \big\|e^k_{\rm pr}\big\|^2_X + \frac{2\Delta t}{\alpha_{LB}(\mu)} \big\|r^k_{\rm du}(\cdot; \mu)\big\|^2_{X'} \right). \end{aligned} \tag{61}$$

The result (42) follows from (41). ∎

### A.2. Proof of Theorem 8 .

*Proof.* We first note from (15a) and (36) that

$$e^J(\mu) = \Delta t \sum_{k=1}^K \big\{ d\big(u^k(\mu), u^k(\mu)\big) - d\big(u^k_N(\mu), u^k_N(\mu)\big) + \ell\big(e^k_{\rm pr}(\mu)\big) \big\} \tag{62}$$

Adding and subtracting $d\big(2u^k_N(\mu), e^k_{\rm pr}(\mu)\big)$ within the sum, we obtain

$$\begin{aligned} e^J(\mu) = \Delta t \sum_{k=1}^K \{ &d\big(2u^k_N(\mu), e^k_{\rm pr}(\mu)\big) + \ell\big(e^k_{\rm pr}(\mu)\big) \\ &+ d\big(u^k(\mu), u^k(\mu)\big) - d\big(u^k_N(\mu), u^k_N(\mu)\big) - d\big(2u^k_N(\mu), e^k_{\rm pr}(\mu)\big) \} \end{aligned} \tag{63}$$

and note that

$$\begin{aligned} & d\big(u^k(\mu),u^k(\mu)\big) - d\big(u_N^k(\mu),u_N^k(\mu)\big) - d\big(2u_N^k(\mu),e_{\rm pr}^k(\mu)\big) \\ & \qquad = d\big(u^k(\mu),u^k(\mu)\big) - d\big(2u_N^k(\mu),u^k(\mu)\big) + d\big(u_N^k(\mu),u_N^k(\mu)\big) \\ & \qquad = d\big(e_{\rm pr}^k(\mu),e_{\rm pr}^k(\mu)\big) \end{aligned} \tag{64}$$

Thus, combining (63) and (64), we have

$$e^J(\mu) = \Delta t \sum_{k=1}^K \{d\big(2u_N^k(\mu),e_{\rm pr}^k(\mu)\big) + \ell\big(e_{\rm pr}^k(\mu)\big) + d\big(e_{\rm pr}^k(\mu),e_{\rm pr}^k(\mu)\big)\}. \tag{65}$$

Taking (39) with $v = e_{\rm pr}^k(\mu)$, we can express the first two terms of (65) as follows:

$$\begin{aligned} \sum_{k=1}^K \{d\big(2u_N^k(\mu),e_{\rm pr}^k(\mu)\big) + \ell\big(e_{\rm pr}^k(\mu)\big)\} = & \sum_{k=1}^K \{r_{\rm du}^k\big(e_{\rm pr}^k(\mu);\mu\big) + a\big(e_{\rm pr}^k(\mu),p^k(\mu);\mu\big) \\ & + \frac{1}{\Delta t} m\big(e_{\rm pr}^k(\mu),p_N^k(\mu) - p_N^{k+1}(\mu)\big)\}. \end{aligned} \tag{66}$$

Since $e_{\rm pr}^0(\mu) = 0$ and $p_N^{k+1}(\mu) = 0$, we observe that

$$\sum_{k=1}^K m\big(e_{\rm pr}^k(\mu),p_N^k(\mu) - p_N^{k+1}(\mu)\big) = \sum_{k=1}^K m\big(e_{\rm pr}^k(\mu) - e_{\rm pr}^{k+1}(\mu),p_N^k(\mu)\big). \tag{67}$$

We may then substitute the primal error-residual relationship,

$$r_{\rm pr}^k(v;\mu) = a\big(e_{\rm pr}^k(\mu),v;\mu\big) + \frac{1}{\Delta t} m\big(e_{\rm pr}^k(\mu) - e_{\rm pr}^{k+1}(\mu),v\big), \quad \forall v \in X,$$

with $v = p_N^k(\mu)$ into (66) to obtain

$$\sum_{k=1}^K \{d\big(2u_N^k(\mu),e_{\rm pr}^k(\mu)\big) + \ell\big(e_{\rm pr}^k(\mu)\big)\} = \sum_{k=1}^K \{r_{\rm du}^k\big(e_{\rm pr}^k(\mu);\mu\big) + r_{\rm pr}^k\big(p_N^k(\mu);\mu\big)\}. \tag{68}$$

Substituting (68) into (65), we get

$$e^J(\mu) = \Delta t \sum_{k=1}^K \{r_{\rm du}^k\big(e_{\rm pr}^k(\mu);\mu\big) + r_{\rm pr}^k\big(p_N^k(\mu);\mu\big) + d\big(e_{\rm pr}^k(\mu),e_{\rm pr}^k(\mu)\big)\}. \tag{69}$$

It then follows from the continuity of $d(\cdot,\cdot)$ that

$$e^J(\mu) \le \Delta t \sum_{k=1}^K \{\big\|r_{\rm du}^k(\cdot;\mu)\big\|_{X'} \big\|e_{\rm pr}^k(\mu)\big\|_X + \gamma_d \big\|e_{\rm pr}^k(\mu)\big\|_X^2 + r_{\rm pr}^k\big(p_N^k(\mu);\mu\big)\}. \tag{70}$$

We now invoke the Cauchy-Schwarz inequality to the first term on the right hand side which yields

$$\begin{aligned} e^J(\mu) \le & \left(\Delta t \sum_{k=1}^K \big\|r_{\rm du}^k(\cdot;\mu)\big\|_{X'}^2\right)^{\frac{1}{2}} \left(\Delta t \sum_{k=1}^K \big\|e_{\rm pr}^k(\mu)\big\|_X^2\right)^{\frac{1}{2}} \\ & + \gamma_d \Delta t \sum_{k=1}^K \big\|e_{\rm pr}^k(\mu)\big\|_X^2 + \Delta t \left|\sum_{k=1}^K r_{\rm pr}^k\big(p_N^k(\mu);\mu\big)\right|. \end{aligned} \tag{71}$$

Finally, we note that

$$\Delta t \sum_{k=1}^{K} \|e_{\mathrm{pr}}^k(\mu)\|_X^2 \leq \frac{\Delta t}{\alpha_{LB}(\mu)} \sum_{k=1}^{K} a(e_{\mathrm{pr}}^k(\mu), e_{\mathrm{pr}}^k(\mu); \mu) \leq \frac{1}{\alpha_{LB}(\mu)} |||e_{\mathrm{pr}}^K(\mu)|||_{\mathrm{pr}}^2 \tag{72}$$

to arrive at the desired results by invoking (41). ∎

### A.3. Proof of Theorem 9 .

*Proof.* Similar to the elliptic case, we define $e^{\nabla_{\mu_i} J}(\mu)$ to be the error in the derivative of the cost with respect to $\mu_i$, the $i$th element of the paramector vector $\mu$. It then follows from (17) and (37) that

$$\begin{aligned} e^{\nabla_{\mu_i} J}(\mu) = \Delta t \sum_{k=1}^{K} \Big\{ & f_{\mu_i}\big(e_{\mathrm{du}}^k(\mu); \mu\big) \\ & - \big[a_{\mu_i}\big(u^k(\mu), p^k(\mu); \mu\big) - a_{\mu_i}\big(u_N^k(\mu), p_N^k(\mu); \mu\big)\big] \Big\}. \end{aligned} \tag{73}$$

Follow the same steps as in the proof of Theorem 4 we arrive at

$$\begin{aligned} e^{\nabla_{\mu_i} J}(\mu) = & \Delta t \sum_{k=1}^{K} \Big\{ \|f_{\mu_i}(\cdot; \mu)\|_{X'} \big\|e_{\mathrm{du}}^k(\mu)\big\|_X + \gamma_{a_{\mu_i}}^{UB}(\mu) \big\|e_{\mathrm{pr}}^k(\mu)\big\|_X \big\|e_{\mathrm{du}}^k(\mu)\big\|_X \\ & + \gamma_{a_{\mu_i}}^{UB}(\mu) \big\|e_{\mathrm{pr}}^k(\mu)\big\|_X \big\|p_N^k(\mu)\big\|_X + \gamma_{a_{\mu_i}}^{UB}(\mu) \big\|u_N^k(\mu)\big\|_X \big\|e_{\mathrm{du}}^k(\mu)\big\|_X \Big\} \end{aligned} \tag{74}$$

Invoking the Cauchy-Schwarz inequality and Lemma 7, we obtain (45). ∎

## REFERENCES


[1] N. M. Alexandrov, J. E. Dennis, Jr., R. M. Lewis, and V. Torczon, *A trust-region framework for managing the use of approximation models in optimization*, Structural optimization, 15 (1998), pp. 16–23.

[2] E. Arian, M. Fahl, and E. Sachs, *Managing POD models by optimization methods*, in Proceedings of the 41st IEEE Conference on Decision and Control, Las Vegas, NV, 2002, pp. 3300–3305.

[3] R. H. Byrd, J. C. Gilbert, and J. Nocedal, *A trust region method based on interior point techniques for nonlinear programming*, Mathematical Programming, 89 (2000), pp. 149–185.

[4] R. H. Byrd, M. E. Hribar, and J. Nocedal, *An interior point algorithm for large-scale nonlinear programming*, SIAM Journal on Optimization, 9 (1999), pp. 877–900.

[5] K. Carlberg and C. Farhat, *A compact proper orthogonal decomposition basis for optimization-oriented reduced-order models*, in Proceedings of the 12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2008. AIAA Paper 2008-5964.

[6] R. G. Carter, *Numerical optimization in Hilbert space using inexact function and gradient evaluations*, ICASE report no. 89-45, NASA Langley Research Center, 1989.

[7] R. G. Carter, *On the global convergence of trust region algorithms using inexact gradient information*, SIAM Journal on Numerical Analysis, 28 (1991), pp. 251–265.

[8] R. G. Carter, *Numerical experience with a class of algorithms for nonlinear optimization using inexact function and gradient information*, SIAM Journal on Scientific Computing, 14 (1993), pp. 368–388.

[9] A. Conn, N. Gould, and P. Toint, *Trust-Region Methods*, MPS/SIAM Series on Optimization, SIAM, 2000.

[10] D. Ryckelynck, *A priori hyperreduction method: an adaptive approach*, Journal of Computational Physics, 202 (2005), pp. 346–366.

[11] A. I. Forrester and A. J. Keane, *Recent advances in surrogate-based optimization*, Progress in Aerospace Sciences, 45 (2009), pp. 50–79.

[12] M. A. Grepl and A. T. Patera, A posteriori *error bounds for reduced-basis approximation of parametrized parabolic partial differential equations*, ESAIM: Mathematical Modelling and Numerical Analysis, 39 (2005), pp. 157–181.

[13] M. Heinkenschloss and L. N. Vicente, *Analysis of inexact trust-region SQP algorithms*, SIAM Journal on Optimization, 12 (2002), pp. 283–302.

[14] D. B. P. Huynh, J. Peraire, A. T. Patera, and G. R. Liu, *Real-time reliable prediction of linear-elastic mode-i stress intensity factors for failure analysis*, Proceedings of the 6th Singapore-MIT Alliance Annual Symposium, (2006).

[15] D. B. P. Huynh, G. Rozza, S. Sen, 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 (2007), pp. 473–478.

[16] D. P. Kouri, M. Heinkenschloss, D. Ridzal, and B. G. V. B. Waanders, *Inexact objective function evaluations in a trust region algorithm for PDE-constrained optimization under uncertainty*, SIAM J. Sci. Comput., 36 (2014), pp. A3011–A3029.

[17] K. Kunisch and S. Volkwein, *Proper orthogonal decomposition for optimality systems*, ESAIM: Math. Model. Num., 42 (2008), pp. 1–23.

[18] S. Li, L. Feng, P. Benner, and A. Seidel-Morgenstern, *Using surrogate models for efficient optimization of simulated moving bed chromatography*, Computers & Chemical Engineering, 67 (2014), pp. 121–132.

[19] L. Machiels, Y. Maday, A. T. Patera, and D. V. Rovas, *A blackbox reduced-basis output bound method for shape optimization*, in 12th International Conference on Domain Decomposition Methods, T. Chan, T. Kako, H. Kawarada, and O. Pironneau, eds., 2001.

[20] A. Paul-Dubois-Taine and D. Amsallem, *An adaptive and efficient greedy procedure for the optimal training of parametric reduced-order models*, International Journal for Numerical Methods in Engineering, 102 (2015), pp. 1262–1292.

[21] B. Peherstorfer and K. Willcox, *Online adaptive model reduction for nonlinear systems via low-rank updates*, SIAM J. Sci. Comput., 37 (2015), pp. A2123–A2150.

[22] L. Peng and K. Mohseni, *An online manifold learning approach for model reduction of dynamical systems*, SIAM J. Numer. Anal., 52 (2014), pp. 1928–1952.

[23] N. A. Pierce and M. B. Giles, *Adjoint recovery of superconvergent functionals from PDE approximations*, SIAM Rev., 42 (2000), pp. 247–264.

[24] 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 (2008), pp. 229–275.
[25] P. L. Toint, *Global convergence of a class of trust-region methods for nonconvex minimization in Hilbert space*, IMA Journal of Numerical Analysis, 8 (1988), pp. 231–252.
[26] F. Tröltzsch, *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*, vol. 112 of Graduate Studies in Mathematics, American Mathematical Society, 2010.
[27] 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*, Internat. J. Numer. Methods Fluids, 47 (2005), pp. 773–788.
[28] 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*, in Proceedings of the 16th AIAA Computational Fluid Dynamics Conference, 2003. AIAA Paper 2003-3847.
[29] R. A. Waltz, J. L. Morales, J. Nocedal, and D. Orban, *An interior algorithm for nonlinear optimization that combines line search and trust region steps*, Mathematical Programming, 107 (2006), pp. 391–408.
[30] Y. Yue and K. Meerbergen, *Accelerating optimization of parametric linear systems by model order reduction*, SIAM J. Optim., 23 (2013), pp. 1344–1370.
[31] M. J. Zahr and C. Farhat, *Progressive construction of a parameteric reduced-order model for PDE-constrained optimization*, Int. J. Numer. Meth. Engng., 102 (2015), pp. 1111–1135.