MULTIOBJECTIVE OPTIMAL CONTROL OF A NON-SMOOTH SEMILINEAR ELLIPTIC PARTIAL DIFFERENTIAL EQUATION[*]

CONSTANTIN CHRISTOF[1,**] AND GEORG MÜLLER[2]

Abstract. This paper is concerned with the derivation and analysis of first-order necessary optimality conditions for a class of multiobjective optimal control problems governed by an elliptic non-smooth semilinear partial differential equation. Using an adjoint calculus for the inverse of the non-linear and non-differentiable directional derivative of the solution map of the considered PDE, we extend the concept of strong stationarity to the multiobjective setting and demonstrate that the properties of weak and proper Pareto stationarity can also be characterized by suitable multiplier systems that involve both primal and dual quantities. The established optimality conditions imply in particular that Pareto stationary points possess additional regularity properties and that mollification approaches are – in a certain sense – exact for the studied problem class. We further show that the obtained results are closely related to rather peculiar hidden regularization effects that only reveal themselves when the control is eliminated and the problem is reduced to the state. This observation is also new for the case of a single objective function. The paper concludes with numerical experiments that illustrate that the derived optimality systems are amenable to numerical solution procedures.

---

[1] Technische Universität München, Chair of Optimal Control, Center for Mathematical Sciences, M17, Boltzmannstraße 3, 85748 Garching, Germany.
[2] Universität Konstanz, Department of Mathematics and Statistics, WG Numerical Optimization, Universitätsstraße 10, 78457 Konstanz, Germany.

[**] Corresponding author: christof@ma.tum.de

1. Introduction

The aim of this paper is to study first-order necessary optimality conditions for multiobjective optimal control problems of the form

$$\text{Minimize} \quad \begin{pmatrix} J_1(y,u) := j_1(y) + \dfrac{\nu_1}{2}\|u\|_{L^2}^2 \\ \vdots \\ J_N(y,u) := j_N(y) + \dfrac{\nu_N}{2}\|u\|_{L^2}^2 \end{pmatrix} \tag{P}$$

$$\text{w.r.t.} \quad u \in L^2(\Omega), \quad y \in H_0^1(\Omega) \cap H^2(\Omega),$$

$$\text{s.t.} \quad -\Delta y + \max(0, y) = u \text{ a.e. in } \Omega.$$

Here, $\Omega \subset \mathbb{R}^d$, $d \geq 1$, is a bounded domain with a sufficiently regular boundary, ν_n, $n = 1, \ldots, N$, are non-negative Tikhonov parameters with $\nu_N > 0$, and j_n, $n = 1, \ldots, N$, are given objective functions with suitable mapping and smoothness properties. For the precise assumptions on the quantities in (P), we refer to Section 2. Problems of the type (P) – *i.e.*, multicriteria optimization problems governed by semilinear partial differential equations involving non-smooth Nemytskii operators - arise, for instance, in mechanics, plasma physics, and the context of certain combustion processes when the state of the system is supposed to meet several, potentially conflicting design goals. See, *e.g.*, [33, 47, 53, 55] for some examples of possible application areas.

From the mathematical point of view, problems of the type (P) are challenging for a number of reasons. The first and probably most obvious one is the presence of the non-smooth Nemytskii operator in the governing partial differential equation. Because of this term, the control-to-state mapping $S : u \mapsto y$ of (P) does not possess a Gâteaux derivative, but is only directionally differentiable, and it is not possible to apply standard results to derive, for instance, first-order necessary optimality conditions or to devise efficient numerical solution algorithms. The strategy that is most commonly used in the literature to handle such a lack of smoothness is to work with elements of appropriately defined subdifferentials (*e.g.*, those of Clarke, Dini, Mordukhovich, or Fréchet) instead of the non-existing gradients of the involved (reduced) objective functions. See, for example, [17, 23, 43, 50, 52] for an overview of these concepts and references on the use of subgradients in the single-objective context, [9, 10, 26, 27, 30, 31, 36, 39, 42, 56] for applications in the multiobjective setting, and [3, 6, 35] for results that additionally also rely on regularization techniques. Unfortunately, for problems of the type (P), a standard subgradient-based analysis turns out to be not very rewarding either. Since the non-smoothness enters (P) only indirectly *via* the non-differentiable Nemytskii operator in the governing PDE, a full characterization of quantities like the Clarke subdifferentials of the reduced objective functions $J_n(S(\cdot), \cdot)$, $n = 1, \ldots, N$, of (P) is only available in pathological situations. Optimality conditions involving these generalized differentials are thus rather academic and barely usable in practice. Note that this is a major difference to situations, in which the non-smoothness stems from the "outer" functions J_n and not from the governing PDE, and in which, as a consequence, classical chain rules for subgradients can be applied, *cf.* [7]. A possible way around this difficulty, that has recently been employed in [14, 48, 49] in the single-objective setting, is to work with notions of generalized derivatives on the level of the solution operator $S : u \mapsto y$. Such approaches typically result in necessary optimality conditions of intermediate strength that can indeed be solved with standard solution algorithms. For more details on this topic and further remarks on how the findings of [14, 48, 49] are related to the present paper, we refer to [14], Section 4 and Sections 4 and 5. For results on smooth multiobjective optimization and optimal control problems, see also [4, 8, 22, 29, 44–46].

A second factor that significantly complicates the derivation of necessary optimality conditions for problems of the type (P) and that is completely absent in the single-objective setting is that - even in the smooth case – there are several sensible, purely primal optimality and stationarity concepts for multiobjective optimization problems. Compare, for instance, with the notions of weak, ordinary, and proper Pareto optimality and stationarity in Definitions 2.3 and 3.2 in this context. Even worse, the multiobjective aspect of (P) also turns out to add an additional layer of non-smoothness to the problem. To see this, consider, for example, the notion of weak Pareto

optimality for (P), *i.e.*, the optimality condition given by

$$\nexists u \in L^2(\Omega): \qquad J_n(S(u), u) < J_n(S(\bar{u}), \bar{u}) \quad \forall n = 1, \ldots, N.$$

This condition is clearly equivalent to $\bar{u}$ being a global optimum of the scalar problem

$$\min_{u \in L^2(\Omega)} \max \big(J_1(S(u), u) - J_1(S(\bar{u}), \bar{u}), \ldots, J_N(S(u), u) - J_N(S(\bar{u}), \bar{u}) \big), \tag{1.1}$$

whose objective function not only contains the non-smooth solution map S, but also the non-differentiable maximum-value function on $\mathbb{R}^N$. We thus indeed have to deal with two sources of non-smoothness here which are even nested. Note that, in first-order necessary optimality conditions for weak Pareto optima, the presence of the maximum-value function in (1.1) typically manifests itself in the form of additional multipliers. Compare, for instance, with the results for multiobjective problems with smooth objective functions in [39], Theorems 3.1.1, 3.1.5 in this context and also with the stationarity systems established in Theorem 4.5.

The main goal of the present paper is to demonstrate that, despite all of the above difficulties, it is indeed possible to derive very rigorous first-order necessary optimality conditions for multiobjective optimal control problems of the type (P). To be more precise, we will show how to prove so-called strong stationarity conditions for this problem class. In the single-objective setting, such conditions are well-known, *e.g.*, for the optimal control of elliptic variational inequalities of the first and the second kind and the optimal control of non-smooth elliptic and parabolic partial differential equations, see [13, 14, 18, 19, 38, 40, 41]. In the multiobjective context, strong stationarity conditions have, at least to the best of the authors' knowledge and with some exceptions for finite-dimensional MPECs in [56], not been considered so far (most likely because of the doubly non-smooth behavior in (1.1) that is generally hard to handle). Recall that the distinguishing feature of a strong stationarity system is its equivalence to the first-order necessary optimality condition in primal form. In the single-objective case, this means that a point is strongly stationary if and only if it is Bouligand stationary in the sense of [20], Definition 5.4, *i.e.*, stationary in the sense that the directional derivative is non-negative in all directions. When considering the problem (P) with its multiple sensible purely primal necessary optimality conditions, one, of course, has to differentiate at this point. We will thus establish not only one but even two strong stationarity systems for (P) – one equivalent to weak Pareto stationarity and one equivalent to both ordinary and proper Pareto stationarity, see Definition 3.2 and Theorem 4.5. (Note that this implies in particular that the concepts of ordinary and proper Pareto stationarity are the same for the problem (P).) A main ingredient of our analysis is the - at first glance rather surprising – fact that the non-linear and non-differentiable inverse $S'(u; \cdot)^{-1}: H_0^1(\Omega) \cap H^2(\Omega) \to L^2(\Omega)$, $z \mapsto -\Delta z + \mathbb{1}_{\{S(u)=0\}} \max(0, z) + \mathbb{1}_{\{S(u)>0\}} z$, of the directional derivative $S'(u; \cdot)$ of the control-to-state mapping $S: u \mapsto y$ of (P) is self-adjoint in the sense that

$$\big(u, S'(u; \cdot)^{-1}(z) \big)_{L^2} = \big\langle S'(u; \cdot)^{-1}(u), z \big\rangle_{H_0^1 \cap H^2} \tag{1.2}$$

holds for all $u \in L^2(\Omega)$ and all $z \in H_0^1(\Omega) \cap H^2(\Omega)$, where, on the right-hand side of (1.2), the map $S'(u; \cdot)^{-1}$ is interpreted as a function from $L^2(\Omega)$ to $(H_0^1(\Omega) \cap H^2(\Omega))^*$. This behavior makes it possible to resolve the difficulties related to the additional layer of non-smoothness in (1.1) that normally prevent the derivation of strong stationarity conditions in the multiobjective context, *cf.* Lemma 4.2 and Theorem 4.5.

Note that the self-adjointness of the operator $S'(u; \cdot)^{-1}$ in (1.2) is also of relevance for scalar Tikhonov regularized optimal control problems governed by non-smooth semilinear partial differential equations as it allows to employ an adjoint calculus similar to that available in the smooth setting in the non-smooth case. Quite interestingly, the self-adjointness property in (1.2) is also directly related to certain non-standard regularization effects that have apparently not been documented so far in the literature – neither in the single- nor in the multiobjective context. These effects cause the problem (P) to be Gâteaux differentiable when it is reduced to the state y although it is not Gâteaux when reduced to the control u. For further details on this topic, see Section 6. We remark that the results of Section 6 particularly imply that the problems considered in [18],

Section 5, [14], Section 5, and [21], Section 5, Case 1 all admit a Gâteaux differentiable reformulation. As we will see in Section 5, they further yield that regularization approaches are exact for the problem (P) in the sense that weak L^2-accumulation points of weakly Pareto stationary points of the regularized multiobjective optimal control problems are weakly Pareto stationary for the non-smooth limit problem. Note that this behavior is again quite surprising as such effects can typically not even be observed in simple, one-dimensional examples. Compare, for instance, with the situation, where the function $f(x) := -|x|$ is approximated by the sequence $f_\varepsilon(x) := -\sqrt{x^2 + \varepsilon}$, $\varepsilon > 0$, and where the point $\bar{x} = 0$ is Bouligand stationary for all f_ε but not for the limit function f, in this context.

Before we begin with our analysis, we would like to point out that we consider the problem (P) as a model problem in this paper. It is easy to check that our arguments can be extended straightforwardly to cases where the governing PDE contains a more general elliptic second-order partial differential operator or a more complicated piecewise C^1-function with properties similar to that of $\max(0, \cdot)$. An extension to the parabolic setting, $cf.$ [38], is also possible by invoking the results in the appendix of [15]. We restrict our attention to the setting in (P) to avoid obscuring the basic ideas of our analysis with unnecessary technicalities.

To help the reader navigate the paper, we conclude this introduction with a brief overview of the content of the upcoming sections:

Sections 1.1 and 2 deal with preliminaries. Here, we comment on the notation used, state our standing assumptions on j_n, ν_n, and Ω (see Asm. 2.1), and recall known results on the properties of the control-to-state mapping $S : u \mapsto y$ of (P), notions of optimality for multiobjective optimization problems, and the existence of Pareto optimal controls.

Section 3 is concerned with first-order necessary optimality conditions in purely primal form (*i.e.*, conditions based on directional derivatives). We remark that, in the finite-dimensional setting, such stationarity concepts have already been discussed, *e.g.*, in [26], Section 4 and [30], Section 3.

In Section 4, we prove the already mentioned strong stationarity conditions for the multiobjective optimal control problem (P). This section contains the main result of the paper, Theorem 4.5. It further addresses in detail the self-adjointness property in (1.2) (see Lem. 4.2), the consequences that our findings have for the regularity properties of Pareto stationary points (see Cor. 4.7), and the relationship of our results to the notion of subdifferential studied in [14, 48, 49] and classical scalarization techniques (see Rem. 4.6).

Section 5 demonstrates that smoothing methods are indeed exact for problems of the type (P) in the sense that they allow to determine weakly Pareto stationary points when the regularization parameter is driven to zero, see Theorem 5.3. Corollary 5.4 in this section moreover shows that the concepts of strong and C-stationarity are the same for the problem (P) in the single-objective case $N = 1$.

In Section 6, we discuss the non-standard regularization effects that are responsible for the results of the previous sections and that cause the problem (P) to be Gâteaux differentiable when reduced to the state y. Here, we further give an alternative interpretation of the strong stationarity conditions in Theorem 4.5 and prove an auxiliary result on non-smooth Nemytskii operators that is also interesting on its own, see Theorem 6.1.

Section 7 of the paper finally contains numerical experiments which demonstrate that the multiplier systems in Sections 4 and 5 are amenable to numerical solution procedures and allow to compute approximations of the Pareto front of (P).

1.1. Remarks on the notation

In what follows, we use the standard symbols $L^q(\Omega)$, $C^{k,\gamma}(\Omega)$, $H_0^k(\Omega)$, $H^k(\Omega)$, and $W^{k,q}(\Omega)$, $1 \le q \le \infty$, $k \in \mathbb{N}$, $0 < \gamma \le 1$, to denote the Lebesgue-, Hölder-, and Sobolev spaces on a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, $d \ge 1$, respectively. For details on these spaces, we refer to [1, 2, 24, 28]. Given two Banach spaces X and Y, we further define X^* to be the dual space of X and $L(X, Y)$ to be the space of linear and continuous functions from X to Y. In the special case $X = H_0^1(\Omega)$, we also set $H^{-1}(\Omega) := H_0^1(\Omega)^*$. As usual, we interpret $H_0^1(\Omega)$, $L^2(\Omega)$, and $H^{-1}(\Omega)$ as a Gelfand triple, *i.e.*, $H_0^1(\Omega) \hookrightarrow L^2(\Omega) \cong L^2(\Omega)^* \hookrightarrow H^{-1}(\Omega)$. The same convention is used for the space $H_0^1(\Omega) \cap H^2(\Omega)$, *cf.* [32], Section 1.9. Norms, scalar products, and dual pairings are denoted by the symbols $\| \cdot \|$, $(\cdot, \cdot)$, and $\langle \cdot, \cdot \rangle$ in this paper, and the modes of weak and strong convergence by the arrows $\rightharpoonup$ and

$\rightarrow$. If we want to specify the space/topology that we are referring to, then we add suitable sub- or superscripts and write, $e.g.$, $\|\cdot\|_{L^2}$. With Δ, ∇, ∂_c, and ∂_i, $i = 1, \ldots, d$, we denote the (distributional) Laplacian, the (weak) gradient, the convex subdifferential, and the first (weak) partial derivatives of a function, respectively. For higher-order derivatives, we also use the multi-index notation ∂^α, $\alpha \in \mathbb{N}_0^d$. Gâteaux-, Fréchet-, and directional derivatives in the functional analytic sense are denoted by a prime in the usual way. Given a function $v : \Omega \to \mathbb{R}$ and a measurable set $D \subset \Omega$, we finally define $\{v * 0\}$, $* \in \{=, \neq, <, >, \leq, \geq\}$, to be the set $\{x \in \Omega \mid v(x) * 0\}$ and $\mathbb{1}_D$ to be the indicator function of D (with values in $\{0, 1\}$). If v is an element of an L^q-space and D is only defined up to sets of measure zero, then we consider $\{v * 0\}$ to be defined up to sets of measure zero as well and identify $\mathbb{1}_D$ with an element of $L^\infty(\Omega)$.

Note that, in the remainder of this paper, new symbols $etc.$ are introduced whenever necessary. For the sake of readability, this additional notation is defined where it first appears in the text.

2. Problem setting and preliminaries

As already mentioned in the introduction, the aim of this paper is to study multiobjective optimal control problems of the type

$$\text{Minimize} \quad \begin{pmatrix} J_1(y, u) := j_1(y) + \dfrac{\nu_1}{2}\|u\|_{L^2}^2 \\ \vdots \\ J_N(y, u) := j_N(y) + \dfrac{\nu_N}{2}\|u\|_{L^2}^2 \end{pmatrix} \tag{P}$$
$$\text{w.r.t.} \quad u \in L^2(\Omega), \quad y \in H_0^1(\Omega) \cap H^2(\Omega),$$
$$\text{s.t.} \quad -\Delta y + \max(0, y) = u \text{ a.e. in } \Omega.$$

Our standing assumptions on the quantities in (P) are as follows:

Assumption 2.1 (Standing assumptions for the study of problem (P))**.**

- i) $\Omega \subset \mathbb{R}^d$, $d \geq 1$, is a bounded domain that is convex or possesses a $C^{1,1}$-boundary (in the sense of [25], Sect. 6.2).
- ii) $j_n : H_0^1(\Omega) \cap H^2(\Omega) \to \mathbb{R}$, $n = 1, \ldots, N$, $N \in \mathbb{N}$, are functions that are weakly lower semicontinuous, continuously differentiable, and bounded from below.
- iii) ν_n, $n = 1, \ldots, N$, are given non-negative real numbers and ν_N is positive.

Note that the PDE in (P) is uniquely solvable for all $u \in L^2(\Omega)$ by the theorem of Browder and Minty, [51], Theorem 3-1.5. To be more precise, we have:

Proposition 2.2 (Properties of the PDE in (P))**.** *For every right-hand side* $u \in L^2(\Omega)$*, there exists a unique solution* $y \in H_0^1(\Omega) \cap H^2(\Omega)$ *of the partial differential equation*

$$-\Delta y + \max(0, y) = u \text{ a.e. in } \Omega. \tag{2.1}$$

Further, the solution operator $S : L^2(\Omega) \to H_0^1(\Omega) \cap H^2(\Omega)$*,* $u \mapsto y$*, associated with the PDE (2.1) satisfies:*

- i) *S is globally Lipschitz continuous, i.e., there exists an absolute constant $C > 0$ with*

$$\|S(u_1) - S(u_2)\|_{H^2} \leq C\|u_1 - u_2\|_{L^2} \qquad \forall u_1, u_2 \in L^2(\Omega).$$

- ii) *S is weakly continuous, i.e., for every $u \in L^2(\Omega)$ it holds*

$$u_k \xrightarrow{L^2} u \quad \implies \quad S(u_k) \xrightarrow{H^2} S(u). \tag{2.2}$$

iii) S *is strongly and weakly Hadamard directionally differentiable in every point $u \in L^2(\Omega)$ in every direction $v \in L^2(\Omega)$, i.e., for all $u, v \in L^2(\Omega)$, there exists a unique $S'(u; v) \in H_0^1(\Omega) \cap H^2(\Omega)$ such that the implications*

$$v_k \xrightarrow{L^2} v, \; t_k \to 0^+ \quad \Longrightarrow \quad \frac{S(u + t_k v_k) - S(u)}{t_k} \xrightarrow{H^2} S'(u; v)$$

and

$$v_k \xrightarrow{L^2} v, \; t_k \to 0^+ \quad \Longrightarrow \quad \frac{S(u + t_k v_k) - S(u)}{t_k} \xrightarrow{H^2} S'(u; v)$$

hold. Moreover, the directional derivative $\delta_v := S'(u; v) \in H_0^1(\Omega) \cap H^2(\Omega)$ in a point u with state $y := S(u)$ in a direction v is uniquely characterized by the partial differential equation

$$-\Delta \delta_v + \mathbb{1}_{\{y=0\}} \max(0, \delta_v) + \mathbb{1}_{\{y>0\}} \delta_v = v \; \text{a.e. in } \Omega. \tag{2.3}$$

iv) S *is Gâteaux differentiable in a point $u \in L^2(\Omega)$ with state $y := S(u)$ (i.e., $S'(u; \cdot)$ is an element of the space $L(L^2(\Omega), H_0^1(\Omega) \cap H^2(\Omega))$) if and only if the set $\{y = 0\}$ has measure zero.*

Proof. All assertions of the proposition have been proved in [14], Proposition 2.1, Theorem 2.2, Corollary 2.3, 3.8. Note that, under our assumptions on Ω, the space $(Y, \|\cdot\|_Y)$ used in [14] is isomorphic to $(H_0^1(\Omega) \cap H^2(\Omega), \|\cdot\|_{H^2})$ by [25], Theorem 9.15, Lemma 9.17 and [28], Theorem 3.2.1.2. The results of [14] thus indeed yield the asserted mapping properties of the solution operator S. $\qquad\square$

As usual in the context of multiobjective optimization, in the remainder of this paper, we are interested in finding controls $u \in L^2(\Omega)$ that yield – at least in some sense – an optimal compromise between the different objective functions $J_n(S(\cdot), \cdot)$, $n = 1, \ldots, N$, of (P). The notions of optimality that we are mainly concerned with in our analysis are the following (*cf.* [30], Defs. 3.1, 3.2 and also [22, 39]):

Definition 2.3 (Notions of Pareto optimality)**.** A control $\bar{u} \in L^2(\Omega)$ with associated state $\bar{y} := S(\bar{u})$ is called:

i) a local weak Pareto optimum of (P) if there exists an $r > 0$ such that there is no $u \in L^2(\Omega)$ satisfying

$$\|u - \bar{u}\|_{L^2} < r, \qquad J_n(S(u), u) < J_n(\bar{y}, \bar{u}) \quad \forall n = 1, \ldots, N.$$

ii) a local Pareto optimum of (P) (in the ordinary sense) if there exists an $r > 0$ such that there is no $u \in L^2(\Omega)$ satisfying

$$\|u - \bar{u}\|_{L^2} < r, \qquad J_n(S(u), u) \leq J_n(\bar{y}, \bar{u}) \quad \forall n = 1, \ldots, N,$$
$$J_n(S(u), u) < J_n(\bar{y}, \bar{u}) \text{ for at least one } n \in \{1, \ldots, N\}.$$

iii) a local proper Pareto optimum of (P) (in the sense of Geoffrion) if there exist constants $r, C > 0$ such that, for every control $u \in L^2(\Omega)$ satisfying $\|u - \bar{u}\|_{L^2} < r$ and $J_l(S(u), u) < J_l(\bar{y}, \bar{u})$ for some $l \in \{1, \ldots, N\}$, there exists an index $m \in \{1, \ldots, N\}$ with

$$J_l(\bar{y}, \bar{u}) - J_l(S(u), u) \leq C\big(J_m(S(u), u) - J_m(\bar{y}, \bar{u})\big).$$

iv) a global weak/ordinary/proper Pareto optimum, respectively, of (P) if the condition in i)/ii)/iii), respectively, holds with $r = \infty$.

Note that we trivially have

$$\bar{u} \text{ properly Pareto optimal} \quad \Rightarrow \quad \bar{u} \text{ Pareto optimal} \quad \Rightarrow \quad \bar{u} \text{ weakly Pareto optimal}, \tag{2.4}$$

and that the concepts of local (respectively, global) weak, ordinary, and proper Pareto optimality coincide with each other and the classical notion of local (respectively, global) optimality in the single-objective case $N = 1$. Using our standing assumptions and the properties of the map S in Proposition 2.2, it is easy to prove:

Theorem 2.4 (Existence of proper Pareto optima)**.** *There exists at least one global proper Pareto optimum $\bar{u} \in L^2(\Omega)$ of the problem (P).*

Proof. We use a scalarization approach, *cf.* [22], Theorem 3.11 and [39]: Consider the auxiliary problem

$$\min_{u \in L^2(\Omega)} \sum_{n=1}^{N} \left(j_n(S(u)) + \frac{\nu_n}{2} \|u\|_{L^2}^2 \right). \tag{2.5}$$

Then, it follows from our assumptions on j_n and ν_n, the weak continuity of the map S in Proposition 2.2ii), and the weak lower semicontinuity of convex and continuous functions that the objective of (2.5) is weakly lower semicontinuous, bounded from below, and radially unbounded as a function from $L^2(\Omega)$ to $\mathbb{R}$. These properties imply, in combination with the direct method of calculus of variations, that (2.5) admits at least one global minimum $\bar{u} \in L^2(\Omega)$, *i.e.*, at least one $\bar{u} \in L^2(\Omega)$ with associated state $\bar{y} := S(\bar{u})$ such that

$$\sum_{n=1}^{N} J_n(\bar{y}, \bar{u}) \leq \sum_{n=1}^{N} J_n(S(u), u) \qquad \forall u \in L^2(\Omega). \tag{2.6}$$

We claim that this $\bar{u}$ is also a global proper Pareto optimum of the problem (P). To see this, suppose that we are given a $u \in L^2(\Omega)$ that satisfies $J_l(S(u), u) < J_l(\bar{y}, \bar{u})$ for some $l \in \{1, \dots, N\}$. Then, (2.6) yields

$$0 < J_l(\bar{y}, \bar{u}) - J_l(S(u), u) \leq \sum_{n \neq l} \left(J_n(S(u), u) - J_n(\bar{y}, \bar{u}) \right) \leq N \max_{n=1,\dots,N} \left(J_n(S(u), u) - J_n(\bar{y}, \bar{u}) \right), \tag{2.7}$$

and we may deduce that there exists an $m \in \{1, \dots, N\}$ with

$$J_l(\bar{y}, \bar{u}) - J_l(S(u), u) \leq N \left(J_m(S(u), u) - J_m(\bar{y}, \bar{u}) \right).$$

This shows that the condition in Definition 2.3iii) holds (with $r = \infty$ and $C = N$) and completes the proof. $\qquad \square$

3. First-order necessary optimality conditions in primal form

Having discussed the properties of the PDE (2.1) and the solvability of the problem (P), we now turn our attention to first-order necessary optimality conditions. We begin with "purely primal" optimality conditions that rely only on directional derivatives and do not involve additional multipliers. In the finite-dimensional setting, such conditions have already been discussed, for instance, in [26], Section 4 and [30], Section 3.

Theorem 3.1 (First-order necessary optimality conditions in primal form)**.**

    i) If $\bar{u} \in L^2(\Omega)$ is a local weak Pareto optimum of (P) with associated state $\bar{y} := S(\bar{u})$, then there exists no $v \in L^2(\Omega)$ satisfying

$$\langle j_n'(\bar{y}), S'(\bar{u}; v) \rangle_{H_0^1 \cap H^2} + \nu_n (\bar{u}, v)_{L^2} < 0 \qquad \forall n = 1, \dots, N. \tag{3.1}$$

ii) *If $\bar{u} \in L^2(\Omega)$ is a local proper Pareto optimum of (P) with state $\bar{y} := S(\bar{u})$ and constants $r, C > 0$ as in Definition 2.3iii), then, for every direction $v \in L^2(\Omega)$ satisfying $\langle j'_l(\bar{y}), S'(\bar{u}; v)\rangle_{H_0^1 \cap H^2} + \nu_l\,(\bar{u}, v)_{L^2} < 0$ for some $l \in \{1, \ldots, N\}$, there exists an index $m \in \{1, \ldots, N\}$ with*

$$- \left(\langle j'_l(\bar{y}), S'(\bar{u}; v)\rangle_{H_0^1 \cap H^2} + \nu_l\,(\bar{u}, v)_{L^2} \right) \leq C \left(\langle j'_m(\bar{y}), S'(\bar{u}; v)\rangle_{H_0^1 \cap H^2} + \nu_m\,(\bar{u}, v)_{L^2} \right). \qquad (3.2)$$

Proof. From our assumptions on the functions j_n, the differentiability properties of the solution map S in Proposition 2.2, and the chain rule, [11], Proposition 2.47, it follows straightforwardly that

$$\lim_{t \to 0^+} \frac{J_n(S(u + tv), u + tv) - J_n(S(u), u)}{t} = \langle j'_n(S(u)), S'(u; v)\rangle_{H_0^1 \cap H^2} + \nu_n\,(u, v)_{L^2} \qquad (3.3)$$

holds for all $u, v \in L^2(\Omega)$ and all $n = 1, \ldots, N$. Suppose now that we are given a local weak Pareto optimum $\bar{u} \in L^2(\Omega)$ of (P) such that there exists a direction $v \in L^2(\Omega)$ with (3.1). Then, (3.3) yields that we can find arbitrarily small numbers $t > 0$ with $J_n(S(\bar{u} + tv), \bar{u} + tv) - J_n(S(\bar{u}), \bar{u}) < 0$ for all $n = 1, \ldots, N$. This contradicts the local weak Pareto optimality of $\bar{u}$, shows that every local weak Pareto optimum has to satisfy the condition in i), and proves the first part of the theorem. To establish ii), we can proceed along similar lines: If we are given a local proper Pareto optimum $\bar{u} \in L^2(\Omega)$ with state $\bar{y} := S(\bar{u})$ and constants $r, C > 0$ as in Definition 2.3iii) and a $v \in L^2(\Omega)$ satisfying $\langle j'_l(\bar{y}), S'(\bar{u}; v)\rangle_{H_0^1 \cap H^2} + \nu_l\,(\bar{u}, v)_{L^2} < 0$ for some $l \in \{1, \ldots, N\}$, then (3.3) yields that $J_l(S(\bar{u} + tv), \bar{u} + tv) - J_l(\bar{y}, \bar{u}) < 0$ holds for all sufficiently small $t > 0$, and we may use the condition in Definition 2.3iii) to deduce that there exist an $m \in \{1, \ldots, N\}$ and a sequence $\{t_k\} \subset \mathbb{R}^+$ with $t_k \to 0^+$ and

$$0 < J_l(\bar{y}, \bar{u}) - J_l(S(\bar{u} + t_k v), \bar{u} + t_k v) \leq C\big(J_m(S(\bar{u} + t_k v), \bar{u} + t_k v) - J_m(\bar{y}, \bar{u})\big)$$

for all $k \in \mathbb{N}$. Due to (3.3) and the properties of l, the above implies

$$0 < - \left(\langle j'_l(\bar{y}), S'(\bar{u}; v)\rangle_{H_0^1 \cap H^2} + \nu_l\,(\bar{u}, v)_{L^2} \right) \leq C \left(\langle j'_m(\bar{y}), S'(\bar{u}; v)\rangle_{H_0^1 \cap H^2} + \nu_m\,(\bar{u}, v)_{L^2} \right).$$

This establishes (3.2) and completes the proof.   □

Theorem 3.1 motivates the following definition:

Definition 3.2 (Notions of stationarity). A control $\bar{u} \in L^2(\Omega)$ with associated state $\bar{y} := S(\bar{u})$ is called:

i) a weakly Pareto stationary point of (P) if there is no $v \in L^2(\Omega)$ satisfying

$$\langle j'_n(\bar{y}), S'(\bar{u}; v)\rangle_{H_0^1 \cap H^2} + \nu_n\,(\bar{u}, v)_{L^2} < 0 \qquad \forall n = 1, \ldots, N. \qquad (3.4)$$

ii) a Pareto stationary point of (P) if there is no $v \in L^2(\Omega)$ satisfying

$$\langle j'_n(\bar{y}), S'(\bar{u}; v)\rangle_{H_0^1 \cap H^2} + \nu_n\,(\bar{u}, v)_{L^2} \leq 0 \qquad \forall n = 1, \ldots, N,$$
$$\langle j'_n(\bar{y}), S'(\bar{u}; v)\rangle_{H_0^1 \cap H^2} + \nu_n\,(\bar{u}, v)_{L^2} < 0 \text{ for at least one } n \in \{1, \ldots, N\}. \qquad (3.5)$$

iii) a properly Pareto stationary point of (P) if there exists a $C > 0$ such that, for every $v \in L^2(\Omega)$ satisfying $\langle j'_l(\bar{y}), S'(\bar{u}; v)\rangle_{H_0^1 \cap H^2} + \nu_l\,(\bar{u}, v)_{L^2} < 0$ for some $l \in \{1, \ldots, N\}$, there is an index $m \in \{1, \ldots, N\}$ with

$$- \left(\langle j'_l(\bar{y}), S'(\bar{u}; v)\rangle_{H_0^1 \cap H^2} + \nu_l\,(\bar{u}, v)_{L^2} \right) \leq C \left(\langle j'_m(\bar{y}), S'(\bar{u}; v)\rangle_{H_0^1 \cap H^2} + \nu_m\,(\bar{u}, v)_{L^2} \right). \qquad (3.6)$$

Some remarks are in order regarding the concepts in Definition 3.2:

Remark 3.3. i) Pareto optima satisfying the condition in Definition 3.2ii) are sometimes also referred to as "efficient in Kuhn and Tucker's sense" or "KT-proper", see [22], Definition 2.49 and [30], Definition 3.3.

ii) Analogously to the hierarchy in (2.4), we have

$$\bar{u} \text{ properly stationary} \quad \Rightarrow \quad \bar{u} \text{ stationary} \quad \Rightarrow \quad \bar{u} \text{ weakly stationary.}$$

iii) The conditions i), ii), and iii) in Definition 3.2 are equivalent to the global weak, global ordinary, and global proper Pareto optimality of the function $\bar{v} = 0$ in the first-order approximation of (P) defined by

$$\text{Minimize} \quad \begin{pmatrix} J_1(\bar{y}, \bar{u}) + \langle j_1'(\bar{y}), S'(\bar{u}; v) \rangle_{H_0^1 \cap H^2} + \nu_1 (\bar{u}, v)_{L^2} \\ \vdots \\ J_N(\bar{y}, \bar{u}) + \langle j_N'(\bar{y}), S'(\bar{u}; v) \rangle_{H_0^1 \cap H^2} + \nu_N (\bar{u}, v)_{L^2} \end{pmatrix} \quad \text{w.r.t.} \quad v \in L^2(\Omega),$$

respectively. This shows in particular that the concepts in Definition 3.2 extend the notion of Bouligand stationarity, see [20], Definition 5.4, to the multiobjective setting. (It is easy to check that, in the single-objective case $N = 1$, all of the conditions in Definition 3.2 coincide with each other and with the notion of Bouligand stationarity.)

iv) While the conditions i) and iii) in Definition 3.2 are always necessary for weak and proper Pareto optimality, respectively, by Theorem 3.1, the property in Definition 3.2ii) is typically not a necessary condition for ordinary Pareto optimality. Compare, *e.g.*, with the simple bi-criterial optimization problem $\min(f_1(x), f_2(x))$ with $f_1(x) := -x$ and $f_2(x) := x^3$ in this context, where the point $\bar{x} := 0$ is a Pareto optimum but does not satisfy ii). In the finite-dimensional setting, the necessity of ii) can be recovered under a generalized Abadie constraint qualification, see [26], Theorem 4.1.

Due to their reliance on the non-linear directional derivative $S'(\bar{u}; \cdot)$ of the solution map S and their formulation as variational inequalities, the necessary optimality conditions in Theorem 3.1 and the stationarity concepts in Definition 3.2 are typically not very useful in practical applications. In the following Sections 4 and 5, we will derive multiplier systems that are easier to work with and more suitable as starting points for the development of numerical solution algorithms.

4. Strong stationarity conditions for the problem (P)

The aim of this section is to establish so-called strong stationarity conditions for the multiobjective optimal control problem (P), *i.e.*, multiplier systems that are equivalent to the purely primal necessary optimality conditions in Definition 3.2. As already pointed out in the introduction, in the single-objective setting, such stationarity systems are well-known for various problem classes. See, *e.g.*, [13, 14, 18, 19, 38, 40, 41] and the references therein for some examples. In the multiobjective context, the situation is different since the additional layer of non-smoothness in (1.1) and the non-differentiability of the control-to-state mapping S create a nested structure that is generally hard to handle analytically. In the following, we will show that, for the problem (P), the difficulties arising from the doubly non-smooth behavior in (1.1) can be resolved by exploiting the self-adjointness property of the operator $S'(u; \cdot)^{-1}$ in (1.2). The starting point of our investigation is:

Lemma 4.1 (Behavior of higher-order weak derivatives on level sets). *Suppose that a function $w \in W^{k,1}(\Omega)$, $k \in \mathbb{N}$, is given. Then, for every $\alpha \in \mathbb{N}_0^d$ with $1 \leq |\alpha| \leq k$ and every $b \in \mathbb{R}$, it holds $\partial^\alpha w = 0$ a.e. in $\{w = b\}$.*

Proof. We use induction w.r.t. the absolute value $l := |\alpha|$ of the multi-index α to establish the claim: Suppose that an arbitrary but fixed $k \in \mathbb{N}$ and a $w \in W^{k,1}(\Omega)$ are given. Then, for every $i = 1, \ldots, d$, the classical lemma of Stampacchia, see [2], Proposition 5.8.2, implies that $(\partial_i w) \mathbb{1}_{\{w=b\}} = 0$ holds as an identity in $L^2(\Omega)$. This proves the assertion for $l = 1$. It remains to perform the induction step $l \mapsto l + 1$. To this end, let us assume

that a multi-index $\alpha \in \mathbb{N}_0^d$ with $|\alpha| = l + 1 \leq k$, $l \geq 1$, is given. Then, we can find multi-indices $\beta, \gamma \in \mathbb{N}_0^d$ with $|\beta| = l$, $|\gamma| = 1$, and $\alpha = \beta + \gamma$. From the induction hypothesis, we obtain that $\partial^\beta w$ vanishes a.e. on $\{w = b\}$, *i.e.*, it holds $(\partial^\beta w)\mathbb{1}_{\{w=b\}} = 0 \in L^2(\Omega)$ and, as a consequence, $\mathbb{1}_{\{w=b\}} = \mathbb{1}_{\{\partial^\beta w=0\}}\mathbb{1}_{\{w=b\}} \in L^2(\Omega)$. Using again the classical version of Stampacchia's lemma, we further obtain that $(\partial^{\beta+\gamma}w)\mathbb{1}_{\{\partial^\beta w=0\}} = 0 \in L^2(\Omega)$. Combining the last two identities yields that

$$(\partial^\alpha w)\mathbb{1}_{\{w=b\}} = (\partial^{\beta+\gamma}w)\mathbb{1}_{\{\partial^\beta w=0\}}\mathbb{1}_{\{w=b\}} = 0 \in L^2(\Omega).$$

Thus, $\partial^\alpha w = 0$ a.e. in $\{w = b\}$ and the induction step is complete. This proves the claim of the lemma. $\qquad\square$

We remark that, in a less general format, the above result has already been used in [15], proof of Lemma A.1 and [16], proof of Theorem 2.2. By applying Lemma 4.1 to the PDE (2.3), it is straightforward to check that the directional derivative $S'(u; \cdot)$ indeed satisfies the identity (1.2). To be more precise, we have:

Lemma 4.2 (Properties of $S'(u; \cdot)$ and $S'(u; \cdot)^{-1}$)**.** *Consider an arbitrary but fixed control $u \in L^2(\Omega)$ with state $y := S(u)$. Then, the following is true:*

  i) *The map $S'(u; \cdot) : L^2(\Omega) \to H_0^1(\Omega) \cap H^2(\Omega)$ is bi-Lipschitz and its inverse is given by*

$$S'(u; \cdot)^{-1} : H_0^1(\Omega) \cap H^2(\Omega) \to L^2(\Omega), \qquad w \mapsto -\Delta w + \mathbb{1}_{\{y=0\}}\max(0, w) + \mathbb{1}_{\{y>0\}}w. \qquad (4.1)$$

  ii) *The map $S'(u; \cdot)^{-1} : H_0^1(\Omega) \cap H^2(\Omega) \to L^2(\Omega)$ admits a unique, globally Lipschitz continuous extension $S'(u; \cdot)^{-1} : L^2(\Omega) \to (H_0^1(\Omega) \cap H^2(\Omega))^*$, and this extension satisfies*

$$\left\langle S'(u; \cdot)^{-1}(w), z \right\rangle_{H_0^1 \cap H^2} = \int_\Omega w(-\Delta z) + \mathbb{1}_{\{y=0\}}\max(0, w)z + \mathbb{1}_{\{y>0\}}wz \, \mathrm{d}x$$
$$\forall z \in H_0^1(\Omega) \cap H^2(\Omega) \quad \forall w \in L^2(\Omega). \qquad (4.2)$$

  iii) *For every $z \in H_0^1(\Omega) \cap H^2(\Omega)$, it holds*

$$\left(u, S'(u; \cdot)^{-1}(z)\right)_{L^2} = \left\langle -\Delta u + \mathbb{1}_{\{y>0\}}u, z \right\rangle_{H_0^1 \cap H^2}. \qquad (4.3)$$

  Here, $\Delta u \in (H_0^1(\Omega) \cap H^2(\Omega))^*$ denotes the "very weak" Dirichlet Laplacian of the function $u \in L^2(\Omega)$, i.e., the unique element of $(H_0^1(\Omega) \cap H^2(\Omega))^*$ that satisfies the adjoint relation $(u, \Delta z)_{L^2} = \langle \Delta u, z \rangle_{H_0^1 \cap H^2}$ for all $z \in H_0^1(\Omega) \cap H^2(\Omega)$.

  iv) *For every $z \in H_0^1(\Omega) \cap H^2(\Omega)$, it holds*

$$\left(u, S'(u; \cdot)^{-1}(z)\right)_{L^2} = \left\langle S'(u; \cdot)^{-1}(u), z \right\rangle_{H_0^1 \cap H^2}.$$

Proof. Part i) of the lemma is a trivial consequence of Proposition 2.2. To prove ii), we note that the right-hand side of (4.2) defines a globally Lipschitz continuous map from $L^2(\Omega)$ to $(H_0^1(\Omega) \cap H^2(\Omega))^*$ that coincides with $S'(u; \cdot)^{-1}$ on $H_0^1(\Omega) \cap H^2(\Omega)$ by Green's formula. (Recall that $H_0^1(\Omega) \cap H^2(\Omega)$, $L^2(\Omega)$, and $(H_0^1(\Omega) \cap H^2(\Omega))^*$ are interpreted as a Gelfand triple, *i.e.*, $H_0^1(\Omega) \cap H^2(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow (H_0^1(\Omega) \cap H^2(\Omega))^*$.) The function $S'(u; \cdot)^{-1}$ thus admits an extension with the desired properties. Since $H_0^1(\Omega) \cap H^2(\Omega)$ is dense in $L^2(\Omega)$, we further know that there can only be one L^2-Lipschitz continuous extension of $S'(u; \cdot)$. This establishes ii). It remains to prove iii) and iv). To this end, we note that Lemma 4.1 and the definition of S imply

$$\int_\Omega \mathbb{1}_{\{y=0\}}\max(0, z)u \, \mathrm{d}x = \int_\Omega \mathbb{1}_{\{y=0\}}\max(0, z)(-\Delta y + \max(0, y))\mathrm{d}x = 0$$

and, analogously,

$$\int_\Omega \mathbb{1}_{\{y=0\}} z \max(0, u) \mathrm{d}x = \int_\Omega \mathbb{1}_{\{y=0\}} z \max(0, -\Delta y + \max(0, y)) \mathrm{d}x = 0$$

for all $z \in H_0^1(\Omega) \cap H^2(\Omega)$. Combining the above with (4.1) and (4.2) yields

$$\left(u, S'(u; \cdot)^{-1}(z)\right)_{L^2} = \int_\Omega u \left(-\Delta z + \mathbb{1}_{\{y=0\}} \max(0, z) + \mathbb{1}_{\{y>0\}} z\right) \mathrm{d}x = \int_\Omega u \left(-\Delta z + \mathbb{1}_{\{y>0\}} z\right) \mathrm{d}x$$

$$= \int_\Omega u(-\Delta z) + \mathbb{1}_{\{y=0\}} \max(0, u)z + \mathbb{1}_{\{y>0\}} uz \, \mathrm{d}x = \left\langle S'(u; \cdot)^{-1}(u), z\right\rangle_{H_0^1 \cap H^2}$$

for all $z \in H_0^1(\Omega) \cap H^2(\Omega)$. This establishes iii) and iv) and completes the proof. $\square$

Remark 4.3. An argument based on Stampacchia's lemma similar to that in the proof of Lemma 4.2 has also been used in [18], Section 5 for the analysis of the solution map of a quasilinear partial differential equation involving a term of the form $g(y)\nabla y$ with a piecewise smooth $g : \mathbb{R} \to \mathbb{R}$. For this PDE, the lemma of Stampacchia for first weak derivatives allows to show that all terms that could possibly prevent the expression $g'(y; z)\nabla y$ from being linear in z are negligible, and to establish that the solution map of the considered quasilinear partial differential equation is Gâteaux differentiable in spite of the fact that it contains the non-smooth Nemytskii operator g. Compare also with Theorem 6.1 and Remark 6.3 in this context. We point out that, for the PDE (2.1), the situation is different since the maps S, $S'(u; \cdot)$, and $S'(u; \cdot)^{-1}$ are not Gâteaux differentiable but contain "proper" non-differentiabilities, *cf.* Proposition 2.2iv) and also the example in Section 6.

As we will see below, Lemma 4.2 makes it possible to handle the non-linearity of the directional derivative $S'(u; \cdot)$ in the necessary optimality conditions (3.4), (3.5), and (3.6). To deal with the multiobjective aspect in these conditions, we need the following infinite-dimensional version of Tucker's/Motzkin's theorem of the alternative (*cf.* [22], Thms. 3.22, 3.24 and also [26], Prop. 2.2):

Lemma 4.4 (Existence of multipliers in general Hilbert spaces)**.** *Suppose that V is a real Hilbert space and that $w_1^*, \ldots, w_N^*$, $N \in \mathbb{N}$, are given elements of V^*. Then, it holds*

$$\nexists z \in V : \quad \langle w_n^*, z\rangle_V < 0 \quad \forall n = 1, \ldots, N$$

$$\Longleftrightarrow \quad \exists \lambda \in \mathbb{R}^N : \quad \lambda_n \geq 0 \ \ \forall n = 1, \ldots, N, \ \ \sum_{n=1}^N \lambda_n = 1, \ \ \sum_{n=1}^N \lambda_n w_n^* = 0 \tag{4.4}$$

and

$$\nexists z \in V : \quad \langle w_n^*, z\rangle_V \leq 0 \ \ \forall n = 1, \ldots, N, \ \ \langle w_n^*, z\rangle_V < 0 \ \text{ for at least one } n$$

$$\Longleftrightarrow \quad \exists \lambda \in \mathbb{R}^N : \quad \lambda_n > 0 \ \ \forall n = 1, \ldots, N, \ \ \sum_{n=1}^N \lambda_n = 1, \ \ \sum_{n=1}^N \lambda_n w_n^* = 0. \tag{4.5}$$

Proof. We begin with (4.4): If we assume that the right-hand side of (4.4) holds and that there exists a $z \in V$ with $\langle w_n^*, z\rangle_V < 0$ for all $n = 1, \ldots, N$, then we arrive at the contradiction

$$0 = \left\langle \sum_{n=1}^N \lambda_n w_n^*, z\right\rangle_V < 0.$$

This proves "$\Leftarrow$". To establish "$\Rightarrow$", we note that the condition on the left-hand side of (4.4) can be recast as

$$\max\left(\langle w_1^*, z\rangle_V, \ldots, \langle w_N^*, z\rangle_V\right) \geq 0 \qquad \forall z \in V,$$

or, equivalently, as $0 \in \partial_c(g \circ F)(0)$, where ∂_c denotes the convex subdifferential and where g and F are the functions defined by

$$g: \mathbb{R}^N \to \mathbb{R}, \quad g(x_1, \ldots, x_N) := \max(x_1, \ldots, x_N) \qquad \text{and} \qquad F: V \to \mathbb{R}^N, \quad F(z) := (\langle w_n^*, z\rangle_V)_{n=1,\ldots,N}.$$

Using the chain rule for the convex subdifferential, see [23], Proposition I-5.7, and the formula for the subdifferential of the vector-maximum ([50], Ex. 8.26), we may now deduce that

$$0 \in F^* \partial_c g(F(0)) = F^* \partial_c g(0) = F^* \left\{ \mu \in \mathbb{R}^N \, \middle| \, \mu_n \geq 0 \; \forall n = 1, \ldots, N, \; \sum_{n=1}^{N} \mu_n = 1 \right\}.$$

This proves the existence of a $\lambda \in \mathbb{R}^N$ with the properties on the right-hand side of (4.4) and establishes "$\Rightarrow$".

It remains to show (4.5). As in the case of (4.4), the implication "$\Leftarrow$" in (4.5) follows immediately by contradiction. To prove "$\Rightarrow$", we note that, by the Riesz representation theorem, we can find elements $w_n \in V$, $n = 1, \ldots, N$, such that $\langle w_n^*, \cdot\rangle_V = (w_n, \cdot)_V$ holds for all n. Suppose that at least one of these w_n is not zero (else the proof is trivial) and denote the subspace spanned by the elements w_n with W. Then, W is obviously finite-dimensional and possesses a $(\cdot, \cdot)_V$-orthonormal basis $e_1, \ldots, e_M$, $1 \leq M \leq N$. Let α_m^n, $n = 1, \ldots, N$, $m = 1, \ldots, M$, be the coordinates of w_n w.r.t. the basis $\{e_m\}$. Then, it holds

$$\left\langle w_n^*, \sum_{m=1}^{M} \beta_m e_m \right\rangle_V = \left(\sum_{l=1}^{M} \alpha_l^n e_l, \sum_{m=1}^{M} \beta_m e_m \right)_V = \sum_{m=1}^{M} \alpha_m^n \beta_m$$

for all $\beta \in \mathbb{R}^M$, and we may use the left-hand side of (4.5) to conclude that

$$\nexists \beta \in \mathbb{R}^M : \quad A\beta \in (-\infty, 0]^N, \quad A\beta \neq 0,$$

where $A \in \mathbb{R}^{N \times M}$ is the matrix defined by $A := (\alpha_m^n)_{n=1,\ldots,N, m=1,\ldots,M}$. From the classical, finite-dimensional version of Tucker's theorem of the alternative, see [37], Theorem II-4.3, we may now deduce that there exists a $\lambda \in \mathbb{R}^N$ satisfying $A^T \lambda = 0$, $\sum_{n=1}^{N} \lambda_n = 1$, and $\lambda_n > 0$ for all $n = 1, \ldots, N$, and from the definition of A that

$$\sum_{n=1}^{N} \lambda_n w_n = \sum_{m=1}^{M} \sum_{n=1}^{N} \lambda_n \alpha_m^n e_m = 0.$$

This establishes the implication "$\Rightarrow$" in (4.5) and completes the proof. $\qquad\square$

We are now in the position to prove the main result of this section and the paper as a whole:

Theorem 4.5 (Strong stationarity conditions).

i) *A control $\bar{u} \in L^2(\Omega)$ with state $\bar{y} := S(\bar{u})$ is weakly Pareto stationary for (P) if and only if there exist an adjoint state $\bar{p}$ and a multiplier $\bar{\lambda}$ such that $\bar{u}$, $\bar{y}$, $\bar{p}$, and $\bar{\lambda}$ satisfy the system*

$$\bar{u}, \bar{p} \in L^2(\Omega), \qquad \bar{y} \in H_0^1(\Omega) \cap H^2(\Omega), \qquad \bar{\lambda} \in \mathbb{R}^N,$$

$$\bar{\lambda}_n \geq 0 \ \ \forall n = 1, \ldots, N, \qquad \sum_{n=1}^N \bar{\lambda}_n = 1, \tag{4.6}$$

$$-\Delta \bar{y} + \max(0, \bar{y}) = \bar{u}, \qquad -\Delta \bar{p} + \mathbb{1}_{\{\bar{y}>0\}} \bar{p} = \sum_{n=1}^N \bar{\lambda}_n j_n'(\bar{y}), \qquad \bar{p} + \sum_{n=1}^N \bar{\lambda}_n \nu_n \bar{u} = 0.$$

ii) *A control $\bar{u} \in L^2(\Omega)$ with state $\bar{y} := S(\bar{u})$ is Pareto stationary for (P) (in the ordinary sense) if and only if there exist an adjoint state $\bar{p}$ and a multiplier $\bar{\lambda}$ such that $\bar{u}$, $\bar{y}$, $\bar{p}$, and $\bar{\lambda}$ satisfy the system*

$$\bar{u}, \bar{p} \in L^2(\Omega), \qquad \bar{y} \in H_0^1(\Omega) \cap H^2(\Omega), \qquad \bar{\lambda} \in \mathbb{R}^N,$$

$$\bar{\lambda}_n > 0 \ \ \forall n = 1, \ldots, N, \qquad \sum_{n=1}^N \bar{\lambda}_n = 1, \tag{4.7}$$

$$-\Delta \bar{y} + \max(0, \bar{y}) = \bar{u}, \qquad -\Delta \bar{p} + \mathbb{1}_{\{\bar{y}>0\}} \bar{p} = \sum_{n=1}^N \bar{\lambda}_n j_n'(\bar{y}), \qquad \bar{p} + \sum_{n=1}^N \bar{\lambda}_n \nu_n \bar{u} = 0.$$

iii) *A control $\bar{u} \in L^2(\Omega)$ is Pareto stationary for (P) (in the ordinary sense) if and only if it is properly Pareto stationary for (P).*

Proof. We begin with i): From Definition 3.2i), equation (4.3), and Lemma 4.2i), it follows that a control $\bar{u} \in L^2(\Omega)$ with state $\bar{y} := S(\bar{u})$ is weakly Pareto stationary for (P) if and only if

$$\nexists z \in H_0^1(\Omega) \cap H^2(\Omega):$$
$$\langle j_n'(\bar{y}), z \rangle_{H_0^1 \cap H^2} + \nu_n \left(\bar{u}, S'(\bar{u}; \cdot)^{-1}(z) \right)_{L^2} = \left\langle j_n'(\bar{y}) + \nu_n \left(-\Delta \bar{u} + \mathbb{1}_{\{\bar{y}>0\}} \bar{u} \right), z \right\rangle_{H_0^1 \cap H^2} < 0 \quad \forall n = 1, \ldots, N. \tag{4.8}$$

Here, the Laplacian $\Delta \bar{u}$ is again understood in the very weak sense. Due to Lemma 4.4, we further know that (4.8) is equivalent to the statement

$$\exists \bar{\lambda} \in \mathbb{R}^N: \quad \bar{\lambda}_n \geq 0 \ \ \forall n = 1, \ldots, N, \quad \sum_{n=1}^N \bar{\lambda}_n = 1,$$

$$\sum_{n=1}^N \bar{\lambda}_n \left(j_n'(\bar{y}) + \nu_n \left(-\Delta \bar{u} + \mathbb{1}_{\{\bar{y}>0\}} \bar{u} \right) \right) = 0 \in \left(H_0^1(\Omega) \cap H^2(\Omega) \right)^*. \tag{4.9}$$

If we now define $\bar{p} := -\sum_{n=1}^N \bar{\lambda}_n \nu_n \bar{u}$, then the assertion in i) follows immediately. To establish ii), we can use exactly the same arguments as for i) (with (4.4) replaced by (4.5)). It remains to prove iii). To this end, let us suppose that $\bar{u}$ is Pareto stationary in the ordinary sense, that $\bar{p}$ and $\bar{\lambda}$ are as in (4.7), and that we are given a $v \in L^2(\Omega)$ with associated directional derivative $z := S'(\bar{u}; v) \in H_0^1(\Omega) \cap H^2(\Omega)$ such that $\langle j_l'(\bar{y}), S'(\bar{u}; v) \rangle_{H_0^1 \cap H^2} + \nu_l (\bar{u}, v)_{L^2} < 0$ holds for some $l \in \{1, \ldots, N\}$. Then, Lemma 4.2iii) yields

$$\langle j_l'(\bar{y}), S'(\bar{u}; v) \rangle_{H_0^1 \cap H^2} + \nu_l (\bar{u}, v)_{L^2} = \left\langle j_l'(\bar{y}) + \nu_l \left(-\Delta \bar{u} + \mathbb{1}_{\{\bar{y}>0\}} \bar{u} \right), z \right\rangle_{H_0^1 \cap H^2},$$

and we may use the strong stationarity system (4.7) and the same arguments as in (4.9) to deduce that

$$\sum_{n=1}^{N} \bar{\lambda}_n \langle j_n'(\bar{y}) + \nu_n \left(-\Delta \bar{u} + \mathbb{1}_{\{\bar{y}>0\}} \bar{u} \right), z \rangle_{H_0^1 \cap H^2} = 0.$$

The above implies (completely analogously to (2.7), again by Lemma 4.2iii), and since $\bar{\lambda}_n > 0$ for all n) that

$$
\begin{aligned}
0 &< - \left(\langle j_l'(\bar{y}), S'(\bar{u}; v) \rangle_{H_0^1 \cap H^2} + \nu_l \left(\bar{u}, v \right)_{L^2} \right) \\
&= \frac{1}{\bar{\lambda}_l} \sum_{n \neq l} \bar{\lambda}_n \langle j_n'(\bar{y}) + \nu_n \left(-\Delta \bar{u} + \mathbb{1}_{\{\bar{y}>0\}} \bar{u} \right), z \rangle_{H_0^1 \cap H^2} \\
&\leq \frac{1}{\bar{\lambda}_l} \max_{n=1,\dots,N} \langle j_n'(\bar{y}) + \nu_n \left(-\Delta \bar{u} + \mathbb{1}_{\{\bar{y}>0\}} \bar{u} \right), z \rangle_{H_0^1 \cap H^2} \\
&\leq \left(\min_{n=1,\dots,N} \bar{\lambda}_n \right)^{-1} \max_{n=1,\dots,N} \left(\langle j_n'(\bar{y}), S'(\bar{u}; v) \rangle_{H_0^1 \cap H^2} + \nu_n \left(\bar{u}, v \right)_{L^2} \right).
\end{aligned}
$$

The proper Pareto stationarity of $\bar{u}$ now follows immediately, see Definition 3.2iii). Since the reverse implication is trivial, this completes the proof. $\qquad\square$

Several things are noteworthy regarding the last result:

Remark 4.6.

i) The structure of the strong stationarity systems in Theorem 4.5 is completely analogous to that of classical first-order necessary optimality conditions for smooth, finite-dimensional multiobjective optimization problems, see, *e.g.*, [22], Section 3.3, [39], Section 3.1.1, and [27], Section 4. In particular, (4.6) and (4.7) do not contain any kind of evaluation of the subdifferential of the function $\max(0, \cdot)$. Section 6 will give an explanation for this behavior.

ii) From [14], Theorem 3.18, it follows straightforwardly that the lines

$$-\Delta \bar{p} + \mathbb{1}_{\{\bar{y}>0\}} \bar{p} = \sum_{n=1}^{N} \bar{\lambda}_n j_n'(\bar{y}) \qquad \text{and} \qquad \bar{p} + \sum_{n=1}^{N} \bar{\lambda}_n \nu_n \bar{u} = 0$$

in the strong stationarity conditions (4.6) and (4.7) imply the existence of a generalized derivative $G \in \partial_B^{ss} S(\bar{u})$ with

$$\sum_{n=1}^{N} \bar{\lambda}_n \left(G^* j_n'(\bar{y}) + \nu_n \bar{u} \right) = 0. \tag{4.10}$$

Here, $\partial_B^{ss} S(\bar{u}) \subset L(L^2(\Omega), H_0^1(\Omega) \cap H^2(\Omega))$ is the strong-strong Bouligand subdifferential of the solution map S associated with (2.1) in the sense of [14], Definition 3.1, and $G^* \in L((H_0^1(\Omega) \cap H^2(\Omega))^*, L^2(\Omega))$ denotes the adjoint of G. Due to the chain rule for the Bouligand subdifferential ∂_B and the fact that the Bouligand subdifferential is smaller than the subdifferential of Clarke (which we denote by ∂_C in the following), (4.10) further yields that

$$0 \in \sum_{n=1}^{N} \bar{\lambda}_n \left(\partial_B (j_n \circ S)(\bar{u}) + \nu_n \bar{u} \right) \subset \sum_{n=1}^{N} \bar{\lambda}_n \partial_B \left(J_n(S(\cdot), \cdot) \right)(\bar{u}) \subset \sum_{n=1}^{N} \bar{\lambda}_n \partial_C \left(J_n(S(\cdot), \cdot) \right)(\bar{u}).$$

The strong stationarity conditions (4.6) and (4.7) thus imply that there exist elements of the Clarke sub-differentials $\partial_C J_n(S(\cdot), \cdot)$ of the reduced objective functions $J_n(S(\cdot), \cdot)$, $n = 1, \ldots, N$, of (P) at $\bar{u}$ such that a linear combination of these elements with the coefficients $\bar{\lambda}_n$ vanishes. Note that this is precisely the "standard" first-order necessary optimality condition for non-smooth multiobjective optimization problems, see, *e.g.*, [10, 30, 39]. The systems (4.6) and (4.7) are thus more rigorous than ordinary optimality conditions based on Clarke's generalized differential (and also stronger than conditions which exploit the subdifferential calculus of [14]). This, along with the equivalence to the purely primal stationarity concepts in Definition 3.2, justifies calling (4.6) and (4.7) *strong* stationarity conditions.

iii) Theorem 4.5 is closely related to classical scalarization approaches (as found, *e.g.*, in [22], Sects. 3 and 4) in the sense that the systems (4.6) and (4.7) can be identified with strong stationarity conditions (in the ordinary, single-objective sense) of optimal control problems governed by (2.1) with objective functions of the form $u \mapsto \sum_{n=1}^{N} \bar{\lambda}_n J_n(S(u), u)$. Compare, *e.g.*, with [14], Theorem 4.12 in this context. The main insight provided by Theorem 4.5 is, of course, that every weak, ordinary, or proper stationary point of (P) is a strongly stationary point of a suitably scalarized auxiliary problem.

iv) We would like to point out that the proof of Theorem 4.5 relies heavily on the fact that the non-smooth Nemytskii operator $\max(0, \cdot)$ in the PDE (2.1) interacts in a special way with the Tikhonov regularization terms in the objective functions of (P) and that the control space of (P) is all of $L^2(\Omega)$. If more general non-smooth Nemytskii operators or objective functions $J_n : H_0^1(\Omega) \cap H^2(\Omega) \times L^2(\Omega) \to \mathbb{R}$ are considered or if the controls are subject to additional constraints, then it is still often possible to establish stationarity systems similar to those in (4.6) and (4.7) by using a scalarization approach and regularization. However, in this general situation, it is typically completely unclear how the resulting optimality conditions are related to the notions of stationarity in Definition 3.2 (or if they are of a sensible strength at all). Compare, for instance, with the optimality conditions derived by regularization for scalar optimal control problems with general objective functions governed by non-smooth semilinear PDEs in [14], Section 4.1 in this context and also with the analysis of Section 5.

As a direct consequence of Theorem 4.5, we obtain:

Corollary 4.7 (Increased regularity of Pareto stationary points)**.**

i) *If the maps j_n, $n = 1, \ldots, N$, are continuously differentiable as functions from $H_0^1(\Omega)$ to $\mathbb{R}$, i.e., if there exist continuously differentiable $\hat{j}_n : H_0^1(\Omega) \to \mathbb{R}$ such that $j_n = \hat{j}_n \circ E_{H_0^1}$ holds for all n, where $E_{H_0^1} : H_0^1(\Omega) \cap H^2(\Omega) \to H_0^1(\Omega)$ denotes the canonical embedding of $H_0^1(\Omega) \cap H^2(\Omega)$ into $H_0^1(\Omega)$, then every Pareto stationary control $\bar{u}$ of (P) is an element of $H_0^1(\Omega)$.*

ii) *If the maps j_n, $n = 1, \ldots, N$, are continuously differentiable as functions from $L^2(\Omega)$ to $\mathbb{R}$, i.e., if there exist continuously differentiable $\hat{j}_n : L^2(\Omega) \to \mathbb{R}$ such that $j_n = \hat{j}_n \circ E_{L^2}$ holds for all n, where $E_{L^2} : H_0^1(\Omega) \cap H^2(\Omega) \to L^2(\Omega)$ denotes the canonical embedding of $H_0^1(\Omega) \cap H^2(\Omega)$ into $L^2(\Omega)$, then every Pareto stationary control $\bar{u}$ of (P) is an element of $H_0^1(\Omega) \cap H^2(\Omega)$.*

Proof. From Theorem 4.5, we obtain that, for every Pareto stationary $\bar{u} \in L^2(\Omega)$ with state $\bar{y} := S(\bar{u})$, we can find a multiplier $\bar{\lambda} \in \mathbb{R}^N$ and an adjoint state $\bar{p} \in L^2(\Omega)$ with

$$\bar{\lambda}_n > 0 \ \ \forall n = 1, \ldots, N, \qquad \sum_{n=1}^{N} \bar{\lambda}_n = 1, \qquad \bar{p} + \sum_{n=1}^{N} \bar{\lambda}_n \nu_n \bar{u} = 0, \qquad -\Delta \bar{p} + \mathbb{1}_{\{\bar{y}>0\}} \bar{p} = \sum_{n=1}^{N} \bar{\lambda}_n j_n'(\bar{y}). \quad (4.11)$$

If we assume that there exist continuously differentiable functions $\hat{j}_n : H_0^1(\Omega) \to \mathbb{R}$ such that $j_n = \hat{j}_n \circ E_{H_0^1}$ holds for all n, then it follows from (4.11), the chain rule, and the definitions of the very weak Laplacian and

the dual pairings in $H_0^1(\Omega)$ and $H_0^1(\Omega) \cap H^2(\Omega)$ that $\bar{p}$ satisfies

$$(\bar{p}, -\Delta z)_{L^2} = \langle -\Delta \bar{p}, z \rangle_{H_0^1 \cap H^2} = \left\langle -\mathbb{1}_{\{\bar{y}>0\}}\bar{p} + \sum_{n=1}^N \bar{\lambda}_n j_n'(\bar{y}), z \right\rangle_{H_0^1 \cap H^2} = \left\langle -\mathbb{1}_{\{\bar{y}>0\}}\bar{p} + \sum_{n=1}^N \bar{\lambda}_n \hat{j}_n'(E_{H_0^1}\bar{y}), z \right\rangle_{H_0^1}$$

for all $z \in H_0^1(\Omega) \cap H^2(\Omega)$. Let us denote the element of $H^{-1}(\Omega)$ appearing on the right-hand side of the last equation with f. Then, it follows from classical results on the Poisson problem that there exists a unique solution $\tilde{p} \in H_0^1(\Omega)$ of the variational identity $(\nabla \tilde{p}, \nabla z)_{L^2} = \langle f, z \rangle_{H_0^1}$ for all $z \in H_0^1(\Omega)$. From Green's formula, we obtain that this $\tilde{p}$ also satisfies $(\tilde{p}, -\Delta z)_{L^2} = \langle f, z \rangle_{H_0^1} = (\bar{p}, -\Delta z)_{L^2}$ for all $z \in H_0^1(\Omega) \cap H^2(\Omega)$ and, as a consequence, that $(\tilde{p} - \bar{p}, -\Delta z)_{L^2} = 0$ for all $z \in H_0^1(\Omega) \cap H^2(\Omega)$. Using our assumptions on Ω and classical results on the H^2-regularity of solutions of the Poisson problem, see [25], Theorem 9.15, Lemma 9.17 and [28], Theorem 3.2.1.2, we may further deduce that there exists a unique $z \in H_0^1(\Omega) \cap H^2(\Omega)$ satisfying $-\Delta z = \tilde{p} - \bar{p} \in L^2(\Omega)$. Choosing this z as a test function yields $(\tilde{p} - \bar{p}, -\Delta z)_{L^2} = (\tilde{p} - \bar{p}, \tilde{p} - \bar{p})_{L^2} = 0$. We thus have $\bar{p} = \tilde{p} \in H_0^1(\Omega)$ and, since the number $\sum_{n=1}^N \bar{\lambda}_n \nu_n$ is positive by our assumption $\nu_N > 0$ and due to the properties of $\bar{\lambda}$, $\bar{u} \in H_0^1(\Omega)$. This establishes i). The proof of ii) is completely along the same lines. □

We conclude this section with an existence result that shows that it makes sense to use the strong stationarity conditions in Theorem 4.5 for the calculation of Pareto stationary points:

Lemma 4.8 (Solvability of the strong stationarity system). *Suppose that a vector $\bar{\lambda} \in \mathbb{R}^N$ satisfying*

$$\bar{\lambda}_n \geq 0 \ \ \forall n = 1, \dots, N, \qquad \sum_{n=1}^N \bar{\lambda}_n = 1, \quad and \quad \bar{\lambda}_N > 0$$

is given. Then, there exists at least one solution $(\bar{u}, \bar{y}, \bar{p})$ of the system

$$\bar{u}, \bar{p} \in L^2(\Omega), \qquad \bar{y} \in H_0^1(\Omega) \cap H^2(\Omega),$$

$$-\Delta \bar{y} + \max(0, \bar{y}) = \bar{u}, \qquad -\Delta \bar{p} + \mathbb{1}_{\{\bar{y}>0\}}\bar{p} = \sum_{n=1}^N \bar{\lambda}_n j_n'(\bar{y}), \qquad \bar{p} + \sum_{n=1}^N \bar{\lambda}_n \nu_n \bar{u} = 0.$$

Proof. Since $\bar{\lambda}_N$ is positive and since $\nu_N > 0$ holds by our standing assumptions, it follows completely analogously to the proof of Theorem 2.4 that there exists at least one global solution $\bar{u} \in L^2(\Omega)$ of the problem

$$\min_{u \in L^2(\Omega)} \sum_{n=1}^N \bar{\lambda}_n \left(j_n(S(u)) + \frac{\nu_n}{2}\|u\|_{L^2}^2 \right), \tag{4.12}$$

and from the first-order necessary optimality condition of (4.12) and the properties of S in Proposition 2.2, we obtain that this $\bar{u}$ and its state $\bar{y} := S(\bar{u})$ have to satisfy

$$\sum_{n=1}^N \bar{\lambda}_n \left(\langle j_n'(\bar{y}), S'(\bar{u}; v) \rangle_{H_0^1 \cap H^2} + \nu_n (\bar{u}, v)_{L^2} \right) \geq 0 \qquad \forall v \in L^2(\Omega). \tag{4.13}$$

Using Lemma 4.2 and exactly the same arguments as in the proof of Theorem 4.5, we can rewrite (4.13) as

$$\sum_{n=1}^N \bar{\lambda}_n \left(\langle j_n'(\bar{y}) + \nu_n \left(-\Delta \bar{u} + \mathbb{1}_{\{\bar{y}>0\}}\bar{u} \right), z \rangle_{H_0^1 \cap H^2} \right) \geq 0 \qquad \forall z \in H_0^1(\Omega) \cap H^2(\Omega),$$

and this, in turn, entails

$$\sum_{n=1}^{N} \bar{\lambda}_n \left(j_n'(\bar{y}) + \nu_n \left(-\Delta \bar{u} + \mathbb{1}_{\{\bar{y}>0\}} \bar{u} \right) \right) = 0 \in \left(H_0^1(\Omega) \cap H^2(\Omega) \right)^*.$$

If we now define $\bar{p} := -\sum_{n=1}^{N} \bar{\lambda}_n \nu_n \bar{u}$, then the claim follows immediately. $\qquad\square$

Note that Theorem 4.5 and Lemma 4.8 show that, to calculate Pareto stationary points of the problem (P), it is a good strategy to fix vectors $\bar{\lambda} \in \mathbb{R}^N$ with $\bar{\lambda}_n \geq 0$ for all $n = 1, \ldots, N$ and $\sum_{n=1}^{N} \bar{\lambda}_n = 1$, and to subsequently try to solve the remaining equations in the strong stationarity condition (4.6) for $(\bar{u}, \bar{y}, \bar{p})$. Due to Lemma 4.8, we know that the system that is obtained along these lines has at least one solution for all $\bar{\lambda}$ with $\bar{\lambda}_N > 0$, and from Theorem 4.5 it follows that every point that we calculate in this way is weakly or, in the case $\bar{\lambda}_n > 0$ for all n, properly Pareto stationary. Moreover, Theorem 4.5 yields that we can determine all weak (respectively, proper, respectively, ordinary) Pareto stationary points of (P) by solving systems of the form (4.6) (respectively, (4.7)). We will get back to this topic in Section 7, where we will use a similar approach in the discrete setting for our numerical experiments.

5. Relation to regularization approaches and C-stationarity

The aim of this section is to demonstrate that the strong stationarity conditions in Theorem 4.5 are not only interesting for their own sake but also have quite surprising consequences for the analysis of regularization techniques for problems of the type (P). To be more precise, in what follows, we will show that, when a problem of the form (P) satisfying the conditions of Corollary 4.7i) is regularized and the regularization parameter is driven to zero, then all weak L^2-accumulation points of the weakly Pareto stationary points of the regularized multiobjective optimal control problems are weakly Pareto stationary for the unregularized limit problem (P). Note that this observation is indeed remarkable since similar effects cannot be observed even in very simple one-dimensional examples. Compare, *e.g.*, with the situation already mentioned in the introduction where the function $f(x) := -|x|$ is approximated by the family $f_\varepsilon(x) := -\sqrt{x^2 + \varepsilon}$, $\varepsilon > 0$, in this context.

To regularize the problem (P), we follow the lines of [14] and replace the max-function in (2.1) with a suitably chosen differentiable approximation $\max_\varepsilon : \mathbb{R} \to \mathbb{R}$. This gives rise to a family of regularized multiobjective optimal control problems of the form

$$\text{Minimize} \quad \begin{pmatrix} J_1(y_\varepsilon, u) := j_1(y_\varepsilon) + \dfrac{\nu_1}{2} \|u\|_{L^2}^2 \\ \vdots \\ J_N(y_\varepsilon, u) := j_N(y_\varepsilon) + \dfrac{\nu_N}{2} \|u\|_{L^2}^2 \end{pmatrix} \tag{P_ε}$$

$$\text{w.r.t.} \quad u \in L^2(\Omega), \quad y_\varepsilon \in H_0^1(\Omega) \cap H^2(\Omega),$$

$$\text{s.t.} \quad -\Delta y_\varepsilon + \max_\varepsilon(y_\varepsilon) = u \text{ a.e. in } \Omega.$$

Our standing assumptions on the approximations $\max_\varepsilon$ are the same as in [14]:

Assumption 5.1 (Standing assumptions on the functions $\max_\varepsilon$). The functions $\max_\varepsilon : \mathbb{R} \to \mathbb{R}$, $\varepsilon > 0$, satisfy the following:

- i) It holds $\max_\varepsilon \in C^1(\mathbb{R})$ for all $\varepsilon > 0$.
- ii) There is a constant $C > 0$ with $|\max_\varepsilon(x) - \max(0, x)| \leq C\varepsilon$ for all $x \in \mathbb{R}$.
- iii) For all $x \in \mathbb{R}$ and all $\varepsilon > 0$, we have $0 \leq \max_\varepsilon'(x) \leq 1$.
- iv) For every arbitrary but fixed $\delta > 0$, the derivatives $\{\max_\varepsilon'\}_{\varepsilon>0}$ converge uniformly to one in $[\delta, \infty)$ and uniformly to zero in $(-\infty, -\delta]$ for $\varepsilon \to 0^+$.

Note that, *e.g.*, the family $\max_\varepsilon(x) := \frac{1}{2}\left(\sqrt{x^2 + \varepsilon^2} + x\right)$, $\varepsilon > 0$, has all of the above properties. Under Assumption 5.1, the following can be established for the PDE in (P_ε), *cf.* [14], Section 4.1:

Proposition 5.2 (Properties of the regularized PDE)**.** *For every* $u \in L^2(\Omega)$ *and every* $\varepsilon > 0$, *there exists a unique solution* $y_\varepsilon \in H_0^1(\Omega) \cap H^2(\Omega)$ *of the PDE*

$$-\Delta y_\varepsilon + \max_\varepsilon(y_\varepsilon) = u \ \ a.e. \ in \ \Omega. \tag{5.1}$$

Further, the solution operator $S_\varepsilon : L^2(\Omega) \to H_0^1(\Omega) \cap H^2(\Omega)$, $u \mapsto y_\varepsilon$, *associated with the partial differential equation (5.1) satisfies:*

i) S_ε *is weakly continuous (in the sense of (2.2)).*

ii) S_ε *is Fréchet differentiable, and the Fréchet derivative* $S_\varepsilon'(u) \in L(L^2(\Omega), H_0^1(\Omega) \cap H^2(\Omega))$ *of* S_ε *in a point* $u \in L^2(\Omega)$ *is precisely the solution map* $v \mapsto \delta_v$ *of the PDE*

$$-\Delta \delta_v + \max_\varepsilon'(y_\varepsilon)\delta_v = v.$$

iii) *There exists a constant* $C > 0$ *such that, for all* $u \in L^2(\Omega)$, *it holds*

$$\|S(u) - S_\varepsilon(u)\|_{H_0^1 \cap H^2} \leq C\,\varepsilon \quad \forall \varepsilon > 0.$$

iv) *For all* $\{u_k\} \subset L^2(\Omega)$ *and* $\{\varepsilon_k\} \subset (0, \infty)$ *satisfying* $u_k \rightharpoonup u$ *in* $L^2(\Omega)$ *for some* $u \in L^2(\Omega)$ *with state* $y := S(u)$ *and* $\varepsilon_k \to 0^+$, *there exist a subsequence* $\{k_l\}$ *and a function* $\chi \in L^\infty(\Omega)$ *such that*

$$\chi = 0 \ \ a.e. \ in \ \{y < 0\}, \quad \chi = 1 \ \ a.e. \ in \ \{y > 0\}, \quad \chi \in [0, 1] \ \ a.e. \ in \ \{y = 0\}$$

holds and such that the directional derivatives $S_{\varepsilon_{k_l}}'(u_{k_l})v$ *converge weakly in* $H_0^1(\Omega) \cap H^2(\Omega)$ *to the unique solution* η_v *of the partial differential equation*

$$-\Delta \eta_v + \chi \eta_v = v$$

for all $v \in L^2(\Omega)$.

Proof. The assertions of i), ii), and iii) have been proved in [14], Lemmas 4.2, 4.3. The proof of iv) is completely analogous to that of the second part of [14], Lemma 4.3 with the only difference that, in iv), we have to work with weak L^2-convergence instead of strong L^2-convergence. This, however, is not a problem: From the weak convergence $u_k \rightharpoonup u$ in $L^2(\Omega)$, iii), and the weak continuity of S, we obtain that

$$S_{\varepsilon_k}(u_k) = S(u_k) + (S_{\varepsilon_k} - S)(u_k) \rightharpoonup S(u) \quad \text{in } H_0^1(\Omega) \cap H^2(\Omega). \tag{5.2}$$

The above implies in particular that $S_{\varepsilon_k}(u_k)$ converges to $S(u)$ in $L^2(\Omega)$ and, at least after the transition to a subsequence (still denoted by the same symbol), that $S_{\varepsilon_k}(u_k) \to S(u) = y$ pointwise a.e. in Ω. Using this pointwise-a.e. convergence, we can argue exactly as in the second part of the proof of [14], Lemma 4.3 to establish the assertion of iv). This completes the proof. $\square$

We are now in the position to prove that weak L^2-accumulation points of weakly Pareto stationary points of (P_ε) are indeed weakly Pareto stationary for the original problem (P) in the situation of Corollary 4.7i).

Theorem 5.3 (Preservation of weak Pareto stationarity in the limit $\varepsilon \to 0^+$)**.** *Suppose that the maps* j_n *are continuously differentiable as functions from* $H_0^1(\Omega)$ *to* $\mathbb{R}$ *(in the sense of Cor. 4.7i)), and that sequences*

$\{\varepsilon_k\} \subset (0, \infty)$ and $\{\bar{u}_k\} \subset L^2(\Omega)$ are given such that $\varepsilon_k \to 0^+$ holds and such that $\bar{u}_k$ is weakly Pareto stationary for (P_{ε_k}) for all k, i.e., such that there is no $v \in L^2(\Omega)$ satisfying

$$\langle j_n'(S_{\varepsilon_k}(\bar{u}_k)), S_{\varepsilon_k}'(\bar{u}_k)v \rangle_{H_0^1 \cap H^2} + \nu_n (\bar{u}_k, v)_{L^2} < 0 \qquad \forall n = 1, \dots, N. \tag{5.3}$$

Then, every weak L^2-accumulation point of $\{\bar{u}_k\}$ is weakly Pareto stationary for (P).

Proof. Suppose that $\bar{u} \in L^2(\Omega)$ is a weak L^2-accumulation point of the sequence $\{\bar{u}_k\}$ for $k \to \infty$ with state $\bar{y} := S(\bar{u})$. Then, Proposition 5.2iv) yields that we can pass over to subsequences of $\{\bar{u}_k\}$ and $\{\varepsilon_k\}$ (still denoted by the same symbols) such that $\bar{u}_k \rightharpoonup \bar{u}$ holds in $L^2(\Omega)$ and such that there exists a $\chi \in L^\infty(\Omega)$ with

$$\chi = 0 \text{ a.e. in } \{\bar{y} < 0\}, \quad \chi = 1 \text{ a.e. in } \{\bar{y} > 0\}, \quad \chi \in [0, 1] \text{ a.e. in } \{\bar{y} = 0\},$$

and $S_{\varepsilon_k}'(\bar{u}_k)v \rightharpoonup \eta_v$ in $H_0^1(\Omega) \cap H^2(\Omega)$ for all $v \in L^2(\Omega)$, where η_v is again defined by

$$-\Delta \eta_v + \chi \eta_v = v.$$

Since the left-hand side of (5.3) is linear in v, we may further invoke Lemma 4.4 to deduce that, for all $k \in \mathbb{N}$, we can find a $\bar{\lambda}^k \in \mathbb{R}^N$ with

$$\bar{\lambda}_n^k \geq 0 \;\; \forall n = 1, \dots, N, \qquad \sum_{n=1}^N \bar{\lambda}_n^k = 1,$$
$$\sum_{n=1}^N \bar{\lambda}_n^k \left(\langle j_n'(S_{\varepsilon_k}(\bar{u}_k)), S_{\varepsilon_k}'(\bar{u}_k)v \rangle_{H_0^1 \cap H^2} + \nu_n (\bar{u}_k, v)_{L^2} \right) = 0 \quad \forall v \in L^2(\Omega). \tag{5.4}$$

Note that the first line of (5.4) particularly implies that $\{\bar{\lambda}^k\}$ is bounded, and that our assumptions on the functions j_n and (5.2) yield that $j_n'(S_{\varepsilon_k}(\bar{u}_k))$ converges strongly to $j_n'(\bar{y})$ in $(H_0^1(\Omega) \cap H^2(\Omega))^*$ for $k \to \infty$. This allows us to find a further subsequence (again not relabeled) such that $\bar{\lambda}^k \to \bar{\lambda}$ holds for some $\bar{\lambda} \in \mathbb{R}^N$, and to pass to the limit $k \to \infty$ in (5.4) to obtain that there exists a $\bar{\lambda} \in \mathbb{R}^N$ with

$$\bar{\lambda}_n \geq 0 \;\; \forall n = 1, \dots, N, \qquad \sum_{n=1}^N \bar{\lambda}_n = 1, \qquad \sum_{n=1}^N \bar{\lambda}_n \left(\langle j_n'(\bar{y}), \eta_v \rangle_{H_0^1 \cap H^2} + \nu_n (\bar{u}, v)_{L^2} \right) = 0 \quad \forall v \in L^2(\Omega). \tag{5.5}$$

Due to the bijectivity of the map $L^2(\Omega) \ni v \mapsto \eta_v \in H_0^1(\Omega) \cap H^2(\Omega)$ and the identity $\chi \bar{u} = \chi(-\Delta \bar{y} + \max(0, \bar{y})) = \mathbb{1}_{\{\bar{y} > 0\}}(-\Delta \bar{y} + \max(0, \bar{y})) = \mathbb{1}_{\{\bar{y} > 0\}} \bar{u}$, see Lemma 4.1, the last equation in (5.5) can be recast as

$$\sum_{n=1}^N \bar{\lambda}_n \left(\langle j_n'(\bar{y}), z \rangle_{H_0^1 \cap H^2} + \nu_n (\bar{u}, -\Delta z + \chi z)_{L^2} \right)$$
$$= \sum_{n=1}^N \bar{\lambda}_n \langle j_n'(\bar{y}) + \nu_n \left(-\Delta \bar{u} + \mathbb{1}_{\{\bar{y} > 0\}} \bar{u} \right), z \rangle_{H_0^1 \cap H^2} = 0 \quad \forall z \in H_0^1(\Omega) \cap H^2(\Omega).$$

If we now define $\bar{p} := -\sum_{n=1}^N \bar{\lambda}_n \nu_n \bar{u}$, then it follows that $\bar{u}$ satisfies the strong stationarity system (4.6) and, as a consequence, the weak Pareto stationarity condition in Definition 3.2i). This completes the proof. $\square$

We would like to point out that, in the single-objective case $N = 1$, we can use exactly the same arguments as in the proof of Theorem 5.3 to establish that the notion of C-stationarity that is obtained by passing to

the limit with the regularization parameter in the first-order necessary optimality conditions of a regularized version of (P) is equivalent to the concept of strong stationarity. Indeed, we have:

Corollary 5.4 (Equivalence of C- and strong stationarity in the case $N = 1$). *Consider the single-objective case $N = 1$, i.e., the case where (P) has the form*

$$
\begin{aligned}
Minimize \quad & J(y, u) := j(y) + \frac{\nu}{2}\|u\|_{L^2}^2 \\
w.r.t. \quad & u \in L^2(\Omega), \quad y \in H_0^1(\Omega) \cap H^2(\Omega), \\
s.t. \quad & -\Delta y + \max(0, y) = u \ a.e. \ in \ \Omega
\end{aligned}
\tag{5.6}
$$

with a function $j : H_0^1(\Omega) \cap H^2(\Omega) \to \mathbb{R}$ as in Assumption 2.1 and $\nu > 0$. Then, the notions of C-stationarity in the sense of [14], Theorem 4.4, strong stationarity in the sense of [14], Theorem 4.12, and purely primal stationarity in the sense of [14], Proposition 4.10 are the same.

Proof. The equivalence between strong and primal stationarity follows from [14], Proposition 4.13 and the remaining equivalences are straightforward consequences of the systems (26) and (32) in [14], the identity $\bar{p} = -\nu\bar{u}$ in the stationarity conditions, and the fact that the controls $u = -\Delta y + \max(0, y) \in L^2(\Omega)$ in (5.6) always vanish a.e. in the set $\{y = 0\}$ by Lemma 4.1. □

We remark that an effect similar to that in Corollary 5.4 has already been observed for optimal control problems governed by a class of quasilinear PDEs with Gâteaux differentiable solution maps in [18], Section 5, *cf.* Remark 4.3. What is noteworthy about Corollary 5.4 is that, in the case of the problem (P), the notions of strong and C-stationarity are identical in spite of the fact that the solution operator $S : u \mapsto y$ associated with (2.1) typically does not possess a Gâteaux derivative, *cf.* Proposition 2.2iv). As we will see in the next section, this behavior can be explained with "hidden" smoothness properties of the problem (P) that only reveal themselves when (P) is considered as an optimization problem in the variable y.

6. HIDDEN SMOOTHNESS PROPERTIES AND AN ALTERNATIVE VIEW ON STRONG STATIONARITY CONDITIONS

One might ask at this point how the effects observed in the last two sections - the self-adjointness of the non-differentiable and non-linear operator $S'(u; \cdot)^{-1}$ in (1.2), the similarity of the strong stationarity conditions (4.6) and (4.7) in Theorem 4.5 to classical optimality conditions for smooth multiobjective optimal control problems, and the equivalence of the various notions of stationarity, *e.g.*, in Corollary 5.4 – are possible. The answer to this question is that the problem (P) is actually not as non-smooth as it appears at first glance. To see this, let us consider an arbitrary but fixed index $n \in \{1, \ldots, N\}$ and the associated objective

$$
J_n(y, u) := j_n(y) + \frac{\nu_n}{2}\|u\|_{L^2}^2
\tag{6.1}
$$

in (P). Using standard arguments, it is easy to check that the function J_n in (6.1) is typically non-differentiable when we reduce it to the variable u by expressing y in terms of the solution map S of the partial differential equation $-\Delta y + \max(0, y) = u$ in (2.1). Indeed, in the simple example

$$
d := 1, \qquad \Omega := (-1, 1), \qquad j_n(y) := (y(0) + 1)^2, \qquad \nu_n := 1,
\tag{6.2}
$$

a short calculation shows

$$
S(\alpha \mathbb{1}_\Omega)(x) =
\begin{cases}
\left(1 - \dfrac{2\mathrm{e}}{1 + \mathrm{e}^2} \cosh(x)\right)\alpha & \text{if } \alpha \geq 0 \\
\frac{1}{2}(1 - x^2)\alpha & \text{if } \alpha < 0
\end{cases},
\qquad \forall \alpha \in \mathbb{R},
$$

where $\mathbb{1}_\Omega$ denotes the element of $L^2(\Omega)$ that is identical one a.e. in Ω, and we obtain

$$J_n(S(\alpha\mathbb{1}_\Omega), \alpha\mathbb{1}_\Omega) = \begin{cases} \left(\dfrac{(1-\mathrm{e})^2}{1+\mathrm{e}^2}\,\alpha + 1\right)^2 + \alpha^2 & \text{if } \alpha \geq 0 \\ \left(\frac{1}{2}\alpha + 1\right)^2 + \alpha^2 & \text{if } \alpha < 0 \end{cases}, \qquad \forall \alpha \in \mathbb{R}.$$

The reduced objective function $u \mapsto J_n(S(u), u)$ thus has a proper kink at the origin in the situation of (6.2) and is non-differentiable as claimed. Note that this example also demonstrates that it makes sense to refer to the problem (P) as a non-smooth multiobjective optimal control problem. In what follows, we will see that the latter term is, in fact, not entirely appropriate anymore when we reduce J_n not to the control u but to the state y. (A reduction that is rarely used in the context of optimal control since y is, per definition, not the quantity that one has under control in the context of the problem (P).) To begin our analysis, we observe that, by eliminating the variable u from J_n, we arrive at a reduced objective function of the form

$$\begin{aligned} J_n(y, S^{-1}(y)) &= j_n(y) + \frac{\nu_n}{2}\| - \Delta y + \max(0, y)\|_{L^2}^2 \\ &= j_n(y) + \frac{\nu_n}{2}\|\Delta y\|_{L^2}^2 + \frac{\nu_n}{2}\|\max(0, y)\|_{L^2}^2 - \nu_n \int_\Omega \max(0, y)\Delta y \,\mathrm{d}x. \end{aligned} \tag{6.3}$$

Note that the first three terms on the right-hand side of (6.3) are trivially Gâteaux differentiable as functions from $H_0^1(\Omega) \cap H^2(\Omega)$ to $\mathbb{R}$, so that the only remaining potentially non-smooth term in (6.3) is the one involving the expression $\max(0, y)\Delta y$. For this type of product, however, we have the following key result:

Theorem 6.1 (Compensation of non-smoothness by weak derivatives)**.** *Let $k \in \mathbb{N}$ and $r, q \in [1, \infty)$ be given such that $r \in [1, q)$ and $W^{k,q}(\Omega) \hookrightarrow L^{qr/(q-r)}(\Omega)$ holds, and let $g : \mathbb{R} \to \mathbb{R}$ be a globally Lipschitz continuous function. Suppose further that there exists a countable set $\mathcal{N} \subset \mathbb{R}$ such that g possesses a classical derivative at all points $x \in \mathbb{R} \setminus \mathcal{N}$. Then, the map*

$$G : W^{k,q}(\Omega) \to L^r(\Omega), \qquad w \mapsto g(w)\partial^\alpha w,$$

is well-defined and Gâteaux differentiable for all multi-indices $\alpha \in \mathbb{N}_0^d$ with $|\alpha| = k$ (in the sense that directional derivatives exist and depend linearly and continuously on the direction) and the Gâteaux derivative of G in a point w in a direction z is given by

$$G'(w)z = \left(\mathbb{1}_{\{w \notin \mathcal{N}\}} g'(w)\partial^\alpha w\right) z + g(w)\partial^\alpha z. \tag{6.4}$$

Proof. Suppose that k, q, r, and g satisfy the assumptions of the theorem, and assume that an arbitrary but fixed $\alpha \in \mathbb{N}_0^d$ with $|\alpha| = k$ is given. Then, Hölder's inequality, the triangle inequality, and the global Lipschitz continuity of g imply that there exists an absolute constant $C > 0$ with

$$\begin{aligned} \|g(w)\partial^\alpha w\|_{L^r} &\leq \|(g(w) - g(0))\partial^\alpha w\|_{L^r} + \|g(0)\partial^\alpha w\|_{L^r} \\ &\leq C\,\|w\partial^\alpha w\|_{L^r} + C\,\|\partial^\alpha w\|_{L^q} \\ &\leq C\|w\|_{L^{qr/(q-r)}}\,\|\partial^\alpha w\|_{L^q} + C\,\|\partial^\alpha w\|_{L^q} \qquad \forall w \in W^{k,q}(\Omega). \end{aligned} \tag{6.5}$$

This proves that G is well-defined as a function from $W^{k,q}(\Omega)$ to $L^r(\Omega)$. It remains to show that G is Gâteaux differentiable. To this end, we note that the same arguments as in (6.5), the dominated convergence theorem,

and Lemma 4.1 yield that, for all $w, z \in W^{k,q}(\Omega)$, we have

$$
\begin{aligned}
0 &\leq \left\| \frac{g(w+tz)\partial^\alpha(w+tz) - g(w)\partial^\alpha w}{t} - \left(\mathbb{1}_{\{w \notin \mathcal{N}\}} g'(w)\partial^\alpha w \right) z - g(w)\partial^\alpha z \right\|_{L^r} \\
&\leq \left\| \left(\frac{g(w+tz)-g(w)}{t} - \mathbb{1}_{\{w \notin \mathcal{N}\}} g'(w)z \right) \partial^\alpha w \right\|_{L^r} + \left\| (g(w+tz)-g(w))\partial^\alpha z \right\|_{L^r} \\
&= \left\| \left(\frac{g(w+tz)-g(w)}{t} - g'(w)z \right) \mathbb{1}_{\{w \notin \mathcal{N}\}} \partial^\alpha w \right\|_{L^r} + \left\| (g(w+tz)-g(w))\partial^\alpha z \right\|_{L^r} \\
&\leq \left\| \left(\frac{g(w+tz)-g(w)}{t} - g'(w)z \right) \mathbb{1}_{\{w \notin \mathcal{N}\}} \right\|_{L^{qr/(q-r)}} \|\partial^\alpha w\|_{L^q} + Ct \|z\|_{L^{qr/(q-r)}} \|\partial^\alpha z\|_{L^q} \\
&\to 0
\end{aligned}
$$

for $t \to 0^+$, where C again denotes a generic constant. The function G is thus indeed directionally differentiable with the derivative in (6.4). Since the linearity and the continuity of the map $W^{k,q}(\Omega) \ni z \mapsto G'(w; z) \in L^r(\Omega)$ are trivial, the assertion of the theorem now follows immediately. $\qquad \square$

If we consider the special case $k = 2$, $q = 2$, and $r = 1$, then Theorem 6.1 implies in particular that the function

$$
F : H_0^1(\Omega) \cap H^2(\Omega) \to L^1(\Omega), \qquad y \mapsto \max(0, y)\Delta y,
$$

is Gâteaux differentiable with derivative $F'(y)z = \mathbb{1}_{\{y>0\}}\Delta y \, z + \max(0, y)\Delta z$ for all $y, z \in H_0^1(\Omega) \cap H^2(\Omega)$. The map $J_n(\cdot, S^{-1}(\cdot))$ in (6.3) thus possesses a Gâteaux derivative, and we arrive at the – quite counterintuitive – conclusion that, although typically non-smooth as a function of the control u, the objective J_n in (6.1) is always Gâteaux differentiable when reduced to the state y.

Note that the rather peculiar behavior that we observe here is the main reason why the analysis of Section 4 works as well as it does. Indeed, by invoking (6.4), we obtain that the Gâteaux derivative of the map $J_n(\cdot, S^{-1}(\cdot))$ in (6.3) at a point $y \in H_0^1(\Omega) \cap H^2(\Omega)$ is precisely the functional

$$
j_n'(y) + \nu_n \left(\Delta\Delta y + \max(0, y) - \mathbb{1}_{\{y>0\}}\Delta y - \Delta \max(0, y) \right) \in (H_0^1(\Omega) \cap H^2(\Omega))^*,
$$

where the Laplacian is again understood in the very weak sense. If we now assume that we are given a state $\bar{y}$ which is Pareto stationary for the reduced multiobjective optimization problem

$$
\text{Minimize} \quad \begin{pmatrix} J_1(y, S^{-1}(y)) = j_1(y) + \dfrac{\nu_1}{2} \|S^{-1}(y)\|_{L^2}^2 \\ \vdots \\ J_N(y, S^{-1}(y)) = j_N(y) + \dfrac{\nu_N}{2} \|S^{-1}(y)\|_{L^2}^2 \end{pmatrix} \quad \text{w.r.t.} \quad y \in H_0^1(\Omega) \cap H^2(\Omega), \qquad (6.6)
$$

i.e., a $\bar{y} \in H_0^1(\Omega) \cap H^2(\Omega)$ with the property that there is no $z \in H_0^1(\Omega) \cap H^2(\Omega)$ with

$$
\begin{aligned}
\left\langle j_n'(\bar{y}) + \nu_n \left(\Delta\Delta\bar{y} + \max(0, \bar{y}) - \mathbb{1}_{\{\bar{y}>0\}}\Delta\bar{y} - \Delta\max(0, \bar{y}) \right), z \right\rangle_{H_0^1 \cap H^2} &\leq 0 \\
&\forall n = 1, \dots, N, \\
\left\langle j_n'(\bar{y}) + \nu_n \left(\Delta\Delta\bar{y} + \max(0, \bar{y}) - \mathbb{1}_{\{\bar{y}>0\}}\Delta\bar{y} - \Delta\max(0, \bar{y}) \right), z \right\rangle_{H_0^1 \cap H^2} &< 0 \\
&\text{for at least one } n \in \{1, \dots, N\},
\end{aligned}
\qquad (6.7)
$$

then Lemma 4.4 yields that there exists a vector $\bar{\lambda} \in \mathbb{R}^N$ satisfying

$$\bar{\lambda}_n > 0 \ \ \forall n = 1, \dots, N, \qquad \sum_{n=1}^{N} \bar{\lambda}_n = 1,$$

$$\sum_{n=1}^{N} \bar{\lambda}_n \left(j_n'(\bar{y}) + \nu_n \left(\Delta\Delta\bar{y} + \max(0, \bar{y}) - \mathbb{1}_{\{\bar{y}>0\}} \Delta\bar{y} - \Delta\max(0, \bar{y}) \right) \right) = 0. \tag{6.8}$$

With the definitions $\bar{u} := -\Delta\bar{y} + \max(0, \bar{y})$ and $\bar{p} := -\sum_{n=1}^{N} \bar{\lambda}_n \nu_n \bar{u}$, (6.8) takes precisely the form of the strong stationarity system (4.7). We may thus conclude:

Theorem 6.2 (Alternative interpretation of strong stationarity conditions)**.** *The strong stationarity conditions (4.6) and (4.7) are precisely the conditions of weak and ordinary Pareto stationarity of the state $\bar{y}$ in the reduced multiobjective optimal control problem (6.6). In particular, a control $\bar{u}$ is a weakly/ordinarily Pareto stationary point (in the sense of Def. 3.2) for the problem (P) interpreted as a problem in u if and only if the state $\bar{y} := S(\bar{u})$ is a weakly/ordinarily Pareto stationary point (in the sense of (6.7)) for the problem (P) interpreted as a problem in y.*

Proof. The equivalence between the Pareto stationarity conditions of (6.6) and the strong stationarity conditions (4.6) and (4.7) follows straightforwardly from the arguments outlined after the proof of Theorem 6.1. (Note that we indeed obtain an "if and only if" here due to (4.4) and (4.5).) This proves the first part of the theorem. The second one is an immediate consequence of Theorem 4.5. $\qquad\square$

Note that, in view of the above observations, it indeed makes sense that the strong stationarity conditions (4.6) and (4.7) have a structure similar to that of classical necessary optimality conditions for smooth multiobjective optimal control problems, *cf.*, *e.g.*, the results in [29], Theorem 2.6 and [22, 39].

We would like to point out that the identification results in Theorem 6.2 also suggest an alternative approach to the strong stationarity conditions in Theorem 4.5: Instead of the strategy pursued in Section 4, which consists of starting with the purely primal concepts in Definition 3.2 and subsequently applying the adjoint calculus of Lemma 4.2 for the derivation of (4.6) and (4.7), one can also start with the problem (6.6) formulated in the variable y, use the arguments after the proof of Theorem 6.1 to show that the necessary conditions for weak and ordinary Pareto stationarity in (6.6) are equivalent to (4.6) and (4.7), and subsequently prove that a state $\bar{y}$ is a weak/ordinary Pareto stationary point of (6.6) if and only if the associated control $\bar{u}$ is a weak/ordinary Pareto stationary point for the problem (P) when it is reduced to the control u. We remark that the last step in this alternative argumentation requires roughly the same effort as the analysis of Section 4. The amount of work needed for the derivation of Theorems 6.2 and 4.5 is thus essentially independent of the approach that one chooses to take here.

We conclude this section with some further comments on how the results of the last three sections are related to the known literature:

Remark 6.3.

i) The differentiability properties in Theorem 6.1 are also the reason for the Gâteaux differentiability of the solution map of the partial differential equation studied in [18], Section 5. Note that, in contrast to the problem in [18], where the weak formulation of the governing PDE already contains a term of the form $g(y)\nabla y$ with a piecewise smooth $g : \mathbb{R} \to \mathbb{R}$, in the case of the problem (P), the smoothness effects that we have explored in this section only occur since the objective functions J_n in (P) have a composite structure, *i.e.*, can be split into an observation part and an L^2-Tikhonov term. Compare, for instance, with the calculation in (6.3) in this context. Without such a structure, the Gâteaux differentiability of the functions $J_n(\cdot, S^{-1}(\cdot))$ cannot be guaranteed and the situation is much less clear. (When considering general objective functions, one can, of course, still employ the subdifferential calculus of [14].)

ii) The analysis of this section and, in particular, the Gâteaux differentiability of the reduced objective function in (6.3) are also new and relevant in the single-objective setting. Using Theorem 6.1 and a calculation analogous to that in (6.3), we obtain, *e.g.*, that the problems considered in [14], Section 5 and [21], Section 5, Case 1 are all Gâteaux differentiable when reduced to the state y. To the best of the authors' knowledge, this observation has not been made so far in the literature.

7. Numerical experiments

The results of the last sections suggest (at least) two different approaches for the numerical approximation of the set of weak/ordinary Pareto stationary points of a problem of the type (P):

First, we can try to tackle the strong stationarity conditions (4.6) and (4.7) in Theorem 4.5 directly with a generalized Newton method or a comparable algorithm. Recall that, by Lemma 4.8, we know that the systems (4.6) and (4.7) have at least one solution for all $\bar{\lambda}$ with $\bar{\lambda}_N > 0$, and that Theorem 4.5 guarantees that all controls $\bar{u}$ that we determine by solving a system of the form (4.6) or (4.7), are Pareto stationary in the weak or ordinary sense, depending on the choice of the vector $\bar{\lambda}$. It is thus indeed a reasonable strategy to try to solve the problems (4.6) or (4.7), respectively, for an *a priori* fixed selection of multipliers $\bar{\lambda}$ to get an approximation of the set of Pareto stationary points of (P) and the associated pseudo-Pareto front

$$\left\{ (J_n(\bar{y}, \bar{u}))_{n=1,\dots,N} \,\middle|\, \bar{u} \in L^2(\Omega) \text{ (weakly) Pareto stationary for (P) and } \bar{y} = S(\bar{u}) \right\}. \tag{7.1}$$

Note that this approach has the particular advantage that we can control precisely whether we calculate a weakly Pareto stationary or an ordinarily Pareto stationary point by choosing a $\bar{\lambda}$ with non-negative or positive components. This feature comes at the cost of having to deal with the indicator function $\mathbb{1}_{\{\bar{y}>0\}}$ in the strong stationarity conditions, which is typically not easy to handle.

Second, we can also simply mollify the problem (P) as described in Section 5 and then use the set of weakly Pareto stationary points of the regularized multiobjective optimal control problem (P_ε) as an approximation of that of (P). Note that the convergence result in Theorem 5.3 suggests that the set that is obtained in this way can indeed be expected to approximate the set of weakly Pareto stationary points of (P) for small regularization parameters ε (provided the conditions in Corollary 4.7i) are satisfied). However, in contrast to our first approach, this strategy does not allow to guarantee that a calculated point is Pareto stationary in the ordinary sense, as Theorem 5.3 only ensures weak stationarity in the limit $\varepsilon \to 0^+$. Further, one has to deal with an additional regularization error. On the other hand, computing a weakly Pareto stationary point of (P_ε) by solving the system (5.4) is, of course, much easier than the solution of one of the strong stationarity conditions in Theorem 4.5.

In what follows, we will explore and compare both of the above approaches in numerical experiments. Before we begin with our investigation, we would like to emphasize that the subsequent analysis should be understood as a feasibility study. In particular, we postpone a detailed discussion of the solvability and the approximation properties of the discrete counterparts of the systems (4.6), (4.7), and (5.4) as well as the convergence behavior and reliability of the used numerical solution method to future research. For related results in the single-objective setting, see [14, 18].

As a model problem for our numerical tests, we consider a simple tri-criterial, tracking-type optimization problem of the form

$$
\begin{aligned}
\text{Minimize} \quad & \left(\tfrac{1}{2}\|y - y_{D,1}\|_{L^2}^2, \quad \tfrac{1}{2}\|y - y_{D,2}\|_{L^2}^2, \quad \tfrac{\nu}{2}\|u\|_{L^2}^2 \right) \\
\text{w.r.t.} \quad & u \in L^2(\Omega), \quad y \in H_0^1(\Omega) \cap H^2(\Omega), \\
\text{s.t.} \quad & -\Delta y + \max(0, y) = u \text{ a.e. in } \Omega
\end{aligned}
\tag{M}
$$

on the unit square $\Omega := (0,1)^2$ with two given desired states $y_{D,1}, y_{D,2} \in C(\mathrm{cl}(\Omega))$ and a Tikhonov parameter $\nu > 0$ (to be specified below). Note that this problem trivially satisfies all of the conditions in Assumption 2.1

and Corollary 4.7i). Theorems 4.5 and 5.3 are thus applicable here. For the discretization of (M) (or its necessary optimality conditions, to be more precise), we use Friedrichs–Keller triangulations $\mathcal{T}_h$ of Ω with widths $h = 1/n$, $n \in \mathbb{N}$, and the finite element spaces

$$V_h := \left\{ v \in C(\mathrm{cl}(\Omega)) \,\big|\, v|_\tau \text{ is affine for all cells } \tau \in \mathcal{T}_h \text{ and } v|_{\partial\Omega} = 0 \right\} \subset H_0^1(\Omega).$$

Recall that the stiffness and the mass matrix associated with V_h are given by

$$A := \left(\int_\Omega \nabla\varphi_i \cdot \nabla\varphi_k \mathrm{d}x \right)_{i,k=1,\ldots,M} \quad \text{and} \quad B := \left(\int_\Omega \varphi_i\,\varphi_k \mathrm{d}x \right)_{i,k=1,\ldots,M},$$

where $\{\varphi_i\}_{i=1}^M$ denotes the nodal basis of V_h and $M := \dim(V_h)$. (Here and in what follows, we often suppress the dependency on the mesh width h in the notation for the sake of readability.) For later reference, we further define the lumped mass matrix

$$L := \mathrm{diag}\left(\frac{1}{3} |\mathrm{supp}(\varphi_i)|\,, i = 1, \ldots, M \right).$$

Note that this matrix arises from an approximation of the L^2-scalar product on Ω by means of a three point subdivided quadrature rule subordinate to $\mathcal{T}_h$. Indeed, for all $v_1, v_2 \in C(\mathrm{cl}(\Omega))$, we have

$$\int_\Omega v_1 v_2 \mathrm{d}x = \sum_{\tau \in \mathcal{T}_h} \int_\tau v_1 v_2 \mathrm{d}x \approx \sum_{\tau \in \mathcal{T}_h} \frac{1}{3}|\tau| \sum_{x \in \Omega \text{ node of } \tau} v_1(x) v_2(x) = \mathbf{v}_1^T L \mathbf{v}_2, \tag{7.2}$$

where $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^M$ are the coordinate vectors w.r.t. the basis $\{\varphi_i\}_{i=1}^M$ of the Lagrange interpolants $I_h(v_1), I_h(v_2) \in V_h$ of the functions v_1, v_2, respectively, i.e., the vectors containing the function values of v_1 and v_2 at the interior nodes of the mesh $\mathcal{T}_h$.

Let us now first consider the strong stationarity conditions (4.6) and (4.7) of (M), i.e., the system

$$\bar{u}, \bar{p} \in L^2(\Omega), \quad \bar{y} \in H_0^1(\Omega) \cap H^2(\Omega),$$

$$-\Delta\bar{y} + \max(0, \bar{y}) = \bar{u}, \quad -\Delta\bar{p} + \mathbb{1}_{\{\bar{y}>0\}}\bar{p} = \sum_{n=1}^2 \bar{\lambda}_n(\bar{y} - y_{D,n}), \quad \bar{p} + \bar{\lambda}_3\nu\bar{u} = 0 \tag{7.3}$$

with a multiplier $\bar{\lambda} \in \mathbb{R}^3$ that satisfies $\sum_{n=1}^3 \bar{\lambda}_n = 1$ and either $\bar{\lambda}_n \geq 0$ for $n = 1, 2, 3$ in the case of (4.6) or $\bar{\lambda}_n > 0$ for $n = 1, 2, 3$ in the case of (4.7). Then, by formulating the PDEs in (7.3) weakly, by passing over to the finite element space V_h, by applying the nodal quadrature rule in (7.2) to the L^2-scalar products involving the terms $\max(0, \cdot)$ and $\mathbb{1}_{\{\bar{y}>0\}}$ in the resulting variational identities (thus discretizing them by means of the lumped mass matrix L), and by expressing the remaining L^2- and H^1-scalar products with the mass matrix B and the stiffness matrix A, respectively, we arrive at the discrete system of equations

$$A\bar{\mathbf{y}} + L\max(0, \bar{\mathbf{y}}) = B\bar{\mathbf{u}}, \qquad A\bar{\mathbf{p}} + L\,\mathrm{diag}(H(\bar{\mathbf{y}}))\bar{\mathbf{p}} = \sum_{n=1}^2 \bar{\lambda}_n B(\bar{\mathbf{y}} - \mathbf{y}_{D,n}), \qquad \bar{\mathbf{p}} + \bar{\lambda}_3\nu\bar{\mathbf{u}} = 0. \tag{7.4}$$

Here, $\bar{\mathbf{y}}, \bar{\mathbf{p}}, \bar{\mathbf{u}} \in \mathbb{R}^M$ and $\mathbf{y}_{D,1}, \mathbf{y}_{D,2} \in \mathbb{R}^M$ are the coordinate vectors w.r.t. $\{\varphi_i\}_{i=1}^M$ of the V_h-counterparts of the quantities $\bar{y}, \bar{p}, \bar{u}$ in (7.3) and the Lagrange interpolants of $y_{D,1}$ and $y_{D,2}$, respectively, H denotes the Heaviside function (with the convention $H(0) = 0$), and the notation $\max(0, \bar{\mathbf{y}})$ and $H(\bar{\mathbf{y}})$ is understood componentwise.

Note that, by eliminating $\bar{\mathbf{u}}$ and by splitting $\bar{\mathbf{y}}$ into a positive and a negative part, for all multipliers $\bar{\lambda}$ with $\bar{\lambda}_3 > 0$, (7.4) can also be recast as

$$A\bar{\mathbf{y}}_+ - A\bar{\mathbf{y}}_- + L\bar{\mathbf{y}}_+ + (\nu\bar{\lambda}_3)^{-1}B\bar{\mathbf{p}} = 0,$$

$$A\bar{\mathbf{p}} + L\operatorname{diag}(H(\bar{\mathbf{y}}_+ - \bar{\mathbf{y}}_-))\bar{\mathbf{p}} - \sum_{n=1}^{2} \bar{\lambda}_n B(\bar{\mathbf{y}}_+ - \bar{\mathbf{y}}_- - \mathbf{y}_{D,n}) = 0, \qquad \min(\bar{\mathbf{y}}_+, \bar{\mathbf{y}}_-) = 0. \tag{7.5}$$

This reformulation has the advantage that the non-smoothness is completely removed from the state equation and transferred into a separate, standard complementarity constraint. Since the system (7.5) can – at least heuristically – be tackled with a standard Newton-type method with the pseudo-Jacobian

$$\begin{pmatrix} A+L & -A & (\nu\bar{\lambda}_3)^{-1}B \\ -(\bar{\lambda}_1 + \bar{\lambda}_2)B & (\bar{\lambda}_1 + \bar{\lambda}_2)B & A + L\operatorname{diag}(H(\bar{\mathbf{y}}_+ - \bar{\mathbf{y}}_-)) \\ I - \operatorname{diag}(H(\bar{\mathbf{y}}_+ - \bar{\mathbf{y}}_-)) & \operatorname{diag}(H(\bar{\mathbf{y}}_+ - \bar{\mathbf{y}}_-)) & 0 \end{pmatrix} \in \mathbb{R}^{3M \times 3M}, \tag{7.6}$$

we may now follow the first approach outlined at the beginning of this section and try to solve (7.5) for various choices of the multiplier $\bar{\lambda} \in \mathbb{R}^3$ to compute an approximation of the Pareto front of (P). Before we demonstrate how this solution method performs in practice, we would like to remark the following:

Remark 7.1.

i) A strategy similar to the one above has also been used in [18], Section 6 for the calculation of first-order stationary points of single-objective optimal control problems governed by quasilinear elliptic PDEs. The finite element discretization and the mass-lumping scheme that we have employed for the derivation of (7.4) are further the same as those in [14], Section 5.

ii) It is important to realize that the transition from (7.3) to (7.4) follows a pure "first optimize then discretize"-philosophy. Since the effects studied in Sections 4 and 6 rely on Stampacchia's lemma, which is inherently infinite-dimensional, it is *a priori* completely unclear if (7.4) can be identified with the necessary optimality condition of an appropriately defined discrete problem or if this system has a solution at all. As (7.4) arises from the continuous optimality condition (7.3) (which is known to possess a solution) by passing over to the finite element spaces V_h and by applying quadrature rules, one can only guarantee that there exist vectors $\bar{\mathbf{y}}, \bar{\mathbf{p}}, \bar{\mathbf{u}} \in \mathbb{R}^M$ which satisfy (7.4) up to a certain error level $r(h)$ that vanishes in the limit $h \to 0^+$. As we will see below, in practical applications, this deficit of the discretization (7.4) is rather unproblematic as $r(h)$ turns out to be typically much smaller than the tolerances that are normally used in numerical solution algorithms (in our experiments this was 10^{-8}). In fact, it can be observed that a Newton method based on the pseudo-Jacobian (7.6) only fails in exceptionally rare cases, and that the percentage of these cases decreases when the tolerance is fixed and h is driven to zero, see Table 1. However, even in view of the fact that calculating approximate solutions of the discrete system (7.4) works very well in practice, one should keep in mind that this approach (and, in extension, also the numerical procedure in [18], Sect. 6) is a heuristic. (Working with the matrix (7.6) is, of course, heuristically motivated as well as the system of equations in (7.5) is non-smooth and thus not covered by classical convergence results for Newton-type methods.)

iii) As an alternative to the approach in (7.3), one could also first reduce the problem (M) to the state y and subsequently discretize the associated first-order necessary optimality conditions in (6.7). Note that, in this case, the obtained system of equations involves the bi-Laplacian so that some care has to be taken regarding the choice of the used finite element spaces. Compare, *e.g.*, with [12] in this context where a similar technique is applied to state-constrained problems. We do not discuss this alternative discretization here to avoid overloading this paper.

TABLE 1. Performance of our Newton-type algorithm in the situation of (7.7). The first line shows the width h of the considered triangulation, the second one the average number of Newton iterations needed for a successful step over the course of the calculation of the front (7.1) for the given h, the third one the number/percentage of steps in which the Newton method failed to converge for the given h in less than ten iterations, and the fourth one the average number of seconds needed for one successful solution of the stationarity system (7.5). The width of the discretization of Λ in (7.8) was chosen as $1/K = 1/100$ so that, for each h, (7.5) was solved 5050 times. The linear systems of equations arising in our Newton-type method have been solved by means of the "backslash" operator implemented in Matlab, version R2020a, on a computer with eight IntelCore i7-7700 CPUs with 3.60GHz and 32GB of RAM.

h	1/16	1/32	1/64	1/128	1/256	1/512
Average number of Newton iterations	2.0684	2.2321	2.5135	2.8294	2.9899	3.1407
Number/percentage of unsuccessful steps	19 0.376%	9 0.178%	10 0.198%	10 0.198%	5 0.099%	5 0.099%
Average number of seconds needed for a successful step	0.0050 s	0.0257 s	0.1737 s	0.8559 s	4.5399 s	27.767 s

Table 1 and Figures 1 and 2 show the results that we have obtained by solving the system (7.5) in the case

$$
\begin{aligned}
y_{D,1}(x_1, x_2) &:= \mathbb{1}_{\left(\frac{1}{4}, 1\right)^2}(x_1, x_2) \sin\left(\frac{4}{3}\pi\left(x_1 - \frac{1}{4}\right)\right) \sin\left(\frac{4}{3}\pi\left(x_2 - \frac{1}{4}\right)\right), \\
y_{D,2}(x_1, x_2) &:= -\mathbb{1}_{\left(0, \frac{3}{4}\right)^2}(x_1, x_2) \sin\left(\frac{4}{3}\pi x_1\right) \sin\left(\frac{4}{3}\pi x_2\right), \qquad \nu := 10^{-4},
\end{aligned}
\tag{7.7}
$$

with a Newton-type method based on the pseudo-Jacobian (7.6) up to the tolerance tol $= 10^{-8}$ for various multipliers $\bar{\lambda}$. Here, the parameter space for $\bar{\lambda}$, *i.e.*, the simplex $\Lambda := \left\{\lambda \in \mathbb{R}^3 \mid \sum_{n=1}^3 \lambda_n = 1, \lambda_n \geq 0\right\}$ was discretized with an equidistant mesh of the form

$$
\Lambda_K := \left\{\frac{1}{K}(m, n, K - m - n) \in \mathbb{R}^3 \,\Big|\, m, n = 0, \dots, K, \, m + n < K\right\}, \quad K \in \mathbb{N}.
\tag{7.8}
$$

Note that the vectors with $\lambda_3 = 0$ are excluded in (7.8) as for these multipliers the solvability of (7.3) cannot be guaranteed, *cf.* Lemma 4.8. We would like to emphasize that the choice of the equidistant mesh Λ_K in (7.8) does not imply that the computed points in the front (7.1) are distributed equidistantly as well since it is *a priori* unclear how the set (7.1) is parametrized by Λ (see Fig. 1). To achieve a more homogeneous resolution of (7.1), one can employ adaptive discretization techniques or reference point methods analogous to those used for classical scalarization approaches, *cf.* [5, 34, 54] and also Remark 4.6iii). We omit a detailed discussion of this topic to avoid overloading the paper.

As the results in Table 1 show, our Newton-type algorithm requires on average between two and three iterations to reduce the residue of (7.5) below the given threshold tol $= 10^{-8}$. (The starting point was always chosen as zero here.) We further see that the average number of iterations remains nearly constant as h is decreased. This mesh-independence of the solution procedure makes sense as our algorithm is based on the necessary optimality condition of the continuous problem (M) in (7.3). In the third line of Table 1, we can further see that our method only fails to solve the system (7.5) in exceptionally rare situations. Moreover, the number of unsuccessful solution steps (here defined as steps in which the Newton-type algorithm failed to converge for a given λ in less than ten iterations) decreases as the mesh width h tends to zero. This confirms the theoretical considerations made in Remark 7.1.
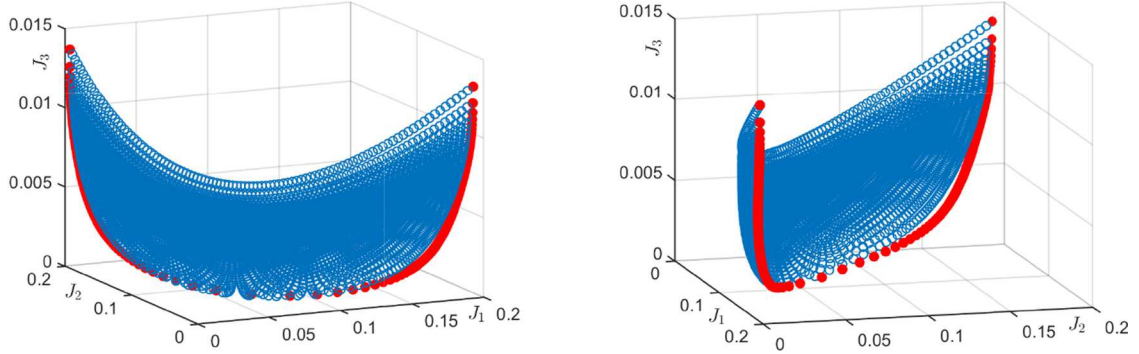
FIGURE 1. Approximation of the set (7.1) for (M) in the situation (7.7) obtained by solving the system (7.5) for 5050 values of the multiplier $\bar{\lambda}$. The spatial mesh width h was chosen as $1/512$ and the width of the equidistant discretization of the simplex Λ in (7.8) as $1/K = 1/100$. The points in the scatter plot associated with multipliers $\bar{\lambda}$ that have a vanishing component and thus correspond to weakly but not ordinarily Pareto stationary points are depicted in red.
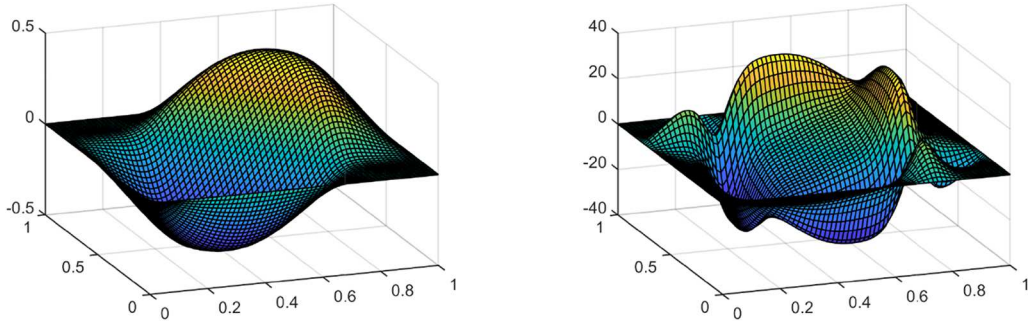

FIGURE 2. State (*left*) and control (*right*) associated with the solution of (7.5) in the situation of (7.7) for the particular multiplier $\bar{\lambda} = (0.49, 0.5, 0.01)$. The width h was chosen as $1/64$ here.

For comparison, let us now consider the regularization approach that we have outlined at the beginning of this section: By starting from the necessary optimality condition (5.4) for the regularized multiobjective optimal control problem (P_ε) and by proceeding along exactly the same lines as for (7.4), we arrive at the equations

$$A\bar{y} + L \max_\varepsilon(\bar{y}) = B\bar{u}, \qquad A\bar{p} + L \operatorname{diag}(\max'_\varepsilon(\bar{y}))\bar{p} = \sum_{n=1}^{2} \bar{\lambda}_n B(\bar{y} - y_{D,n}), \qquad \bar{p} + \bar{\lambda}_3 \nu \bar{u} = 0, \qquad (7.9)$$

where $\bar{\lambda}$ is again an arbitrary but fixed element of Λ, and where we again suppress the dependence on h and ε in the notation for $\bar{y}$, A, B, *etc.* for the sake of readability. Note that, in contrast to (7.4), the system (7.9) is smooth so that a standard Newton algorithm can be used for its solution. Further, it is straightforward to check that (7.9) corresponds to the necessary optimality condition for weak Pareto optimality of a discrete version of (P_ε), and that (7.9) admits at least one solution for all arbitrary but fixed $\bar{\lambda} \in \Lambda$ with $\bar{\lambda}_3 > 0$ (*cf.* the proof of Lem. 4.8 and also the results in [22], Sect. 3.3). The processes of discretization and optimization (or calculating the first-order necessary optimality conditions, to be more precise) thus commute in the case of (P_ε) and the solvability of (7.9) is not an issue at all. This is an advantage over approaches based on the system (7.4) that, of course, again comes at the price of the additional regularization error.

TABLE 2. Average number of Newton iterations needed for the approximation of the set (7.1) by successive solution of the system (7.9) for various h and ε and Λ_K as in (7.8). The width of the mesh Λ_K was again chosen as $1/K := 1/100$ here so that (7.9) was solved 5050 times in each configuration. The number of unsuccessfully solved systems and the average number of seconds needed for a successful step are denoted in parentheses, separated by a comma. The computations have been carried out on the same system as in Table 1.

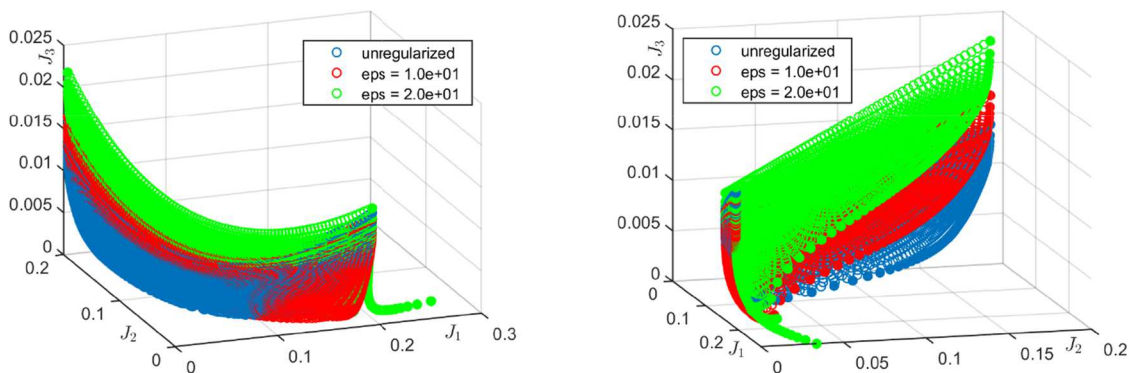| ε \ h	1/16	1/32	1/64	1/128	1/256
10^{-3}	2.2771	2.2283	2.1671	2.0736	2.0392
	$(0,\,0.0047\,\mathrm{s})$	$(0,\,0.0193\,\mathrm{s})$	$(0,\,0.0771\,\mathrm{s})$	$(0,\,0.3927\,\mathrm{s})$	$(0,\,1.9095\,\mathrm{s})$
10^{-6}	2.0849	2.2063	2.3895	2.5165	2.4755
	$(9,\,0.0044\,\mathrm{s})$	$(7,\,0.0194\,\mathrm{s})$	$(3,\,0.0848\,\mathrm{s})$	$(0,\,0.4731\,\mathrm{s})$	$(0,\,2.3120\,\mathrm{s})$
10^{-9}	2.0588	2.1789	2.3645	2.4907	2.4085
	$(19,\,0.0043\,\mathrm{s})$	$(9,\,0.0194\,\mathrm{s})$	$(10,\,0.0839\,\mathrm{s})$	$(8,\,0.4689\,\mathrm{s})$	$(0,\,2.2505\,\mathrm{s})$

FIGURE 3. Approximations of the pseudo-Pareto front (7.1) obtained from the unregularized system (7.5) (*blue*) and the regularized system (7.9) with $\varepsilon = 10$ (*red*) and $\varepsilon = 20$ (*green*) for $h = 1/256$ and $K = 100$. The regularization parameter is chosen very large here because for smaller ε the sets are visually indistinguishable. Points on the front that correspond to weakly but not ordinarily Pareto stationary points are colored solidly. It can be seen that the sets calculated by solving (7.9) approximate that obtained from (7.5) as ε tends to zero.

Table 2 and Figure 3 depict the results that are obtained when the function $x \mapsto \frac{1}{2}(x + \sqrt{x^2 + \varepsilon^2})$ is chosen as $\max_\varepsilon(\cdot)$ in (P_ε), the system (7.9) is solved with a classical Newton method, and the set Λ is discretized as in (7.8). As Table 2 shows, the solution of (7.9) again requires between two and three Newton iterations on average over the course of a single approximation of the front (7.1), and, similarly to the behavior in Table 1, this number is largely independent of the mesh width h and the regularization parameter ε. Iterations, in which the Newton algorithm fails to converge, are encountered here as well (though less frequently than in the unregularized case if ε is moderate), and it can be seen that the behavior of our solution method for the regularized system (7.9) emulates that observed for the unregularized problem in Table 1 when ε tends to zero (with a notable exception for $h = 1/256$ most likely due to a favorable relationship between h and ε). Figure 3 further shows that the front calculated by solving (7.9) approximates that obtained from the system (7.5) as ε decreases. This underlines that the solutions obtained from (7.5) are sensible, that the system (7.5) allows to calculate an approximation of the front (7.1) that does not suffer from a regularization error, and that the strong stationarity conditions derived in Section 4 are indeed not only interesting for their own sake and theoretical purposes but also for numerical solution algorithms.

References

[1] R.A. Adams, Sobolev Spaces. Academic Press, New York (1975).

[2] H. Attouch, G. Buttazzo and G. Michaille, Variational Analysis in Sobolev and BV Spaces. SIAM, Philadelphia (2006).

[3] H. Attouch, G. Garrigos and X. Goudou, A dynamic gradient approach to Pareto optimization with nonsmooth convex objective functions. *J. Math. Anal. Appl.* **422** (2015) 741–771.

[4] S. Banholzer, S. Beermann and S. Volkwein, POD-based bicriterial optimal control by the reference point method. 2nd IFAC Workshop on Control of Systems Governed by Partial Differential Equations CPDE 2016. *IFAC-PapersOnLine* **49** (2016) 210–215.

[5] S. Banholzer and S. Volkwein, Hierarchical convex multiobjective optimization by the Euclidean reference point method. Preprint SPP1962-117 (2019).

[6] V. Barbu, Optimal Control of Variational Inequalities. Research Notes in Mathematics. Pitman (1984).

[7] D. Beermann, M. Dellnitz, S. Peitz and S. Volkwein, POD-based multiobjective optimal control of PDEs with non-smooth objectives. *PAMM* **17** (2017) 51–54.

[8] D. Beermann, M. Dellnitz, S. Peitz and S. Volkwein, Set-oriented multiobjective optimal control of PDEs using proper orthogonal decomposition, in Reduced-Order Modeling (ROM) for Simulation and Optimization: Powerful Algorithms as Key Enablers for Scientific Computing, edited by W. Keiper, A. Milde and S. Volkwein. Springer (2018) 47–72.

[9] S. Bellaassali and A. Jourani, Necessary optimality conditions in multiobjective dynamic optimization. *SIAM J. Control Optim.* **42** (2004) 2043–2061.

[10] J. Bienvenido and N. Vicente, Optimality conditions in directionally differentiable Pareto problems with a set constraint via tangent cones. *Numer. Funct. Anal. Optim.* **24** (2003) 557–574.

[11] J.F. Bonnans and A. Shapiro, Perturbation Analysis of Optimization Problems. *Springer Series in Operations Research.* Springer, New York (2000).

[12] S.C. Brenner, Finite element methods for elliptic distributed optimal control problems with pointwise state constraints (survey), in *Advances in Mathematical Sciences: AWM Research Symposium, Houston, TX, April 2019*, edited by B. Acu, D. Danielli, M. Lewicka, A. Pati, R. V. Saraswathy and M. Teboh-Ewungkem. Springer (2020) 3–16.

[13] C. Christof, *Sensitivity Analysis of Elliptic Variational Inequalities of the First and the Second Kind*. Ph.D. thesis, Technische Universität Dortmund (2018).

[14] C. Christof, C. Meyer, S. Walther and C. Clason, Optimal control of a non-smooth semilinear elliptic equation. *Math. Control Relat. Fields* **8** (2018) 247–276.

[15] C. Christof and B. Vexler, New regularity results and finite element error estimates for a class of parabolic optimal control problems with pointwise state constraints. To appear in: *ESAIM: COCV* (2020). Available from: https://doi.org/10.1051/cocv/2020059

[16] C. Christof and G. Wachsmuth, On second-order optimality conditions for optimal control problems governed by the obstacle problem. To appear in: *Optimization* (2020). Available from: https://doi.org/10.1080/02331934.2020.1778686

[17] F.H. Clarke, Optimization and Nonsmooth Analysis. *SIAM's Classics in Applied Mathematics.* SIAM, Philadelphia, PA (1990).

[18] C. Clason, V.H. Nhu and A. Rösch, Optimal control of a non-smooth quasilinear elliptic equation. Preprint arXiv:1810.08007 (2018).

[19] J.C. De los Reyes and C. Meyer, Strong stationarity conditions for a class of optimization problems governed by variational inequalities of the second kind. *J. Optim. Theory Appl.* **168** (2016) 375–409.

[20] S. Dempe. Foundations of Bilevel Programming. Vol. 61 of Nonconvex Optimization and Its Applications. Kluwer (2002).

[21] O. Ebel, S. Schmidt and A. Walther, Solving non-smooth semi-linear optimal control problems with abs-linearization. Preprint SPP1962-093 (2018).

[22] M. Ehrgott, Multicriteria Optimization, 2nd edn. Springer (2005).

[23] I. Ekeland and R. Temam, Convex Analysis and Variational Problems. North-Holland (1976).

[24] L.C. Evans, Partial Differential Equations, 2nd edn. AMS, Providence, RI (2010).

[25] D. Gilbarg and N.S. Trudinger, Elliptic Partial Differential Equations of Second Order, reprint of the 1998 edn. Springer, Berlin/Heidelberg/New York (2001).

[26] G. Giorgi, B. Jiménez and V. Novo, On constraint qualifications in directionally differentiable multiobjective optimization problems. *RAIRO: OR* **38** (2004) 255–274.

[27] G. Giorgi, B. Jiménez and V. Novo, Strong Kuhn-Tucker conditions and constraint qualifications in locally Lipschitz multiobjective optimization problems. *TOP* **17** (2008) 288–304.

[28] P. Grisvard, Elliptic Problems in Nonsmooth Domains. Pitman (1985).

[29] L. Iapichino, S. Ulbrich and S. Volkwein, Multiobjective PDE-constrained optimization using the reduced-basis method. *Adv. Comput. Math.* **43** (2017) 945–972.

[30] Y. Ishizuka, Optimality conditions for directionally differentiable multi-objective programming problems. *J. Optim. Theory Appl.* **72** (1992) 91–111.

[31] Y. Ishizuka and K. Shimizu, Necessary and sufficient conditions for the efficient solutions of nondifferentiable multiobjective problems. *IEEE Trans. Syst. Man Cybernet.* **SMC-14** (1984) 625–629.

[32] B.N. Khoromskij and G. Wittum, Numerical Solution of Elliptic Differential Equations by Reduction to the Interface. *Lecture Notes in Computational Science and Engineering.* Springer, Berlin/Heidelberg (2012).

[33] F. Kikuchi, K. Nakazato and T. Ushijima, Finite element approximation of a nonlinear eigenvalue problem related to MHD equilibria. *Japan J. Appl. Math.* **1** (1984) 369–403.

[34] I.Y. Kim and O.L. de Weck, Adaptive weighted-sum method for bi-objective optimization: Pareto front generation. *Struct. Multidiscip. Optim.* **29** (2005) 149–158.

[35] Y.-B. Lü and W. Zhong-Ping, A smoothing method for solving bilevel multiobjective programming problems. *J. Oper. Res. Soc. China* **2** (2014) 511–525.

[36] M.M. Mäkelä, V.P. Eronen and N. Karmitsa, On nonsmooth multiobjective optimality conditions with generalized Convexities, in *Optimization in Science and Engineering: In Honor of the 60th Birthday of Panos M. Pardalos*, edited by T.M. Rassias, C.A. Floudas, and S. Butenko. Springer, New York, NY (2014) 333–357.

[37] O.L. Mangasarian, Nonlinear Programming. Vol 10 of *SIAM's Classics in Applied Mathematics*, 2nd edn. SIAM (1994).

[38] C. Meyer and L. Susu, Optimal control of nonsmooth, semilinear parabolic equations. *SIAM J. Control Optim.* **55** (2017) 2206–2234.

[39] K.M. Miettinen, Nonlinear Multiobjective Optimization. Kluwer, Boston/London/Dordrecht (1999).

[40] F. Mignot, Contrôle dans les inéquations variationelles elliptiques. *J. Funct. Anal.* **22** (1976) 130–185.

[41] F. Mignot and J.P. Puel, Optimal control in some variational inequalities. *SIAM J. Control Optim.* **22** (1984) 466–476.

[42] B.S. Mordukhovich, Multiobjective optimization problems with equilibrium constraints. *Math. Program.* **117** (2009) 331–354.

[43] B.S. Mordukhovich, Variational Analysis and Applications. Springer Monographs in Mathematics. Springer (2018).

[44] S. Ober-Blöbaum and K. Padberg-Gehle, Multiobjective optimal control of fluid mixing. *PAMM* **15** (2015) 639–640.

[45] S. Peitz and M. Dellnitz, Gradient-based multiobjective optimization with uncertainties, in *NEO 2016: Results of the Numerical and Evolutionary Optimization Workshop NEO 2016 and the NEO Cities 2016 Workshop held on September 20-24, 2016 in Tlalnepantla, Mexico*, edited by Y. Maldonado, L. Trujillo, O. Schütze, A. Riccardi, and M. Vasile. Springer (2018) 159–182.

[46] S. Peitz and M. Dellnitz, A survey of recent trends in multiobjective optimal control - surrogate models, feedback control and objective reduction. *Math. Comput. Appl.* **23** (2018) 1–33.

[47] J. Rappaz, Approximation of a nondifferentiable nonlinear problem related to MHD equilibria. *Numer. Math.* **45** (1984) 117–133.

[48] A.-T. Rauls and S. Ulbrich, Computation of a Bouligand generalized derivative for the solution operator of the obstacle problem. *SIAM J. Optim.* **57** (2019) 3223–3248.

[49] A.-T. Rauls and G. Wachsmuth, Generalized derivatives for the solution operator of the obstacle problem. *Set-Valued Var. Anal.* (2019) 1–27.

[50] R.T. Rockafellar and R.J.-B. Wets, Variational analysis. Vol. 317 of *Grundlehren der Mathematischen Wissenschaften*. Springer (1998).

[51] M. Ruzicka, Nichtlineare Funktionalanalysis. Springer, Berlin/Heidelberg (2004).

[52] W. Schirotzek, Nonsmooth Analysis. Springer, Berlin/Heidelberg (2007).

[53] R. Temam, A non-linear eigenvalue problem: the shape at equilibrium of a confined plasma. *Arch. Rational Mech. Anal.* **60** (1976) 51–73.

[54] A.P. Wierzbicki, The use of reference objectives in multiobjective optimization, in *Multiple Criteria Decision Making Theory and Application*, edited by G. Fandel and T. Gal. Springer (1980) 468–486.

[55] J. Xin, An Introduction to Fronts in Random Media. Vol. 5 of *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer (2009).

[56] J.J. Ye, Necessary optimality conditions for multiobjective bilevel programs. *Math. Oper. Res.* **36** (2011) 165–184.