Greedy Sampling using Nonlinear Optimization

Karsten Urban, Stefan Volkwein und Oliver Zeeb

Greedy Sampling using Nonlinear Optimization

Karsten Urban and Stefan Volkwein and Oliver Zeeb

Abstract We consider the reduced basis generation in the offline stage. As an alternative for standard Greedy-training methods based upon a-posteriori error estimates on a training subset of the parameter set, we consider a nonlinear optimization combined with a Greedy method. We define an optimization problem for selecting a new parameter value on a given reduced space. This new parameter is then used –in a Greedy fashion– to determine the corresponding snapshot and to update the reduced basis. We show the well-posedness of this nonlinear optimization problem and derive first- and second-order optimality conditions. Numerical comparisons with the standard Greedy-training method are shown.

1 Introduction

Reduced Basis Methods (RBM) are nowadays a well-known tool to solve parametric partial differential equations (PPDEs) in cases, where the PPDE has to be solved for various values of the parameters (the so-called *multi-query* context, e.g. in optimization) or when the solution for different parameter values has to be com-

Karsten Urban

Universität Ulm, Institute for Numerical Mathematics, Helmholtzstraße 20, D-89069 Ulm, Germany, e-mail: Karsten.Urban@uni-ulm.de

Stefan Volkwein

University of Constance, Department of Mathematics and Statistics, Universitätsstraße 10, D-78457 Konstanz, Germany e-mail: Stefan.Volkwein@uni-konstanz.de

Oliver Zeeb

Universität Ulm, Institute for Numerical Mathematics, Helmholtzstraße 22, D-89069 Ulm, Germany, e-mail: Oliver.Zeeb@uni-ulm.de

puted extremely efficient (the *realtime* context), see e.g. [12]. A key ingredient is an *offline-online-decomposition*. In the offline stage, detailed and thus expensive simulations (sometimes called *truth*) are computed for a moderate number of the parameters, $\mu_1, \ldots, \mu_N$. The arising solutions $u(\mu_i)$, $i = 1, \ldots, N$, of the PPDE (sometimes called *snapshots*) are stored and are used to form a low-dimensional linear space spanned by the reduced basis. In the online stage, an approximation $u_N(\mu)$ for a new parameter $\mu \neq \mu_i$ is determined as the Galerkin projection onto the reduced space $V_N = \text{span}\{u(\mu_i) : i = 1, \ldots, N\}$. A whole variety of results for all sorts of problems has been published in the last years so that an even only halfway complete review including a reference list is far beyond the scope of this paper.

The topic of this paper is the generation of the reduced basis in the offline stage, namely the selection of $\mu_1, \ldots, \mu_N$ above. It is nowadays basically standard to use a Greedy method, see e.g. [9]. The starting point is an a-posteriori error estimator $\Delta_N(\mu)$ for the quantity of interest on a current reduced space V_N. Such an estimator can often be constructed in such a way that the evaluation for a given parameter μ is highly efficient (in particular independent of the size of the truth system). A training set Ξ_{train} is defined and the error estimator $\Delta_N(\mu)$ is maximized over Ξ_{train}. The arising maximizer μ_{N+1} is used to compute the next snapshot $u(\mu_{N+1})$ in order to form the reduced space V_{N+1} of the next higher dimension. We refer to this approach as *Greedy-training*.

Even though this approach obviously has the advantage of being efficiently realizable, it may also suffer from the following fact: The training set Ξ_{train} needs to be defined. This may be a delicate task since Ξ_{train} should be small for efficiency reasons and at the same time sufficiently large in order to represent the whole parameter range as good as possible. The performance of the RBM crucially depends on the choice of Ξ_{train}.

This is the starting point of the present paper. Instead of maximizing the error estimator $\Delta_N(\mu)$ over Ξ_{train}, we develop a nonlinear optimization problem w.r.t. μ on V_N based upon the residual of the primal (and possibly the dual) problem. We show the well-posedness of this optimization problem and derive first-order optimality conditions. The optimization problem is solved numerically by a gradient-type method. This method suffers from the fact that we can only determine local but not global solutions. To overcome this problem we combine the optimization strategy with a Greedy training on a coarse training set Ξ_{train}.

Let us refer to the recent work [2, 3, 4, 8], where adaptive strategies are suggested for the Greedy-training to overcome the problem with high-dimensional parameter spaces. In the context of the method of proper orthogonal decomposition (POD) nonlinear optimization is utilized in [7] to determine optimal snapshot locations in order to control the number of snapshots and minimize the error in the POD reduced-order model.

The remainder of the paper is organized as follows. In Section 2, we review the basic ingredients of the RBM and develop the nonlinear optimization problem (which, in fact, is a minimization problem). We also prove the existence of a solution (Theorem 2.1). Section 3 is devoted to the derivation of first order optimality conditions (Theorem 3.1) while second-order conditions are discussed in Section 4.

Finally, in Section 5 we report on numerical experiments in which we compare the optimization method with the known Greedy-training approach.

2 Problem formulation

In this section we introduce our minimization problem and discuss the existence of optimal solutions.

2.1 The exact variational problem

Let $\mathscr{D} \subset \mathbb{R}^P$ be a given nonempty, closed, bounded and convex parameter domain and V a separable Hilbert space. For given $\ell \in V'$ (V' denotes the space of all bounded and linear functionals defined on V with norm $\| \cdot \|_{V'}$ and scalar product $(\cdot, \cdot)_{V'}$), the goal is to find the scalar output

$$s(\mu) := \langle \ell, u(\mu) \rangle_{V',V}, \quad \mu \in \mathscr{D}, \tag{1a}$$

where $u(\mu) \in V$ satisfies the variational problem ($f \in V'$ given)

$$a(u(\mu), \varphi; \mu) = \langle f, \varphi \rangle_{V',V} \quad \text{for all } \varphi \in V. \tag{1b}$$

In (1a), we denote by $\langle \cdot, \cdot \rangle_{V',V}$ the dual pairing of the spaces V' and V. Furthermore, in (1b) the parameter-dependent, bilinear form $a(\cdot, \cdot; \mu) : V \times V \to \mathbb{R}$ is assumed to have the affine form

$$a(\varphi, \psi; \mu) = \sum_{q=1}^{Q} \vartheta^q(\mu) \, a^q(\varphi, \psi) \quad \text{for } \varphi, \psi \in V \text{ and } \mu \in \mathscr{D}$$

with (twice) continuously differentiable coefficient functions $\vartheta^q : \mathscr{D} \to \mathbb{R}$ and with parameter-independent bounded bilinear forms $a^q : V \times V \to \mathbb{R}$, $1 \le q \le Q$. Moreover, that the parameter-dependent bilinear form a is uniformly bounded and coercive, i.e., there exist constants $\alpha_0 > 0$ and $\gamma > 0$ such that

$$\alpha(\mu) := \inf_{\varphi \in V \setminus \{0\}} \frac{a(\varphi, \varphi; \mu)}{\|\varphi\|_V^2} \ge \alpha_0 > 0 \qquad \text{for all } \mu \in \mathscr{D}, \tag{2a}$$

$$\left| a(\varphi, \phi; \mu) \right| \le \gamma \|\varphi\|_V \|\phi\|_V \qquad\qquad \text{for all } \varphi, \phi \in V \text{ and } \mu \in \mathscr{D}. \tag{2b}$$

Since the bilinear forms a^q are bounded we assume that

$$\left| a^q(\varphi, \phi) \right| \le \gamma \|\varphi\|_V \|\phi\|_V \quad \text{for all } \varphi, \phi \in V \text{ and for } 1 \le q \le Q. \tag{3}$$

Notice that (2a) implies

$$a(\varphi,\varphi;\mu) \geq \alpha_0 \|\varphi\|_V^2 \quad \text{for all } \varphi \in V \text{ and for all } \mu \in \mathscr{D}. \tag{4}$$

Let us mention that we suppose that both f and ℓ do not depend on μ in the affine form only for simplifying the presentation. From (2a) it follows by standard arguments that (1b) has a unique solution $u(\mu) \in V$ for any $\mu \in \mathscr{D}$.

Due to (1a) we require the following dual problem: for given $\mu \in \mathscr{D}$ find $p(\mu) \in V$ solving

$$a(\varphi, z(\mu); \mu) = -\langle \ell, \varphi \rangle_{V',V} \quad \text{for all } \varphi \in V. \tag{5}$$

Since the bilinear form $a(\cdot, \cdot; \mu)$ is bounded and uniformly coercive, the dual problem (5) possesses a unique solution $z(\mu) \in V$ for any $\mu \in \mathscr{D}$.

2.2 The truth approximation

Next we introduce a so-called truth approximation for (1). For that purpose let $V^{\mathscr{N}} = \text{span}\{\varphi_1, \ldots, \varphi_{\mathscr{N}}\} \subset V$ be a finite dimensional subspace with linearly independent functions φ_i. The subspace $V^{\mathscr{N}}$ is endowed with the topology of V. We think of $\mathscr{N} \gg 1$ being 'large'. Then, for any $\mu \in \mathscr{D}$ we consider the 'truth' output

$$s^{\mathscr{N}}(\mu) := \langle \ell, u^{\mathscr{N}}(\mu) \rangle_{V',V}, \tag{6a}$$

where $u^{\mathscr{N}}(\mu) \in V^{\mathscr{N}}$ satisfies the variational equation

$$a(u^{\mathscr{N}}(\mu), \varphi_i; \mu) = \langle f, \varphi_i \rangle_{V',V} \quad \text{for } 1 \leq i \leq \mathscr{N}. \tag{6b}$$

We define the discrete coercivity constant

$$\alpha^{\mathscr{N}}(\mu) := \inf_{\varphi^{\mathscr{N}} \in V^{\mathscr{N}} \setminus \{0\}} \frac{a(\varphi^{\mathscr{N}}, \varphi^{\mathscr{N}}; \mu)}{\|\varphi^{\mathscr{N}}\|_V^2}, \quad \mu \in \mathscr{D}.$$

Unsing $V^{\mathscr{N}} \subset V$ and (2a) we find

$$\alpha^{\mathscr{N}}(\mu) \geq \inf_{\varphi \in V \setminus \{0\}} \frac{a(\varphi, \varphi; \mu)}{\|\varphi\|_V^2} \geq \alpha_0 \quad \text{for all } \mu \in \mathscr{D}.$$

Thus, (6b) has a unique solution $u^{\mathscr{N}}(\mu) \in V^{\mathscr{N}}$ for every $\mu \in \mathscr{D}$.

2.3 The reduced-order modelling

Let us introduce a reduced-order scheme for (6). For chosen linearly independent elements $\{\psi_i\}_{i=1}^{N^{\text{pr}}}$ in $V^{\mathscr{N}}$ we define $V_{N^{\text{pr}}} := \text{span}\{\psi_1, \ldots, \psi_{N^{\text{pr}}}\}$. Analogously, for linearly independent $\{\phi_i\}_{i=1}^{N^{\text{du}}}$ in $V^{\mathscr{N}}$ we set $\tilde{V}_{N^{\text{du}}} := \text{span}\{\phi_1, \ldots, \phi_{N^{\text{du}}}\}$. We have

that $\max(N^{\mathrm{pr}}, N^{\mathrm{du}}) \leq \mathscr{N}$. In the context of reduced-order modeling, $\max(N^{\mathrm{pr}}, N^{\mathrm{du}})$ is much smaller than $\mathscr{N}$.

For any $\mu \in \mathscr{D}$ we consider the scalar output

$$\langle \ell, u_N(\mu) \rangle_{V',V}, \tag{7a}$$

where $u_N(\mu) \in V_{N^{\mathrm{pr}}}$ satisfies the variational equation

$$a(u_N(\mu), \psi_i; \mu) = \langle f, \psi_i \rangle_{V',V} \quad \text{for } 1 \leq i \leq N^{\mathrm{pr}}. \tag{7b}$$

For notational convenience, we just write u_N instead of $u_{N^{\mathrm{pr}}}$ (also for other quantities) since there should be no misunderstanding. We collect some more or less known facts for later reference.

Lemma 2.1. *Suppose that the bilinear form $a(\cdot, \cdot; \mu)$ satisfies (2). Further, $f \in V'$ holds. Then, there exists a unique solution $u_N(\mu) \in V_{N^{\mathrm{pr}}}$ to (7b) for every $\mu \in \mathscr{D}$ with*

$$\|u_N(\mu)\|_V \leq \frac{\|f\|_{V'}}{\alpha_0} \quad \text{for all } \mu \in \mathscr{D}. \tag{8}$$

Proof. By assumption, the bilinear form $a(\cdot, \cdot; \mu)$ is bounded for every $\mu \in \mathscr{D}$. Since $V_{N^{\mathrm{pr}}} \subset V$, the form $a(\cdot, \cdot; \mu)$ is also uniformly coercive on $V_{N^{\mathrm{pr}}}$. Thus, it follows from the Lax-Milgram theorem that (7b) possesses a unique solution $u_N \in V_{N^{\mathrm{pr}}}$ for every $\mu \in \mathscr{D}$. Utilizing (4) and (7b) and the uniform coercivity, we obtain

$$\|u_N(\mu)\|_V^2 \leq \frac{a(u_N(\mu), u_N(\mu); \mu)}{\alpha_0} = \frac{\langle f, u_N(\mu) \rangle_{V',V}}{\alpha_0} \leq \frac{\|f\|_{V'}}{\alpha_0} \|u_N(\mu)\|_V,$$

which gives (8). ∎

Remark 2.1. 1) Due to Lemma 2.1 we can define the primal (non-linear) solution operator $\mathscr{S}_N^{\mathrm{pr}} : \mathscr{D} \to V_{N^{\mathrm{pr}}}$, where $u_N(\mu) = \mathscr{S}_N^{\mathrm{pr}}(\mu)$ denotes the unique solution to (7b).

2) Let us consider a specific case. Suppose that the bilinear form is given by $a(\cdot, \cdot; \mu) = \vartheta^1(\mu) a^1(\cdot, \cdot)$ (i.e., $Q = 1$) and $\vartheta^1(\mu) \neq 0$ holds for all $\mu \in \mathscr{D}$. Let $u_N^1 = u_N(\mu_1)$ be a solution to (7b) for given $\mu_1 \in \mathscr{D}$. Then, the function $u_N^2 = \vartheta^1(\mu_1) u_N^1 / \vartheta^1(\mu_2) \in V^N$ solves (7b) for $\mu_2 \in \mathscr{D}$. In fact, we have

$$a(u_N^2, \psi_i; \mu_2) = \vartheta^1(\mu_2) a^1(u_N^2, \psi_i) = \vartheta^1(\mu_1) a^1(u_N^1, \psi_i) = a(u_N^1, \varphi_i; \mu_1)$$
$$= \langle f, \psi_i \rangle_{V',V} \quad \text{for } 1 \leq i \leq N.$$

Consequently, solutions to different parameter values are linearly dependent. ◇

For given $\mu \in \mathscr{D}$ the associated dual variable $z_N(\mu)$ solves the dual problem [1], namely

$$a(\phi_i, z_N(\mu); \mu) = -\langle \ell, \phi_i \rangle_{V',V}, \quad 1 \leq i \leq N^{\mathrm{du}}. \tag{9}$$

Remark 2.2. 1) If the bilinear form satisfies (2) and $\ell \in V'$ holds, it follows by similar arguments as in the proof of Lemma 2.1 that (9) admits a unique solution $z_N(\mu) \in \tilde{V}_{N^{\mathrm{du}}}$ satisfying

$$\|z_N(\mu)\|_V \leq \frac{\|\ell\|_{V'}}{\alpha_0} \quad \text{for all } \mu \in \mathscr{D}. \tag{10}$$

2) We define the (non-linear) solution operator $\mathscr{S}_N^{\mathrm{du}} : \mathscr{D} \to \tilde{V}_{N^{\mathrm{du}}}$, where $z_N = \mathscr{S}_N^{\mathrm{du}}(\mu)$ is the unique solution to (9). $\qquad\qquad\qquad\qquad\qquad\qquad\diamond$

Next we define the residuals $r_N^{\mathrm{pr}}(\cdot;\mu), r_N^{\mathrm{du}}(\cdot;\mu) \in (V^{\mathcal{N}})'$ by

$$r_N^{\mathrm{pr}}(\varphi^{\mathcal{N}};\mu) := \langle f, \varphi^{\mathcal{N}} \rangle_{V',V} - a(u_N(\mu), \varphi^{\mathcal{N}};\mu) \quad \text{for } \varphi \in V^{\mathcal{N}} \text{ and } \mu \in \mathscr{D},$$

$$r_N^{\mathrm{du}}(\varphi^{\mathcal{N}};\mu) := \langle \ell, \varphi^{\mathcal{N}} \rangle_{V',V} + a(\varphi^{\mathcal{N}}, z_N(\mu);\mu) \quad \text{for } \varphi \in V^{\mathcal{N}} \text{ and } \mu \in \mathscr{D}.$$

It has turned out that the primal-dual output defined as

$$s_N(\mu) := \langle \ell, u_N(\mu) \rangle_{V',V} - r_N^{\mathrm{pr}}(z_N(\mu);\mu),$$

gives rise to favorable output error estimates which take the form (see [12], for instance)

$$\left| s^{\mathcal{N}}(\mu) - s_N(\mu) \right| \leq \Delta_N^s(\mu) = \frac{\|r_N^{\mathrm{pr}}(\cdot;\mu)\|_{(V^{\mathcal{N}})'}}{\alpha_0^{1/2}} \frac{\|r_N^{\mathrm{du}}(\cdot;\mu)\|_{(V^{\mathcal{N}})'}}{\alpha_0^{1/2}}. \tag{11}$$

Remark 2.3. 1) From

$$u_N(\mu) = \sum_{j=1}^{N^{\mathrm{pr}}} \mathrm{u}_{N,j}(\mu)\,\psi_j \quad \text{and} \quad z_N(\mu) = \sum_{j=1}^{N^{\mathrm{du}}} \mathrm{z}_{N,j}(\mu)\,\phi_j$$

we infer that

$$r_N^{\mathrm{pr}}(\varphi_i;\mu) = \langle f, \varphi_i \rangle_{V',V} - \sum_{j=1}^{N^{\mathrm{pr}}} \mathrm{u}_{N,j}(\mu)\,a(\psi_j, \varphi_i;\mu)$$

$$= \langle f, \varphi_i \rangle_{V',V} - \sum_{j=1}^{N^{\mathrm{pr}}} \mathrm{u}_{N,j}(\mu) \sum_{q=1}^{Q} \vartheta^q(\mu)\,a^q(\psi_j, \varphi_i),$$

$$r_N^{\mathrm{du}}(\varphi_i;\mu) = \langle \ell, \varphi_i \rangle_{V',V} + a(\varphi_i, z_N(\mu);\mu)$$

$$= \langle \ell, \varphi_i \rangle_{V',V} + \sum_{j=1}^{N^{\mathrm{du}}} \mathrm{z}_{N,j}(\mu) \sum_{q=1}^{Q} \vartheta^q(\mu)\,a^q(\varphi_i, \phi_j)$$

for $1 \leq i \leq \mathcal{N}$. These representations of the residuals are utilized to realize an efficient offline-online decomposition for the reduced-order approach, see e.g. [9, 12].

2) Suppose that the bilinear form is given by $a(\cdot,\cdot;\mu) = \vartheta^1(\mu)a^1(\cdot,\cdot)$ (i.e., $Q = 1$) and $\vartheta^1(\mu) \neq 0$ holds for all $\mu \in \mathscr{D}$. Then, solutions to different parameter values are linearly dependent; see Remark 2.1-2). Let $\mu_1, \mu_2 \in \mathscr{D}$ be chosen arbitrarily. By u_N^i, $i = 1, 2$, we denote the solutions to (7b) for parameter $\mu = \mu_i$. From $u_N^2 = \vartheta^1(\mu_1)u_N^1/\vartheta^1(\mu_2)$ we infer that

$$V' \ni a(u_N^2,\cdot;\mu_2) - f = \frac{\vartheta^1(\mu_1)}{\vartheta^1(\mu_2)}a(u_N^1,\cdot;\mu_2) - f = a(u_N^1,\cdot;\mu_1) - f.$$

Hence, the norm $\|a(u_N(\mu),\cdot;\mu) - f\|_{(V^{\mathscr{N}})'}$ is constant for all $\mu \in \mathscr{D}$, where $u_N(\mu)$ denotes the solution to (7b) for the parameter μ. Analogously, we can prove that the norm $\|a(\cdot,z_N(\mu);\mu) + \ell\|_{(V^{\mathscr{N}})'}$ is constant for all $\mu \in \mathscr{D}$, where $z_N(\mu)$ denotes the solution to (9) for the parameter μ. $\diamond$

2.4 The minimization problem

Let $N := (N^{\mathrm{pr}}, N^{\mathrm{du}})$, $Y_N := V_{N^{\mathrm{pr}}} \times \tilde{V}_{N^{\mathrm{du}}}$, $X_N = Y_N \times \mathbb{R}^P$ and $X_N^{ad} = Y_N \times \mathscr{D}$. We endow X_N with the natural product topology. In the Greedy algorithm a new reduced-basis solution $u_N(\bar{\mu})$ associated with a certain parameter value $\bar{\mu}$ is added to the already computed set of ansatz functions provided an a-posteriori error measure $\Delta_N^s(\bar{\mu})$ in (11) is maximal. The idea here is to avoid the Greedy method and to determine $\bar{\mu}$ as the solution of a minimization problem. Thus, we introduce the cost functional $J : X_N \to \mathbb{R}$ for $x_N = (u_N, z_N, \mu) \in X_N$ by

$$J(x_N) = -\frac{1}{2}\left(\|f - a(u_N,\cdot;\mu)\|_{(V^{\mathscr{N}})'}^2 + \|\ell + a(\cdot,z_N;\mu)\|_{(V^{\mathscr{N}})'}^2\right).$$

If $J(x_N(\mu)) \geq -\varepsilon\alpha_0$ holds true for $x_N(\mu) := (u_N(\mu), z_N(\mu), \mu)$, we infer by using Young's inequality that

$$\left|s^{\mathscr{N}}(\mu) - s_N(\mu)\right|^2 \leq \frac{\|r_N^{\mathrm{pr}}(\cdot;\mu)\|_{(V^{\mathscr{N}})'}^2 + \|r_N^{\mathrm{du}}(\cdot;\mu)\|_{(V^{\mathscr{N}})'}^2}{2\alpha_0} = -\frac{J(x_N(\mu))}{\alpha_0} \leq \varepsilon.$$

Now we consider the following optimization problem:

$$\min_{x_N \in X_N^{ad}} J(x_N) \quad \text{subject to (s.t.)} \quad x_N = (y_N, \mu), \ y_N = \mathscr{S}_N(\mu), \qquad \textbf{(P)}$$

where we have set $\mathscr{S}_N = (\mathscr{S}_N^{\mathrm{pr}}, \mathscr{S}_N^{\mathrm{du}}) : \mathscr{D} \to Y_N$, i.e., $y_N = \mathscr{S}_N(\mu)$ means that $y_N = (u_N(\mu), z_N(\mu))$. Introducing the reduced cost functional

$$\hat{J}(\mu) := J(\mathscr{S}_N(\mu), \mu) \quad \text{for } \mu \in \mathscr{D},$$

we can express (**P**) equivalently in the reduced form

$$\min_{\mu \in \mathscr{D}} \hat{J}(\mu). \tag{$\hat{\mathbf{P}}$}$$

If ($\hat{\mathbf{P}}$) has a local solution $\bar{\mu} \in \mathscr{D}$, then $\bar{x}_N := (\bar{y}_N, \bar{\mu})$ is a local solution to ($\mathbf{P}$), where we set $\bar{y}_N = (\bar{u}_N, \bar{p}_N) := \mathscr{S}_N(\bar{\mu})$. We now give a general existence result.

Theorem 2.1. *Suppose that the bilinear form $a(\cdot, \cdot; \mu)$ satisfies (2). Further, f and ℓ belong to V'. Then, there exists at least one optimal solution $\bar{x}_N = (\bar{y}_N, \bar{\mu})$, $\bar{y}_N = (\bar{u}_N, \bar{z}_N) \in Y_N$, to ($\mathbf{P}$).*

Proof. Since $\mathscr{D}$ is assumed to be nonempty and $\mathscr{S}_N : \mathscr{D} \to Y_N$ is well-defined, the set of admissible solutions

$$\mathcal{F}(\mathbf{P}) = \left\{ x_N = (y_N, \mu) \in X_N^{ad} \, \middle| \, y_N = \mathscr{S}_N(\mu) \right\}$$

is nonempty. Let $\{x_N^{(n)}\}_{n \in \mathbb{N}} \subset \mathcal{F}(\mathbf{P})$, $x_N^{(n)} = (y_N^{(n)}, \mu^{(n)})$ and $y_N^{(n)} = (u_N^{(n)}, z_N^{(n)})$, be a minimizing sequence for J:

$$\inf_{x_N \in \mathcal{F}(\mathbf{P})} J(x_N) = \lim_{n \to \infty} J(x_N^{(n)}).$$

Since $\mathscr{D}$ is bounded and the a-priori bounds (8), (10) hold, $\inf_{x_N \in \mathcal{F}(\mathbf{P})} J(x_N)$ is bounded from below. Moreover, from $\mu^{(n)} \in \mathscr{D} \subset \mathbb{R}^P$ for every n we infer that there exists a subsequence $\{\mu^{(n_k)}\}_{k \in \mathbb{N}}$ in $\mathscr{D}$ and an element $\bar{\mu} \in \mathscr{D}$ so that

$$\lim_{k \to \infty} \mu^{(n_k)} = \bar{\mu} \quad \text{in } \mathbb{R}^P.$$

It follows from the a-priori estimates (8) and (10) that the sequence $\{(u_N^{(n)}, z_N^{(n)})\}_{n \in \mathbb{N}}$ is bounded in Y_N. Consequently, there exist a subsequence $\{y_N^{(n_k)}\}_{k \in \mathbb{N}}$ and a pair $\bar{y}_N = (\bar{u}_N, \bar{p}_N) \in Y_N$ such that

$$u_N^{(n_k)} \rightharpoonup \bar{u}_N \text{ for } k \to \infty \text{ in } V_{N^{\mathrm{pr}}} \quad \text{and} \quad z_N^{(n_k)} \rightharpoonup \bar{z}_N \text{ for } k \to \infty \text{ in } \tilde{V}_{N^{\mathrm{du}}}. \tag{12}$$

Next we prove that $\bar{y}_N = \mathscr{S}_N(\bar{\mu})$ holds. For $1 \le i \le N^{\mathrm{pr}}$ we have

$$\begin{aligned}
\langle f, \psi_i \rangle_{V', V} - a(\bar{u}_N, \psi_i; \bar{\mu}) &= a(u_N^{(n_k)}, \psi_i; \mu^{(n_k)}) - a(\bar{u}_N, \psi_i; \bar{\mu}) = \\
&= a(u_N^{(n_k)}, \psi_i; \mu^{(n_k)}) - a(u_N^{(n_k)}, \psi_i; \bar{\mu}) + a(u_N^{(n_k)} - \bar{u}_N, \psi_i; \bar{\mu}) \\
&= \sum_{q=1}^{Q} \left(\left(\vartheta^q(\mu^{(n_k)}) - \vartheta^q(\bar{\mu}) \right) a^q(u_N^{(n_k)}, \psi_i) \right) + a(u_N^{(n_k)} - \bar{u}_N, \psi_i; \bar{\mu}).
\end{aligned}$$

Let us define the functionals $F_i \in V' \subset V_N'$ by $\langle F_i, \varphi \rangle_{V', V} := a(\varphi, \psi_i; \bar{\mu})$ for $\varphi \in V$ and $1 \le i \le N^{\mathrm{pr}}$. From (12) we infer that

$$a(u_N^{(n_k)} - \bar{u}_N, \psi_i; \bar{\mu}) = F_i(u_N^{(n_k)} - \bar{u}_N) \to 0 \quad \text{for } k \to \infty \text{ and } 1 \le i \le N^{\mathrm{pr}}.$$

Moreover, $\|u_N^{(n_k)}\|_V$ is uniformly bounded and the ϑ^q's are continuous. Thus,

$$\sum_{q=1}^{Q} \left(\left(\vartheta^q(\mu^{(n_k)}) - \vartheta^q(\bar{\mu}) \right) a^q(u_N^{(n_k)}, \psi_i) \right) \to 0 \quad \text{for } k \to \infty \text{ and } 1 \leq i \leq Q.$$

Consequently, $\bar{u}_N = \mathscr{S}_N^{\mathrm{pr}}(\bar{\mu})$ holds. Analogously, we find that $\bar{z}_N = \mathscr{S}^{\mathrm{du}}(\bar{\mu})$ holds true. Thus, $\bar{x}_N = (\bar{y}_N, \bar{\mu}) \in \mathcal{F}(\mathbf{P})$ is satisfied. Next, we show that $\bar{x}_N$ is a minimizer for J. Note that with the above arguments

$$\|a(u_N^{(n_k)}, \cdot; \bar{\mu}) - a(u_N^{(n_k)}, \cdot; \mu^{(n_k)})\|_{(V^{\mathcal{N}})'}$$
$$\leq \sum_{q=1}^{Q} \left| \vartheta^q(\bar{\mu}) - \vartheta^q(\mu^{(n_k)}) \right| \|a^q(u_N^{(n_k)}, \cdot)\|_{(V^{\mathcal{N}})'} \xrightarrow{k \to \infty} 0.$$

This and (12) imply

$$\lim_{k \to \infty} \|f - a(u_N^{(n_k)}, \cdot; \mu^{(n_k)})\|_{(V^{\mathcal{N}})'} =$$
$$= \lim_{k \to \infty} \|f - a(u_N^{(n_k)}, \cdot; \bar{\mu})\|_{(V^{\mathcal{N}})'} + \lim_{k \to \infty} \|a(u_N^{(n_k)}, \cdot; \bar{\mu}) - a(u_N^{(n_k)}, \cdot; \mu^{(n_k)})\|_{(V^{\mathcal{N}})'}$$
$$= \|f - a(\bar{u}_N, \cdot; \bar{\mu})\|_{(V^{\mathcal{N}})'}.$$

Analogously, $\lim_{k \to \infty} \|\ell + a(\cdot, z_N^{(n_k)}; \mu^{(n_k)})\|_{(V^{\mathcal{N}})'} = \|\ell + a(\cdot, \bar{z}_N; \bar{\mu})\|_{(V^{\mathcal{N}})'}$ and therefore

$$\inf_{x_N \in \mathcal{F}(\mathbf{P})} J(x_N) = \lim_{k \to \infty} J(x_N^{(n_k)}) = J(\bar{x}_N),$$

i.e., $\bar{x}_N$ is a solution to $(\mathbf{P})$. $\diamond$

Before we continue, let us collect some notation that will be needed in the sequel. Let $\bar{x}_N = (\bar{y}_N, \bar{\mu})$, $\bar{y}_N = (\bar{u}_N, \bar{z}_N)$, be an optimal solution to $(\mathbf{P})$ according to Theorem 2.1. Then, define corresponding (optimal) primal and dual residuals as

$$\bar{r}_N^{\mathrm{pr}}(\varphi^{\mathcal{N}}) := \langle f, \varphi^{\mathcal{N}} \rangle_{V', V} - a(\bar{u}_N, \varphi^{\mathcal{N}}; \bar{\mu}) \qquad \text{for } \varphi^{\mathcal{N}} \in V^{\mathcal{N}},$$
$$\bar{r}_N^{\mathrm{du}}(\varphi^{\mathcal{N}}) := \langle \ell, \varphi^{\mathcal{N}} \rangle_{V', V} + a(\varphi^{\mathcal{N}}, \bar{z}_N; \bar{\mu}) \qquad \text{for } \varphi^{\mathcal{N}} \in V^{\mathcal{N}}.$$

We define the corresponding Riesz representations $\bar{\rho}_N^{\mathrm{pr}}, \bar{\rho}_N^{\mathrm{du}} \in V^{\mathcal{N}}$, i.e.,

$$(\bar{\rho}_N^{\mathrm{pr}}, \varphi^{\mathcal{N}})_V = \bar{r}_N^{\mathrm{pr}}(\varphi^{\mathcal{N}}) = \langle f, \varphi^{\mathcal{N}} \rangle_{V', V} - a(\bar{u}_N, \varphi^{\mathcal{N}}; \bar{\mu}) \quad \text{for all } \varphi^{\mathcal{N}} \in V^{\mathcal{N}},$$
$$(\bar{\rho}_N^{\mathrm{du}}, \varphi^{\mathcal{N}})_V = \bar{r}_N^{\mathrm{du}}(\varphi^{\mathcal{N}}) = \langle \ell, \varphi^{\mathcal{N}} \rangle_{V', V} + a(\varphi^{\mathcal{N}}, \bar{z}_N; \bar{\mu}) \quad \text{for all } \varphi^{\mathcal{N}} \in V^{\mathcal{N}}.$$

This in particular implies that

$$(g, \bar{r}_N^{\mathrm{pr}})_{(V^{\mathcal{N}})'} = \langle g, \bar{\rho}_N^{\mathrm{pr}} \rangle_{(V^{\mathcal{N}})', V^{\mathcal{N}}} \quad \text{for all } g \in (V^{\mathcal{N}})',$$

which will be used later. It is noticable to mention that we have in general $\bar{\rho}_N^{\mathrm{pr}} \notin V_{N^{\mathrm{pr}}}$ and $\bar{\rho}_N^{\mathrm{du}} \notin \tilde{V}_{N^{\mathrm{du}}}$.

3 First-order necessary optimality conditions

First we write the equality constraints in (**P**) in a compact from. For that purpose we introduce the nonlinear mapping $e = (e_1, e_2) : X_N \to Y_N'$ by

$$\langle e(x_N), \lambda_N \rangle_{Y_N', Y_N} = \langle e_1(x_N), \lambda_N^1 \rangle_{V_{N^{\mathrm{pr}}}', V_{N^{\mathrm{pr}}}} + \langle e_2(x_N), \lambda_N^2 \rangle_{\tilde{V}_{N^{\mathrm{du}}}', \tilde{V}_{N^{\mathrm{du}}}}$$

for $x_N = (u_N, z_N, \mu) \in X_N^{ad}$ and $\lambda_N = (\lambda_N^1, \lambda_N^2) \in Y_N$. Here, we identify the dual Y_N' with $V_{N^{\mathrm{pr}}}' \times \tilde{V}_{N^{\mathrm{du}}}'$ and we put

$$\langle e_1(x_N), \lambda_N^1 \rangle_{V_{N^{\mathrm{pr}}}', V_{N^{\mathrm{pr}}}} = \langle f, \lambda_N^1 \rangle_{V_{N^{\mathrm{pr}}}', V_{N^{\mathrm{pr}}}} - a(u_N, \lambda_N^1; \mu),$$

$$\langle e_2(x_N), \lambda_N^2 \rangle_{\tilde{V}_{N^{\mathrm{du}}}', \tilde{V}_{N^{\mathrm{du}}}} = \langle \ell, \lambda_N^2 \rangle_{\tilde{V}_{N^{\mathrm{du}}}', \tilde{V}_{N^{\mathrm{du}}}} + a(\lambda_N^2, z_N; \mu).$$

Using (2b) we infer that

$$\begin{aligned}
\|e(x_N)\|_{Y_N'} &= \sup_{\|\lambda_N\|_{Y_N} = 1} \langle e(x_N), \lambda_N \rangle_{Y_N', Y_N} \\
&= \sup_{\|\lambda_N^1\|_V = 1} \langle e_1(x_N), \lambda_N^1 \rangle_{V_{N^{\mathrm{pr}}}', V_{N^{\mathrm{pr}}}} + \sup_{\|\lambda_N^2\|_V = 1} \langle e_2(x_N), \lambda_N^2 \rangle_{\tilde{V}_{N^{\mathrm{du}}}', \tilde{V}_{N^{\mathrm{du}}}} \\
&\le C_e \left(1 + \|u_N\|_V + \|z_N\|_V \right)
\end{aligned}$$

with $C_e = \max(\|f\|_{V'} + \|\ell\|_{V'}, \gamma)$.

To derive first-order optimality conditions for (**P**) we have to ensure that the mapping e is continuously (Fréchet) differentiable and satisfies a standard constraint qualification; see, e.g., [5, 13].

Proposition 3.1. *Suppose that the bilinear form $a(\cdot, \cdot; \mu)$ satisfies (2). Further, $f, \ell \in V'$ holds and the functions ϑ^q are continuously differentiable for $1 \le q \le Q$. Then, the mapping e is continuously (Fréchet) differentiable and its (Fréchet) derivative at $x_N = (y_N, \mu) \in X_N^{ad}$, $y_N = (u_N, z_N)$, is given by*

$$\langle e'(x_N) x_N^\delta, \lambda_N \rangle_{Y_N', Y_N} = \langle e_1'(x_N) x_N^\delta, \lambda_N^1 \rangle_{V_{N^{pr}}', V_{N^{pr}}} + \langle e_2'(x_N) x_N^\delta, \lambda_N^2 \rangle_{\tilde{V}_{N^{du}}', \tilde{V}_{N^{du}}}$$

for any direction $x_N^\delta = (u_N^\delta, z_N^\delta, \mu^\delta) \in X_N$ and for $\lambda_N = (\lambda_N^1, \lambda_N^2) \in Y_N$, where

$$\langle e_1'(x_N) x_N^\delta, \lambda_N^1 \rangle_{\tilde{V}_{N^{pr}}', \tilde{V}_{N^{pr}}} = -a(u_N^\delta, \lambda_N^1; \mu) - \sum_{q=1}^Q a^q(u_N, \lambda_N^1) \nabla \vartheta^q(\mu)^\top \mu^\delta,$$

$$\langle e_2'(x_N) x_N^\delta, \lambda_N^2 \rangle_{V_{N^{du}}', V_{N^{du}}} = a(\lambda_N^2, z_N^\delta; \mu) + \sum_{q=1}^Q a^q(\lambda_N^2, z_N) \nabla \vartheta^q(\mu)^\top \mu^\delta$$

with $\nabla \vartheta^q(\mu) = (\vartheta_{\mu_1}^q(\mu), \dots, \vartheta_{\mu_P}^q(\mu))^\top \in \mathbb{R}^P$ and $\vartheta_{\mu_i}^q = \frac{\partial \vartheta^q}{\partial \mu_i}$. Furthermore, the (Fréchet) derivative $e'(x_N) : X_N \to Y_N'$ is a surjective operator for every $x_N \in X_N^{ad}$.

Proof. It follows by standard arguments that e is (Fréchet) differentiable for every $x_N \in X_N^{ad}$. Therefore, we only prove that the linear operator $e'(x_N)$ is onto. Let $F_N = (F_N^1, F_N^2) \in Y_N'$ be chosen arbitrarily. Then, $e'(x_N)$ is surjective if there exists an element $x_N^\delta = (u_N^\delta, z_N^\delta, \mu^\delta) \in X_N$ satisfying

$$e'(x_N)x_N^\delta = F_N \quad \text{in } Y_N'. \tag{13}$$

Equation (13) is equivalent with

$$e_1'(x_N)x_N^\delta = F_N^1 \text{ in } V_{N^{pr}}' \quad \text{in} \quad e_2'(x_N)x_N^\delta = F_N^2 \quad \text{in } \tilde{V}_{N^{du}}'. \tag{14}$$

Choosing $\mu^\delta = 0$ we obtain from (14) that

$$\begin{aligned}
a(u_N^\delta, \lambda_N^1; \mu) &= -\langle F_N^1, \lambda_N^1 \rangle_{V_{N^{pr}}', V_{N^{pr}}} \quad \text{for all } \lambda_N^1 \in V_{N^{pr}}, \\
a(\lambda_N^2, z_N^\delta; \mu) &= \langle F_N^2, \lambda_N^2 \rangle_{\tilde{V}_{N^{du}}', \tilde{V}_{N^{du}}} \quad \text{for all } \lambda_N^2 \in \tilde{V}_{N^{du}}.
\end{aligned} \tag{15}$$

Since the bilinear form $a(\cdot, \cdot; \mu)$ is bounded and coercive, there exists a unique pair $y_N^\delta = (u_N^\delta, z_N^\delta) \in Y_N$ solving (15). Summarizing, $x_N^\delta = (y_N^\delta, 0)$ solves (13) which implies that $e'(x_N)$ is surjective.

Next let us introduce the Lagrange functional $\mathscr{L} : X_N \times Y_N \to \mathbb{R}$ for $x_N = (x_N^1, x_N^2, \mu) \in X_N$ and $\lambda_N = (\lambda_N^1, \lambda_N^2) \in Y_N$ as

$$\begin{aligned}
\mathscr{L}(x_N, \lambda_N) &= J(x_N) + \langle e(x_N), \lambda_N \rangle_{Y_N', Y_N} \\
&= -\frac{1}{2} \left(\|f - a(u_N, \cdot; \mu)\|_{(V^{\mathcal{N}})'}^2 + \|a(\cdot, z_N; \mu) + \ell\|_{(V^{\mathcal{N}})'}^2 \right) \\
&\quad + \langle (f, \ell), \lambda_N \rangle_{Y_N', Y_N} - a(u_N, \lambda_N^1; \mu) + a(\lambda_N^2, z_N; \mu).
\end{aligned}$$

We infer from Proposition 3.1 that first-order necessary optimality conditions are given as follows [5, 13]: Let $\bar{x}_N = (\bar{y}_N, \bar{\mu}) \in X_N^{ad}$, $\bar{y}_N = (\bar{u}_N, \bar{z}_N) \in Y_N$, be a local solution to (**P**). Then, there exists a Lagrange multiplier $\bar{\lambda}_N = (\bar{\lambda}_N^1, \bar{\lambda}_N^2) \in Y_N$ solving the following system

$$\begin{aligned}
\mathscr{L}_{u_N}(\bar{x}_N, \bar{\lambda}_N)u_N^\delta &= 0 && \text{for all } u_N^\delta \in V_{N^{pr}}, & \text{(16a)} \\
\mathscr{L}_{z_N}(\bar{x}_N, \bar{\lambda}_N)z_N^\delta &= 0 && \text{for all } z_N^\delta \in \tilde{V}_{N^{du}}, & \text{(16b)} \\
\mathscr{L}_{\mu}(\bar{x}_N, \bar{\lambda}_N)(\mu^\delta - \bar{\mu}) &\geq 0 && \text{for all } \mu^\delta \in \mathscr{D}, & \text{(16c)}
\end{aligned}$$

where, for instance, $\mathscr{L}_{u_N}$ denote the (Fréchet) derivative of the Lagrangian with respect to the argument u_N. First we study (16a). For $u_N^\delta \in V_{N^{pr}}$ we find

$$\mathscr{L}_{u_N}(\bar{x}_N, \bar{\lambda}_N)u_N^\delta = (f - a(\bar{u}_N, \cdot; \bar{\mu}), a(u_N^\delta, \cdot; \bar{\mu}))_{(V^{\mathcal{N}})'} - a(u_N^\delta, \bar{\lambda}_N^1; \bar{\mu}).$$

Using the Riesz representation $\bar{\rho}_N^{pr} \in V^{\mathcal{N}}$ of $\bar{r}_N^{pr} \in (V^{\mathcal{N}})'$, we get

$$
\begin{aligned}
\mathscr{L}_{u_N}(\bar{x}_N, \bar{\lambda}_N) u_N^\delta &= (\bar{r}_N^{\mathrm{pr}}, a(u_N^\delta, \cdot; \bar{\mu}))_{(V^{\mathcal{N}})'} - a(u_N^\delta, \bar{\lambda}_N^1; \bar{\mu}) \\
&= a(u_N^\delta, \bar{\rho}_N^{\mathrm{pr}}; \bar{\mu}) - a(u_N^\delta, \bar{\lambda}_N^1; \bar{\mu}) = a(u_N^\delta, \bar{\rho}_N^{\mathrm{pr}} - \bar{\lambda}_N^1; \bar{\mu}).
\end{aligned}
\tag{17}
$$

From (16a) and (17) we infer the first adjoint equation:

$$
a(u_N^\delta, \bar{\lambda}_N^1; \bar{\mu}) = a(u_N^\delta, \bar{\rho}_N^{\mathrm{pr}}; \bar{\mu}) \quad \text{for all } u_N^\delta \in V_{N^{\mathrm{pr}}}.
\tag{18}
$$

Remark 3.4. Since in general $\bar{\rho}_N^{\mathrm{pr}} \notin V_{N^{\mathrm{pr}}}$ holds, we obtain in general $\bar{\lambda}_N^1 \neq \bar{\rho}_N^{\mathrm{pr}}$. Rather, $\bar{\lambda}_N^1$ is the a-orthogonal projection of $\bar{\rho}_N^{\mathrm{pr}} \in V$ onto $\bar{\lambda}_N^1 \in V_{N^{\mathrm{pr}}}$. $\qquad \diamond$

Further, we have

$$
\mathscr{L}_{z_N}(\bar{x}_N, \bar{\lambda}_N) z_N^\delta = -(\ell + a(\cdot, \bar{z}_N; \bar{\mu}), a(\cdot, z_N^\delta; \bar{\mu}))_{(V^{\mathcal{N}})'} + a(\bar{\lambda}_N^2, z_N^\delta; \bar{\mu})
\tag{19}
$$

for any direction $z_N^\delta \in \tilde{V}_{N^{\mathrm{du}}}$. Using the Riesz representation $\bar{\rho}_N^{\mathrm{du}} \in V^{\mathcal{N}}$ of $\bar{r}_N^{\mathrm{du}} \in (V^{\mathcal{N}})'$, combining (16b) and (19) we get

$$
\mathscr{L}_{z_N}(\bar{x}_N, \bar{\lambda}_N) z_N^\delta = a(\bar{\lambda}_N^2 - \bar{\rho}_N^{\mathrm{du}}, z_N^\delta; \bar{\mu}) = 0 \quad \text{for all } z_N^\delta \in V_{N^{\mathrm{du}}}
$$

which gives the second adjoint equation

$$
a(\bar{\lambda}_N^2, z_N^\delta; \bar{\mu}) = a(\bar{\rho}_N^{\mathrm{du}}, z_N^\delta; \bar{\mu}) \quad \text{for all } z_N^\delta \in V_{N^{\mathrm{du}}}.
\tag{20}
$$

Remark 3.5. Analogous to Remark 3.4 we infer that $\bar{\lambda}_N^2$ is the a-orthogonal decomposition of $\bar{\rho}_N^{\mathrm{du}}$ onto $\tilde{V}_{N^{\mathrm{du}}}$. $\qquad \diamond$

Next we consider (16c). Using the Riesz representations $\bar{\rho}_N^{\mathrm{pr}}, \bar{\rho}_N^{\mathrm{du}} \in V^{\mathcal{N}}$ of $\bar{r}_N^{\mathrm{pr}}, \bar{r}_N^{\mathrm{du}} \in (V^{\mathcal{N}})'$, respectively, it follows that

$$
\begin{aligned}
\mathscr{L}_\mu(\bar{x}_N, \bar{\lambda}_N) \mu^\delta &= \sum_{q=1}^Q \nabla \vartheta^q(\bar{\mu})^\top \mu^\delta \, (\bar{r}_N^{\mathrm{pr}}, a^q(\bar{u}_N, \cdot))_{(V^{\mathcal{N}})'} \\
&\quad + \sum_{q=1}^Q \nabla \vartheta^q(\bar{\mu})^\top \mu^\delta \left((-\bar{r}_N^{\mathrm{du}}, a^q(\cdot, \bar{z}_N))_{V'} + a^q(\bar{\lambda}_N^2, \bar{z}_N) - a^q(\bar{u}_N, \bar{\lambda}_N^1) \right) \\
&= \sum_{q=1}^Q \left(a^q(\bar{u}_N, \bar{\rho}_N^{\mathrm{pr}} - \bar{\lambda}_N^1) + a^q(\bar{\lambda}_N^2 - \bar{\rho}_N^{\mathrm{du}}, \bar{z}_N) \right) \nabla \vartheta^q(\bar{\mu})^\top \mu^\delta
\end{aligned}
\tag{21}
$$

for any direction $\mu^\delta \in \mathbb{R}^P$. We define the Jacobi matrix

$$
D\vartheta(\bar{\mu}) = \begin{pmatrix} \nabla \vartheta^1(\bar{\mu})^\top \\ \vdots \\ \nabla \vartheta^Q(\bar{\mu})^\top \end{pmatrix} \in \mathbb{R}^{Q \times P}
$$

with $\nabla \vartheta^q(\mu) = (\vartheta_{\mu_1}^q(\mu), \dots, \vartheta_{\mu_P}^q(\mu))^\top \in \mathbb{R}^P$ and $\vartheta_{\mu_i}^q = \frac{\partial \vartheta^q}{\partial \mu_i}$. Further, we set $\bar{\xi} = \bar{\xi}(\bar{x}_N, \bar{\lambda}_N) = (\bar{\xi}_1, \dots, \bar{\xi}_Q)^\top \in \mathbb{R}^Q$ with

$$\bar{\xi}_q = a^q(\bar{u}_N, \bar{\rho}_N^{\text{pr}} - \bar{\lambda}_N^1) + a^q(\bar{\lambda}_N^2 - \bar{\rho}_N^{\text{du}}, \bar{z}_N) \quad \text{for } 1 \le q \le Q.$$

Then, we derive from (16c) and (21)

$$\left(D\vartheta(\bar{\mu})^\top \bar{\xi}\right)^\top (\mu^\delta - \bar{\mu}) \ge 0 \quad \text{for all } \mu^\delta \in \mathscr{D}. \tag{22}$$

Summarizing we have proved the following result.

Theorem 3.1. *Suppose that the bilinear form $a(\cdot, \cdot; \mu)$ satisfies (2). Further, $f, \ell \in V'$ holds and the functions ϑ^q are continuously differentiable for $1 \le q \le Q$. Let $\bar{x}_N = (\bar{y}_N, \bar{\mu}) \in X_N^{ad}$, $\bar{y}_N = (\bar{u}_N, \bar{z}_N) \in Y_N$, be a local solution to (**P**). Then, there exists a unique associated Lagrange multiplier pair $\bar{\lambda}_N = (\bar{\lambda}_N^1, \lambda_N^2) \in Y_N$ satisfying together with $\bar{x}_N$ the first-order necessary optimality conditions (18), (20) and (22).*

The gradient $\nabla \hat{J}$ of the reduced cost functional $\hat{J}$ at a point $\mu \in \mathscr{D}$ is given by the formula [5, 13]

$$\nabla \hat{J}(\mu) = D\vartheta(\mu)^\top \xi \in \mathbb{R}^P, \tag{23}$$

where the components of the vector $\xi \in \mathbb{R}^Q$ are

$$\xi_q = a^q(u_N, \bar{\rho}_N^{\text{pr}} - \lambda_N^1) + a^q(\lambda_N^2 - \bar{\rho}_N^{\text{du}}, z_N) \quad \text{for } 1 \le q \le Q,$$

$(u_N, z_N) = \mathscr{S}(\mu)$ holds and $\lambda_N = (\lambda_N^1, \lambda_N^2) \in Y_N$ solves the dual system

$$\begin{aligned}
a(u_N^\delta, \lambda_N^1; \mu) &= a(u_N^\delta, \rho_N^{\text{pr}}; \mu) && \text{for all } u_N^\delta \in V_{N^{\text{pr}}}, \\
a(\lambda_N^2, z_N^\delta; \mu) &= a(\rho_N^{\text{du}}, z_N^\delta; \mu) && \text{for all } z_N^\delta \in \tilde{V}_{N^{\text{du}}}.
\end{aligned}$$

Here, $\rho_N^{\text{pr}}, \rho_N^{\text{du}} \in V^{\mathscr{N}}$ are the Riesz representants of the residuals $r_N^{\text{pr}}(\cdot; \mu)$, $r_N^{\text{du}}(\cdot; \mu) \in (V^{\mathscr{N}})'$, respectively.

Remark 3.6. Suppose that the bilinear form is given by $a(\cdot, \cdot; \mu) = \vartheta^1(\mu) a^1(\cdot, \cdot)$ (i.e., $Q = 1$) and $\vartheta^1(\mu) \ne 0$ holds for all $\mu \in \mathscr{D}$. Then, solutions to different parameter values are linearly dependent; see Remark 2.1-2) and Remark 2.3-2). Then, it follows from $\vartheta^1(\mu) \ne 0$, (18) and (20) that

$$\begin{aligned}
a^1(u_N^\delta, \bar{\lambda}_N^1) &= a^1(u_N^\delta, \bar{\rho}_N^{\text{pr}}) && \text{for all } u_N^\delta \in V_{N^{\text{pr}}}, \\
a^1(\bar{\lambda}_N^2, z_N^\delta) &= a^1(\bar{\rho}_N^{\text{du}}, z_N^\delta) && \text{for all } z_N^\delta \in \tilde{V}_{N^{\text{du}}}.
\end{aligned}$$

In particular, $a^1(\bar{u}_N, \bar{\rho}_N^{\text{pr}} - \bar{\lambda}_N^1) = a^1(\bar{\lambda}_N^2 - \bar{\rho}_N^{\text{du}}, \bar{z}_N) = 0$ holds true, which gives $\xi_1 = 0$. Therefore, $\nabla \hat{J}(\mu) = 0$ is satisfied. This coincides with the observation in Remark 2.3-2) that the mappings

$$\mu \mapsto \|a(\mathscr{S}_N^{\text{pr}}(\mu), \cdot; \mu) - f\|_{(V^{\mathscr{N}})'} \quad \text{and} \quad \mu \mapsto \|a(\cdot, \mathscr{S}_N^{\text{du}}(\mu); \mu) + \ell\|_{(V^{\mathscr{N}})'}$$

are constant. $\diamond$

4 Second-order derivatives

To solve (**P**) in our numerical experiments we apply a globalized sequential quadratic programming (SQP) method which is makes use of second-order derivatives of the Lagrange functional; see [10], for example. For that reason we address second-order optimality conditions in this section. We restrict ourselves to simple bounds, i.e., we assume that the bounded and convex parameter set $\mathscr{D}$ is given by

$$\mathscr{D} = \underbrace{\big[\mu_{a,1},\mu_{b,1}\big] \times \ldots \times \big[\mu_{a,P},\mu_{b,P}\big]}_{P\text{-times}} \subset \mathbb{R}^P$$

with lower and upper bounds $\mu_{a,i} \leq \mu_{b,i}$, $1 \leq i \leq P$. Let $\bar{x}_N = (\bar{y}_N, \bar{\mu}) \in X_N^{ad}$, $\bar{y}_N = (\bar{u}_N, \bar{z}_N) \in Y_N$, be a solution to the first-order necessary optimatity conditions for (**P**); see Theorem 3.1. Moreover, the pair $\bar{\lambda}_N = (\bar{\lambda}_N^1, \bar{\lambda}_N^2) \in Y_N$ denotes for the associated unique Lagrange multiplier. We suppose that the functions ϑ^q are twice continuously differentiable. For $u_N^\delta, \tilde{u}_N^\delta \in V_{N^{\mathrm{pr}}}$ we deduce

$$\mathscr{L}_{u_N u_N}(\bar{x}_N, \bar{\lambda}_N)(u_N^\delta, \tilde{u}_N^\delta) = -(a(\tilde{u}_N^\delta, \cdot\,; \bar{\mu}), a(u_N^\delta, \cdot\,; \bar{\mu}))_{(V^{\mathscr{N}})'}. \tag{24}$$

Analogously, we find for $z_N^\delta, \tilde{z}_N^\delta \in \tilde{V}_{N^{\mathrm{du}}}$

$$\mathscr{L}_{z_N z_N}(\bar{x}_N, \bar{\lambda}_N)(z_N^\delta, \tilde{z}_N^\delta) = -(a(\cdot, \tilde{z}_N^\delta; \bar{\mu}), a(\cdot, z_N^\delta; \bar{\mu}))_{(V^{\mathscr{N}})'}. \tag{25}$$

Further, it follows that

$$\mathscr{L}_{u_N z_N}(\bar{x}_N, \bar{\lambda}_N)(u_N^\delta, z_N^\delta) = \mathscr{L}_{z_N u_N}(\bar{x}_N, \bar{\lambda}_N)(z_N^\delta, u_N^\delta). \tag{26}$$

for $u_N^\delta \in V_{N^{\mathrm{pr}}}$ and $z_N^\delta \in \tilde{V}_{N^{\mathrm{du}}}$. Using $\bar{r}_N^{\mathrm{pr}} = f - a(\bar{u}_N, \cdot\,; \bar{\mu}) \in V'$ and the Riesz representant $\bar{\rho}_N^{\mathrm{pr}} \in V$ of $\bar{r}_N^{\mathrm{pr}}$ we observe that

$$\mathscr{L}_{\mu u_N}(\bar{x}_N, \bar{\lambda}_N)(u_N^\delta, \mu^\delta) = \mathscr{L}_{u_N \mu}(\bar{x}_N, \bar{\lambda}_N)(u_N^\delta, \mu^\delta)$$

$$= \sum_{q=1}^{Q} \nabla \vartheta^q(\bar{\mu})^\top \mu^\delta \left(a^q(u_N^\delta, \bar{\rho}_N^{\mathrm{pr}} - \bar{\lambda}_N^1) - (a^q(\bar{u}_N, \cdot), a(u_N^\delta, \cdot\,; \bar{\mu}))_{(V^{\mathscr{N}})'} \right).$$

for $u_N^\delta \in V_{N^{\mathrm{pr}}}$ and $\mu^\delta \in \mathbb{R}^P$. Let $\bar{\zeta}_N^{\mathrm{pr},q} \in V^{\mathscr{N}}$, $1 \leq q \leq Q$, denote the Riesz representants of $a^q(\bar{u}_N, \cdot) \in (V^{\mathscr{N}})'$, i.e.

$$\langle \zeta_N^{\mathrm{pr},q}, \varphi^{\mathscr{N}} \rangle_V = a^q(\bar{u}_N, \varphi^{\mathscr{N}}) \quad \text{for all } \varphi^{\mathscr{N}} \in V^{\mathscr{N}}.$$

Then, we derive that

$$\mathscr{L}_{u_N \mu}(\bar{x}_N, \bar{\lambda}_N)(u_N^\delta, \mu^\delta)$$

$$= \sum_{q=1}^{Q} \nabla \vartheta^q(\bar{\mu})^\top \mu^\delta \left(a^q(u_N^\delta, \bar{\rho}_N^{\mathrm{pr}} - \bar{\lambda}_N^1) - a(u_N^\delta, \bar{\zeta}_N^{\mathrm{pr},q}; \bar{\mu}) \right) \tag{27}$$

for $u_N^\delta \in V_{N^{\mathrm{pr}}}$ and $\mu^\delta \in \mathbb{R}^P$. As above we apply $\bar{r}_N^{\mathrm{du}} = \ell + a(\cdot, \bar{z}_N; \bar{\mu}) \in (V^{\mathcal{N}})'$ and the Riesz representant $\bar{\rho}_N^{\mathrm{du}} \in V^{\mathcal{N}}$ of $\bar{r}_N^{\mathrm{du}}$ we observe that

$$\mathscr{L}_{\mu z_N}(\bar{x}_N, \bar{\lambda}_N)(\mu^\delta, z_N^\delta) = \mathscr{L}_{z_N \mu}(\bar{x}_N, \bar{\lambda}_N)(z_N^\delta, \mu^\delta)$$
$$= \sum_{q=1}^Q \nabla \vartheta^q(\bar{\mu})^\top \mu^\delta \left(a^q(\bar{\lambda}_N^2 - \rho_N^{\mathrm{du}}, z_N^\delta) - (a^q(\cdot, \bar{z}_N), a(\cdot, z_N^\delta; \bar{\mu}))_{(V^{\mathcal{N}})'} \right)$$

for $z_N^\delta \in \tilde{V}_{N^{\mathrm{du}}}$ and $\mu^\delta \in \mathbb{R}^P$. Let $\bar{\omega}_N^{\mathrm{du},q} \in V^{\mathcal{N}}$, $1 \le q \le Q$, denote the Riesz representants of $a^q(\cdot, \bar{z}_N) \in (V^{\mathcal{N}})'$, i.e.

$$\langle \bar{\omega}_N^{\mathrm{du},q}, \varphi^{\mathcal{N}} \rangle_V = a^q(\varphi^{\mathcal{N}}, \bar{z}_N) \quad \text{for all } \varphi^{\mathcal{N}} \in V^{\mathcal{N}}.$$

Then, we conclude that

$$\mathscr{L}_{z_N \mu}(\bar{x}_N, \bar{\lambda}_N)(z_N^\delta, \mu^\delta)$$
$$= \sum_{q=1}^Q \nabla \vartheta^q(\bar{\mu})^\top \mu^\delta \left(a^q(\bar{\lambda}_N^2 - \rho_N^{\mathrm{du}}, z_N^\delta) - a(\bar{\omega}_N^{\mathrm{du},q}, z_N^\delta; \bar{\mu}) \right) \tag{28}$$

for $z_N^\delta \in \tilde{V}_{N^{\mathrm{du}}}$ and $\mu^\delta \in \mathbb{R}^P$. Finally, we find for $\mu^\delta, \tilde{\mu}^\delta \in \mathbb{R}^P$

$$\mathscr{L}_{\mu\mu}(\bar{x}_N, \bar{\lambda}_N)(\mu^\delta, \tilde{\mu}^\delta)$$
$$= \tilde{\mu}^{\delta,\top} \left(\left(\sum_{q=1}^Q \left(a^q(\bar{u}_N, \bar{\rho}_N^{\mathrm{pr}}) - a^q(\bar{\rho}_N^{\mathrm{du}}, \bar{z}_N) \right) \right) \nabla^2 \vartheta^q(\bar{\mu}) \right) \mu^\delta$$
$$- \sum_{q=1}^Q \nabla \vartheta^q(\bar{\mu})^\top \mu^\delta \nabla \vartheta^q(\bar{\mu})^\top \tilde{\mu}^\delta \left(\|\bar{\zeta}_N^{\mathrm{pr},q}\|_V^2 + \|\bar{\omega}_N^{\mathrm{du},q}\|_V^2 \right) \tag{29}$$

with $\tilde{\mu}^{\delta,\top} = (\tilde{\mu}^\delta)^\top$.

The convergence of the SQP method relies on second-order sufficient optimality conditions for (**P**). For an arbitrary $\tau \ge 0$ let us define the set of *strongly active constraints* for the parameter $\bar{\mu}$ by

$$\mathscr{A}_\tau(\bar{\mu}) = \left\{ i \in \{1, \ldots, P\} \,\middle|\, |(\nabla \hat{J}(\bar{\mu}))_i| \ge \tau \right\}$$
$$= \left\{ i \in \{1, \ldots, P\} \,\middle|\, |(\nabla D\vartheta(\bar{\mu})^\top \bar{\xi})_i| \ge \tau \right\},$$

where $(\nabla \hat{J}(\bar{\mu}))_i$ denotes the i-th component of the vector $\nabla \hat{J}(\bar{\mu}) \in \mathbb{R}^P$. *Second-order sufficient optimality conditions* for (**P**) are as follows [13]: Let $\bar{x}_N = (\bar{y}_N, \bar{\mu}) \in X_N^{ad}$, $\bar{y}_N = (\bar{u}_N, \bar{z}_N) \in Y_N$, be a solution to the first-order necessary optimatity conditions for (**P**); see Theorem 3.1. Moreover, the pair $\bar{\lambda}_N = (\bar{\lambda}_N^1, \bar{\lambda}_N^2) \in Y_N$ are the associated Lagrange multiplier. If there exists a $\kappa > 0$ such that

$$\mathscr{L}_{x_N x_N}(\bar{x}_N, \bar{\lambda}_N)(x_N^\delta, x_N^\delta) \ge \kappa \left(\|u_N^\delta\|_V^2 + \|z_N^\delta\|_V^2 + \|\mu^\delta\|_{\mathbb{R}^P}^2 \right) \tag{30}$$

for all $x_N^\delta = (y_N^\delta, \mu^\delta) \in X_N$, $y_N^\delta = (u_N^\delta, z_N^\delta)$, satisfying $y_N^\delta \in \ker e'(\bar{x}_N)$ and

$$(\mu^\delta)_i \begin{cases} = 0 & \text{if } i \in \mathscr{A}_\tau(\bar{\mu}), \\ \geq 0 & \text{if } \bar{\mu}_i = \mu_{a,i} \text{ and } i \notin \mathscr{A}_\tau(\bar{\mu}), \\ \leq 0 & \text{if } \bar{\mu}_i = \mu_{b,i} \text{ and } i \notin \mathscr{A}_\tau(\bar{\mu}). \end{cases}$$

then $\bar{x}_N$ is a strictly local solution to (**P**).

Suppose that $x_N^\delta = (y_N^\delta, \mu^\delta) \in \ker e'(\bar{x}_N)$ with $y_N^\delta = (u_N^\delta, z_N^\delta) \in Y_N$. Then we have

$$a(u_N^\delta, \psi; \bar{\mu}) = -\sum_{q=1}^{Q} \nabla \vartheta^q(\bar{\mu})^\top \mu^\delta a^q(\bar{u}_N, \psi) \qquad \text{for all } \psi \in V_{N^{\text{pr}}}, \qquad (31a)$$

$$a(\phi, z_N^\delta; \bar{\mu}) = -\sum_{q=1}^{Q} \nabla \vartheta^q(\bar{\mu})^\top \mu^\delta a^q(\phi, \bar{z}_N) \qquad \text{for all } \phi \in \tilde{V}_{N^{\text{du}}}. \qquad (31b)$$

Utilizing (2a), (3) and (31a) we find

$$\alpha_0 \|u_N^\delta\|_V^2 \leq a(u_N^\delta, u_N^\delta; \bar{\mu}) \leq \gamma \|\bar{u}_N\|_V \sum_{q=1}^{Q} \|\nabla \vartheta^q(\bar{\mu})\|_{\mathbb{R}^P} \|u_N^\delta\|_V \|\mu^\delta\|_{\mathbb{R}^P}$$

which implies

$$\|u_N^\delta\|_V \leq \bar{C}_1 \|\mu^\delta\|_{\mathbb{R}^P} \quad \text{for all } x_N^\delta = (y_N^\delta, \mu^\delta) \in \ker e'(\bar{x}_N). \qquad (32a)$$

with $\bar{C}_1 = \gamma \|\bar{u}_N\|_V \sum_{q=1}^{Q} \|\nabla \vartheta^q(\bar{\mu})\|_{\mathbb{R}^P}$. Analogously, we derive from (2a), (3) and (31a)

$$\|z_N^\delta\|_V \leq \bar{C}_2 \|\mu^\delta\|_{\mathbb{R}^P} \quad \text{for all } x_N^\delta = (y_N^\delta, \mu^\delta) \in \ker e'(\bar{x}_N). \qquad (32b)$$

with $\bar{C}_2 = \gamma \|\bar{z}_N\|_V \sum_{q=1}^{Q} \|\nabla \vartheta^q(\bar{\mu})\|_{\mathbb{R}^P}$. From (2b), (32) and (32b) we infer that

$$\begin{aligned} -\|a(u_N^\delta, \cdot; \bar{\mu})\|_{(V^{\mathcal{N}})'}^2 - \|a(\cdot, z_N^\delta; \bar{\mu})\|_{(V^{\mathcal{N}})'}^2 &\geq -\gamma^2 \left(\|u_N^\delta\|_V^2 + \|z_N^\delta\|_V^2 \right) \\ &\geq -\gamma^2 \left(\bar{C}_1^2 + \bar{C}_2^2 \right) \|\mu\|_{\mathbb{R}^P}^2. \end{aligned} \qquad (33)$$

We set $\bar{C}_3 = \gamma^2 (\bar{C}_1^2 + \bar{C}_2^2)$. Then, we derive from (24)-(29) and (33) that

$$\mathscr{L}_{x_N x_N}(\bar{x}_N, \bar{\lambda}_N)(u_N^\delta, u_N^\delta) =$$

$$\geq -\bar{C}_3 \|\mu^\delta\|_{\mathbb{R}^P}^2 + 2 \sum_{q=1}^{Q} \nabla \vartheta^q(\bar{\mu})^\top \mu^\delta \left(a^q(u_N^\delta, \bar{\rho}_N^{\mathrm{pr}} - \bar{\lambda}_N^1) - a(u_N^\delta, \bar{\zeta}_N^{\mathrm{pr},q}; \bar{\mu}) \right)$$

$$+ 2 \sum_{q=1}^{Q} \nabla \vartheta^q(\bar{\mu})^\top \mu^\delta \left(a^q(\bar{\lambda}_N^2 - \rho_N^{\mathrm{du}}, z_N^\delta) - a(\bar{\omega}_N^{\mathrm{du},q}, z_N^\delta; \bar{\mu}) \right)$$

$$+ \mu^{\delta,\top} \left(\left(\sum_{q=1}^{Q} \left(a^q(\bar{u}_N, \bar{\rho}_N^{\mathrm{pr}}) - a^q(\bar{\rho}_N^{\mathrm{du}}, \bar{z}_N) \right) \right) \nabla^2 \vartheta^q(\bar{\mu}) \right) \mu^\delta$$

$$- \sum_{q=1}^{Q} \left| \nabla \vartheta^q(\bar{\mu})^\top \mu^\delta \right|^2 \left(\|\bar{\zeta}_N^{\mathrm{pr},q}\|_V^2 + \|\bar{\omega}_N^{\mathrm{du},q}\|_V^2 \right)$$

for all $x_N^\delta \in \ker e'(\bar{x}_N)$. Since

$$-\bar{C}_3 \|\mu^\delta\|_{\mathbb{R}^P}^2 - \sum_{q=1}^{Q} \left| \nabla \vartheta^q(\bar{\mu})^\top \mu^\delta \right|^2 \left(\|\bar{\zeta}_N^{\mathrm{pr},q}\|_V^2 + \|\bar{\omega}_N^{\mathrm{du},q}\|_V^2 \right) \leq 0$$

holds and the matrix

$$\left(\sum_{q=1}^{Q} \left(a^q(\bar{u}_N, \bar{\rho}_N^{\mathrm{pr}}) - a^q(\bar{\rho}_N^{\mathrm{du}}, \bar{z}_N) \right) \right) \nabla^2 \vartheta^q(\bar{\mu})$$

need not be positive definite, the second-order sufficient optimality condition (30) is not obvious in our case.

Remark 4.7. If $\bar{\mu}$ is strongly active in all P components, it follows that $\mathscr{A}_\tau(\bar{\mu}) = \{1, \ldots, P\}$. Thus, $\mu^\delta = 0$ is satisfied. From (32) and (32b) we conclude that $y_N^\delta = 0$ holds. This imply the second-order necessary optimality conditions at $\bar{x}_N$. ◇

5 Numerical experiments

In this section we present some numerial results for the described theory. We use two versions of the well-known Thermal-Block-Model (see e.g. [12]) as a model example. Model 1 consists of two blocks (i.e. $B_1 = 2, B_2 = 1$) while Model 2 consists of four blocks (i.e. $B_1 = 2, B_2 = 2$), see Figure 1. The parameter domain is chosen as $\mathscr{D} = [0.2, 2]^P$, where P again denotes the number of parameters, i.e., $P = 2$ for Model 1 and $P = 4$ for Model 2, see Figure 1. We choose $\varepsilon_{\mathrm{stop}} = 1e-5$ as stopping criteria for the Greedy-algorithm. Since $P = 2$ for Model 1, we can easily visualize the reduced cost functional $\hat{J}(\mu)$ in that case, see Figure 2.
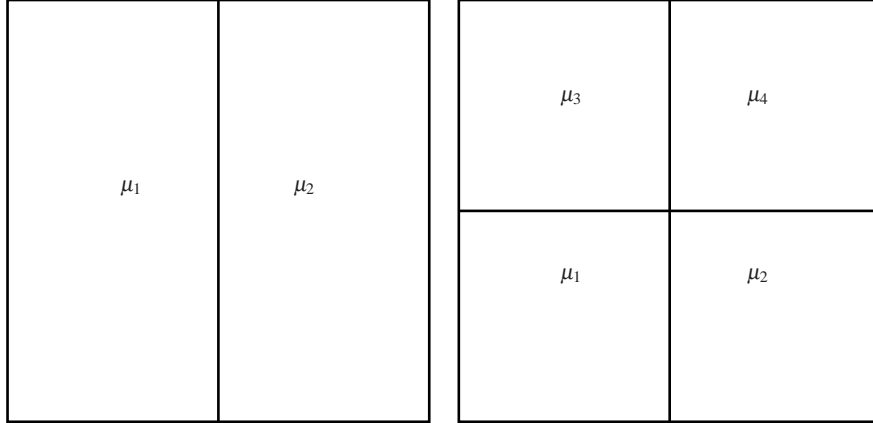
Fig. 1 Left: Model 1 ($B_1 = 2, B_2 = 1$), right: Model 2 ($B_1 = 2, B_2 = 2$).

As we can deduce from the shape of the cost functional, the appropriate choice for an initial value for the optimization scheme[1] is crucial in order to avoid determining a local minimum only. Let us clarify this in Figure 2 (b): Choosing an initial parameter μ_{init}^N in the left half of the plane will lead to a local minimum whereas an initial value located in the right half of the plane will yield the global optimum $(0.2, 2)$. In order to avoid the output of a local minimum, we have used three different strategies:

1. `euclidian_mu`: μ_{init}^N is chosen by maximizing the Euclidian distance to the barycenter of the previously determined parameter values μ_i, $1 \leq i \leq N-1$.
2. `random_mu` with "safety zone": μ_{init}^N is chosen randomly in $\mathscr{D}$, but ensuring a minimal distance (measured in the Euclidian norm) to all μ_i, $1 \leq i \leq N-1$. This "safety zone" is chosen adaptively, i.e., the radius of the circular zone is decreased with increasing N. If we would not do that, we would get an N^{max}, where no additional feasible points could be found.
3. `coarse_grid_mu`: An equidistant coarse parameter-mesh consisting of $M = 3^P$ ($M = 9$ for Model 1 and $M = 81$ for Model 2) points is used. We choose that parameter as initial value μ_{init}^N whose cost functional is minimal on that grid.

Best results were obtained using `coarse_grid_mu` and all figures correspond to this strategy. We used an SQP algorithm as optimization scheme and compare the results to a classical training set strategy, using equidistant training sets consisting of $3^2 = 81$ respectively $10^2 = 100$ parameter values. Figure 3 (left) shows decay of the error estimator during the Greedy-process (i.e., with increasing N). We choose

---

[1] We used MATLAB's function `fmincon` for this. We set `options.TolCon=1e-6;` `options.TolFun=1e-6;` and `options.Algorithm='sqp'`, i.e., we used a MATLAB internal SQP algorithm.

$\mu = (1,1)$ as initial snapshot-parameter. As expected (see [11]) the Greedy stops after two steps with $\mu_1 = (\mu_{min}, \mu_{max})$ and $\mu_2 = (\mu_{max}, \mu_{min})$. In this example there is no difference between using an optimization algorithm and using a training set strategy since the optimal parameter values μ_1 and μ_2 are contained in the training set. Hence, our optimization procedure is consistent with the known theory.

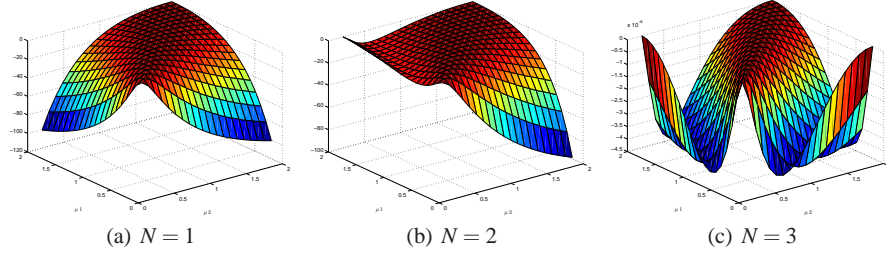(a) $N = 1$ (b) $N = 2$ (c) $N = 3$

Fig. 2 Reduced cost functional $\hat{J}(\mu)$ for Model 1. Note that the range for the y-axis for $N = 3$ is 10^{-6}.

In Figure 3 (right) the decay of the error estimator is shown for Model 2. The training sets consisted of $\{3^4, 5^4, 7^4, 10^4\}$ equidistant parameter values and the desired tolerance of 1e-5 is reached at a basis size of $N = 20$ for all strategies. We observe the expected exponential decay and our optimization strategy performs as good as the classical training set strategies. This is remarkable since in our model cases the distribution of the optimal parameters is known (the so-called 'magic points'). From the point of view of the optimization method, this is the worst case which is a strong indication that the optimization approach also works when no a-priori knowledge for the choice of the training set is available.

In Table 1 we show the overall number of evaluations of $\hat{J}(\mu)$ - i.e., the number of reduced simulations - during the Greedy process in the offline phase for Model 1 and Model 2. Especially for Model 2 the Greedy algorithm combined with the optimization scheme needs much less function calls than the Greedy algorithm combined with a training set strategy. This can be an advantage in order to overcome the curse of dimension which prohibits to choose the training set arbitrarily large especially in high dimensions.

References

1. R. Becker and R. Rannacher An optimal control approach to a posteriori error estimation in finite element methods. Acta Numerica 2001, 1-102.
2. M. Drohmann, B. Haasdonk, and M. Ohlberger. Adaptive reduced basis methods for nonlinear convection-diffusion equations. Finite Volumes for Complex Applications VI – Problems & Perspectives, J. Fort et al. (eds.), Springer Proceedings in Mathematics 4, Volume 1, page 369-377, 2011
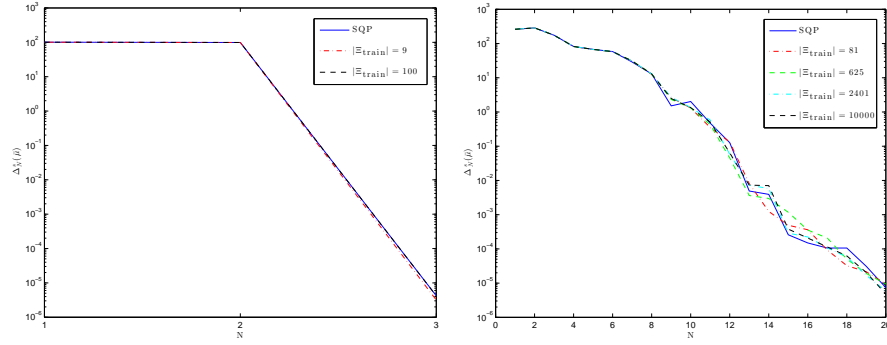
Fig. 3 Error estimator $\Delta_N^s(\bar\mu)$ for the optimal value $\bar\mu$ in Model 1 (left) and Model 2 (right).

Model 1	# RB simulations	Model 2	# RB simulations
SQP	79	SQP	1.511
$|\Xi_{\text{train}}| = 9$	18	$|\Xi_{\text{train}}| = 81$	1.439
$|\Xi_{\text{train}}| = 100$	200	$|\Xi_{\text{train}}| = 625$	11.875
		$|\Xi_{\text{train}}| = 2401$	45.619
		$|\Xi_{\text{train}}| = 10000$	190.000

Table 1 Number of reduced simulations during the offline phase for Model 1 (left) and Model 2 (right).

3. B. Haasdonk, and M. Ohlberger. Adaptive basis enrichment for the reduced basis method applied to finite volume schemes. Submitted, 2012.
4. J.S. Hesthaven, B. Stamm, and S. Zhang. Efficient greedy algorithms for high-dimensional parameter spaces with applications to empirical interpolation and reduced basis methods. Submitted, 2012.
5. M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE Constraints*. Mathematical Modelling: Theory and Applications, vol. 23, Springer Verlag, 2009.
6. C.T. Kelley. *Iterative Methods for Optimization*. Frontiers in Applied Mathematics, SIAM, Philadelphia, PA, 1999.
7. O. Lass and S. Volkwein. Adaptive POD basis computation for parametrized nonlinear systems using optimal snapshot location. Submitted, 2012.
8. Y. Maday and B. Stamm. Locally adaptive greedy approximations for anisotropic parameter reduced basis spaces. Submitted, 2012.
9. N.C. Nguyen, K. Veroy, and A.T. Patera. *Certified real-time solution of parametrized partial differential equations*. In: Handbook of Materials Modeling, (S. Yip, editor), Springer, pp. 1523-1558, 2005.
10. J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer Series in Operation Research, Second Edition (2006).
11. A.T. Patera and G.Rozza, *Reduced Basis Approximation and a Posteriori Error Estimation for Parametrized Partial Differential Equations*. MIT, 2007
12. 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. Application to transport and continuum mechanics. *Archives of Computational Methods in Engineering*, 15:229-275, 2008.

13. F. Tröltzsch. *Optimal Control of Partial Differential Equations. Theory, Methods and Applications*. American Mathematical Society, Providence, 2010.