DFG Deutsche Forschungsgemeinschaft
Priority Programme 1962

Stationarity Conditions and Scalarization in Multiobjective Optimal Control of a Nonsmooth PDE

Marco Bernreuther, Georg Müller, Stefan Volkwein

SPP 1962 | Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization

Preprint Number SPP1962-167

received on April 9, 2021

Edited by
SPP1962 at Weierstrass Institute for Applied Analysis and Stochastics (WIAS)
Leibniz Institute in the Forschungsverbund Berlin e.V.
Mohrenstraße 39, 10117 Berlin, Germany
E-Mail: spp1962@wias-berlin.de

World Wide Web: http://spp1962.wias-berlin.de/

Computational Optimization and Applications manuscript No.
(will be inserted by the editor)

Stationarity Conditions and Scalarization in Multiobjective Optimal Control of a Nonsmooth PDE

Marco Bernreuther · Georg Müller · Stefan Volkwein

Received: date / Accepted: date

Abstract This work deals with the numerical characterization of Pareto stationary fronts for multiobjective optimal control problems with moderately many cost functionals with a mildly nonsmooth, elliptic, semilinear PDE constraint. We extend known stationarity conditions for ample controls to weaker conditions in the case where the controls are taken from a finite dimensional space and thus not "ample". The conditions associated with strong stationarity remain meaningful when numerically characterizing the fronts. We compare the performance of the weighted-sum and the (Euclidean) reference point method for this application using quantifiable measures for approximation quality of the fronts. The subproblems of either method are solved with a line search globalized pseudo semismooth Newton method that seems to remove the degenerate behavior of the local version employed previously. When solving the subproblems of the reference point method, memory and computation complexity related issues appear and are tackled using a matrix-free, iterative approach and comparing multiple preconditioning tactics.

Keywords multiobjective optimal control · nonsmooth optimization · stationarity conditions · Pareto optimality · scalarization methods · pseudo semismooth Newton method.

This research was supported by the German Research Foundation (DFG) under grant number VO 1658/5-2 within the priority program *Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization* (SPP 1962).

M. Bernreuther, G. Müller, S. Volkwein
University of Konstanz
Department of Mathematics and Statistics
D-78457 Konstanz, Germany
E-mail: {marco.bernreuther, georg.mueller, stefan.volkwein}@uni-konstanz.de

1 Problem Formulation

The aim of this paper is the numerical characterization of the Pareto fronts of nonsmooth bicriterial optimal control problems of the form

$$\min_{(y,u)} \mathcal{J}(y,u) = \begin{pmatrix} \mathcal{J}_1(y,u) \\ \mathcal{J}_2(y,u) \end{pmatrix} = \begin{pmatrix} j_1(y) + \dfrac{\sigma_1}{2}\,\|u\|_U^2 \\ j_2(y) + \dfrac{\sigma_2}{2}\,\|u\|_U^2 \end{pmatrix} \tag{\mathbf{P}}$$
$$\text{s.t. } (y,u) \in V \times U \text{ satisfies } -\Delta y + \kappa \max\{0,y\} = \mathcal{B}(u) \text{ in } V'$$

based on generalized stationarity conditions and the comparison of the performance of a weighted-sum approach and a reference point method in terms of quantifiable discretization quality. In (**P**), the symbols y and u denote the state and control variables in the corresponding state space V and (possibly finite dimensional) control space U, respectively, the j_i denote suitably well behaved scalar cost functionals, the σ_i are nonnegative regularization parameters and $\mathcal{B}\colon U \to L^2(\Omega)$ denotes a control-to-right-hand-side mapping. For the detailed assumptions on the problem, we refer to Assumption 11. Note that we restrict ourselves to two objective functions in this work for ease of presentation, especially with respect to the reference point method, however the scope of this paper can readily be extended to moderately many objective functions using hierarchical approaches, see [3,5,19].

Multicriterial optimization problems with nonsmooth PDE constraints like (**P**) arise in various physical applications with conflicting objectives, see, e.g., [17,21,25,27]. The combination of generally only (Hadamard) directionally differentiable Nemytski operators in the constraint and the inherently nonsmooth structure of multiobjective optimization makes sensitivity and stationarity analysis and the numerical solution of these problems rather delicate and requires specialized stationarity concepts and approaches that do not follow standard procedure for Gateaux differentiable problems, cf. the detailed introduction in [10]. The particular case of the PDE constraint in (**P**) is rather well understood in terms of existence and regularity of solutions and differentiability properties of the solution operator and has previously been addressed as a constraint in optimization problems in, e.g., [9,10]. The specific structure of the PDE even allows for the derivation of strong stationarity systems when the control space is sufficiently rich, which has been considered in both scalar and multiobjective optimization with an arbitrary number of objectives, see [9,10]. Stationarity conditions of intermediate strength based on the characterization of the subdifferentials of the solution operator to the constraining PDE have been addressed in [9] and the considerations in [10] show that C-stationarity and strong stationarity in fact coincide for ample controls. Numerically, the Pareto front of the problem (**P**) has been characterized for two and three cost functionals and $L^2(\Omega)$ controls [10] by using a first-optimize-then-discretize approach and a pseudo semismooth Newton (PSN) method and by using a regularization approach. Especially in the multiobjective case, computation times can quickly become large with increased fineness of discretizations of

the domain and the Pareto front. In [7], the authors discussed a standard offline/online greedy based *reduced basis* approach for (**P**) with a single scalar objective function and both low and high dimensional control/parameter space and compared the results to an adaptive way of generating the reduced basis along the solution process of the PSN.

In this paper, we will build mostly on the results in [10] with the goal of characterizing the Pareto front of the bicriterial problem (**P**) using a weighted-sum and a reference point approach, which we will compare in terms of approximation quality of the front using quantifiable quality measures. Since our setting includes finite dimensional controls (with specific structure), we have to extend the strong stationarity systems used in [10] to conditions that will turn out to be weaker, as is well known to be expected for controls that are not ample. We show that the numerical procedure in [10] based on the strong stationarity system for the multiobjective setting corresponds to solving stationarity systems for the weighted-sum method and that the subproblems appearing in the reference point method have similar structure that can also be treated by the PSN. Additionally, we will see that the systems solved in the PSN remain usable in the numerical approximation even when the controls are finite dimensional. We further show that the rare, mesh-dependent nonconvergence issue for some subproblems observed for the undamped PSN in [7,10] can be effectively eliminated using a line search globalization strategy.

The structure of this paper is as follows: We will shortly comment on the assumptions for this paper in Subsection 1.1. Then, we recall the required notions of Pareto optimality and Pareto stationarity and state the respective first order stationarity systems for (non-)ample controls in Section 2. In the case $U = H$ the analytical results are analogous to those in [10] and the strong stationarity results are simply recalled. The case $U = \mathbb{R}^p$ requires additional considerations and yields weaker stationarity conditions. We will further show that the system of strong stationarity conditions remains a viable system to solve numerically when characterizing the Pareto stationary front for either control space. In Section 3, we will recall the weighted-sum method and the reference point method and address their roles in characterizing the Pareto front. The numerical implementation is explained in detail in Section 4, where we present a matrix-free preconditioned limited-memory generalized minimal residual (L-GMRES) method for the line search globalized pseudo semismooth Newton (gPSN) method that will be used to handle the density of the discretization matrices in the reference point method and the convergence issues arising from the nonsmoothness of the PDE-constraint, respectively. We further present two numerical examples – one with (FE-discretized) L^2-controls and one with inherently finite dimensional controls in Section 5. The interpretation of the numerical results is specifically focused on the effects of different preconditioning strategies for the reference point method. We introduce two quantities to measure the approximation quality of the two methods and use them as a basis for a performance comparison of the two scalarization approaches.

1.1 Notation and Assumptions on the Data

We endow $V = H_0^1(\Omega)$ with the inner product $\langle \varphi, \phi \rangle_V = \int_\Omega \nabla\varphi \cdot \nabla\phi + \varphi\phi \, \mathrm{d}\boldsymbol{x}$ for $\varphi, \phi \in V$ and the induced norm $\|\cdot\|_V = \langle \cdot, \cdot \rangle_V^{1/2}$. Its topological dual space is written as $V' = H^{-1}(\Omega)$. The space Y denotes $V \cap H^2(\Omega)$ with topological dual space Y'. We also set $H = L^2(\Omega)$. For functions in Y, the Laplacian is understood in the non-variational sense and the Dirichlet Laplacian $\Delta\colon H \to Y'$ is understood in the very weak sense (see [16], Section 1.9). Our assumptions on the data are as follows:

Assumption 11 *1. $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N} \setminus \{0\}$ is a bounded domain that is either convex or possesses a $C^{1,1}$-boundary (cf.* [14, Section 6.2]*),*
2. $j_1, j_2\colon Y \to \mathbb{R}$ are weakly lower semicontinuous, twice continuously Fréchet-differentiable and bounded from below,
3. $\sigma_1 \geq 0$, $\sigma_2 > 0$, $\kappa \geq 0$,
4. $U = \mathbb{R}^p$ for $p \in \mathbb{N} \setminus \{0\}$ and $\|\cdot\|_U$ denotes the Euclidean norm or $U = H$ and $\|\cdot\|_U$ denotes the L^2-norm,
5. $\mathcal{B}\colon U \to H$ possesses the following property:
(a) If $U = \mathbb{R}^p$ the operator $\mathcal{B} : U \to H$ is linear and bounded and the pairwise intersection of the sets $\{b_i \neq 0\} \subset \Omega$ of $b_i := \mathcal{B}(e_i) \in H$, where e_i, $i = 1, \ldots, p$ denote the unit vectors in U, are Lebesgue nullsets and none of the b_i are zero.
(b) If $U = H$ the operator $\mathcal{B}\colon U \to H$ is unitary.

2 Stationarity Conditions

Structurally, this section follows [10] closely, where the case $U = H = L^2(\Omega)$ carries over immediately, but the results for the case $U = \mathbb{R}^p$ are novel. Note that stationarity concepts of intermediate strength for the scalar optimization case have been presented in [9].

The results in this section will be the basis for showing that the strong stationarity system obtained for the infinite dimensional controls remains a meaningful system to treat numerically for the case of finite dimensional controls. The main issue in the analysis is the well-known fact that strong stationarity conditions typically require ample controls. Note that the results presented in this section can readily be generalized to an arbitrary finite number of objective functionals.

We start by summarizing the main properties of the solution operator $\mathcal{S}$ to the PDE-constraint in the next lemma as a slight extension to [10, Lemma 4.2]. Note that, since we will mostly focus on the case $U = \mathbb{R}^p$, we do not obtain all results compared to the case $U = H$, where further results are possible. Again, see [10, Lemma 4.2].

Lemma 1 (Properties of the solution operator $\mathcal{S}$) *Let $u \in U$ be a control with associated state $y = \mathcal{S}(u)$. Then:*

1. *There is a solution operator $\mathcal{S}\colon U \to Y$ that is Lipschitz continuous and Hadamard directionally differentiable, where the derivative $\mathcal{S}'[u](h) = w \in Y$ for given direction $h \in U$ is the unique solution to*

$$-\Delta w + \kappa \mathbb{1}_{\{y=0\}} \max\{0, w\} + \kappa \mathbb{1}_{\{y>0\}} w = \mathcal{B}(h) \quad \text{in } V'. \tag{1}$$

   This especially implies the Y-regularity of the state variable y.
2. *If $U = \mathbb{R}^p$, then the map*

$$\tilde{\mathcal{B}}^{-\ell}\colon H \to U, \quad v \mapsto \nu = (\nu_i)_{1\le i\le p} \quad \text{with } \nu_i = \frac{\langle v, b_i\rangle_H}{\|b_i\|_H^2} \tag{2}$$

   is a linear and bounded left inverse of $\mathcal{B}\colon U \to H$.
3. *The map $\mathcal{S}'[u]\colon U \to Y$ is Lipschitz continuous and allows for a Lipschitz continuous left inverse given by*

$$\mathcal{S}'[u]^{-\ell}\colon Y \to U, \quad w \mapsto \mathcal{B}^{-\ell}[u]\big(\underbrace{-\Delta w + \kappa \mathbb{1}_{\{y=0\}} \max\{0, w\} + \kappa \mathbb{1}_{\{y>0\}} w}_{\in H}\big),$$

   where $\mathcal{B}^{-\ell}[u]$ is any linear, bounded left inverse of $\mathcal{B}$ that may depend on u.
4. *There exists a linear and bounded left inverse $\mathcal{B}^{-\ell}$ of $\mathcal{B}$ that does not depend on u such that*

$$\big\langle u, \mathcal{S}'[u]^{-\ell}(w)\big\rangle_U = \big\langle (-\Delta + \kappa \mathbb{1}_{\{y>0\}})\mathcal{B}^{-\ell,*}(u), w\big\rangle_{Y',Y}$$

   for every $w \in Y$, where $\mathcal{B}^{-\ell,*}\colon U \to H$ is the Hilbert adjoint of the left inverse $\mathcal{B}^{-\ell}$.

Proof Let $u \in U$ be arbitrarily given and $y = \mathcal{S}(u) \in Y$. Notice that the Y-regularity also follows from Proposition 2.1 in [9].

1 The linearity and boundedness of $\mathcal{B}$ imply Lipschitz continuity and Hadamard differentiability analogously to Proposition 2.1 and Theorem 2.2 in [9] and due to the chain rule, the directional derivative of the solution operator $w = \mathcal{S}'[u](h)$ solves

$$-\Delta w + \kappa \mathbb{1}_{\{y=0\}} \max\{0, w\} + \kappa \mathbb{1}_{\{y>0\}} w = \mathcal{B}(h) \quad \text{in } V'.$$

2 Note that the operator $\tilde{\mathcal{B}}^{-\ell}$ is well defined due to Assumption 11-*5*. Clearly, the operator is linear and bounded. The left inverse quality remains to be proved. Let $\tilde{u} \in U = \mathbb{R}^p$ and $v = \mathcal{B}\tilde{u} = \sum_{j=1}^p b_j \tilde{u}_j \in H$. For every $i \in \{1, \dots, p\}$, we have that

$$\big(\tilde{\mathcal{B}}^{-\ell}(v)\big)_i = \frac{\langle v, b_i\rangle_H}{\|b_i\|_H^2} = \sum_{j=1}^p \tilde{u}_j \frac{\langle b_j, b_i\rangle_H}{\|b_i\|_H^2} = \tilde{u}_i \frac{\langle b_i, b_i\rangle_H}{\|b_i\|_H^2} = \tilde{u}_i,$$

   where the second to last equality holds due to the H-orthogonality of the b_i induced by Assumption 11-*5*.

3 The Lipschitz continuity of the linearized solution operator is implied by the form of the linearization (1). Existence of a left inverse is clear due to part *2* if $U = \mathbb{R}^p$ and since $\mathcal{B}$ is unitary if $U = H$. Thus existence of a left inverse of $\mathcal{S}'[u]$ and its Lipschitz continuity are obvious from the explicit definition.

4 First assume that $U = H$ and that $\mathcal{B}$ is a unitary operator. Then, $\mathcal{B}^{-1}$ exists, and we choose $\mathcal{B}^{-\ell} = \mathcal{B}^{-1}$. Thus we have

$$\mathcal{B}^{-\ell,*} = (\mathcal{B}^{-1})^* = (\mathcal{B}^*)^* = \mathcal{B}.$$

Consequently, for $u \in U$ and $y = \mathcal{S}(u)$,

$$\begin{aligned}\langle \mathcal{B}^{-\ell,*}(u), \kappa \mathbb{1}_{\{y=0\}} \max\{0,w\}\rangle_H &= \langle \mathcal{B}(u), \kappa \mathbb{1}_{\{y=0\}} \max\{0,w\}\rangle_H \\ &= \langle \underbrace{-\Delta y + \kappa \max\{0,y\}}_{=0 \text{ a.e. on } \{y=0\}}, \kappa \mathbb{1}_{\{y=0\}} \max\{0,w\}\rangle_H = 0,\end{aligned}$$

where $\Delta y = 0$ a.e. on $\{y = 0\}$ is a consequence of [10, Lemma 4.1]). Thus, using part 3), we find that

$$\begin{aligned}\big\langle u, \mathcal{S}'[u]^{-\ell}(w)\big\rangle_U &= \big\langle u, \mathcal{B}^{-\ell}\big(-\Delta w + \kappa \mathbb{1}_{\{y=0\}} \max\{0,w\} + \kappa \mathbb{1}_{\{y>0\}} w\big)\big\rangle_U \\ &= \big\langle \mathcal{B}^{-\ell,*}(u), \big(-\Delta + \kappa \mathbb{1}_{\{y>0\}}\big) w\big\rangle_H + \big\langle \mathcal{B}^{-\ell,*}(u), \kappa \mathbb{1}_{\{y=0\}} \max\{0,w\}\big\rangle_H \\ &= \big\langle \mathcal{B}^{-\ell,*}(u), \big(-\Delta + \kappa \mathbb{1}_{\{y>0\}}\big) w\big\rangle_H \\ &= \big\langle \big(-\Delta + \kappa \mathbb{1}_{\{y>0\}}\big) \mathcal{B}^{-\ell,*}(u), w\big\rangle_{Y',Y}\end{aligned}$$

for every $w \in Y$, where the last line follows due to the definition of the very weak Dirichlet Laplacian.
Now assume that $U = \mathbb{R}^p$. We show that $\tilde{\mathcal{B}}^{-\ell}$ is the desired left inverse. For $u \in U$, $y = \mathcal{S}(u)$ and any $i \in \{1, \dots, p\}$, we have that $\{y = 0\} \cap \{b_i u_i \neq 0\}$ is a nullset, because

$$\{b_i u_i \neq 0\} \cap \{y = 0\} \subset \{b_i u_i \neq 0\} \cap \{\mathcal{B}(u) = 0\} \subset \{b_i u_i \neq 0\} \cap \{b_i u_i = 0\},$$

where the first inclusion is again a consequence of [10, Lemma 4.1]) and the second inclusion is due to Assumption 11-*5*. Thus, for any $i \in \{1, \dots, p\}$, we infer

$$\frac{\langle \mathbb{1}_{\{y=0\}} \max\{0,w\}, b_i u_i\rangle_H}{\|b_i\|_H^2} = 0.$$

Due to part 2), we obtain that

$$\begin{aligned}&\big\langle u, \tilde{\mathcal{B}}^{-\ell}\big(\kappa \mathbb{1}_{\{y=0\}} \max\{0,w\}\big)\big\rangle_U \\ &= \sum_{i=1}^p u_i \frac{\langle \kappa \mathbb{1}_{\{y=0\}} \max\{0,w\}, b_i\rangle_H}{\|b_i\|_H^2} = \kappa \sum_{i=1}^p \frac{\langle \mathbb{1}_{\{y=0\}} \max\{0,w\}, u_i b_i\rangle_H}{\|b_i\|_H^2} = 0\end{aligned}$$

for every $w \in Y$. Consequently,

$$\begin{aligned}\langle u, \mathcal{S}'[u]^{-\ell}(w)\rangle_U &= \big\langle u, \tilde{\mathcal{B}}^{-\ell}\big(-\Delta w + \kappa \mathbb{1}_{\{y=0\}} \max\{0, w\} + \kappa \mathbb{1}_{\{y>0\}} w\big)\big\rangle_U \\ &= \big\langle \big(-\Delta + \kappa \mathbb{1}_{\{y>0\}}\big)\tilde{\mathcal{B}}^{-\ell,*}(u), w\big\rangle_{Y',Y},\end{aligned}$$

for every $w \in Y$. □

As usual, we denote the reduced cost functional as $\hat{\mathcal{J}}\colon U \to \mathbb{R}^2$, $\hat{\mathcal{J}}(u) = \mathcal{J}(\mathcal{S}(u), u)$. Having established the properties of the solution operator, we are ready to review the different notions of Pareto optimality and the optimality conditions that will play a role later on.

Definition 1 (Pareto Optimality) Let $\bar{y}, \bar{u}$ with $\bar{y} = \mathcal{S}(\bar{u})$ and $\bar{u} \in U$. The control $\bar{u}$ is called:

1) a *local weak Pareto optimal point* of (**P**) if an $r > 0$ exists such that there is no $u \in U$ satisfying

$$\|u - \bar{u}\|_U < r, \qquad \mathcal{J}_i(\mathcal{S}(u), u) < \mathcal{J}_i(\bar{y}, \bar{u}) \text{ for } i = 1, 2;$$

the set of all local weak Pareto optimal points and the corresponding Pareto front are denoted $\mathcal{P}_s^{lw}$ and $\mathcal{P}_f^{lw}$, respectively;

2) a *local Pareto optimal point* of (**P**) if an $r > 0$ exists such that there is no $u \in U$ satisfying

$$\|u - \bar{u}\|_U < r, \qquad \mathcal{J}_i(\mathcal{S}(u), u) \leq \mathcal{J}_i(\bar{y}, \bar{u}) \text{ for } i = 1, 2,$$

where the latter inequality is strict for at least one i; the set of all local Pareto optimal points and the corresponding Pareto front are denoted $\mathcal{P}_s^{l}$ and $\mathcal{P}_f^{l}$, respectively;

3) a *local proper Pareto optimal point* of (**P**) if there are $r, C > 0$ such that for every $u \in U$ satisfying $\|u - \bar{u}\|_U < r$ and $\mathcal{J}_i(\mathcal{S}(u), u) \leq \mathcal{J}_i(\bar{y}, \bar{u})$ for some index $i \in \{1, 2\}$, there exists an index $m \in \{1, 2\} \setminus \{i\}$ with

$$\mathcal{J}_i(\bar{y}, \bar{u}) - \mathcal{J}_i(\mathcal{S}(u), u) \leq C(\mathcal{J}_m(\mathcal{S}(u), u) - \mathcal{J}_m(\bar{y}, \bar{u}));$$

the set of all local proper Pareto optimal points and the corresponding Pareto front are denoted $\mathcal{P}_s^{lp}$ and $\mathcal{P}_f^{lp}$, respectively;

4) a *global (weak/proper) Pareto optimal point* of (**P**) if the previous conditions hold with $r = \infty$; the corresponding Pareto fronts and sets are denoted using the letter g instead of l in the superscript.

Analogously to [10], we obtain the following corresponding primal optimality conditions.

Theorem 1 (Optimality Conditions – Primal Form)

1) *If $\bar{u} \in U$ with associated state $\bar{y} = \mathcal{S}(\bar{u})$ is a local weak Pareto optimal point of* (**P**)*, then there exists no direction $h \in U$ satisfying*

$$\langle j_i'(\bar{y}), \mathcal{S}'[\bar{u}](h)\rangle_{Y',Y} + \sigma_i \langle \bar{u}, h\rangle_U < 0 \quad \text{for } i = 1, 2. \tag{3}$$

2) *If $\bar{u} \in U$ with associated state $\bar{y} = \mathcal{S}(\bar{u})$ is a local proper Pareto optimal point of* (**P**) *with constants $r, C > 0$, then for every $h \in U$ with $\langle j_i'(\bar{y}), \mathcal{S}'[\bar{u}](h)\rangle_{Y',Y} + \sigma_i \langle \bar{u}, h\rangle_U < 0$ for some $i \in \{1,2\}$, there exists an $m \in \{1,2\} \setminus \{i\}$ with*

$$-\big(\langle j_i'(\bar{y}), \mathcal{S}'[\bar{u}](h)\rangle_{Y',Y} + \sigma_i \langle \bar{u}, h\rangle_U\big) \leq C\big(\langle j_m'(\bar{y}), \mathcal{S}'[\bar{u}](h)\rangle_{Y',Y} + \sigma_m \langle \bar{u}, h\rangle_U\big).$$

Proof See [10, Theorem 3.1]. □

Analogously, we have the corresponding notions of Pareto stationarity.

Definition 2 (Pareto Stationarity) Let $\bar{u} \in U$ and $\bar{y} = \mathcal{S}(\bar{u})$. The control $\bar{u}$ is called:

1) a *weak Pareto stationary point* of (**P**) if there is no $h \in U$ satisfying

$$\langle j_i'(\bar{y}), \mathcal{S}'[\bar{u}](h)\rangle_{Y',Y} + \sigma_i \langle \bar{u}, h\rangle_U < 0 \quad \text{for } i = 1,2;$$

the set of all weak Pareto stationary points and the corresponding Pareto front are denoted $\mathcal{P}_s^{sw}$ and $\mathcal{P}_f^{sw}$, respectively;

2) a *Pareto stationary point* of (**P**) if there is no $h \in U$ satisfying

$$\langle j_i'(\bar{y}), \mathcal{S}'[\bar{u}](h)\rangle_{Y',Y} + \sigma_i \langle \bar{u}, h\rangle_U \leq 0 \quad \text{for } i = 1,2,$$

where the latter inequality is strict for at least one i; the set of all Pareto stationary points and the corresponding Pareto front are denoted $\mathcal{P}_s^{s}$ and $\mathcal{P}_f^{s}$, respectively;

3) a *proper Pareto stationary point* of (**P**) if there is a $C > 0$ such that for all $h \in U$ with $\langle j_i'(\bar{y}), \mathcal{S}'[\bar{u}](h)\rangle_{Y',Y} + \sigma_i \langle \bar{u}, h\rangle_U < 0$ for some $i \in \{1,2\}$, there exists an $m \in \{1,2\} \setminus \{i\}$ with

$$-\big(\langle j_i'(\bar{y}), \mathcal{S}'[\bar{u}](h)\rangle_{Y',Y} + \sigma_i \langle \bar{u}, h\rangle_U\big) \leq C\big(\langle j_m'(\bar{y}), \mathcal{S}'[\bar{u}](h)\rangle_{Y',Y} + \sigma_m \langle \bar{u}, h\rangle_U\big).$$

The set of all proper Pareto stationary points and the corresponding Pareto front are denoted $\mathcal{P}_s^{sp}$ and $\mathcal{P}_f^{sp}$, respectively.

Remark 1 We want to mention some of the connections between the introduced sets.

1) By definition, we have that $\mathcal{P}_s^{lp} \subset \mathcal{P}_s^{l} \subset \mathcal{P}_s^{lw}$. The same holds for the global Pareto optima and for the Pareto fronts.
2) As a consequence of Theorem 1, weak Pareto stationarity is a necessary condition for local weak Pareto optimality, i.e., $\mathcal{P}_s^{lw} \subset \mathcal{P}_s^{sw}$, and proper Pareto stationarity is a necessary condition for local proper Pareto optimality, i.e., $\mathcal{P}_s^{lp} \subset \mathcal{P}_s^{sp}$. Corresponding results hold for the Pareto fronts. However, Pareto stationarity is generally not necessary for local Pareto optimality.

For more details we refer to [10]. ◊

Next we derive adjoint-based necessary optimality conditions. We will extend a version of the infinite dimensional case of Tucker's/Motzkin's theorem of the alternative (see [12, Theorem 3.22] for the finite dimensional case and [10, Lemma 4.4] for a general Hilbert space setting) to subsets of Hilbert spaces. The proof is an adaptation of the proof of [10, Lemma 4.4] to the case of subsets. Note that in the latter result the equivalence between the positivity of the multipliers and (4) is stated, whereas in the following lemma there is only an implication.

Lemma 2 *Suppose $\mathcal{W}$ is a nonempty subset of a real Hilbert space $\mathcal{V}$ and $v_1', \dots, v_N' \in \mathcal{V}'$ are given. Then the following are equivalent:*

1. *There exists no $z \in \mathcal{W}$ such that*

$$\langle v_i', z\rangle_{\mathcal{V}} < 0 \qquad \text{for all } i = 1, \dots, N.$$

2. *There exists $\lambda \in \mathbb{R}^N$ with $\lambda_i \geq 0$ for $i = 1, \dots, N$ such that*

$$\sum_{i=1}^{N} \lambda_i = 1 \qquad \text{and} \qquad \sum_{i=1}^{N} \lambda_i \langle v_i', w\rangle_{\mathcal{V}} \geq 0 \qquad \text{for all } w \in \mathcal{W}.$$

Furthermore, if $\lambda_i > 0$ for all $i = 1, \dots, N$ in 2, then there exists no $z \in \mathcal{W}$ such that

$$\langle v_i', z\rangle_{\mathcal{V}} \leq 0 \qquad \text{for all } i = 1, \dots, N, \tag{4}$$

with the inequality holding strictly for at least one i.

Proof First, we show that part *2* implies part *1*:
Assume that part *2* holds but there exists a $z \in \mathcal{W}$ with $\langle v_i', z\rangle_{\mathcal{V}} < 0$ for all $i = 1, \dots, N$. This would imply that

$$\sum_{i=1}^{N} \lambda_i \langle v_i', z\rangle_{\mathcal{V}} < 0,$$

which is a contradiction to *2*. Analogously, if $\lambda_i > 0$ for all $i = 1, \dots, N$, then there exists no $z \in \mathcal{W}$ such that $\langle v_i', z\rangle_{\mathcal{V}} \leq 0$ for all $i = 1, \dots, N$ with the inequality holding strictly for at least one i, which shows the claim.
Now we show that part *1* implies part *2*: Note that part *1* is equivalent to

$$\max\left\{\langle v_1', w\rangle_{\mathcal{V}}, \dots, \langle v_N', w\rangle_{\mathcal{V}}\right\} \geq 0 \quad \text{for all } w \in \mathcal{W}. \tag{5}$$

We introduce

$$g\colon \mathbb{R}^N \to \mathbb{R}, \qquad g(x_1, \dots, x_N) \coloneqq \max\{x_1, \dots, x_N\},$$
$$F\colon \mathcal{V} \to \mathbb{R}^N, \qquad F(v) \coloneqq \left(\langle v_i', v\rangle_{\mathcal{V}}\right)_{i=1,\dots,N}.$$

Then we can rewrite (5) as $(g \circ F)(w) \geq 0$ for all $w \in \mathcal{W}$ and the chain rule for Hadamard differentiable functions implies

$$(g \circ F)'[0](w) \geq 0 \quad \text{for all } w \in \mathcal{W},$$

since both g and F are Hadamard differentiable (see Example 2.2.8 in [11] for the directional derivative of g) and since (5) equivalently holds for all $v \in \{tw \colon w \in \mathcal{W}, t \geq 0\}$. Note that $g \circ F$ is also convex, since F is linear and g convex. Thus [11, Proposition 2.27] and [22, Example 8.26] imply

$$\begin{aligned}\partial(g \circ F)[0] &= \partial_c(g \circ F)[0] \\ &= F^* \left\{ \lambda \in \mathbb{R}^N : \lambda_i \geq 0 \text{ for } i = 1, \dots, N, \ \sum_{i=1}^{N} \lambda_i = 1 \right\},\end{aligned}$$

where F^* denotes the Hilbert adjoint of the linear, bounded mapping F, ∂ denotes the generalized gradient (in the sense of Clarke) and ∂_c denotes the convex subdifferential. Proposition 2.2.2 in [11] implies that $(g \circ F)'[0] \in \partial(g \circ F)[0]$. This proves existence of a multiplier λ with the desired properties. □

Modifying the result and the proof in [10, Theorem 4.5], we now obtain the following stationarity conditions that are strong in the sense that we obtain an adjoint based system that is equivalent to the primal stationarity defined above.

Theorem 2 (Strong Stationarity Conditions)

1. *The following are equivalent:*
   (a) *A control $\bar{u} \in U$ with associated state $\bar{y} = \mathcal{S}(\bar{u}) \in Y$ is a weak Pareto stationary point of* (**P**)*, i.e. satisfies* (3)*.*
   (b) *There exists an adjoint state $\bar{p}$ and a multiplier $\bar{\alpha}$ such that $\bar{u}, \bar{y}, \bar{p}, \bar{\alpha}$ satisfy the system*

$$\bar{u} \in U, \quad \bar{y} \in Y, \quad \bar{p} \in H, \quad \bar{\alpha} \in \mathbb{R}^2, \tag{6a}$$

$$\bar{\alpha}_i \geq 0 \quad \text{for } i = 1, 2, \ \sum_{i=1}^{2} \bar{\alpha}_i = 1, \tag{6b}$$

$$-\Delta \bar{y} + \kappa \max\{0, \bar{y}\} = \mathcal{B}(\bar{u}) \text{ in } V', \tag{6c}$$

$$\begin{aligned}\left\langle -\Delta \bar{p} + \kappa \mathbb{1}_{\{\bar{y}>0\}} \bar{p}, w \right\rangle_{Y',Y} &\leq \sum_{i=1}^{2} \bar{\alpha}_i \langle j_i'(\bar{y}), w \rangle_{Y',Y} \\ &\qquad \text{for all } w \in \operatorname{Im}(\mathcal{S}'[\bar{u}]),\end{aligned} \tag{6d}$$

$$\bar{p} + \sum_{i=1}^{2} \bar{\alpha}_i \sigma_i \mathcal{B}^{-\ell,*}(\bar{u}) = 0 \qquad \text{in } H. \tag{6e}$$

2. *Assume that $\bar{u}, \bar{y}, \bar{p}, \bar{\alpha}$ satisfy the system* (6)*, where the inequality in* (6b) *is strict, i.e. $\bar{\alpha}_i > 0$ for $i = 1, 2$. Then $\bar{u}$ is a proper Pareto stationary point of* (**P**) *(and thus also a Pareto stationary point).*

Proof As the image of $\mathcal{S}'[\bar{u}]^{-\ell}$ is U, weak Pareto stationarity is equivalent to

$$(\langle j_i'(\bar{y}), \mathcal{S}'[\bar{u}](\mathcal{S}'[\bar{u}]^{-\ell}(w))\rangle)_{Y',Y} + \sigma_i \langle \bar{u}, \mathcal{S}'[\bar{u}]^{-\ell}(w)\rangle_U < 0 \quad \text{for } i = 1, 2$$

being valid for no $w \in \mathrm{Im}(\mathcal{S}'[\bar{u}]) \subset Y$. Note that $\mathcal{S}'[\bar{u}](\mathcal{S}'[\bar{u}]^{-\ell}(w)) = w$ for all $w \in \mathrm{Im}(\mathcal{S}'[\bar{u}])$. This, together with the explicit definition of $\mathcal{S}'[\bar{u}]^{-\ell}$ in Lemma 1-*4*, gives the equivalent reformulation of

$$\langle j_i'(\bar{y}), w\rangle_{Y',Y} + \sigma_i \langle(-\Delta + \kappa \mathbb{1}_{\{\bar{y}>0\}})\mathcal{B}^{-\ell,*}(\bar{u}), w\rangle_{Y',Y} < 0$$

being valid for no $w \in \mathrm{Im}(\mathcal{S}'[\bar{u}])$. Due to Lemma 2-*1*, this is equivalent to the existence of $\bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_2) \in \mathbb{R}^2$ with $\bar{\alpha}_i \geq 0$, $\sum_{i=1}^2 \bar{\alpha}_i = 1$ and

$$\sum_{i=1}^{2} \bar{\alpha}_i \langle j_i'(\bar{y}) + \sigma_i(-\Delta + \kappa \mathbb{1}_{\{\bar{y}>0\}})\mathcal{B}^{-\ell,*}(\bar{u}), w\rangle_{Y',Y} \geq 0 \quad \text{for all } w \in \mathrm{Im}(\mathcal{S}'[\bar{u}]).$$

If we now define $\bar{p} = -\sum_{i=1}^2 \bar{\alpha}_i \sigma_i \mathcal{B}^{-\ell,*}(\bar{u})$, part *1* follows immediately.
The proof of part *2* can be found in [10], Theorem 4.5-iii). The only difference is that one equality needs to be replaced by an inequality. □

As already mentioned above, system (6) can be tightened for the case of $U = H$.

Remark 2 If $U = H$ and $\mathcal{B}$ is unitary, then because of surjectivity of $\mathcal{S}'[\bar{u}]$, (6d) becomes an equality in V' and (6e) is equivalent to

$$\mathcal{B}^*(\bar{p}) + \sum_{i=1}^{2} \bar{\alpha}_i \sigma_i \bar{u} = 0 \quad \text{in } U. \tag{7}$$

In this case, the reverse implication in Theorem 2, part *2* is true as well. This is essentially the result in [10, Theorem 4.5]. ◊

Additionally, we make the following observations.

Corollary 1 *Consider Theorem 2 and the case where $U = \mathbb{R}^p$.*

1) *If* (6e) *is replaced by* (7) *in Theorem* 2-1)*, then the resulting system is necessary (but generally not sufficient) for weak Pareto stationarity.*
2) *If* (6e) *is replaced by* (7) *in Theorem* 2-1) *or* -2) *and the sign condition*

$$\langle \mathbb{1}_{\{\bar{y}=0\}} \max\{0, w\}, \bar{p}\rangle_H \leq 0 \quad \textit{for all } w \in \mathrm{Im}(\mathcal{S}'[\bar{u}]), \tag{8}$$

is added, then the resulting system is sufficient (but generally not necessary) for weak/proper Pareto stationarity.

Proof Part *1* is easy to see because the adjoint of a left inverse is a right inverse of the adjoint, so (6e) implies (7). For part *1*, assume that $\bar{u}, \bar{y}, \bar{p}, \bar{\alpha}$ satisfy the

system (6a)-(6d), (7) and (8) with $\bar{\alpha}_i \geq 0$ for $i = 1, 2$. For arbitrary $h \in U$ set $w = \mathcal{S}'[\bar{u}](h)$. It follows that

$$\begin{aligned}
-\sum_{i=1}^{2} \bar{\alpha}_i \sigma_i \langle \bar{u}, h \rangle_U = \langle \mathcal{B}^*(\bar{p}), h \rangle_U &= \langle \bar{p}, \mathcal{B}(h) \rangle_H \\
&= \langle -\Delta w + \kappa \mathbb{1}_{\{\bar{y}>0\}} w + \kappa \mathbb{1}_{\{\bar{y}=0\}} \max\{0, w\}, \bar{p} \rangle_H \\
&\leq \langle -\Delta w + \kappa \mathbb{1}_{\{\bar{y}>0\}} w, \bar{p} \rangle_H = \langle -\Delta \bar{p} + \kappa \mathbb{1}_{\{\bar{y}>0\}} \bar{p}, w \rangle_{Y',Y} \\
&\leq \sum_{i=1}^{2} \bar{\alpha}_i \, \langle j_i'(\bar{y}), w \rangle_{Y',Y}.
\end{aligned}$$

Thus since $\bar{\alpha}_i \geq 0$ and $\sum_{i=1}^{2} \bar{\alpha}_i = 1$ the inequality

$$\langle j_i'(\bar{y}), w \rangle_{Y',Y} + \sigma_i \, \langle \bar{u}, h \rangle_U < 0$$

cannot be true for all $i = 1, 2$. This implies the desired weak Pareto stationarity. (Proper) Pareto stationarity can be shown analogously. □

3 Scalarization Methods

As shown in Section 2, Pareto stationarity can be equivalently described by (strong) stationarity conditions. We want to characterize the Pareto stationary points and the corresponding front numerically. Therefore, we explain briefly how two well-known scalarization methods – the weighted-sum method (cf., e.g., [12]) and the reference-point method (cf., e.g., [18,23]) – can be used to that end.

3.1 Weighted-Sum Method (WSM)

For weights $\alpha_1, \alpha_2 \geq 0$ with $\alpha_1 + \alpha_2 = 1$, the optimization problem

$$\begin{aligned}
&\min_{(y,u)} \alpha_1 \mathcal{J}_1(y, u) + \alpha_2 \mathcal{J}_2(y, u) \\
&\text{s.t. } (y, u) \in V \times U \text{ satisfies } -\Delta y + \kappa \max\{0, y\} = \mathcal{B}(u) \text{ in } V',
\end{aligned} \tag{$\mathbf{P}_\alpha$}$$

is called the *weighted-sum problem* (with non-negative weights α_1, α_2) corresponding to ($\mathbf{P}$). We will employ the following notation:

$$\begin{aligned}
\mathcal{W}_s^{g\geq} &:= \{u \in U : u \text{ glob. sol. to } (\mathbf{P}_\alpha),\ \alpha \geq 0,\ \alpha_1 + \alpha_2 = 1\}, \quad \mathcal{W}_f^{g\geq} := \hat{\mathcal{J}}(\mathcal{W}_s^{g\geq}), \\
\mathcal{W}_s^{g>} &:= \{u \in U : u \text{ glob. sol. to } (\mathbf{P}_\alpha),\ \alpha > 0,\ \alpha_1 + \alpha_2 = 1\}, \quad \mathcal{W}_f^{g>} := \hat{\mathcal{J}}(\mathcal{W}_s^{g>}), \\
\mathcal{W}_s^{l\geq} &:= \{u \in U : u \text{ loc. sol. to } (\mathbf{P}_\alpha),\ \alpha \geq 0,\ \alpha_1 + \alpha_2 = 1\}, \quad \mathcal{W}_f^{l\geq} := \hat{\mathcal{J}}(\mathcal{W}_s^{l\geq}), \\
\mathcal{W}_s^{l>} &:= \{u \in U : u \text{ loc. sol. to } (\mathbf{P}_\alpha),\ \alpha > 0,\ \alpha_1 + \alpha_2 = 1\}, \quad \mathcal{W}_f^{l>} := \hat{\mathcal{J}}(\mathcal{W}_s^{l>}),
\end{aligned}$$

where the subscripts s and f stand for "set" and "front", respectively, as for the Pareto optimal/stationary sets in the previous section. The primal optimality conditions for the WSM are given in the following theorem.

Theorem 3 *Let the control $\bar{u} \in U$ be locally optimal for $(\mathbf{P})$ with associated state $\bar{y} = \mathcal{S}(\bar{u}) \in Y$. Then*

$$\sum_{i=1}^{2} \alpha_i \big(\langle j_i'(\bar{y}), \mathcal{S}'[\bar{u}](h) \rangle_{Y',Y} + \sigma_i \langle u, h \rangle_U \big) \geq 0 \quad \text{for all } h \in U. \tag{9}$$

Proof The claim follows analogously to [10, Theorem 3.1]. □

Definition 3 A control $\bar{u} \in U$ with associated $\bar{y} = \mathcal{S}(\bar{u})$ is called a *stationary point of* $(\mathbf{P}_\alpha)$ if (9) is satisfied. We set

$$\mathcal{W}_s^{s\geq} := \{u \in U : u \text{ stat. pt. of } (\mathbf{P}_\alpha),\ \alpha \geq 0,\ \alpha_1 + \alpha_2 = 1\}, \quad \mathcal{W}_f^{s\geq} := \hat{\mathcal{J}}(\mathcal{W}_s^{s\geq}),$$
$$\mathcal{W}_s^{s>} := \{u \in U : u \text{ stat. pt. of } (\mathbf{P}_\alpha),\ \alpha > 0,\ \alpha_1 + \alpha_2 = 1\}, \quad \mathcal{W}_f^{s>} := \hat{\mathcal{J}}(\mathcal{W}_s^{s>}).$$

Corollary 2 *Let $\alpha_1, \alpha_2 \geq 0$ with $\alpha_1 + \alpha_2 = 1$ and denote $\alpha = (\alpha_1, \alpha_2)$. Then the following statements are equivalent:*

1. *A control $\bar{u} \in U$ with associated state $\bar{y} = \mathcal{S}(\bar{u}) \in Y$ is a stationary point of $(\mathbf{P}_\alpha)$.*
2. *There exists $\bar{p}$ such that $\bar{u}, \bar{y}, \bar{p}$ satisfy the system (6) with $\bar{\alpha} = \alpha$.*

Proof The corollary follows analogously to the proof of Theorem 2. □

Remark 3 1) Corollary 2 especially implies that $\mathcal{W}_s^{s\geq} = \mathcal{P}_s^{sw}$ and $\mathcal{W}_s^{s>} \subset \mathcal{P}_s^{s}$. This means that the WSM is a good choice to characterize (weakly) Pareto stationary points.

1. 2)] For the sets of optimal points, only the inclusions $\mathcal{W}_s^{l>} \subset \mathcal{P}_s^{l}$, $\mathcal{W}_s^{l\geq} \subset \mathcal{P}_s^{lw}$, $\mathcal{W}_s^{g>} \subset \mathcal{P}_s^{g}$ and $\mathcal{W}_s^{g\geq} \subset \mathcal{P}_s^{gw}$ hold. The proof of this result is the same as for the finite dimensional case shown in [12]. ◊

In the algorithm, we can now set $\alpha_1 = 1 - \alpha_2$ and solve the stationarity system of the WSM for varying $\alpha_2 \in [0, 1]$, where $\alpha_2 \neq 0$ for solvability requirements. Specifically, we introduce an additional small parameter $\alpha_{\text{tol}} > 0$ and choose α_2 in $[\alpha_{\text{tol}}, 1 - \alpha_{\text{tol}}]$. The procedure of the WSM is summarized in Algorithm 1.

Algorithm 1: Weighted-sum method (WSM)

Require: Number $k_{\max} \in \mathbb{N}$ of discretization points, $\alpha_{\text{tol}} > 0$;
Return : Discrete approximations $\tilde{\mathcal{P}}_s^{sw}$ and $\tilde{\mathcal{P}}_f^{sw}$ of $\mathcal{P}_s^{sw}$ and $\mathcal{P}_f^{sw}$;
for $i = 1, \ldots, k_{\max}$ **do**
  Set $\alpha_2 = \alpha_{\text{tol}} + (1 - 2\alpha_{\text{tol}}) \frac{i-1}{k_{\max}-1}$;
  Solve (6) with weight $(1 - \alpha_2, \alpha_2)$ and save a solution as u^i;
  Set $\tilde{\mathcal{P}}_s^{sw} = \tilde{\mathcal{P}}_s^{sw} \cup \{u^i\}$, $\tilde{\mathcal{P}}_f^{sw} = \tilde{\mathcal{P}}_f^{sw} \cup \{\hat{\mathcal{J}}(u^i)\}$;

The result of Algorithm 1 is a discrete approximation of the set of weakly Pareto stationary points $\mathcal{P}_s^{sw}$ and the corresponding front $\mathcal{P}_f^{sw}$. However, system (6) may have multiple solutions and (in the case of the finite dimensional controls) we cannot solve it numerically because of the variational inequality (6d) on the possibly unknown and nonlinear set $\mathrm{Im}(\mathcal{S}'[\bar{u}])$. In practice, we therefore modify (6d)-(6e) and instead consider the system

$$\bar{u} \in U, \quad \bar{y} \in Y, \qquad \bar{p} \in H, \quad \bar{\alpha} \in \mathbb{R}^2 \tag{10a}$$

$$\bar{\alpha}_i \geq 0 \quad \text{for } i = 1, 2, \qquad \sum_{i=1}^{2} \bar{\alpha}_i = 1, \tag{10b}$$

$$-\Delta \bar{y} + \kappa \max\{0, \bar{y}\} = \mathcal{B}(\bar{u}) \qquad \text{in } V', \tag{10c}$$

$$-\Delta \bar{p} + \kappa \mathbb{1}_{\{\bar{y}>0\}} \bar{p} = \sum_{i=1}^{2} \bar{\alpha}_i j_i'(\bar{y}) \text{ in } V', \tag{10d}$$

$$\mathcal{B}^*(\bar{p}) + \sum_{i=1}^{2} \bar{\alpha}_i \sigma_i \bar{u} = 0 \qquad \text{in } U, \tag{10e}$$

which coincides with the strong stationarity system in the case $U = H$; cf. Remark 2. If $U = \mathbb{R}^p$, then $\bar{u}$ satisfying (10) is generally only stationary in a weaker sense. In that case, we additionally check the sign condition

$$\bar{p} \leq 0 \quad \text{a.e. in } \{\bar{y} = 0\} \tag{11}$$

for $\bar{p}$ a posteriori. If it is satisfied, the solution is still strongly stationary and therefore a weak Pareto stationary point, see Corollaries 1 and 2.

3.2 Reference Point Method (RPM)

In this section, we will show that the same favourable set inclusions mentioned in Remark 3 for the WSM hold for the RPM, which is well known to yield higher approximation qualities than the WSM for smooth, convex problems, see, e.g., [4]. The *reference point problem* with Euclidean norm $\|\cdot\|_2$ for a reference point $z \in \mathbb{R}^2$ is given by

$$\begin{aligned} &\min_{(y,u)} \mathcal{F}_z(y,u) = \frac{1}{2} \|\mathcal{J}(y,u) - z\|_2^2 \\ &\text{s.t. } (y,u) \in V \times U \text{ satisfies } -\Delta y + \kappa \max\{0, y\} = \mathcal{B}(u) \text{ in } V'. \end{aligned} \tag{$\mathbf{P}_z$}$$

Analogously to the previous subsection, we employ the following notation:

$$\begin{aligned} \mathcal{R}_s^{g\geq} &:= \{u \in U : u \text{ glob. sol. to } (\mathbf{P}_z),\ z \in \mathbb{R}^2,\ 0 \neq \hat{\mathcal{J}}(u) - z \geq 0\}, \\ \mathcal{R}_s^{g>} &:= \{u \in U : u \text{ glob. sol. to } (\mathbf{P}_z),\ z \in \mathbb{R}^2,\ \hat{\mathcal{J}}(u) - z > 0\}, \\ \mathcal{R}_s^{l\geq} &:= \{u \in U : u \text{ loc. sol. to } (\mathbf{P}_z),\ z \in \mathbb{R}^2,\ 0 \neq \hat{\mathcal{J}}(u) - z \geq 0\}, \\ \mathcal{R}_s^{l>} &:= \{u \in U : u \text{ loc. sol. to } (\mathbf{P}_z),\ z \in \mathbb{R}^2,\ \hat{\mathcal{J}}(u) - z > 0\}, \\ \mathcal{R}_f^{g\geq} &:= \hat{\mathcal{J}}(\mathcal{R}_s^{g\geq}), \quad \mathcal{R}_f^{g>} := \hat{\mathcal{J}}(\mathcal{R}_s^{g>}), \quad \mathcal{R}_f^{l\geq} := \hat{\mathcal{J}}(\mathcal{R}_s^{l\geq}), \quad \mathcal{R}_f^{l>} := \hat{\mathcal{J}}(\mathcal{R}_s^{l>}). \end{aligned}$$

Theorem 4 *We have that* $\mathcal{R}_s^{l>} \subset \mathcal{P}_s^l$ *and* $\mathcal{R}_s^{g>} \subset \mathcal{P}_s^g$.

Proof We assume that $\bar{u} \in \mathcal{R}_s^{l>}$ with $\bar{y} = \mathcal{S}(\bar{u})$, i.e., there exists an $r_1 > 0$ such that for all $u \in U$ with $\|\bar{u} - u\|_U < r_1$ the inequality $\mathcal{F}_z(\bar{y}, \bar{u}) \leq \mathcal{F}_z(\mathcal{S}(u), u)$ is satisfied. Now, we assume that $\bar{u} \notin \mathcal{P}_s^l$, which implies that for every $r_2 > 0$ there exists $u_{\mathrm{end}} \in U$ with $\|u_{\mathrm{end}} - \bar{u}\| < r_2$ and $\mathcal{J}_i(\mathcal{S}(u_{\mathrm{end}}), u_{\mathrm{end}}) \leq \mathcal{J}_i(\bar{y}, \bar{u})$ for $i = 1, 2$, where the latter inequality is strict for at least one i. Since we can choose r_2 arbitrarily small and since $\mathcal{S}$ and $\mathcal{J}_i$ are continuous and $\mathcal{J}_i(\bar{y}, \bar{u}) - z_i > 0$ for $i = 1, 2$ by assumption, this implies that

$$z_i \leq \mathcal{J}_i(\mathcal{S}(u_{\mathrm{end}}), u_{\mathrm{end}}) \leq \mathcal{J}_i(\bar{y}, \bar{u}), \quad i = 1, 2,$$

where the second inequality is strict for at least one i. Since the Euclidean norm is strictly monotone [12, Definition 4.19], this implies the contradiction $\mathcal{F}_z(\mathcal{S}(u_{\mathrm{end}}), u_{\mathrm{end}}) < \mathcal{F}_z(\bar{y}, \bar{u})$ and proves the first inclusion. The second inclusion now immediately follows from the first by choosing $r_1 = \infty$. □

Next we introduce primal stationarity conditions for $(\mathbf{P}_z)$.

Theorem 5 *Let the control* $\bar{u} \in U$ *with associated state* $\bar{y} = \mathcal{S}(\bar{u}) \in Y$ *be locally optimal for* $(\mathbf{P}_z)$. *Then, we have for all* $h \in U$

$$\sum_{i=1}^{2} \left(\mathcal{J}_i(\bar{y}, \bar{u}) - z_i\right) \left(\langle j_i'(\bar{y}), \mathcal{S}'[\bar{u}](h)\rangle_{Y',Y} + \sigma_i \langle \bar{u}, h\rangle_U\right) \geq 0. \tag{12}$$

Proof Due to the chain rule for Hadamard differentiable functions, this follows analogously to [10, Theorem 3.1]. □

Definition 4 A control $\bar{u} \in U$ with associated state $\bar{y} = \mathcal{S}(\bar{u})$ is called a *stationary point of* $(\mathbf{P}_z)$ for $z \in \mathbb{R}^2$ if (12) is satisfied. We set

$$\begin{aligned}
\mathcal{R}_s^{s\geq} &:= \{u \in U : u \text{ stat. pt. of } (\mathbf{P}_z),\ z \in \mathbb{R}^2,\ 0 \neq \hat{\mathcal{J}}(u) - z \geq 0\},\\
\mathcal{R}_s^{s>} &:= \{u \in U : u \text{ stat. pt. of } (\mathbf{P}_z),\ z \in \mathbb{R}^2,\ \hat{\mathcal{J}}(u) - z > 0\},\\
\mathcal{R}_f^{s\geq} &:= \hat{\mathcal{J}}(\mathcal{R}_s^{s\geq}), \quad \mathcal{R}_f^{s>} := \hat{\mathcal{J}}(\mathcal{R}_s^{s>}).
\end{aligned}$$

Corollary 3 *Let the control* $\bar{u} \in U$ *with associated state* $\bar{y} = \mathcal{S}(\bar{u})$ *be given.*

1) *The following are equivalent:*
   (a) *The control* $\bar{u}$ *is a stationary point of* $(\mathbf{P}_z)$.
   (b) *There exists an adjoint state* $\bar{p}$ *such that* $\bar{u}, \bar{y}, \bar{p}$ *satisfy*

$$\bar{u} \in U, \quad \bar{y} \in Y, \quad \bar{p} \in H \tag{13a}$$

$$-\Delta \bar{y} + \kappa \max\{0, \bar{y}\} = \mathcal{B}(\bar{u}) \text{ in } V', \tag{13b}$$

$$\langle -\Delta \bar{p} + \kappa \mathbb{1}_{\{\bar{y}>0\}} \bar{p}, w\rangle_{Y',Y} \leq \sum_{i=1}^{2} \left(\mathcal{J}_i(\bar{y}, \bar{u}) - z_i\right) \langle j_i'(\bar{y}), w\rangle_{Y',Y} \quad \text{for all } w \in \mathrm{Im}(\mathcal{S}'[\bar{u}]), \tag{13c}$$

$$\bar{p} + \sum_{i=1}^{2} \left(\mathcal{J}_i(\bar{y}, \bar{u}) - z_i\right) \sigma_i B^{-\ell,*}(\bar{u}) = 0 \quad \text{in } H. \tag{13d}$$

2) *The following are equivalent:*
   (a) *There exists* $z \in \mathbb{R}^2$ *such that the control* $\bar{u}$ *is a stationary point of* $(\mathbf{P}_z)$ *with* $0 \neq \mathcal{J}(\bar{y}, \bar{u}) - z \geq 0$ *(or* $\mathcal{J}(\bar{y}, \bar{u}) - z > 0$*).*
   (b) *There exists* $\alpha \in \mathbb{R}^2$ *with* $\alpha \geq 0$ *for* $i = 1, 2$ *(or* $\alpha > 0$*) and* $\alpha_1 + \alpha_2 = 1$ *such that the control* $\bar{u}$ *is a stationary point of* $(\mathbf{P}_\alpha)$*.*

Proof Part 1) follows analogously to the proof of Theorem 2 executed for one cost functional.
To show part 2), let $\bar{u}$ be a stationary point of $(\mathbf{P}_z)$ with associated state $\bar{y} = \mathcal{S}(\bar{u})$ such that $0 \neq \mathcal{J}(\bar{y}, \bar{u}) - z \geq 0$. Then part *3* implies that there exists an adjoint state $\bar{p}$ such that $\bar{u}, \bar{y}, \bar{p}$ solve (13). Accordingly, with normalized weight α and adjoint $\tilde{p}$ given by

$$\alpha_i = \tilde{\alpha}_i / \sum_{j=1}^{2} \tilde{\alpha}_j \text{ with } \tilde{\alpha}_i = \mathcal{J}_i(\bar{y}, \bar{u}) - z_i \geq 0, \quad \tilde{p} = \bar{p} / \sum_{i=1}^{2} \tilde{\alpha}_i,$$

system (6) is also satisfied. Thus Corollary 2 implies that $\bar{u}$ is a stationary point of $(\mathbf{P}_\alpha)$ with $\alpha \geq 0$ and $\alpha_1 + \alpha_2 = 1$. The other implication follows analogously without normalization by choosing the reference point $z = \mathcal{J}(\bar{y}, \bar{u}) - \alpha$, since then $0 \neq \mathcal{J}(\bar{y}, \bar{u}) - z = \alpha \geq 0$. The cases with strict inequalities follow analogously. □

Remark 4 As a direct consequence of Corollaries 2 and 3, we get $\mathcal{R}_s^{s>} = \mathcal{W}_s^{s>} \subset \mathcal{P}_s^s$ and $\mathcal{R}_s^{s\geq} = \mathcal{W}_s^{s\geq} = \mathcal{P}_s^{sw}$, which means that the RPM is also a reasonable choice to characterize (weak) Pareto stationary points. ◊

A central question for the RPM is how suitable reference points can be chosen in the numerical implementation. To this end, we follow the approach presented in [2]. Let $k_{\max}$ denote the maximal number of Pareto stationary points in the numerical implementation and let (y^1, u^1) denote an initial starting point with u^1 being a stationary point of the weighted-sum problem with weights $\alpha_1 = 1 - \alpha_{\text{tol}}$ and $\alpha_2 = \alpha_{\text{tol}} \ll 1$. Then the first reference point z^2 (corresponding to the second point on the front) is chosen as

$$z^2 = \mathcal{J}(y^1, u^1) - \begin{pmatrix} h^\perp \\ h^\parallel \end{pmatrix}, \tag{14}$$

where $h^\perp, h^\parallel > 0$ are scaling parameters. For $i = 2, \ldots, k_{\max} - 2$ the reference point z^{i+1} is chosen as

$$z^{i+1} = \mathcal{J}(y^i, u^i) + h^\parallel \cdot \frac{\varphi^\parallel}{\|\varphi^\parallel\|} + h^\perp \cdot \frac{\varphi^\perp}{\|\varphi^\perp\|}, \tag{15}$$

with $\varphi^\perp = z^i - \mathcal{J}(y^i, u^i)$ and $\varphi^\parallel = (-\varphi_2^\perp, \varphi_1^\perp)^T$. Note that due to the strong weighting of $\mathcal{J}_1$ at (y^1, u^1), the Pareto front is approximately vertical in the area of the first reference point. This motivates the initial choice $\varphi^\parallel = (0, -1)^T$ and $\varphi^\perp = (-1, 0)^T$. Now, we can formulate the reference point method in Algorithm 2.

Algorithm 2: Reference point method (RPM)

Require: Maximal number $k_{\max} \in \mathbb{N}$ of Pareto stationary points, recursive parameters $h^{\parallel}, h^{\perp} > 0$, weighted-sum parameter $0 < \alpha_{\text{tol}} \ll 1$;
Return : Discrete approximations $\tilde{\mathcal{P}}_s^{sw}$ and $\tilde{\mathcal{P}}_f^{sw}$ of $\mathcal{P}_s^{sw}$ and $\mathcal{P}_f^{sw}$;
Compute solution (y^1, u^1) to (6) with $(1-\alpha_{\text{tol}}, \alpha_{\text{tol}})$;
Compute solution $(y_{\text{end}}, u_{\text{end}})$ to (6) with $(\alpha_{\text{tol}}, 1-\alpha_{\text{tol}})$;
Set $\tilde{\mathcal{P}}_s^{sw} = \{u^1\}$, $\tilde{\mathcal{P}}_f^{sw} = \{\hat{\mathcal{J}}(u^1)\}$ and $i = 2$;
Compute reference point z^i using (14);
while $z_1^{i+1} < \mathcal{J}_1(y_{\text{end}}, u_{\text{end}})$ **and** $i \leq k_{\max} - 1$ **do**
  Compute solution (y^i, u^i) to (13) with reference point z^i;
  Set $i = i+1$, $\tilde{\mathcal{P}}_s^{sw} = \tilde{\mathcal{P}}_s^{sw} \cup \{u^{i-1}\}$, $\tilde{\mathcal{P}}_f^{sw} = \tilde{\mathcal{P}}_f^{sw} \cup \{\hat{\mathcal{J}}(u^{i-1})\}$;
  Compute reference point z^i using (15);
Set $\tilde{\mathcal{P}}_s^{sw} = \tilde{\mathcal{P}}_s^{sw} \cup \{u_{\text{end}}\}$ and $\tilde{\mathcal{P}}_f^{sw} = \tilde{\mathcal{P}}_f^{sw} \cup \{\hat{\mathcal{J}}(u_{\text{end}})\}$;

Note that the stopping criterion implies that if $k_{\max}$ is large enough, then the upper left as well as the lower right corner points of the Pareto front coincide with those of the WSM. If $0 \neq \mathcal{J}(\mathcal{S}(\bar{u}), \bar{u}) - z \geq 0$ holds for all $\bar{u} \in \tilde{\mathcal{P}}_s^{sw}$, then the result of Algorithm 2 is a discrete approximation of the set of weak Pareto stationary points $\mathcal{P}_s^{sw}$ with the corresponding Pareto front $\mathcal{P}_f^{sw}$. If one wants to ensure this condition a priori, it is possible to, e.g., choose fixed reference points on shifted coordinate axes. The shift has to be performed such that all reference points are below the lower bounds on $\mathcal{J}_i$, cf. [3].

It should be pointed out that in both Algorithm 1 and Algorithm 2, there might be multiple solutions to (6) for one weight $\alpha = (\alpha_1, \alpha_2)$ or to (13) for one reference point z. Numerically, the gPSN method that we use to solve these systems later on can only produce one of these solutions, which can generally depend on the initial value, which in turn will typically be chosen as the previous subproblem's solution. This means that even if the number of used weights / reference points is large, the algorithms might struggle to give a good approximation of the Pareto front. Again, we generally cannot solve (13) directly in the implementation when $U = \mathbb{R}^p$ because of the variational inequality on an possibly unknown and nonlinear image set. Therefore, we will proceed as for the WSM and solve the following system.

$$\bar{u} \in U, \quad \bar{y} \in Y, \quad \bar{p} \in H, \tag{16a}$$

$$-\Delta \bar{y} + \kappa \max\{0, \bar{y}\} = \mathcal{B}(\bar{u}) \qquad \text{in } V', \tag{16b}$$

$$-\Delta \bar{p} + \kappa \mathbb{1}_{\{\bar{y}>0\}} \bar{p} = \sum_{i=1}^{2} \left(\mathcal{J}_i(\bar{y}, \bar{u}) - z_i\right) j_i'(\bar{y}) \qquad \text{in } V', \tag{16c}$$

$$\mathcal{B}^*(\bar{p}) + \sum_{i=1}^{2} \left(\mathcal{J}_i(\bar{y}, \bar{u}) - z_i\right) \sigma_i \bar{u} = 0 \qquad \text{in } U, \tag{16d}$$

cf. the modified system (10) for the WSM. For $U = L^2(\Omega)$, this again coincides with the strong stationarity system from [10], see Remark 2. For $U = \mathbb{R}^p$ we test for the sign condition

$$\bar{p} \leq 0 \quad \text{a.e. in } \{\bar{y} = 0\}, \tag{17}$$

a posteriori. By exactly the same arguments as in Corollary 3 and as a consequence of Corollary 1, the control $\bar{u}$ solving (16) generally satisfies a somewhat weaker stationarity system. When (17) is satisfied as well, $\bar{u}$ is strongly stationary and therefore a weak Pareto stationary point of (**P**).

4 Numerical Implementation

For the numerical realization and tests of the algorithms, we will assume that $j_1(y) = \frac{1}{2}\|y - y^d\|_H^2$, $j_2(y) = 0$ and $\sigma_1 = 0$. We fix the domain $\Omega = (0,1)^2$ and consider P_1-type finite elements (FE) on a Friedrichs-Keller triangulation of the domain. The measure of fineness of the grids will be $h > 0$, which denotes the inverse number of square cells per dimension – i.e., the grid will have $2/h^2$ triangles. We write the coefficient vector of the piecewise linear interpolant of a function $w\colon \Omega \to \mathbb{R}$ on the grid vertices in typewriter font (i.e., $\mathtt{w} \in \mathbb{R}^N$) and use the same font for the matrices in the discretized settings. We resort to mass lumping for the nonlinear max-term in order to be able to evaluate it componentwisely. Inevitably, this introduces a numerical discretization error. Its effects decrease with increasing fineness of the discretization but increase with the coefficient κ that scales the nonlinearity. The corresponding stiffness matrix $\mathsf{K} \in \mathbb{R}^{N\times N}$, mass matrix $\mathsf{M} \in \mathbb{R}^{N\times N}$ and lumped mass matrix $\tilde{\mathsf{M}} \in \mathbb{R}^{N\times N}$ are given from the FE ansatz functions φ_i, $i = 1, \dots, N$ as

$$\mathsf{K}_{ij} = \langle \nabla\varphi_i, \nabla\varphi_j\rangle_H,\ \mathsf{M}_{ij} = \langle \varphi_i, \varphi_j\rangle_H,\ \tilde{\mathsf{M}} = \operatorname{diag}\left(\frac{|\operatorname{supp}(\varphi_i)|}{3} : i = 1, \dots, N\right).$$

Thus the FE approximation of (10c)-(10e) introduced for the WSM is

$$\begin{pmatrix} \mathsf{K}\bar{\mathsf{y}}_h + \kappa\tilde{\mathsf{M}}\max\{0, \bar{\mathsf{y}}_h\} - \mathsf{B}\bar{u} \\ \mathsf{K}\bar{\mathsf{p}}_h + \kappa\tilde{\mathsf{M}}\Theta(\bar{\mathsf{y}}_h)\bar{\mathsf{p}}_h - \alpha_1\mathsf{M}(\bar{\mathsf{y}}_h - \mathsf{y}^d) \\ \mathsf{B}^T\bar{\mathsf{p}}_h + \alpha_2\sigma_2\mathsf{A}\bar{u} \end{pmatrix} = 0 \tag{18}$$

for some given $\alpha \in \mathbb{R}^2$ that satisfies (10a)-(10b), where B is the FE-discretized version of the linear operator $\mathcal{B}$ and $\Theta \coloneqq \Theta_0$ where $\Theta_x\colon \mathbb{R}^N \to \mathbb{R}^{N\times N}$ maps a vector to the diagonal matrix that takes the Heaviside function with functional value x at 0 evaluated for each entry of the vector as its diagonal entries. The matrix $\mathsf{A} = \mathsf{I}_p \in \mathbb{R}^{p\times p}$ is the idendity if $U = \mathbb{R}^p$, and $\mathsf{A} = \mathsf{M}$ is the mass matrix if $U = H$. Note that this means that, depending on the space U, sometimes typewrite notation would be appropriate for the (discretized) control u. To avoid any misunderstandings, we will always denote u without typewriter style. These finite dimensional systems are solved with a globalized version of a pseudo semismooth Newton (PSN) method. For more details on

the PSN without globalization, we refer to [8,10]. The FE system matrix at iterates $(\mathsf{y}_h, \mathsf{p}_h, u)$ reads as:

$$\begin{pmatrix} \mathsf{K} + \kappa\tilde{\mathsf{M}}\Theta(\mathsf{y}_h) & 0 & -\mathsf{B} \\ -\alpha_1\mathsf{M} & \mathsf{K} + \kappa\tilde{\mathsf{M}}\Theta(\mathsf{y}_h) & 0 \\ 0 & \mathsf{B}^T & \alpha_2\sigma_2\mathsf{A} \end{pmatrix}. \tag{19}$$

We proceed analogously for the RPM and discretize (16b)-(16d) using finite elements, which yields

$$\begin{pmatrix} \mathsf{K}\bar{\mathsf{y}}_h + \kappa\tilde{\mathsf{M}}\max\{0, \bar{\mathsf{y}}_h\} - \mathsf{B}\bar{u} \\ \mathsf{K}\bar{\mathsf{p}}_h + \kappa\tilde{\mathsf{M}}\Theta(\bar{\mathsf{y}}_h)\bar{\mathsf{p}}_h - \left(\frac{1}{2}(\bar{\mathsf{y}}_h - \mathsf{y}^d)^T\mathsf{M}(\bar{\mathsf{y}}_h - \mathsf{y}^d) - z_1\right)\mathsf{M}(\bar{\mathsf{y}}_h - \mathsf{y}^d) \\ \mathsf{B}^T\bar{\mathsf{p}}_h + \left(\frac{\sigma_2}{2}\bar{u}^T\mathsf{A}\bar{u} - z_2\right)\sigma_2\mathsf{A}\bar{u} \end{pmatrix} = 0. \tag{20}$$

These discretized systems are solved with a PSN method as well. The FE system matrix at iterates $(\mathsf{y}_h, \mathsf{p}_h, u)$ reads as:

$$\begin{pmatrix} \mathsf{K} + \kappa\tilde{\mathsf{M}}\Theta(\mathsf{y}_h) & 0 & -\mathsf{B} \\ \mathsf{C}(\mathsf{y}_h) & \mathsf{K} + \kappa\tilde{\mathsf{M}}\Theta(\mathsf{y}_h) & 0 \\ 0 & \mathsf{B}^T & \mathsf{D}(u) \end{pmatrix} \tag{21}$$

with

$$\mathsf{C}(\mathsf{y}_h) \coloneqq (\mathsf{M}(\mathsf{y}^d - \mathsf{y}_h))(\mathsf{M}(\mathsf{y}_h - \mathsf{y}^d))^T - \left(\tfrac{1}{2}(\mathsf{y}_h - \mathsf{y}^d)^T\mathsf{M}(\mathsf{y}_h - \mathsf{y}^d) - z_1\right)\mathsf{M}, \tag{22}$$

$$\mathsf{D}(u) \coloneqq (\sigma_2\mathsf{A}u)(\sigma_2\mathsf{A}u)^T + \left(\frac{\sigma_2}{2}u^T\mathsf{A}u - z_2\right)\sigma_2\mathsf{A}. \tag{23}$$

Remark 5 We again point out that, for both methods, the sign conditions (11) and (17) are not added into the discretized stationarity system. Instead, they are verified a posteriori if $U = \mathbb{R}^p$. ♢

Compared to the system matrix of the single objective case presented in [8], especially the system matrix of the RPM is more complicated and possesses a non-sparse substructure. This is due to the matrices $\mathsf{C}(\mathsf{y}_h)$ and $\mathsf{D}(u)$, which posses the dense terms $\mathsf{M}(\mathsf{y}_h - \mathsf{y}^d)(\mathsf{M}(\mathsf{y}_h - \mathsf{y}^d))^T$ and $(\sigma_2\mathsf{A}u)(\sigma_2\mathsf{A}u)^T$. These can cause severe memory and runtime problems when the reference point problem is solved on fine finite element grids with a linear solver. Due to these restrictions, the reference point method's subproblems will generally take longer to solve than the WSM, also on coarser grids.

One thing to keep in mind when applying the PSN method is that there is no guarantee of convergence and as seen in the numerical examples in [10]. The method in fact shows failure to converge in practice, with the rate of failed attempts over the subproblems decaying as the grid discretization's fineness is increased. This suggest some sort of degeneration of the undamped search directions that could be countered with a globalization mechanism.

Accordingly, the two questions that we will address in the remainder of this section are the following:

1) Can the non-convergence issue of the PSN method itself be removed?
2) Is it possible to reduce computation times and memory problems of the (linear) PSN steps in the reference point method to make it competitive in terms of computation times?

4.1 Globalized PSN

The numerical experiments in [10] indicate that non-convergence of the PSN is an issue that strongly depends on (insufficiently fine) discretizations. As a stabilization approach independently of the grid fineness, we will present and test a line search globalization of the PSN based on results for semismooth Newton methods as in, e.g., [13] and [15], which will be referred to as the gPSN method. Let us assume that we want to find a root of a function $F : \mathbb{R}^{2N+p} \to \mathbb{R}^{2N+p}$ with system matrix $G : \mathbb{R}^{2N+p} \to \mathbb{R}^{(2N+p)\times(2N+p)}$. This will either be (18) with system matrix (19) for the subproblems in the WSM or (20) with system matrix (21) for the subproblems in the RPM. We will employ a line search globalization with the merit function $\Lambda : \mathbb{R}^{2N+p} \to \mathbb{R}$, $x \mapsto \frac{1}{2}\|F(x)\|^2$.

Remark 6 Note that the norm in the merit function Λ is the discrete equivalence of the norm in $V' \times V' \times U$. Hence it is generally expensive to evaluate, since we need to compute a Riesz representative. However, we will precompute the necessary factorizations to speed-up the computations to some extend. $\Diamond$

The gPSN method is summarized in the following algorithm. Note that we cannot conclude – e.g., from theory on globalized semismooth Newton methods – that the algorithm converges without introducing a maximum number $k_{\max}$ of PSN steps and a minimum step length $\epsilon_2 > 0$.

Algorithm 3: Globalized PSN Method (gPSN)

Require: Initial point $x^0 \in \mathbb{R}^{2N+p}$, tolerances $\epsilon_1, \epsilon_2 > 0$, maximum number of iterations $k_{\max} \in \mathbb{N}$ and line search parameters $\beta \in (0,1)$, $\gamma \in (0, \frac{1}{2})$;
Return : Approximated root $\bar{x} \in \mathbb{R}^{2N+p}$;
Set $i = 0$;
while $\sqrt{2\Lambda(x^i)} > \epsilon_1$ **and** $i < k_{\max}$ **do**
  Compute line search direction d^i by solving $G(x^i)d^i = -F(x^i)$;
  Set $k_i = 0$;
  while $\Lambda(x^i + \beta^{k_i} d^i) > (1 - 2\gamma\beta^{k_i})\Lambda(x^i)$ **and** $\beta^{k_i+1} > \epsilon_2$ **do**
    Set $k_i = k_i + 1$;
  Set $x^{i+1} = x^i + \beta^{k_i} d^i$ and $i = i + 1$;

It turns out that there are examples, where Algorithm 3 converges while a refinement of the grid does not yield convergence, see Section 5. However, it can still happen that the gPSN method does not converge. Obviously, it would

not be a good idea to add the final iterate of the gPSN method to the Pareto set nonetheless. Instead, if within the RPM the PSN does not converge, we update the reference point as follows: If no previous solution to the reference point problem is available, we choose

$$z^{i+1} = z^i - \begin{pmatrix} 0 \\ h^{\|} \end{pmatrix}.$$

If previous solutions to the reference point problem are available, we choose

$$z^{i+1} = z^i + h^{\|} \frac{\varphi^{\|}}{\|\varphi^{\|}\|}.$$

Essentially, previous information is used repeatedly to find a new reference point by going into the same parallel direction. If the gPSN is used within the WSM and does not converge, we can simply proceed to the next discretized weight.

4.2 A Preconditioned Matrix-Free L-GMRES Method

We will now focus on how to speed-up the computation and how to overcome the difficulties arising from the dense terms in the RPM. Notice that both dense terms are rank-1-matrices. Therefore it is easy to implement the matrix-vector-product for some $w \in \mathbb{R}^N$ and $v \in \mathbb{R}^N$ with $v = \mathsf{M}(\mathsf{y}_h - \mathsf{y}^d)$ or $v = \sigma_2 \mathsf{A} u$:

$$\left(vv^T\right) w = \left(v^T w\right) v.$$

This motivates the use of an iterative solver that only relies on matrix-vector-products in each PSN step for solving the reference point subproblem. Since the system is not symmetric positive definite, the CG method is not an alternative and we will use L-GMRES instead. As the performance heavily depends on the condition number of the system matrix (see [24]), which might be very large, especially for very small values of the regularization parameter σ_2, we will precondition the method with one of the following preconditioners:

- **a** The dense terms $\mathsf{M}(\mathsf{y}_h - \mathsf{y}^d)(\mathsf{M}(\mathsf{y}_h - \mathsf{y}^d))^T$ and $(\sigma_2 \mathsf{A} u)(\sigma_2 \mathsf{A} u)^T$ are omitted in an approximated system matrix that is used as a preconditioner.
- **aBJ** A block Jacobi preconditioner is applied with the approximation described for the preconditioner **a**.
- **aBGS** A block Gauss-Seidel preconditioner is applied with the approximation described for the preconditioner **a**.
- **aILU** An incomplete LU factorization together with the approximation described for the preconditioner **a** is applied.

Of course the same iterative, preconditioned approach can be implemented for the WSM, with the only difference being that no approximation – by ignoring dense terms – is necessary for the preconditioner. Note that also standard Jacobi, Gauss-Seidel, block Jacobi and block Gauss-Seidel and incomplete LU

factorization preconditioners were tested. But the first two did not give any speed-up and the last four were still significantly less effective than the preconditioners above due to the dense terms still remaining. As expected, it is quite important to use the block structure of the problem as much as possible and to avoid the dense terms.

Remark 7 As long as $\frac{\sigma_2}{2}u^T\mathsf{A}u - z \neq 0$, the invertibility of the **aBJ** preconditioner is ensured for the RPM. In case of the WSM, the invertibility is always ensured.

For the four different preconditioners above, we propose three different update strategies. Those strategies are:

Never update Only one preconditioner is generated for the first iteration of the first subproblem and then this preconditioner is used for all subproblems and all gPSN iterations.

Update once One preconditioner is generated for each subproblem and then used for all gPSN iterations.

Always update The preconditioner is generated for each gPSN iteration of each subproblem.

5 Numerical Examples

In this section, we present numerical results for two examples – one with finite and one with infinite dimensional control space. First, the focus of our exposition will be on the performance of the RPM and the different preconditioning strategies. After the best update strategy is identified, we will compare RPM and WSM method.

In order to reasonably quantify the quality of the approximation of the respective Pareto (stationary) fronts, we employ two quality measures. The maximal distance between neighboring points on the Pareto front

$$\Delta_{\max} := \max_{a\in\tilde{\mathcal{P}}_f} \min_{b\in\tilde{\mathcal{P}}_f\setminus\{a\}} \|a-b\|_2 \tag{24}$$

will be our first measure. As a second measure of approximation quality, we will consider

$$\Delta_{\text{clust}} = \frac{|\tilde{\mathcal{P}}_f|\Delta_{\max}}{\sum\limits_{a\in\tilde{\mathcal{P}}_f} \min\limits_{b\in\tilde{\mathcal{P}}_f\setminus\{a\}} \|a-b\|_2}, \tag{25}$$

which is the maximum shortest distance between points on the front devided by the average shortest distance and therefore bounded from below by one. If this quantity is small, this indicates that the approximation quality is somewhat uniform across the entire Pareto front, while a large value indicates that some parts of the Pareto front are approximated better than others are, i.e., a localized clustering.

Table 1 Fixed parameters for the two numerical examples

gPSN			RPM		PDE	
β	γ	ϵ_2	k_{max}	α_{tol}	Ω	κ
$5 \cdot 10^{-1}$	10^{-1}	10^{-3}	10^3	10^{-2}	$(0,1)^2$	10

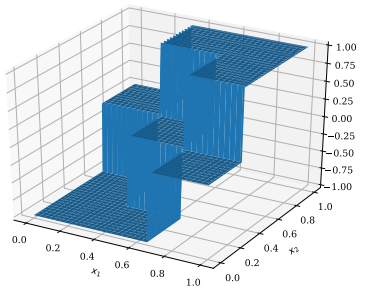

Fig. 1 Example 1. Desired state $y^d(\boldsymbol{x})$ for step size $1/h = 100$

Note that in all results presented here, the sign condition (see Remark 5) is satisfied and the gPSN method always converges.
Our code is implemented in Python3 and uses FEniCS [1] for the matrix assembly. Sparse memory management and computations (especially L-GMRES) are implemented with SciPy [26]. All computations below were run on an Ubuntu 20.04 notebook with 32 GB main memory and an Intel Core i7-8565U CPU.

5.1 The Numerical Examples

First we introduce the two examples. The parameters listed in Table 1 are fixed for the rest of this work.

Example 1 – Infinite Dimensional Controls

For the first numerical example the desired state is chosen as $y^d = \mathbb{1}_{\Omega_1} - \mathbb{1}_{\Omega_2}$ with

$$\Omega_1 = \left\{(x_1, x_2) \in \Omega : x_1, x_2 > \frac{1}{3}\right\}, \quad \Omega_2 = \left\{(x_1, x_2) \in \Omega : x_1, x_2 < \frac{2}{3}\right\}.$$

This desired state is shown in Figure 1. The control space U is chosen as $H = L^2(\Omega)$, the operator $\mathcal{B}$ as the identity on U and A is the mass matrix.

Example 2 – Finite Dimensional Controls

For the second numerical example, we choose $y^d(x) = \left(\frac{1}{2} - x_1\right) \sin(\pi x_1) \sin(\pi x_2)$. The space U is chosen as $\mathbb{R}^2$, $\mathsf{A} = \mathsf{I}_2$ is the idendity matrix in $\mathbb{R}^{2\times 2}$ and the

Table 2 Example 1. Comparison of average L-GMRES iterations (av. it.) and speed-up (s.-up) of the RPM for different preconditioning approaches and different step sizes h for $\sigma_2 = 5 \cdot 10^{-3}$

$1/h$	**a**		**aBJ**		**aBGS**		**aILU**		**none**	
	av. it.	s.-up	av. it.	**s.-up**	av. it.	s.-up	av. it.	s.-up	av. it.	time [s]
Never update										
50	3.61	8.82	3.17	23.51	3.24	16.49	3.57	9.19	83.65	$2.55 \cdot 10^2$
100	3.53	14.64	3.36	42.04	3.21	26.53	3.44	17.62	257.46	$2.02 \cdot 10^3$
200	3.51	20.65	3.38	63.46	3.23	35.28	3.53	23.07	380.42	$1.14 \cdot 10^4$
Update once										
50	2.55	10.37	2.92	17.71	2.88	10.01	2.10	7.92		
100	2.67	12.01	3.02	30.51	2.92	11.10	2.12	10.27		
200	2.85	9.40	3.19	39.75	2.98	8.04	2.34	8.20		
Always update										
50	2.35	7.75	2.93	15.45	2.95	7.43	2.00	5.58		
100	2.43	9.04	3.02	26.94	2.99	8.29	2.03	7.32		
200	2.62	7.20	3.15	35.54	3.05	6.29	2.32	6.28		

operator $\mathcal{B}$ is set to

$$(\mathcal{B}(u))(\boldsymbol{x}) = 10 \begin{cases} u_1 x_1 x_2, & \text{for } \boldsymbol{x} = (x_1, x_2) \text{ and } x_1 \leq \frac{1}{2}, \\ u_2 x_1^2 x_2^2, & \text{otherwise,} \end{cases}$$

where the definition is to be understood as L^2-functions mapping $\boldsymbol{x} \in \Omega$ to $\mathbb{R}$ that are embedded into V'. For plots of the operator $\mathcal{B}$ and the desired state we refer to [8, Figures 1 and 4].

5.2 Preconditioning the Reference Point Method

In this section, we consider the different preconditioning approaches for the RPM. Therefore, we additionally choose the parameter $\sigma_2 = 5 \cdot 10^{-3}$ and the parameters $h^{\perp} = 10$, $h^{\parallel} = 0.1$ in the RPM and $\epsilon_1 = 1 \cdot 10^{-4}$ for the gPSN. Note that the arguably large value of ϵ_1 is necessary for the L-GMRES method to converge without preconditioner, because the (sometimes badly conditioned) problem is numerically difficult to solve. The results for Example 1 are given in Table 2. First of all, we can see that the computation time decreases for all preconditioning approaches. Also the average number of L-GMRES iterations is very small (between 2 and 3.5) and increases only slightly for smaller step sizes h. The latter observation is in contrast to the performance of the non-preconditioned solving, which starts out with a large number of average L-GMRES iterations that nonetheless significantly increases for smaller step sizes h. Furthermore, we can see that cheaper preconditioners lead to a larger speed-up if the preconditioner is updated more often, i.e. with the preconditioning strategy **always update** the preconditioner **aBJ** still gives a significant speed-up of about 35, but all other preconditioners cannot give a speed-up above 9. Nonetheless, the best preconditioning approach surprisingly is the **never update** strategy combined with the preconditioner **aBJ**. This indicates that the problem structure does not change significantly with respect

Table 3 Example 2. Comparison of average L-GMRES iterations (av. it.) and speed-up (s.-up) of the RPM for different preconditioning approaches and different step sizes h for $\sigma_2 = 5 \cdot 10^{-3}$

$1/h$	**a**		**aBJ**		**aBGS**		**aILU**		**none**	
	av. it.	s.-up	av. it.	**s.-up**	av. it.	s.-up	av. it.	s.-up	av. it.	time [s]
Never update										
50	2.79	3.43	2.78	4.16	2.95	3.21	2.79	3.37	9.97	$1.79 \cdot 10^1$
100	2.78	6.46	2.62	11.18	2.81	5.54	7.00	0.55	40.50	$1.91 \cdot 10^2$
200	2.72	14.73	2.38	59.41	2.67	12.83	15.23	0.98	197.35	$3.85 \cdot 10^3$
Update once										
50	2.10	0.94	2.19	2.87	2.05	1.20	2.04	0.70		
100	2.11	0.67	2.20	7.19	2.08	1.00	6.01	0.40		
200	2.18	0.53	2.16	30.25	2.12	0.70	12.95	0.74		
Always update										
50	2.05	0.67	2.17	2.36	2.05	0.88	2.17	0.48		
100	2.09	0.52	2.20	6.08	2.07	0.74	5.99	0.35		
200	2.15	0.41	2.16	25.44	2.11	0.53	12.95	0.65		

to the current reference point and thus a computationally expensive update of the preconditioner is unnecessary.

Next, we consider the results for Example 2, which can be found in Table 3. We basically observe the same behavior as previously and again **never update** and **aBJ** is the best preconditioning approach. Note that this finite dimensional example is inherently better conditioned and thus combinations of more expensive preconditioners such as **a**, **aBGS** and **aILU** and expensive preconditioning strategies such as **update once** and **always update** often lead to larger computation times compared to the performance in the absence of a preconditioner. Furthermore, the preconditioner **aILU** seems to behave unstably, since the number of average L-GMRES iterations increases significantly for smaller step sizes h. Note that **aILU** includes some parameters which could be varied and might improve this behavior, but we will not go into details here.

Next we consider a fixed step size $h = 1/100$ and investigate the behavior of the different preconditioning approaches for varying σ_2. We expect larger condition numbers for smaller values of σ_2 and therefore problems which are harder to solve numerically. Since the strategies **update once** and **always update** and the preconditioner **aILU** did not prove useful, they are excluded from this considerations. The results can be found in Table 4 for Example 1 and in Table 5 for Example 2. In both examples the average number of L-GMRES iterations increases slightly for smaller values of σ_2. This increase is stronger for the first example. If, on the other hand, no preconditioner is used, there is a significant increase for smaller values of σ_2. This is especially true for the first example, which starts with an average number of 18.75 iterations for $\sigma_2 = 1$ and ends with an average number of 501.07 iterations for $\sigma_2 = 10^{-3}$. In the second example there is only an increase from 10.48 to 52.10. Nonetheless in both examples preconditioning pays off and again the preconditioner **aBJ** is the best. It results in a speed-up of 62.84 in the first example and a speed-up of 12.64 in the second example for $\sigma_2 = 10^{-3}$.

Table 4 Example 1. Comparison of average L-GMRES iterations (av. it.) and speed-up (s.-up) of the RPM for different preconditioning approaches and different values of σ_2 for fixed step size $h = 1/100$

σ_2	**a**		**aBJ**		**aBGS**		**none**	
	av. it.	s.-up	av. it.	**s.-up**	av. it.	s.-up	av. it.	time [s]
10^0	2.50	2.95	2.50	6.60	2.86	3.57	18.75	$6.44 \cdot 10^{-1}$
10^{-1}	3.47	5.82	2.98	14.46	3.12	8.59	57.65	$2.34 \cdot 10^{0}$
10^{-2}	3.37	15.08	3.27	40.73	3.12	25.18	201.42	$8.92 \cdot 10^{0}$
10^{-3}	3.53	19.46	3.46	62.84	3.28	40.75	501.07	$1.73 \cdot 10^{1}$

Table 5 Example 2. Comparison of average L-GMRES iterations (av. it.) and speed-up (s.-up) of the RPM for different preconditioning approaches and different values of σ_2 for fixed step size $h = 1/100$

σ_2	**a**		**aBJ**		**aBGS**		**none**	
	av. it.	s.-up	av. it.	**s.-up**	av. it.	s.-up	av. it.	time [s]
10^0	2.53	1.35	2.30	3.71	2.63	1.61	10.48	$2.46 \cdot 10^{-1}$
10^{-1}	2.61	2.03	2.21	5.30	2.11	2.90	13.41	$4.50 \cdot 10^{-1}$
10^{-2}	2.75	5.13	2.59	9.24	2.71	4.60	28.12	$8.91 \cdot 10^{-1}$
10^{-3}	2.68	9.68	2.88	12.64	2.93	6.78	52.10	$1.41 \cdot 10^{0}$

5.3 Comparison of the RPM and the WSM

In this section, we want to compare the performance of the reference point method and the weighted-sum method both in terms of computation times and discretization quality. We choose a step size of $h = 1/100$, a tolerance $\epsilon_1 = 10^{-5}$ in the gPSN and $h^{\perp} = 1$, $h^{\parallel} = 0.2$ in the RPM. We will consider $\sigma_2 = 1$ for both examples. This means that the problems are relatively well conditioned, but the Pareto fronts are harder to approximate than for smaller values of σ_2.

In order to make the results of the RPM and the WSM qualitatively comparable, we first run the RPM, which yields a number of discretization points on the front. Afterwards we run the WSM with $k_{\max}$ (the number of Pareto points) chosen as the number of discretization points generated by the RPM. At this point we have the same number of discretization points on the respective approximated fronts. However, the WSM tends to cluster the discretization points. In order to obtain comparable approximation quality, we then double the number of points in the WSM until the maximal distance for points on the Pareto front (see (24)) is below the maximal distance for the RPM. Afterwards, the parameter $h^{\parallel}$ is halved as often as $k_{\max}$ in the WSM was doubled before to compare the evolution of the quality measures $\Delta_{\max}$ and Δ_{clust} for an increasing size of the approximated Pareto front.

The Pareto fronts are shown in Figure 2. We can see that both methods approximate the same curve. But the WSM shows a clustering behavior in the lower right corner whilst giving only a poor approximation in the remainder of the Pareto front. This already indicates that some refinement of the weights'

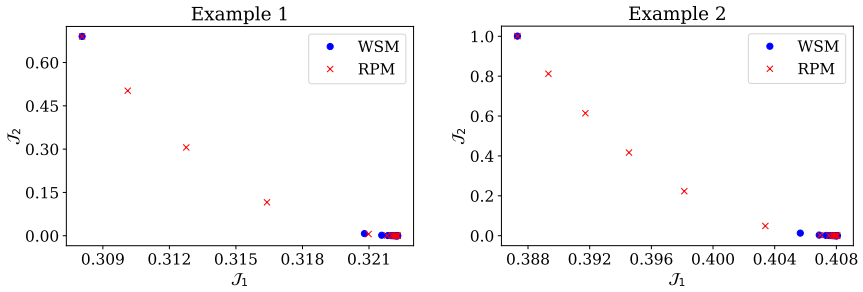

Fig. 2 Pareto fronts from WSM and RPM for step size $1/h = 100$ and $\sigma_2 = 1$

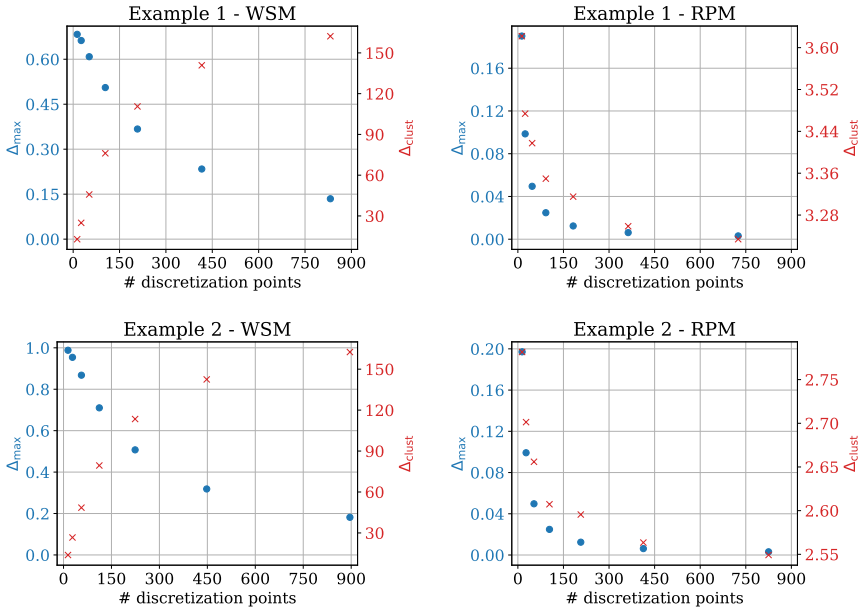

Fig. 3 Evolution of measures of approximation quality ($\Delta_{\max}$ and Δ_{clust}) for the WSM with respect to doubling the number of discretization points on the Pareto front and for the RPM with respect to halving $h^{\|}$ with $\sigma_2 = 1$ and step size $1/h = 100$

distribution is generally well advised for the WSM. In Figure 3, the evolution of the quality measures for the WSM and RPM for the procedure described above is shown. In the left figures, we can see that for the WSM the number of points on the Pareto front needs to be doubled seven times in order to reach a maximal distance of points on the Pareto front that is smaller than that of the RPM. Furthermore the approximation quality decreases every time the size of the Pareto front is doubled. This is due to the clustering behavior in the lower right corner. As a result, an unnecessarily large number of points on the Pareto front is needed to reach a desired maximal distance $\Delta_{\max}$. On the

Table 6 Comparison of computation time and approximation quality for RPM and WSM with $\sigma_2 = 1$. WSM (first) indicates WSM with the size of RPM. WSM (last) indicates WSM after doubling the size until a maximal distance on the Pareto front below that of RPM is reached. The step size is chosen as $h = 1/100$

		time [s]	$\Delta_{\max}$	Δ_{clust}	$\lvert\tilde{\mathcal{P}}_f\rvert$
Ex 1	**RPM**	$2.42 \cdot 10^{0}$	$1.90 \cdot 10^{-1}$	$3.62 \cdot 10^{0}$	13
	WSM (first)	$1.41 \cdot 10^{0}$	$6.83 \cdot 10^{-1}$	$1.28 \cdot 10^{1}$	13
	WSM (last)	$6.99 \cdot 10^{1}$	$1.34 \cdot 10^{-1}$	$1.62 \cdot 10^{2}$	832
Ex 2	**RPM**	$1.61 \cdot 10^{0}$	$1.97 \cdot 10^{-1}$	$2.78 \cdot 10^{0}$	14
	WSM (first)	$1.04 \cdot 10^{0}$	$9.88 \cdot 10^{-1}$	$1.38 \cdot 10^{1}$	14
	WSM (last)	$4.49 \cdot 10^{1}$	$1.81 \cdot 10^{-1}$	$1.62 \cdot 10^{2}$	896

other hand, with the RPM, the approximation quality even decreases whilst $h^{\|}$ is halved. This behavior can be seen in the right figures and is even better than expected. Note that the size of the Pareto front is approximately doubled when $h^{\|}$ is halved.
The question remains, which method performs better in terms of computational cost. A comparison of the results from the RPM and the first and last result from the WSM in the procedure of doubling the number of discretization points is shown in Table 6.

The observation for both examples are similar with respect to sizes of the Pareto fronts and the measures of approximation quality. Also whilst the RPM is slightly slower than the WSM if the same size for the Pareto front is used, it is about 28 times faster than the WSM when an at least equally good approximation quality is desired.

6 Conclusion

If the controls on the right-hand-side of the constraining PDE to (**P**) are finite dimensional, then stationarity conditions that are equivalent to primal stationarity can be found. They are, however, not usable numerically because they contain unknown nonlinearities in the spaces that the conditions are formulated in. Modifying the conditions, we end up with linear systems as in the case of ample controls and a weaker stationarity sense that can be useful in numerical computation. We have shown that both WSM and RPM can be applied to characterize the front of Pareto stationary points for this nonsmooth problem. The reference point method performs significantly better when both approximation quality and computation time are considered, as long as preconditioning is used intelligently in GMRES. In our tests, the preconditioning strategy **aBJ** without updates performs the best. We also saw that the line search globalized version of the PSN method leads to better performance and convergence of the method. A reduction of the step size does not seem to guarantee this numerically.

References

1. Alnæs, M., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M., Wells, G.: The FEniCS Project Version 1.5. Archive of Numerical Software **3**(100) (2015)
2. Banholzer, S.: POD-Based Bicriterial Optimal Control of Convection-Diffusion Equations. Master thesis, University of Konstanz (2017). See `http://nbn-resolving.de/urn:nbn:de:bsz:352-0-421545`
3. Banholzer, S.: ROM-based multiobjective optimization with PDE constraints. Ph.D. thesis, University of Konstanz (2021)
4. Banholzer, S., Beermann, D., Volkwein, S.: POD-based error control for reduced-order bicriterial PDE-constrained optimization. Annual Reviews in Control **44**, 226–237 (2017)
5. Banholzer, S., Volkwein, S.: Hierarchical convex multiobjective optimization by the Euclidean reference point method (2019). See `http://nbn-resolving.de/urn:nbn:de:bsz:352-2-xd965ctqkqax3`
6. Bernreuther, M.: RB-Based PDE-Constrained Non-Smooth Optimization. Master thesis, University of Konstanz (2019). See `http://nbn-resolving.de/urn:nbn:de:bsz:352-2-t4k1djyj77yn3`
7. Bernreuther, M., Müller, G., Volkwein, S.: Reduced basis model order reduction in optimal control of a non-smooth semilinear elliptic PDE. To appear in: Optimization and Control for Partial Differential Equations, Radon Series on Computational and Applied Mathematics. De Gruyter (2020). See `http://nbn-resolving.de/urn:nbn:de:bsz:352-2-1brmrl3906b9p5`
8. Bernreuther, M., Müller, G., Volkwein, S.: Reduced basis model order reduction in optimal control of a nonsmooth semilinear elliptic PDE. To appear in: New Trends in PDE Constrained Optimization (2020)
9. Christof, C., Clason, C., Meyer, C., Walther, S.: Optimal control of a non-smooth semilinear elliptic equation. Mathematical Control and Related Fields (MCRF) **8**, 247–276 (2018)
10. Christof, C., Müller, G.: Multiobjective Optimal Control of a Non-Smooth Semilinear Elliptic Partial Differential Equation. To appar in ESAIM: Control, Optimisation and Calculus of Variations (2020). See `https://spp1962.wias-berlin.de/preprints/130.pdf`
11. Clarke, F.H.: Optimization and Nonsmooth Analysis. Classics in Applied Mathematics. SIAM, Philadelphia (1990)
12. Ehrgott, M.: Multicriteria Optimization. Lecture notes in economics and mathematical systems. Springer-Verlag Berlin Heidelberg (2005)
13. Gerdts, M., Horn, S., Kimmerle, S.J.: Line search globalization of a semismooth Newton method for operator equations in Hilbert spaces with applications in optimal control. Journal of Industrial and Management Optimization **13**, 47–62 (2016)
14. Gilbarg, D., Trudinger, N.: Elliptic Partial Differential Equations of Second Order. Classics in Mathematics. Springer, Berlin, Heidelberg (2001)
15. Ito, K., Kunisch, K.: On a semi-smooth newton method and its globalization. Mathematical Programming **118**, 347–370 (2009)
16. Khoromskij, B.N., Wittum, G.: Numerical Solution of Elliptic Differential Equations by Reduction to the Interface. Lecture Notes in Computational Science and Engineering. Springer-Verlag Berlin Heidelberg (2004)
17. Kikuchi, F., Nakazato, K., Ushijima, T.: Finite element approximation of a nonlinear eigenvalue problem related to MHD equilibria. Japan Journal of Applied Mathematics **1**(2), 369–403 (1984)
18. Miettinen, K.: Nonlinear Multiobjective Optimization. International Series in Operations Research & Management Science. Springer New York (1998)
19. Mueller-Gritschneder, D., Graeb, H., Schlichtmann, U.: A successive approach to compute the bounded Pareto front of practical multiobjective optimization problems. SIAM J. Optim. **20**, 915–934 (2009)
20. Nocedal, J., Wright, S.: Numerical Optimization. Springer Series in Operations Research and Financial Engineering. Springer New York (2000).

21. Rappaz, J.: Approximation of a nondifferentiable nonlinear problem related to MHD equilibria. Numerische Mathematik **45**(1), 117–133 (1984)
22. Rockafellar, R., Wets, R.: Variational Analysis. Springer-Verlag Berlin Heidelberg (2004)
23. Romaus, C., Bocker, J., Witting, K., Seifried, A., Znamenshchykov, O.: Optimal energy management for a hybrid energy storage system combining batteries and double layer capacitors. In: 2009 IEEE Energy Conversion Congress and Exposition, pp. 1640–1647 (2009)
24. Saad, Y.: Iterative Methods for Sparse Linear Systems, SIAM Philadelphia (2003).
25. Temam, R.: A non-linear eigenvalue problem: the shape at equilibrium of a confined plasma. Arch. Rational Mech. Anal. **60**(1), 51–73 (1976)
26. Virtanen, P., Gommers, R., Oliphant, T., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E., Harris, C., Archibald, A., Ribeiro, A., Pedregosa, F., van Mulbregt, P., Contributors, S...: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods **17**, 261–272 (2020)
27. Xin, J.: An Introduction to Fronts in Random Media, *Surveys and Tutorials in the Applied Mathematical Sciences*, vol. 5. Springer (2009)