J Optim Theory Appl (2016) 169:735–758
DOI 10.1007/s10957-015-0799-4

Optimal Control of Constrained Self-Adjoint Nonlinear Operator Equations in Hilbert Spaces

M. A. El-Gebeily[1] · B. S. Mordukhovich[2] · M. M. Alshahrani[1]

Received: 2 February 2015 / Accepted: 17 August 2015 / Published online: 27 August 2015

Abstract This paper deals with the study of a new class of optimal control problems governed by nonlinear self-adjoint operator equations in Hilbert spaces under general constraints of the equality and inequality types on state variables. While the unconstrained version of such problems has been considered in our preceding publication, the presence of constraints significantly complicates the derivation of necessary optimality conditions. Developing a geometric approach based on multineedle control variations and finite-dimensional subspace extensions of unbounded self-adjoint operators, we establish necessary optimality conditions for the constrained control problems under considerations in an appropriate form of the Pontryagin Maximum Principle.

Keywords Optimal control · Constrained self-adjoint nonlinear operator equations in Hilbert spaces · Necessary optimality conditions · Maximum principle

Mathematics Subject Classification 49K20 · 47H15 · 93C30

B. S. Mordukhovich: Research of this author was partly supported by the USA National Science Foundation under Grants DMS-1007132 and DMS-1512846 and by the USA Air Force Office of Scientific Research under Grant No. 15RT0462.

✉ B. S. Mordukhovich
aa1086@wayne.edu

M. A. El-Gebeily
mgebeily@kfupm.edu.sa

M. M. Alshahrani
mshahrani@kfupm.edu.sa

1 Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

2 Department of Mathematics, Wayne State University, Detroit, MI 48202, USA

1 Introduction

The paper concerns optimal control theory to which Professor Elijah (Lucien) Polak has made fundamental contributions; see, e.g., his seminal book [1]. We consider a general class of control systems governed by self-adjoint nonlinear operator equations in Hilbert spaces formulated and discussed in Sect. 2. Self-adjoint operators cover ordinary differential operators, partial differential operators, integral operators, and pseudodifferential operators among others; see, e.g., the books [1–10] and the references therein. However, to the best of our knowledge, the control model for general operator equations, which we first addressed in [11] in the unconstrained setting, has never been studied earlier in the literature. Its unconstrained version for singular ODEs was considered in our previous paper [12].

The major result of [11] establishes necessary optimality conditions for optimal controls of unconstrained self-adjoint operator equations in Hilbert spaces given in the form of the *Maximum Principle*, which is an appropriate operator counterpart of the classical Pontryagin Maximum principle for ordinary differential equations [2]. The main goal of the current paper is to extend this result to control problems with constraints on state variables described finitely by many equalities and inequalities given by Fréchet differentiable functions.

It has been well realized in optimal control theory for any type of state equations that the presence of even simple constraints on state variables dramatically complicates the device of necessary optimality conditions for optimal controls. As shown, e.g., in [10, Section 6.3], the derivation of the Pontryagin Maximum Principle for ODE systems without any constraints on system trajectories/arcs can be done by using a pure analytic technique via the increment formula for the cost functional on single-needle variations of optimal controls. On the other hand, the presence of smooth endpoint constraints on trajectories required in [10] a much more involved geometric technique based on multineedle control variations with the usage of convex separation and delicate fixed-point theorems.

This paper reveals a similar situation in the case of general optimal control problems for nonlinear self-adjoint operator equations with constraints on state variables. We show that developing a geometric technique based on multineedle control variations together with convex separation and fixed-point results allows us to derive an appropriate version of the Maximum Principle for the general constrained operator control systems under consideration. It should be mentioned to this end that the new level of generality encompassed in this paper does not allow us to cover all the specific features of the Maximum Principle for particular types of state equations; see more discussions in the text below.

The rest of the paper is organized as follows. In Sect. 2 we introduce the class of constrained operator control problems of our study and then formulate and discuss the standing assumptions made on their initial data throughout the whole paper. Section 3 recalls some results from [11] on finite-dimensional subspace extensions of self-adjoint operators, which play a significant role in the subsequent considerations. Based on these results and given an arbitrary Fréchet differentiable function $\sigma : H \to \mathbb{R}$ on the state space, we construct in Sect. 4 the so-called σ-auxiliary and σ-adjoint problems to the one under consideration and derive an important formula for representing the increment of this function with respect to control variations along the original and adjoint systems.

Section 5 contains the formulation and discussion of our main result–the Maximum Principle for the optimal control problem governed by the general self-adjoint operator equations with constraints on state variables. We also define here the class of multineedle control variations and establish some of their properties needed in the subsequent proof of the main theorem. The proof in the case of inequality constraints is given in Sect. 6, while Sect. 7 presents a more involved proof in the case of equality constraints. The final Sect. 8 collects concluding remarks and formulates some problems of the future research.

Throughout the paper we use the standard notation from theory of operator equations and theory of optimal control; see, e.g., [3,8,10].

2 Problem Formulation and Basic Assumptions

The paper addresses the following optimal control problem: minimize the functional

$$J[u,x] := \varphi_0(x) \quad \text{subject to} \quad u \in \mathcal{U}, \tag{1}$$

where *admissible control* actions u are selected from the given *control set* $\mathcal{U}$ of a metric space, and where the corresponding *states* $x = x_u \in H$ solve the *nonlinear operator equation*

$$\widetilde{T}x = F(x,u) \tag{2}$$

and satisfy the *constraints* of the inequality and equality types:

$$\varphi_i(x) \le 0 \quad \text{for} \quad i = 1, \ldots, m, \tag{3}$$

$$\varphi_i(x) = 0 \quad \text{for} \quad i = m+1, \ldots, m+r. \tag{4}$$

Here $\widetilde{T}\colon H \to H$ is a given *unbounded* densely defined *self-adjoint* operator on the Hilbert space H considered over the field of complex numbers with the *domain* $\widetilde{D} \subset H$ (the tilde notation will be removed later on for the corresponding *extensions* of $\widetilde{T}$ and $\widetilde{D}$ needed in what follows); $F\colon H \times \mathcal{U} \to H$ is a given *nonlinear operator* on the right-hand side of the operator equation (2); and $\varphi_i\colon H \to \mathbb{R}$ for $i = 0, 1 \ldots, m+r$ are the *cost* and *constraint* real-valued functions in (1) and (3), (4), respectively. The class of problems considered here does not include general first-order initial value problems since the differential operator involved in this case is not always self-adjoint.

By $\mathcal{A}$ we denote the set of pairs (u,x) satisfying the operator equation (2) with $u \in \mathcal{U}$ and $x \in \widetilde{D}$. The symbol $\mathcal{B}$ stands for the collection of pairs $(u,x) \in \mathcal{A}$ such that x satisfies all the constraints in (3) and (4), which is the set of *feasible solutions* to the control problem under consideration. Since this paper focuses on deriving necessary optimality conditions, we assume that there is a pair $(\bar{u},\bar{x}) \in \mathcal{B}$ minimizing the cost functional (1) over all $(u,x) \in \mathcal{B}$ and fix such an optimal pair $(\bar{u},\bar{x})$ in further discussions. Throughout the paper we impose the following assumptions on the initial data in (1)–(4):

(H1) The self-adjoint operator $\widetilde{T}: H \to H$ is *bounded below*, i.e., there is $\mu > 0$ such that

$$\left\langle \widetilde{T}x, x \right\rangle \geq \mu \|x\|^2 \quad \text{for all} \quad x \in \widetilde{D}. \tag{5}$$

(H2) The mapping $F: H \times \mathcal{U} \to H$ is continuous in u and *continuously Fréchet differentiable* in x *around* $\bar{x}$ with the partial derivative operator $F'_x(\bar{x}, u)$ for any $u \in \mathcal{U}$, while the mapping $F(\cdot, u) : H \to H$ is *weakly continuous* and *quasimonotone* in the following sense: there is $\rho < \mu$, with μ taken from (5), such that

$$\left\langle F(x, u) - F(\bar{x}, u), (x - \bar{x}) \right\rangle \leq \rho \|x - \bar{x}\|^2 \quad \text{for all} \quad x \in H. \tag{6}$$

(H3) The cost functional φ_0 and those for the inequality constraint $\varphi_1, \ldots, \varphi_m$ are *Fréchet differentiable at* the optimal point $\bar{x}$.

(H4) The equality constraint functionals $\varphi_{m+1}, \ldots, \varphi_{m+r}$ are *continuous around* $\bar{x}$ and *Fréchet differentiable at* this point.

Let us comment on the assumptions above. Note that condition (5) in (H1) implies that the spectrum of $\widetilde{T}$ is contained in the interval $[\mu, \infty[$. Consequently, any $\lambda \in]-\infty, \mu[$ is a resolvent point of $\widetilde{T}$ meaning that $(\widetilde{T} - \lambda I)^{-1}$ is a bounded operator defined on all of H. In particular, this is true when $\lambda = 0$. The assumption that $\mu > 0$ can be relaxed to $\mu \in \mathbb{R}$. The latter situation can be reduced to the former one by replacing (2) with the equation

$$\left(\widetilde{T} + \delta I\right) x = F(x, u) + \delta x,$$

where $\delta > 0$ is chosen so that $\delta + \mu > 0$, without affecting the results of this paper.

Regarding (H2), observe that ρ in (6) does not need to be positive. The only purpose for assuming that F is weakly continuous with respect to x is to ensure the existence of a solution to (2). It is not used otherwise for the rest of this paper. Other assumptions that ensure the existence of a solution are also possible. For instance, we can suppose that either F is completely continuous in x, or that F is continuous in x, bounded on bounded sets and $\widetilde{T}$ has the compact resolvent. There is a trade-off between requirements on $\widetilde{T}$ and those on F. A general class of such operators satisfying our assumptions arises from nonlinear functionals in coefficients of basis expansions. To illustrate it, let $g_n : \mathbb{C} \to \mathbb{C},\ 1 \leq n \leq N$, be C^1 functions with bounded derivatives, $\{\zeta_n\}_{n=1}^N$ and $\{\psi_n\}_{n=1}^N$ be two sequences in H. Define the operator $G : H \to H$ by

$$G(x) := \sum_{n=1}^{N} g_n\left(\langle x, \zeta_n \rangle\right) \psi_n$$

and consider the operator $F : H \times \mathcal{U} \to H$ given by

$$F(x, u) := G(x) + h(u),$$

where $h : U \to H$ is continuous in u. Then $F(x, u)$ satisfies all the stated assumptions; see below for more details. This class of nonlinear operators includes some filtering operators, which are broadly used in, e.g., signal processing.

Observe that, in contrast to the smoothness (C^1) requirement on the operator F with respect to the state variable x in (H2), the at-point differentiability assumptions on $\varphi_0, \ldots, \varphi_{m+r}$ in (H3) and (H4) are weaker than smoothness (typically required in control theory outside of nonsmooth analysis) even in finite dimensions. Thus the optimal control problem (1)–(4) is not generally smooth with respect to the states x. However, the major source of nonsmoothness in this infinite-dimensional optimization problem comes from the geometric constraint on control actions $u \in \mathcal{U}$; see also Sect. 5.

To conclude this section, we illustrate the validity of the major assumptions (H1) and (H2) for an interesting class of controlled operator equations (2) in Hilbert spaces.

Example 2.1 (Illustrating major assumptions) Consider the differential equation

$$-x''(t) + \alpha^2 x(t) = g(\langle x, \zeta \rangle)\psi + u, \; x(0) = x(1) = 0,$$

where $\alpha > 0$, $\zeta, \psi \in L^2(0, 1)$ and where $g : \mathbb{C} \to \mathbb{C}$ has bounded derivative; say, $|g'| \leq \beta$. Then the operator $\widetilde{T}x := -x'' + \alpha^2 x$ on $L^2(0, 1)$ with

$$\widetilde{D} = \left\{ x \in L^2(0, 1) \,\middle|\, x, x' \text{ are absolutely continuous, } x'' \in L^2(0, 1), x(0) = x(1) = 0 \right\}$$

is self-adjoint. Also $\langle \widetilde{T}x, x \rangle = \|x'\|^2 + \alpha^2 \|x\|^2 \geq \alpha^2 \|x\|^2$, and so $\widetilde{T}$ satisfies (H1).

Denote next $F(x, u) := g(\langle x, \zeta \rangle)\psi + u$ and verify the following properties of this mapping:

- F is *Fréchet differentiable and weakly continuous in* x. To see this, observe that $\frac{\partial}{\partial x} g(\langle x, \zeta \rangle)h = g'(\langle x, \zeta \rangle) \langle h, \zeta \rangle$, and hence $F'_x(x, u)h = g'(\langle x, \zeta \rangle) \langle h, \zeta \rangle \psi$. Assuming now that $x_n \rightharpoonup x$ (weakly converges) gives us $\langle x_n, \zeta \rangle \longrightarrow \langle x, \zeta \rangle$, and the continuity of g ensures that $g(\langle x_n, \zeta \rangle) \longrightarrow g(\langle x, \zeta \rangle)$. Thus for any $v \in H$ we get that

$$\langle g(\langle x_n, \zeta \rangle)\psi, v \rangle = g(\langle x_n, \zeta \rangle) \langle \psi, v \rangle \longrightarrow g(\langle x, \zeta \rangle) \langle \psi, v \rangle = \langle g(\langle x, \zeta \rangle)\psi, v \rangle,$$

which implies the claimed weak continuity of the mapping F.

- F is *monotone*. Indeed, for any $x, y \in H$ we have

$$\begin{aligned}
\langle F(x, u) - F(y, u), x - y \rangle &= \langle (g(\langle x, \zeta \rangle) - g(\langle y, \zeta \rangle))\, \psi, x - y \rangle \\
&\leq \|(g(\langle x, \zeta \rangle) - g(\langle y, \zeta \rangle))\, \psi\| \, \|x - y\| \\
&= |g(\langle x, \zeta \rangle) - g(\langle y, \zeta \rangle)| \, \|\psi\| \, \|x - y\| \\
&\leq \beta \, |\langle x - y, \zeta \rangle| \, \|\psi\| \, \|x - y\| \\
&\leq \beta \, \|\zeta\| \, \|\psi\| \, \|x - y\|^2 .
\end{aligned}$$

Then (H2) is satisfied if $\beta < \alpha^2$. Note also that introducing

$$\Psi(t, \tau) := \frac{1}{\alpha \cosh \alpha} \begin{cases} \sinh \alpha(1 - t) \sinh \alpha\tau, & 0 \leq \tau \leq t \\ \sinh \alpha(1 - \tau) \sinh \alpha t, & t \leq \tau \leq 1 \end{cases},$$

we can rewrite our system as the integral equation

$$x(t) = \widetilde{T}^{-1} F(x,u)(t) = \int_0^1 \Psi(t,\tau) F(x,u)(\tau)\, \mathrm{d}\tau$$

with the Hilbert–Schmidt kernel $\Psi(t,\tau)$ and thus conclude by basic functional analysis that the operator $\widetilde{T}^{-1}$ is compact.

While deriving the main result on necessary optimality conditions in Sects. 5–7, we further specify the state and control spaces, imposing in this way an additional assumption on behavior of the underlying nonlinear operator F in (2) with respect to "needle" variations of the optimal control; see Sect. 5 for more details.

3 Extensions of Self-Adjoint Operators

In this section we recall some notation and results from [11] needed in what follows.

Given natural numbers $k, r \in \mathbb{N}$ and elements $X \in H^k$, $Y \in H^r$ from the power spaces of H, the *matrix/operator inner product* $\langle X, Y\rangle$ is defined by applying the inner product in H to the entries of the formal matrix XY^*, where the symbol * signifies the duality/transposition operation in building the *adjoints*. In other words,

$$\langle X, Y\rangle_{ij} := \left\langle x_i, y_j \right\rangle \quad \text{as} \quad i = 1, \ldots, k \quad \text{and} \quad j = 1, \ldots, r.$$

It is easy to see that $\langle AX, BY\rangle = A\langle X, Y\rangle B^*$ whenever $A \in \mathbb{C}^{m\times k}$ and $B \in \mathbb{C}^{d\times r}$, where $\mathbb{C}$ stands as usual for the collection of complex numbers.

Fixing $n \in \mathbb{N}$, we say that the components of the vector $Z = (z_1, \ldots, z_n) \in H^n$ are *linearly independent modulo* $\widetilde{D}$ if the inclusion $\alpha \in \mathbb{C}^{1\times n}$ with $\alpha Z \in \widetilde{D}$ holds only when $\alpha = 0$. Let us mention that the assumption that $\widetilde{T}$ is self-adjoint implies that it is closed. This together with the unboundedness of $\widetilde{T}$ mean, by the classical closed graph theorem, that $\widetilde{D}$ is a proper subspace of H. We should note here that the construction described below does not apply to the case of bounded operators, since in this case the bilinear form defined in (8) is exactly zero. The bounded operator case is of our ongoing research.

Taking now any number $\lambda \in \mathbb{C}$ with $\mathrm{Im}\lambda \neq 0$ and the complex conjugate $\overline{\lambda}$, define the *Cayley transform* of the operator $\widetilde{T}$ by

$$V := \left(\widetilde{T} - \lambda I\right)\left(\widetilde{T} - \overline{\lambda} I\right)^{-1},$$

where $I \colon H \to H$ is the identity operator on H. Denoting further the *span* of $Z = (z_1, \ldots, z_n)$ by $[Z] := \mathrm{span}(Z)$ and taking its *orthogonal complement* $[Z]^\perp$, we form the new domain

$$D_0 := (I - V)\,[Z]^\perp$$

and have the *domain representation* established in [11]:

$$\widetilde{D} = D_0 \dotplus [W] \quad \text{with} \quad W := (I - V)\, Z. \tag{7}$$

Now we are ready to construct the *n-dimensional extension* T of the original operator $\widetilde{T}$ in (2) in the following two-step way:

$$T_0 := \widetilde{T}\big|_{D_0} \quad \text{and} \quad T := T_0^*.$$

It is easy to see that T_0 and T are closed operators, T_0 is symmetric and relates to $\widetilde{T}$ as

$$T_0 \subset \widetilde{T} \subset T.$$

Furthermore, the *domain* D of the extension T relates to the original one $\widetilde{D}$ as

$$D = \widetilde{D} \dotplus [Z]\,.$$

The operator T generates the *antisymmetric sesquilinear form* $[\cdot,\cdot] : D \times D \to \mathbb{C}$ defined by

$$[x, y] := \langle Tx, y\rangle - \langle x, Ty\rangle\,, \tag{8}$$

which can be extended to the product vectors $X = (x_1, \ldots, x_k) \in H^k$ and $Y = (y_1, \ldots, y_r) \in H^r$ in the same way as the vector inner products above. We have the relationships

$$D_0 = \big\{x \in D \big|\ [x, W] = [x, Z] = 0\big\} \quad \text{and} \quad \widetilde{D} = \big\{x \in D \big|\ [x, W] = 0\big\} \tag{9}$$

for the product vectors W and Z under consideration. Moreover, the complex matrices $[W, Z]$ and $[Z, Z]$ are *isomorphisms* in $\mathbb{C}^n$ allowing us to show that the equation $[W, p] = \alpha$ is *solvable* for any given $\alpha \in \mathbb{C}^n$; see [11] for more details.

4 σ-Adjoint Problem and Increment Formula

Having in hands the results of Sect. 3, we can now proceed with constructing the auxiliary and adjoint operator systems to the original one (2) with taking into account the cost and constraint functions φ_i, $i = 0, \ldots, m + r$, in the control problem (1)–(4) formulated in Sect. 2. In fact, in this section we are going to do it with respect to an arbitrary differentiable function σ on H, which will be specified later on (Sect. 5) via the initial data of the control problem under consideration.

Observe first that the self-adjoint property of the operator $\widetilde{T}$ in (2) and the imposed boundedness from below assumption (5) in (H1) imply that $\lambda = 0$ is a *resolvent point* for $\widetilde{T}$. Moreover, it follows from the results of Sect. 3 and from [3] that the point $\lambda = 0$ is of *regular type* for T_0, and we have the equalities

$$\dim \mathcal{R}_{T_0}^{\perp} = \dim \left(\ker T\right) = n,$$

where the symbol $\mathcal{R}_{T_0}$ stands for the range of the operator T_0.

Let further the vectors $z_1, \ldots, z_n$ form a basis of the kernel subspace $\ker T \subset H$, and let $w_i := \widetilde{T}^{-1} z_i \in \widetilde{D}$ for $i = 1, \ldots, n$. It is easy to see that the relationships in (9) hold with $Z := (z_1, \ldots, z_n)$ and $W := (w_1, \ldots, w_n)$. Denoting by P the *projector operator* onto the closed subspace $\mathcal{R}_{T_0}$, we get that the mapping $(I - P)$ projects onto the orthogonal complement of $\mathcal{R}_{T_0}$, which is precisely $\ker T = [Z]$.

For any $w \in H$ and $\alpha \in \mathbb{C}^n$ consider now the following problems under assumption (H2):

$$Ty = F_x'^* (\overline{x}, \overline{u})\, y, \quad [W, y] = \alpha, \tag{10}$$

$$\widetilde{T} y = F_x'^* (\overline{x}, \overline{u})\, y + w. \tag{11}$$

It is shown in [11] (see Sect. 3) that under the assumptions made problem (10) admits the *unique solution* $p \in D$, while the uniqueness of the solution $q \in \widetilde{T}$ to (11) follows from the coercivity of the operator $\widetilde{T} - F_x'^*(\overline{x}, \overline{u})$, which is a consequence of the inequality $\rho < \mu$; see assumptions (H1) and (H2).

Next we fix a pair $(\bar{u}, \bar{x}) \in \mathcal{A}$ [which is treated later as an optimal solution to the control problem (1)–(4)] and take an arbitrary function $\sigma : H \to \mathbb{R}$, which is *Fréchet differentiable* at $\overline{x}$. Using the function σ and the data of the initial operator equation (2) at $(\bar{u}, \bar{x})$, let us introduce the σ*-auxiliary equation*

$$\widetilde{T} y = F_x'^* (\overline{x}, \overline{u})\, y - F_x'^* (\overline{x}, \overline{u})\, P \widetilde{T}^{-1} \sigma' (\overline{x}), \tag{12}$$

and the σ*-adjoint problem* at $(\bar{u}, \bar{x})$ defined via the operator extension T by

$$Ty = F_x'^* (\overline{x}, \overline{u})\, y, \quad [y, W] = \left\langle \sigma' (\overline{x}), W \right\rangle. \tag{13}$$

Observe that the σ-auxiliary equation (12) is the same as (11) with $w := -F_x'^* (\overline{x}, \overline{u})\, P \widetilde{T}^{-1} \sigma' (\overline{x})$ having hence the unique solution $q \in \widetilde{D}$, while the σ-adjoint problem (13) is the same as (10) with $\alpha := \langle \sigma'(\bar{x}), W \rangle$ and has therefore the unique solution $p \in D$.

The following proposition shows how to obtain a solution to the unified problem constructed upon (12) and (13).

Proposition 4.1 (Unified solution to the σ-auxiliary and σ*-adjoint problems*) *Let $q \in \widetilde{D}$ and $p \in D$ solve the σ-auxiliary equation* (12) *and the σ-adjoint problem* (13)*, respectively. Then $\widehat{p} := p + q$ solves the unified problem*

$$Ty + F_x'^* (\overline{x}, \overline{u})\, P \widetilde{T}^{-1} \sigma' (\overline{x}) - F_x'^* (\overline{x}, \overline{u})\, y = 0, \quad [y, W] = \left\langle \sigma' (\overline{x}), W \right\rangle.$$

Proof It follows from the direct calculations that

$$\begin{aligned} T\widehat{p} &= T\,(p+q) = Tp + \widetilde{T}q \\ &= F_x'^*\,(\overline{x},\overline{u})\,p + F_x'^*\,(\overline{x},\overline{u})\,q - F_x'^*\,(\overline{x},\overline{u})\,P\widetilde{T}^{-1}\sigma'\,(\overline{x}) \\ &= F_x'^*\,(\overline{x},\overline{u})\,\widehat{p} - F_x'^*\,(\overline{x},\overline{u})\,P\widetilde{T}s^{-1}\sigma'\,(\overline{x})\,, \\ [\widehat{p},W] &= [p+q,W] = [p,W] + [q,W] \\ &= [p,W] = \left\langle \sigma'\,(\overline{x})\,,W \right\rangle \end{aligned}$$

since $[q, W] = 0$ by (9). This verifies the conclusion made. □

Next we take an arbitrary pair $(u, x) \in \mathcal{A}$ feasible to the original operator equation (2) and define the following *increment values* in comparison with the reference pair $(\bar{u}, \bar{x})$:

$$\begin{aligned} \triangle\overline{x} &:= (x - \overline{x})\,, \\ \triangle\sigma &:= \sigma\,(x) - \sigma\,(\overline{x})\,, \\ \triangle_x F\,(\overline{x},u) &:= F\,(x,u) - F\,(\overline{x},u)\,, \\ \triangle_u F\,(x,\overline{u}) &:= F\,(x,u) - F\,(x,\overline{u})\,. \end{aligned}$$

It is easy to deduce from the smoothness assumptions in (H2) that we have the representation

$$\begin{aligned} F(x,u) - F(\overline{x},\overline{u}) &= F(x,u) - F(\overline{x},u) + \triangle_u F(\overline{x},\overline{u}) \\ &= F_x'(\overline{x},u)\triangle x + \triangle_u F(\overline{x},\overline{u}) + o\,(\|\triangle x\|) \\ &= F_x'(\overline{x},\overline{u})\triangle x + \triangle_u F_x'(\overline{x},\overline{u})\triangle x + o\,(\|\triangle x\|) \end{aligned} \tag{14}$$

The next important result gives us a constructive formula representing the increment of σ via the corresponding solution of the original and adjoint problems.

Proposition 4.2 (Increment formula) *Let* $p, q, \widehat{p}$ *be taken from Proposition* 4.1*, and let* $z := \widehat{p} - P\widetilde{T}^{-1}\sigma'\,(\overline{x})$*. Then we have the representation*

$$\triangle\sigma = -\left\langle z, \triangle_u F\,(\overline{x},\overline{u})\right\rangle - \left\langle z, \triangle_u F_x'\,(\overline{x},\overline{u})\,\triangle\overline{x}\right\rangle + o\,(\|\triangle\overline{x}\|)\,. \tag{15}$$

Proof It follows from the Fréchet differentiability of σ at $\bar{x}$ that

$$\triangle\sigma = \sigma\,(x) - \sigma\,(\overline{x}) = \left\langle \sigma'\,(\overline{x})\,,\triangle\overline{x}\right\rangle + o\,(\|\triangle\overline{x}\|)\,. \tag{16}$$

Picking $x \in \widetilde{D}$ and using the domain representation in (7), we find $x_0 \in D_0$ and a vector β with n complex components giving us

$$x = x_0 + \beta^* W, \;\; \overline{x} = \overline{x}_0 + \overline{\beta}^* W, \;\; \text{and} \;\; \triangle\overline{x} = \triangle\overline{x}_0 + \triangle\overline{\beta}^* W.$$

Taking into account that $\triangle\overline{x}_0 \in D_0$, using the relationships in (9) and (13), and then employing the standard transformations tell us that

$$\begin{aligned}\left\langle \sigma'(\overline{x}), \triangle\overline{x}\right\rangle &= \left\langle \sigma'(\overline{x}), \triangle\overline{x}_0\right\rangle + \left\langle \sigma'(\overline{x}), W\right\rangle \triangle\overline{\beta} \\ &= \left\langle \sigma'(\overline{x}), \triangle\overline{x}_0\right\rangle + [p, W]\triangle\overline{\beta} \\ &= \left\langle \sigma'(\overline{x}), \triangle\overline{x}_0\right\rangle + \left[\widehat{p}, \triangle\overline{\beta}^* W\right] \\ &= \left\langle \sigma'(\overline{x}), \triangle\overline{x}_0\right\rangle + \left[\widehat{p}, \triangle\overline{x}_0 + \triangle\overline{\beta}^* W\right] \\ &= \left\langle \sigma'(\overline{x}), \triangle\overline{x}_0\right\rangle + [\widehat{p}, \triangle\overline{x}],\end{aligned}$$

where we use the fact that $[\widehat{p}, \triangle x_0] = 0$. To see the latter, observe that

$$\begin{aligned}[\widehat{p}, \triangle x_0] &= \langle T\widehat{p}, \triangle x_0\rangle - \langle \widehat{p}, T\triangle x_0\rangle \\ &= \langle T\widehat{p}, \triangle x_0\rangle - \langle \widehat{p}, T_0\triangle x_0\rangle \ (\text{by } \triangle x_0 \in D_0) \\ &= \ \langle T\widehat{p}, \triangle x_0\rangle - \langle T\widehat{p}, \triangle x_0\rangle = 0.\end{aligned}$$

It follows from (2), (14) due to the smoothness assumptions made that

$$\begin{aligned}[\widehat{p}, \triangle\overline{x}] &= \langle T\widehat{p}, \triangle\overline{x}\rangle - \langle \widehat{p}, T\triangle\overline{x}\rangle = \langle T\widehat{p}, \triangle\overline{x}\rangle - \left\langle \widehat{p}, \widetilde{T}\triangle\overline{x}\right\rangle \ (\text{by } \triangle\overline{x} \in \widetilde{D}) \\ &= \langle T\widehat{p}, \triangle\overline{x}\rangle - \langle \widehat{p}, F(x,u) - F(\overline{x},\overline{u})\rangle . \\ &= \left\langle T\widehat{p} - F_x'^*(\overline{x},\overline{u})\,\widehat{p}, \triangle\overline{x}\right\rangle \\ &\quad - \left\langle \widehat{p}, \triangle_u F_x'(\overline{x},\overline{u})\,\triangle\overline{x} + \triangle_u F(\overline{x},\overline{u})\right\rangle + o(\|\triangle\overline{x}\|).\end{aligned}$$

Inserting now (14) in the representation above gives us the equalities

$$\begin{aligned}[\widehat{p}, \triangle\overline{x}] &= \langle T\widehat{p}, \triangle\overline{x}\rangle - \left\langle \widehat{p}, F_x'(\overline{x},\overline{u})\,\triangle\overline{x} + \triangle_u F_x'(\overline{x},\overline{u})\,\triangle\overline{x} + \triangle_u F(\overline{x},\overline{u})\right\rangle \\ &\quad + o(\|\triangle\overline{x}\|). \\ &= \langle T\widehat{p}, \triangle\overline{x}\rangle - \left\langle \widehat{p}, F_x'(\overline{x},\overline{u})\,\triangle\overline{x}\right\rangle - \left\langle \widehat{p}, \triangle_u F_x'(\overline{x},\overline{u})\,\triangle\overline{x} + \triangle_u F(\overline{x},\overline{u})\right\rangle \\ &\quad + o(\|\triangle\overline{x}\|).\end{aligned}$$

Since $F_x'(\overline{x},\overline{u})$ is a bounded linear operator, the adjoint operator $F_x'^*(\overline{x},\overline{u})$ is well defined and bounded; hence we have $\left\langle \widehat{p}, F_x'(\overline{x},\overline{u})\,\triangle\overline{x}\right\rangle = \left\langle F_x'^*(\overline{x},\overline{u})\,\widehat{p}, \triangle\overline{x}\right\rangle$. Substituting this into the formula above brings us to the expression

$$[\widehat{p}, \triangle\overline{x}] = \left\langle T\widehat{p} - F_x'^*(\overline{x},\overline{u})\,\widehat{p}, \triangle\overline{x}\right\rangle - \left\langle \widehat{p}, \triangle_u F_x'(\overline{x},\overline{u})\,\triangle\overline{x} + \triangle_u F(\overline{x},\overline{u})\right\rangle + o(\|\triangle\overline{x}\|).$$

Employing further the constructions in (12) and (13) as well as Proposition 4.1 yields

$$\begin{aligned}\left\langle \sigma'(\overline{x}), \triangle\overline{x}_0\right\rangle &= \left\langle \sigma'(\overline{x}), \widetilde{T}^{-1}P\,(F(x,u) - F(\overline{x},\overline{u}))\right\rangle \\ &= \left\langle P\widetilde{T}^{-1}\sigma'(\overline{x}), F(x,u) - F(\overline{x},\overline{u})\right\rangle \\ &= \left\langle P\widetilde{T}^{-1}\sigma'(\overline{x}), F_x'(\overline{x},\overline{u})\,\triangle\overline{x} + \triangle_u F_x'(\overline{x},\overline{u})\,\triangle\overline{x} + \triangle_u F(\overline{x},\overline{u})\right\rangle \\ &\quad + o(\|\triangle\overline{x}\|) \\ &= \left\langle F_x'^*(\overline{x},\overline{u})\,P\widetilde{T}^{-1}\sigma'(\overline{x}), \triangle\overline{x}\right\rangle\end{aligned}$$

$$+\left\langle P\widetilde{T}^{-1}\sigma'(\overline{x}),\, \triangle_u F'_x(\overline{x},\overline{u})\,\triangle\overline{x}+\triangle_u F(\overline{x},\overline{u})\right\rangle$$
$$+o(\|\triangle\overline{x}\|).$$

Adding the expressions for $\left\langle\sigma'(\overline{x}),\triangle\overline{x}_0\right\rangle$ and $[\widehat{p},\triangle\overline{x}]$ above and observing from Proposition 4.1 that $T\widehat{p}-F_x'^{*}(\overline{x},\overline{u})\,\widehat{p}+F_x'^{*}(\overline{x},\overline{u})\,P\widetilde{T}^{-1}\sigma'(\overline{x})=0$, we get the representation

$$\left\langle\sigma'(\overline{x}),\triangle\overline{x}\right\rangle=\left\langle P\widetilde{T}^{-1}\sigma'(\overline{x})-\widehat{p},\, \triangle_u F'_x(\overline{x},\overline{u})\,\triangle\overline{x}+\triangle_u F(\overline{x},\overline{u})\right\rangle+o(\|\triangle\overline{x}\|).$$

Substituting finally the latter into (16) justifies the increment formula (15). ◻

5 Maximum Principle for Operator Control Systems with Constraints on State Variables

This section presents the main result of the paper. Having in mind the variational method use in what follows, we consider here a particular structure of the set $\mathcal{U}$ of admissible controls.

Let U be an arbitrary subset of a Banach space, and let I be an interval of $\mathbb{R}$. A control function $u(\cdot)$ is called *admissible* if it is Lebesgue measurable on I and satisfying

$$u(t)\in U\quad\text{for a.e. }\ t\in I. \tag{17}$$

Define $\mathcal{U}$ to be a set of controls satisfying (17).

Given a real Hilbert space X, consider the state space $H:=L^2(I,X)$ of square integrable maps $x:I\to X$ equipped with the inner product and norm

$$\langle x,y\rangle=\int_I\langle x(t),y(t)\rangle_X\,\mathrm{d}t,\quad \|x\|^2=\int_I\|x(t)\|_X^2\,\mathrm{d}t,$$

respectively, where $\langle\cdot,\cdot\rangle_X$, $\|\cdot\|_X$ denote the inner product and norm in X. A typical example of a self-adjoint operator on this space is $\widetilde{T}x:=x_{tt}+\triangle x$ with appropriate boundary conditions and with X being an appropriate Sobolev space. From now on we concentrate on the study of the following optimal control problem (P) for general operator equations: it is (1)–(4) with admissible controls given by (17) and the corresponding trajectories $x(\cdot)\in L^2(I,X)$ of the operator equation (2). Since H is a real Hilbert space, we need to clarify how extensions of $\widetilde{T}$ can be carried out. This is guided by the construction developed in Sect. 4. To clarify it, choose $z_1,\ldots,z_n$ linearly independent modulo $\widetilde{D}$, let $R_0:=[Z]^{\perp}$, $D_0:=\widetilde{T}^{-1}R_0$, and let $W:=\widetilde{T}^{-1}Z$. The extension T of $\widetilde{T}$ is defined as follows. Put $D(T)=D:=\widetilde{D}\dotplus[Z]$, then write $x=\widetilde{x}+\beta Z$ for any $x\in D$ with some $\beta\in\mathbb{C}^{1\times n}$, and define $Tx:=\widetilde{T}\widetilde{x}$. Thus we get $[Z]=\ker T$. As in Sect. 3, define now T_0 by $D(T_0)=D_0$, $T_0x=\widetilde{T}x$ whenever $x\in D_0$. In this setting we say that a pair (u,x) is *feasible* to (P) if $u(\cdot)$ is admissible by (17) and all the relationships in (2)–(4) hold for (u,x). We keep the notation $\mathcal{A}$ and $\mathcal{B}$

from Sect. 2 for the classes of admissible and feasible pairs (u, x) to (P), respectively. It is worth mentioning here the choice of the state space $H = L^2(I, X)$ is not conventional for standard optimal control problem governed, e.g., by ordinary differential equations while it occurs to be the most appropriate for the general operator equation model (P) due the scalar product structure in $L^2(I, X)$.

Our major goal is to establish necessary optimality conditions for the given feasible pair $(\bar{u}, \bar{x})$ minimizing the cost functional (1) over $\mathcal{B}$. To formulate this result, we suppose for convenience that the operator F in (2) is defined via its *pointwise* values $F(x, u)(t)$ with respect to $t \in I$ and introduce the generalized *Hamilton–Pontryagin function*

$$\mathcal{H}\big(x(t), y(t), p(t), u(t)\big) := \Big\langle p(t) - \Big(P\widetilde{T}^{-1}y\Big)(t), F(x,u)(t)\Big\rangle_X \tag{18}$$

constructing entirely in terms of its initial data for any feasible solution (u, x) to (P) and the corresponding adjoint trajectories, where $y(t)$ is specified below. Note that, in particular, the operator F can be defined pointwise by $F(x, u)(t) := f(x(t), u(t))$, which is the case for various classes of differential and other evolution equations; see, e.g., [1,2,4,6,7,9,10,12].

To formulate and prove the desired result, we need to add to (H1)–(H4) one more assumption concerning the required behavior of the operator F with respect to the following class of (single) *needle variations* u_ε of the reference control $\bar{u}$ defined by

$$u_\varepsilon(t) := \begin{cases} v(t), & \text{if } t \in I_\varepsilon, \\ \bar{u}(t), & \text{if } t \notin I_\varepsilon, \end{cases} \tag{19}$$

for $\varepsilon > 0$, where $I_\varepsilon \subset I$ is a measurable subset of I with $\text{mes}(I_\varepsilon) \le \varepsilon$, and where v is a measurable function on I_ε with $v(t) \in U$ a.e. $t \in I_\varepsilon$. Note the first use of needle variations goes back to the Chicago School of the calculus of variations in handling abnormal problems which arose in that field in the late 1930s; see [13]. The broad contemporary usage of control variations of this type for systems governed by ordinary differential equations has started in the Russian School of optimal control in the 1950s while being crucially employed in the proof of the Pontryagin Maximum Principle [2]. Observe also that *multineedle* control variations used below in the proof of Theorem 5.1 can be compared with the so-called distributed spike variations, which play an important role in deriving necessary optimality conditions for optimal control problems governed by partial differential equations and other control systems in infinite dimensions; see, e.g., [7].

The required additional assumption related to (single) needle variations (19) is as follows:

(H5) For every variation u_ε in (19) on a union of disjoint intervals of total length ε we have

$$\|F(\bar{x}, u_\varepsilon) - F(\bar{x}, \bar{u})\| = o(\varepsilon) \text{ as } \varepsilon \downarrow 0.$$

Note that (H5) postulates a natural property of nonlinear operators F smooth with respect to state variables, which holds for a number of control systems considered, e.g., in [1,2,4,6,9,10,12]. It follows from (H5) and the assumptions imposed in (H2) on F that, given any sequence of intervals $\{I_\varepsilon\}$, we have the increment estimate

$$\|\Delta\bar{x}\| = o(\varepsilon) \ \text{ with } \ \Delta\bar{x} = x - \bar{x} \tag{20}$$

for the corresponding state increment in (2) generated by needle control variations. This can be seen from the following transformations:

$$\begin{aligned}(\mu - \rho)\|\Delta\overline{x}\|^2 &= \mu\|\Delta\overline{x}\|^2 - \rho\|\Delta\overline{x}\|^2 \le \left\langle \widetilde{T}\Delta\overline{x}, \Delta\overline{x}\right\rangle \\ &\quad - \langle F(x, u_\varepsilon) - F(\overline{x}, u_\varepsilon), \Delta\overline{x}\rangle \\ &= \langle F(x, u_\varepsilon) - F(\overline{x}, \overline{u}), \Delta\overline{x}\rangle - \langle F(x, u_\varepsilon) - F(\overline{x}, u_\varepsilon), \Delta\overline{x}\rangle \\ &= \langle F(\overline{x}, u_\varepsilon) - F(\overline{x}, \overline{u}), \Delta\overline{x}\rangle \le \|F(\overline{x}, u_\varepsilon) - F(\overline{x}, \overline{u})\| \, \|\Delta\overline{x}\|.\end{aligned}$$

The next theorem is our main *Maximum Principle* for the constrained control problem (P) governed by general operator equations. It asserts that the reference optimal control function $\bar{u}(\cdot)$ gives the pointwise maximum values over U to the appropriate Hamilton–Pontryagin function (18) along the uniquely defined solutions to the original and adjoint operator equations satisfying in the latter case the additional conditions formulated via the given cost and constraint functions φ_i, $i = 0, \ldots, m + r$. In the absence of the imposed constraints (3) and (4) on the state variables, the result obtained reduces to the Maximum Principle from our preceding paper [11] given in a bit different while equivalent form. As the reader can see in what follows, the proof in the constrained case is significantly more involved in comparison with the unconstrained one presented in [11].

Note that the adjoint system in the theorem below is formulated via the extension T of the original operator generated by $W = (w_1, \ldots, w_n)$ as described in Sect. 3 and further specified at the beginning of this section. Furthermore, the auxiliary and adjoint systems in the theorem are the σ-ones from (12) and (13) with

$$\sigma(x) = \phi(x) := \sum_{i=0}^{m+r} \lambda_i \varphi_i(x), \quad x \in H. \tag{21}$$

Theorem 5.1 (Maximum Principle for constrained control systems) *Let* $(\bar{u}, \bar{x})$ *be an optimal solution to problem* (P) *under the assumptions in* (H1)–(H5)*, and let* $z_1, \ldots, z_n$ *be real functions that are linearly independent modulo* $\widetilde{D}$*. Set* $Z := (z_1, \ldots, z_n)$*, define*
$W := \widetilde{T}^{-1}Z$ *and* D_0, T_0, D, T *as above. Then there are multipliers* $(\lambda_0, \ldots, \lambda_{m+r}) \neq 0$ *satisfying the sign and complementary slackness conditions*

$$\lambda_i \ge 0 \ \text{ for } \ i = 0, \ldots, m, \tag{22}$$

$$\lambda_i \varphi_i(\bar{x}) = 0 \ \text{ for } \ i = 1, \ldots, m \tag{23}$$

as well as the unique solutions $q \in \widetilde{D}$ of the auxiliary system

$$\widetilde{T} y = F_x'^{*}(\overline{x}, \overline{u})\, y - F_x'^{*}(\overline{x}, \overline{u})\, P \widetilde{T}^{-1} \phi'(\overline{x}) \tag{24}$$

and $p \in D$ of the adjoint system

$$T y = F_x'^{*}(\overline{x}, \overline{u})\, y, \quad [y, W] = \left\langle \phi'(\overline{x}), W \right\rangle \tag{25}$$

with ϕ from (21) depending on $\lambda_0, \ldots, \lambda_{m+r}$ such that the following maximum condition

$$\mathcal{H}\big(\overline{x}(t), \phi'(\overline{x})(t), \widehat{p}(t), \overline{u}(t)\big) = \max_{u \in U} \left\{\mathcal{H}\big(\overline{x}(t), \phi'(\overline{x})(t), \widehat{p}(t), u\big)\right\} \ \ a.e.\ \ t \in I \tag{26}$$

holds, where $\widehat{p}(t) := p(t) + q(t)$, and where the maximization of the Hamilton–Pontryagin function (18) on the right-hand side of (26) is understood as

$$\mathcal{H}\big(\overline{x}(t), \phi'(\overline{x})(t), \widehat{p}(t), \overline{u}(t)\big) \geq \mathcal{H}\big(\overline{x}(t), \phi'(\overline{x})(t), \widehat{p}(t), v(t)\big) \ \ a.e.\ \ t \in I \tag{27}$$

for any measurable function $v(t) \in U$ admissible by (17).

The proof of this theorem will be given in Sects. 6 and 7 for the two essentially distinct cases: of only the inequality constraints (3) and of only the equality constraints (4), respectively. This is done for simplicity of the arguments by taking into account that the proofs in these two cases are different. The reader can observe that the proof of the theorem in the general case can be obtained as the combination of the arguments in the cases of inequality and equality constraints. Observe that in contrast to the unconstrained problem [11], where the proof is based on using the increment formula for the cost functions (1) on *single*-needle control variations (19), here in both cases of inequality and equality constraints we need to apply the σ-increment formula of Sect. 4 for $\sigma = \phi$ in (21) on more involved *multineedle* variations of the optimal control constructed as follows.

Fix a natural number $M > 0$ and pick M points $\tau_j \in I$ such that $\tau_1 < \cdots < \tau_M$. Let $\{I_i\}_{i=1}^M$ be a collection of subintervals of I such that $I_i := [\tau_i, \upsilon_i[$ for $i = 1, \ldots, M$, while yields $\upsilon_i < \tau_{i+1}$ as $i = 1, \ldots, M-1$. Select further $0 < \varepsilon_0 \leq \min_i |I_i|$ and let $\varepsilon_1 := \frac{1}{M}\varepsilon_0$. For any $i = 1, \ldots, M$ we let $N_i \in \mathbb{N}$, take $N_i + 1$ points $\left\{\gamma_{i,j}\right\}_{j=0}^{N_i}$ with $0 = \gamma_{i,0} < \gamma_{i,1} < \cdots < \gamma_{i,N_i} \leq 1$, and define the subintervals $I_{i,j}$ of I_i by

$$I_{i,j} := \left[\tau_i + \varepsilon_1 \gamma_{i,j-1}, \tau_i + \varepsilon_1 \gamma_{i,j}\right[, \quad 1 \leq j \leq N_i.$$

Then the total number of these subintervals is $N_1 + \cdots + N_M$ and their total length does not exceed $M\varepsilon_1 = \varepsilon_0$. Pick now $\varepsilon \in [0, \varepsilon_1)$, define $I_{i,j}^{\varepsilon} \subset I_{i,j}$ by $I_{i,j}^{\varepsilon} := [\tau_i + \varepsilon\gamma_{i,j-1}, \tau_i + \varepsilon\gamma_{i,j}[$, and let $I_0 := I \backslash \cup I_{i,j}^{\varepsilon}$. Denoting $\alpha_{i,j} := \gamma_{i,j} - \gamma_{i,j-1}$, choose arbitrary elements $v_{i,j} \in U$ for $i = 1, \ldots, M$ and $j = 1, \ldots, N_i$ and then define the *multineedle variation* u of the optimal control $\overline{u}$ with parameters $\left(\tau_i, v_{i,j}, \alpha_{i,j}, \varepsilon\right)$ by

$$u\left(t\right):=\overline{u}(t)\chi_{I_0}\left(t\right)+\sum_{i=1}^{M}\sum_{j=1}^{N_i}v_{i,j}\chi_{I_{i,j}}\left(t\right), \tag{28}$$

where χ_A stands for the characteristic function of the set A. Observe the partition of unity

$$\chi_{I_0}\left(t\right)+\sum_{i=1}^{M}\sum_{j=1}^{N_i}\chi_{I_{i,j}}\left(t\right)=1,\quad t\in I. \tag{29}$$

As preliminary steps of the proof of Theorem 5.1 in both cases of inequality and equality constraints, we present first some straightforward consequences of the increment formula of Sect. 4 on multineedle variations (28) and also some properties of the linearized system to (2) along such variations of the optimal control.

It follows directly from the increment formula (15) in Proposition 4.2, the structure of (28), and the increment estimate in (20) that

$$\triangle\sigma=-\left\langle z,\triangle_u F\left(\overline{x},\overline{u}\right)\right\rangle+o\left(\varepsilon\right). \tag{30}$$

Consider now a multineedle variation (28) with parameters $\left(\tau_i, v_{i,j}, \alpha_{i,j}, \varepsilon\right)$ and denote by $\Lambda_{\tau_i,v_{i,j}}\bar{x}$ the corresponding solution to the linearized equation

$$\left(\widetilde{T}-F_x'(\bar{x},\bar{u})\right)x=\left(\varepsilon\alpha_{i,j}\right)^{-1}\Delta_{v_{i,j}}F(\bar{x},\bar{u})\chi_{I_{i,j}}\left(t\right).$$

Then the quantity $\varepsilon\alpha_{i,j}\Lambda_{\tau_i,v_{i,j}}\bar{x}$ obviously solves the equation

$$\left(\widetilde{T}-F_x'(\bar{x},\bar{u})\right)x=\Delta_{v_{i,j}}F(\bar{x},\bar{u})\chi_{I_{i,j}}\left(t\right). \tag{31}$$

The next technical lemma is useful in the proof of the main theorem.

Lemma 5.1 (State increment along multineedle control variations) *Given a multineedle variation* (28) *with parameters* $\left(\tau_i, v_{i,j}, \alpha_{i,j}, \varepsilon\right)$, *we have the representation*

$$\triangle\overline{x}=\varepsilon\sum_{i=1}^{M}\sum_{j=1}^{N_i}\alpha_{i,j}\Lambda_{\tau_i,v_{i,j}}\bar{x}+o\left(\varepsilon\right).$$

Proof We begin by showing that the operators $(\widetilde{T}-F_x'(\bar{x},u))$ have uniformly bounded inverses independent of u. It follows from the monotonicity assumption (6) that

$$\left\langle F_x'(\bar{x},u)x,x\right\rangle\le\rho\left\|x\right\|^2,\quad x\in H. \tag{32}$$

Therefore, for any $x\in\widetilde{D}$ we have the relationships

$$\begin{aligned}\mu\left\|x\right\|^2&\le\left\langle\widetilde{T}x,x\right\rangle=\left\langle(\widetilde{T}-F_x'(\bar{x},u))x,x\right\rangle+\left\langle F_x'(\bar{x},u)x,x\right\rangle\\&\le\left\langle(\widetilde{T}-F_x'(\bar{x},u))x,x\right\rangle+\rho\left\|x\right\|^2,\end{aligned}$$

which in turn imply the estimates

$$(\mu - \rho)\,\|x\|^2 \le \left\langle (\widetilde{T} - F_x'(\bar{x}, u))x, x \right\rangle \le \left\| (\widetilde{T} - F_x'(\bar{x}, u))x \right\| \|x\|\,,$$
$$\left\| (\widetilde{T} - F_x'(\bar{x}, u))x \right\| \ge (\mu - \rho)\,\|x\|\,.$$

Since the equation $(\widetilde{T} - F_x'^{*}(\bar{x}, u))^{*}x = \widetilde{T} - F_x'^{*}(\bar{x}, u))x = 0$ has only the trivial solution, it follows that the operator $(\widetilde{T} - F_x'(\bar{x}, u))$ is surjective, which establishes the boundedness of the inverse of $(\widetilde{T} - F_x'(\bar{x}, u))$. Note also that condition (32) immediately yields

$$\left\| F_x'(\bar{x}, u) \right\| \le |\rho|\,.$$

Having this in mind, we proceed with the expansion

$$\widetilde{T}\triangle\overline{x} = F\,(x, u) - F\,(\overline{x}, \overline{u}) = F_x'(\bar{x}, \overline{u})\triangle\overline{x} + \triangle_u F\,(\overline{x}, \overline{u})$$
$$+ \triangle_u F_x'(\bar{x}, \overline{u})\triangle\overline{x} + o\,(\|\triangle\overline{x}\|)\,,$$

which gives us the representation

$$(\widetilde{T} - F_x'^{*}(\overline{x}, \overline{u}))\triangle\overline{x} = \triangle_u F\,(\overline{x}, \overline{u}) + \triangle_u F_x'(\bar{x}, \overline{u})\triangle\overline{x} + o\,(\|x\|)\,.$$

Using now construction (28) of the multineedle variation $u(\cdot)$, the partition of unity (29), and the linearized equations (31), we arrive at the equalities

$$\begin{aligned}
\triangle_u F\,(\overline{x}, \overline{u}) &= F\,(\overline{x}, u) - F\,(\overline{x}, \overline{u}) \\
&= F\left(\overline{x}, \overline{u}(t)\chi_{I_0}\,(t) + \sum_{i=1}^{M}\sum_{j=1}^{N_i} v_{i,j}\,\chi_{I_{i,j}}\,(t)\right) \\
&\quad - F\,(\overline{x}, \overline{u})\left(\chi_{I_0}\,(t) + \sum_{i=1}^{M}\sum_{j=1}^{N_i} \chi_{I_{i,j}}\,(t)\right) \\
&= F\,(\overline{x}, \overline{u}(t))\,\chi_{I_0}\,(t) + \sum_{i=1}^{M}\sum_{j=1}^{N_i} F\left(\overline{x}, v_{i,j}\right)\chi_{I_{i,j}}\,(t) - F\,(\overline{x}, \overline{u}(t))\,\chi_{I_0}\,(t) \\
&\quad - \sum_{i=1}^{M}\sum_{j=1}^{N_i} F\,(\overline{x}, \overline{u}(t))\,\chi_{I_{i,j}}\,(t) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{N_i}\left(F\left(\overline{x}, v_{i,j}\right) - F\,(\overline{x}, \overline{u}(t))\right)\chi_{I_{i,j}}\,(t) \\
&= \sum_{i=1}^{M} \triangle_{v_{i,j}} F\,(\overline{x}, \overline{u}(t))\,\chi_{I_{i,j}}\,(t)
\end{aligned}$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{N_i} \varepsilon\alpha_{i,j}(\widetilde{T} - F'_x(\bar{x}, \bar{u}))\Lambda_{\tau_i, v_{i,j}}\bar{x}$$
$$= (\widetilde{T} - F'_x(\bar{x}, \bar{u}))\sum_{i=1}^{M}\sum_{j=1}^{N_i} \varepsilon\alpha_{i,j}\Lambda_{\tau_i, v_{i,j}}\bar{x}.$$

This brings us finally to the equation

$$(\widetilde{T} - F'^{*}_x(\overline{x}, \overline{u}))\triangle\overline{x} = (\widetilde{T} - F'_x(\bar{x}, \bar{u}))\sum_{i=1}^{M}\sum_{j=1}^{N_i} \varepsilon\alpha_{i,j}\Lambda_{\tau_i, v_{i,j}}\bar{x} + \triangle_u F'_x(\bar{x}, \overline{u})\triangle\overline{x} + o\left(\|x\|\right),$$

which gives us, by taking the inverse, the state increment representation

$$\triangle\overline{x} = \sum_{i=1}^{M}\sum_{j=1}^{N_i} \varepsilon\alpha_{i,j}\Lambda_{\tau_i, v_{i,j}}\bar{x} + (\widetilde{T} - F'^{*}_x(\overline{x}, \overline{u}))^{-1}\triangle_u F'_x(\bar{x}, \overline{u})\triangle\overline{x} + o\left(\|x\|\right).$$

To justify the assertion of the lemma, it remains to estimate the term $(\widetilde{T} - F'^{*}_x(\overline{x}, \overline{u}))^{-1}\triangle_u F'_x(\bar{x}, \overline{u})\triangle\overline{x}$ above. Indeed, we have

$$\left\|(\widetilde{T} - F'^{*}_x(\overline{x}, \overline{u}))^{-1}\triangle_u F'_x(\bar{x}, \overline{u})\triangle\overline{x}\right\| \le \left\|(\widetilde{T} - F'^{*}_x(\overline{x}, \overline{u}))^{-1}\right\| \left\|\triangle_u F'_x(\bar{x}, \overline{u})\triangle\overline{x}\right\| \le \frac{2\,|\rho|}{\mu - \rho}\left\|\triangle\overline{x}\right\|,$$

and thus completes the prof by using estimate (20).

For further convenience, we employ the following notation in connection with multineedle variations: $K := N_1 + \cdots + N_M$,

$$\tau := (\tau_i)_{M\times 1}\,,\ \mathrm{v} := \big(v_{ij}\big)_{K\times 1}\,,\ \text{ and }\ \alpha := \big(\alpha_{ij}\big)_{K\times 1}\,,\ \Lambda_{\tau\mathrm{v}}\overline{x} = \big(\Lambda_{\tau_i, v_{ij}}\overline{x}\big)_{K\times 1}\,. \tag{33}$$

In this notation we can rewrite the above representation of $\Delta\bar{x}$ as

$$\Delta\bar{x} = \varepsilon\big\langle\alpha, \Lambda_{\tau,\mathrm{v}}\overline{x}\big\rangle + o(\varepsilon). \tag{34}$$

6 Proof in the Case of Inequality Constraints

Here we consider problem (P) with only the inequality constraints (3) and use the notation

$$\Phi(x) := \big(\varphi_0(x), \varphi_1(x), \ldots, \varphi_m(x)\big).$$

Despite the state and control spaces are infinite-dimensional, the major role in the proof below is played by the following set of the *linearized images* in $\mathbb{R}^{m+1}$:

$$S := \left\{ y = (y_0, \ldots, y_m) \in \mathbb{R}^{m+1} \middle| \; y = \langle \Phi'(\bar{x}), \Lambda_{\tau,\mathrm{v}}\overline{x} \rangle \alpha \;\text{ for some }\; (\tau, \mathrm{v}, \alpha, \varepsilon) \right\} \tag{35}$$

Suppose without loss of generality that all the inequality constraints in (3) are active at the optimal point $\bar{x}$. Denoting by ψ the (unique) solution to the linear equation

$$(\widetilde{T} - F_x'^*)\psi = \Phi'(\bar{x}),$$

for any $y \in S$ we get the representations

$$\begin{aligned} y &= \langle \Phi'(\bar{x}), \Lambda_{\tau,\mathrm{v}}\overline{x} \rangle \alpha = \langle (\widetilde{T} - F_x'^*)\psi, \Lambda_{\tau,\mathrm{v}}\overline{x} \rangle \alpha \\ &= \langle \psi, (\widetilde{T} - F_x')\Lambda_{\tau,\mathrm{v}}\overline{x} \rangle \alpha = \langle \psi, \triangle_{\mathrm{v}} F(\bar{x}, \bar{u}) \rangle \alpha, \end{aligned}$$

with the notation $\triangle_{\mathrm{v}} F(\bar{x}, \bar{u}) := \left(\triangle_{v_{ij}} F(\bar{x}, \bar{u})\right)_{K\times 1}$ in addition to the previous ones.

The next lemma about the *convexity* of S in (35) follows from the above description of S and the construction of *multi*needle variations (due to the possibility to choose admissible partitions with arbitrary small measures, which is a crucial property of the continuous/nonatomic Lebesgue measure on $\mathbb{R}$). It plays an underlying role in the convex separation technique employed below, which is not needed in the absence of state constraints.

Lemma 6.1 (Convexity of linearized images) *The set S in* (35) *is convex.*

Proof Picking any $y, z \in S$ and $\nu \in (0, 1)$, we may regard y and z as generated by the same parameter $(\tau, \mathrm{v}, \varepsilon)$ by refining if necessary the corresponding partitions to a common one. Then we easily come to the equalities

$$\begin{aligned} \nu y + (1-\nu)z &= \nu \langle \Phi'(\bar{x}), \Lambda_{\tau,\mathrm{v}}\overline{x} \rangle \alpha + (1-\nu)\langle \Phi'(\bar{x}), \Lambda_{\tau,\mathrm{v}}\overline{x} \rangle \beta \\ &= \langle \Phi'(\bar{x}), \Lambda_{\tau,\mathrm{v}}\overline{x} \rangle \left(\nu\alpha + (1-\nu)\beta\right). \end{aligned} \tag{36}$$

Recall that the parameter $\alpha = (\alpha_{i,j})$ of the multineedle variation satisfies the condition $\sum_j \alpha_{i,j} < 1$ for all i, and this condition holds true for the combination $\nu\alpha + (1-\nu)\,\beta$. We can see that the last term in Eq. (36) corresponds to the generator

$$(\tau, \mathrm{v}, \nu\alpha + (1-\nu)\beta)$$

of the multineedle variation and so gives us a point of S thus verifying its convexity. □

The following lemma shows that we are in the situation allowing us to separate the set S in (35) from the set of *forbidden points* in (P) due to the optimality of $\bar{x}$ therein under the inequality constraints. This set is the negative orthant of $\mathbb{R}^{m+1}$ given by

$$\mathbb{R}^{m+1}_{<} := \left\{ y \in \mathbb{R}^{m+1} \middle| \; y_i < 0 \;\text{ for all }\; i = 0, \ldots, m \right\}.$$

Lemma 6.2 (On empty intersection) *The linearized image set S from* (35) *does not intersect the negative orthant* $\mathbb{R}^{m+1}_{<}$.

Proof Assume on the contrary that there is a common vector $y \in S \cap \mathbb{R}^{m+1}_{<}$. Then there exists an admissible interval partition and a corresponding multineedle variation such that

$$\langle \Phi'(\bar{x}), \Lambda_{\tau,\mathrm{v}}\overline{x}\rangle\alpha < 0,$$

where the vector inequality is understood in the componentwise sense. Using the Fréchet differentiability of all the functions $\varphi_0, \ldots, \varphi_m$ at $\bar{x}$, we have by (34) that

$$\Phi(x) - \Phi(\bar{x}) = \langle \Phi'(\bar{x}), \Delta\bar{x}\rangle + o(\varepsilon) = \varepsilon\langle \Phi'(\bar{x}), \Lambda_{\tau,\mathrm{v}}\overline{x}\rangle\alpha + o(\varepsilon) < 0$$

for all sufficiently small ε_0. This means that there exists a multineedle variation of the optimal control generating the corresponding trajectory x of (2), which is *feasible* to all the inequality constrains while giving a value smaller than $\varphi_0(\bar{x})$ to the cost function in (1). This is a clear contradiction, which completes the proof of the lemma. □

Now we are in a position to justify Theorem 5.1 in the case of only inequality constraints in (P) under consideration in this section.

Proof of Theorem 5.1 for the control problem with inequality constraints. The classical separation theorem in finite dimensions allows us to separate the convex sets S and $\mathbb{R}^{m+1}_{<}$ in $\mathbb{R}^{m+1}$. This means that there exists a nonzero vector $\lambda = (\lambda_0, \ldots, \lambda_m) \in \mathbb{R}^{m+1}$ *independent* of any multineedle control variation and such that

$$\langle \lambda, y\rangle \geq \langle \lambda, z\rangle \;\text{ for all }\; y \in S \;\text{ and }\; z \in \mathbb{R}^{m+1}_{<}.$$

This immediately implies by the structure of $\mathbb{R}^{m+1}_{<}$ that the vector λ satisfies the sign (22) and complementary slackness (23) conditions (since all the constraints are assumed to be active at $\bar{x}$), and we have furthermore that

$$\langle \lambda, y\rangle \geq 0 \;\text{ for all }\; y \in S.$$

In particular, if y corresponds to a *single*-needle variation of type (19) generated by (τ, v), then it follows from (34) due to assumption (H5) that the inequality

$$\left\langle \phi'(\bar{x}), \triangle_{\tau,v}\bar{x}\right\rangle + o\,(\varepsilon) \geq 0, \tag{37}$$

holds true for all $\tau \in I$, $v \in U$ and all sufficiently small $\varepsilon > 0$, where the function ϕ is defined in (21) along the above multipliers λ_i, $i = 0, \ldots, m$; remember that $r = 0$ in this section. Thus we find $q = q(t) \in \widetilde{D}$ and $p = p(t) \in D$ satisfying the auxiliary and adjoint systems (24) and (25), respectively. Consider now $\widehat{p}(t) = p(t) + q(t)$ and show that the maximum condition (26) is satisfied for a.e. $t \in I$ in the sense specified in (27).

Assume on the contrary that (26) does not hold along $(\bar{u}, \bar{x})$ and then find by (27) a measurable set $E \subset I$ with $\text{mes}(E) \neq 0$ and a measurable function $v(t) \in U$ on E such that

$$\triangle_v \mathcal{H}(t) := \mathcal{H}\big(\overline{x}(t), \phi'(\overline{x})(t), \widehat{p}(t), v(t)\big) - \mathcal{H}\big(\overline{x}(t), \phi'(\overline{x})(t), \widehat{p}(t), \overline{u}(t)\big) > 0, \ \ t \in E. \tag{38}$$

Furthermore, the classical Denjoy theorem of real analysis tells us that for any measurable function on a measurable set the collection of *Lebesgue regular points* (or points of approximate continuity) is of *full measure*. Thus we may suppose without loss of generality that the whole E is such a set for the function $\triangle_v \mathcal{H}(t)$ in (38).

Having this in mind, pick any $\tau \in E$ and $\varepsilon > 0$ sufficiently small and then construct the single-needle variation of the optimal control $\bar{u}(\cdot)$ by

$$u(t) := \begin{cases} v(t), & t \in I_\varepsilon := [\tau, \tau + \varepsilon) \cap E, \\ \overline{u}(t), & t \notin I_\varepsilon, \end{cases} \tag{39}$$

where the function $v(t)$ and the set E are taken from (38). Applying now the σ-increment formula (30)) from Proposition 4.2 on the *single*-needle variation (39) of $\overline{u}$ for the function $\sigma(x) = \phi(x)$ defined in (21) and using the inner product form in $L^2(I, X)$, we arrive at

$$\triangle \phi = -\int_\tau^{\tau+\epsilon} \triangle_u \mathcal{H}(t) \mathrm{d}t - \int_\tau^{\tau+\epsilon} \langle z(t), \triangle_u F(\overline{x}, \overline{u})(t) \rangle_X \, \mathrm{d}t + o(\epsilon), \tag{40}$$

where $z(t) := \widehat{p}(t) - P\widetilde{T}^{-1}\phi'(\overline{x})(t)$ is defined along the optimal trajectory $\bar{x}(t)$ and $\widehat{p}(t) = p(t) + q(t)$, where p, q are the solutions of the auxiliary and adjoint systems (24) and (25), respectively. It follows from (H5) that

$$\int_\tau^{\tau+\epsilon} \langle z(t), \triangle_u F(\overline{x}, \overline{u})(t) \rangle_X \, \mathrm{d}t = o(\epsilon).$$

Since $\tau \in E$ is a Lebesgue regular point of $\triangle_v \mathcal{H}$, we have

$$\int_{I_\varepsilon} \triangle_v \mathcal{H}(t) \mathrm{d}t = -\varepsilon \, \triangle_v \mathcal{H}(\tau) + o(\varepsilon).$$

Substituting the last two equalities into (40) gives us the representation

$$\triangle \phi = -\varepsilon \, \triangle_v \mathcal{H}(\tau) + o(\varepsilon), \quad \varepsilon > 0. \tag{41}$$

Then using (38) with small $\varepsilon > 0$, we conclude from (41) that $\triangle \phi < 0$, which contradicts (37) due to $\langle \Phi'(\bar{x}), \Lambda_{\tau,v}\overline{x} \rangle \alpha \in S$ and $\langle \lambda, y \rangle \geq 0$ for all $y \in S$ by the constructions above. □

7 Proof in the Case of Equality Constraints

In this section we give the proof of Theorem 5.1 in the case of only the equality constraints assuming for simplicity that the first m constraints in (P) are of the equality type

$$\varphi_i(x) = 0 \text{ for } i = 1, \dots, m, \tag{42}$$

where the functions φ_i in (42) satisfy the assumptions in (H4), which add more to (H3) and are needed to apply below the Brouwer fixed-point theorem. Clearly the proof of Theorem 5.1 as formulated is just the unification of the proofs given in Sects. 6 and 7.

We keep the notation S for the set of linearized images (35) and define the vector function

$$\Psi(x) := \big(\varphi_1(x), \dots, \varphi_m(x)\big), \quad x \in H,$$

which is continuous by (H4). The following topological lemma plays a crucial role in the proof of Theorem 5.1 in the case of the equality constraints (42).

Lemma 7.1 (Projection of linearized images) *We have*

$$0 \notin \operatorname{int}(\Omega) \textit{ for } \Omega := \operatorname{proj}_{\mathbb{R}^m} S.$$

Proof Assuming the contrary gives us $0 \in B_\eta \subset \operatorname{int}(\Omega)$ for the closed ball B_η with a sufficiently small radius η. Let $\mathfrak{T}$ be a regular m-hedron inscribed in B_η with vertices $q^{(1)}, \dots, q^{(m+1)} \in B_\eta$. Then there are parameters $\alpha^{(1)}, \alpha^{(2)}, \dots, \alpha^{(m+1)} \in [0,1]^K$ of multineedle variations (33) and a number $\nu < 0$ such that

$$\left\langle \varphi_0'(\overline{x}), \Lambda_{\tau,\mathrm{v}}\overline{x}\right\rangle \alpha^{(s)} \le \nu \text{ and } q^{(s)} = \left\langle \Psi'(\overline{x}), \Lambda_{\tau,\mathrm{v}}\overline{x}\right\rangle \alpha^{(s)}, \; s = 1, \dots, m+1. \tag{43}$$

By refining the partitions if necessary, suppose that they all correspond to the same parameter (τ, v). Forming the matrices

$$Q = \left[q^{(1)}\, q^{(2)} \dots q^{(m+1)}\right] \text{ and } A = \left[\alpha^{(1)}\, \alpha^{(2)} \dots \alpha^{(m+1)}\right],$$

we arrive at the following convenient representation:

$$Q = \left\langle \Psi'(\overline{x}), \Lambda_{\tau,\mathrm{v}}\overline{x}\right\rangle A.$$

Let $\mathcal{P}$ stand for the standard m-simplex. Then any $q \in \mathfrak{T}$ can be written as $q = Q\gamma$ for some $\gamma \in \mathcal{P}$. Fix now $\gamma \in \mathcal{P}$, let $u_{\gamma,\varepsilon}$ be a multineedle variation with parameters $(\tau, \mathrm{v}, A\gamma, \varepsilon)$, and denote by $x_{\gamma,\varepsilon}$ the solution of (2) corresponding to $u_{\gamma,\varepsilon}$. Then we define the vector function $g\colon \mathcal{P} \times [0, \varepsilon_0] \to \mathbb{R}^m$ by

$$g(\gamma, \varepsilon) := \begin{cases} \dfrac{1}{\varepsilon}\Big(\Psi\left(x_{\gamma,\varepsilon}\right) - \Psi(\overline{x})\Big) & \text{for } \varepsilon > 0, \\ \left\langle \Psi'(\overline{x}), \Lambda_{\tau,\mathrm{v}}\overline{x}\right\rangle A\gamma & \text{for } \varepsilon = 0, \end{cases}$$

where $\Psi(\bar{x}) = 0$ since $\bar{x}$ satisfies the equality constraints (42). By (H4) the mapping g is continuous on $\mathcal{P} \times [0, \varepsilon_0]$ when $\varepsilon_0 > 0$ is sufficiently small. We have furthermore that

$$g(\gamma, 0) = \left\langle \Psi'(\overline{x}), \Lambda_{\tau,\mathrm{v}}\overline{x} \right\rangle A\gamma = Q\gamma,$$

and so $g(\mathcal{P}, 0) = Q\mathcal{P} = \mathfrak{T}$. Thus the mapping $g(\cdot, 0)\colon \mathcal{P} \to \mathfrak{T}$ is continuous and one-to-one.

Denote by $p(\cdot)$ the inverse to $g(\cdot, 0)$ and define the mapping $h\colon \mathcal{T} \times [0, \varepsilon_0] \to \mathbb{R}^m$ by

$$h(q, \varepsilon) := g\big(p(q), \varepsilon\big) \ \text{ for } \ q \in \mathcal{T}, \ \varepsilon \in [0, \varepsilon_0].$$

Choosing $\delta > 0$ so small that $B_\delta \subset \mathfrak{T}$, we find $\widehat{\varepsilon} > 0$ such that

$$\|h(q, \varepsilon) - h(q, 0)\| < \delta \ \text{ for all } \ \varepsilon < \widehat{\varepsilon}.$$

This means that $h(\cdot, 0) - h(\cdot, \varepsilon) : B_\delta \to B_\delta$, i.e., this continuous mapping transforms the closed ball in finite dimensions into itself. Therefore, the fundamental Brouwer theorem ensures that the mapping $h(\cdot, 0) - h(\cdot, \varepsilon)$ has a *fixed point* $q^\varepsilon \in B_\delta$, i.e., we get

$$q^\varepsilon = h\left(q^\varepsilon, 0\right) - h\left(q^\varepsilon, \varepsilon\right) = q^\varepsilon - h\left(q^\varepsilon, \varepsilon\right),$$

which can be rewritten as follows: there is $\gamma^\varepsilon \in \mathcal{P}$ such that

$$h\left(q^\varepsilon, \varepsilon\right) = 0 = g\big(p\left(q^\varepsilon\right), \varepsilon\big) = g\left(\gamma^\varepsilon, \varepsilon\right).$$

This implies that the trajectories $x_{\gamma^\varepsilon,\varepsilon}$ corresponding to the controls $u_{\gamma^\varepsilon,\varepsilon}$ in (2) satisfy the equality constraints (42) for all $\varepsilon \in (0, \varepsilon_0)$. Employing now the needle variations (43) along the cost functional (1) we clearly arrive at

$$\left\langle \varphi_0'(\overline{x}), \Lambda_{\tau,\mathrm{v}}\overline{x} \right\rangle A\gamma \le \nu < 0,$$

and so conclude that $\varphi_0\left(x_{\gamma^\varepsilon,\varepsilon}\right) < \varphi_0(\overline{x})$ for all $\varepsilon \in (0, \widehat{\varepsilon})$. This contradicts the optimality of $(\overline{x}, \overline{u})$ in (P) and thus completes the proof of the lemma. □

Now we are prepared to proceed with the final piece of the proof and justify Theorem 5.1 for the optimal control problem (P) with the equality state constraints (42).

Proof of Theorem 5.1. for the control problem with equality constraints The *forbidden set* in this case is described by

$$\mathcal{F} := \left\{ y \in \mathbb{R}^{m+1} \,\middle|\, y_0 < 0, \ y_i = 0 \ \text{ for } \ i = 1, \ldots, m \right\}.$$

Then we have the following alternatives:

(a) either $S \cap \mathcal{F} = \emptyset$,
(b) or $S \cap \mathcal{F} \neq \emptyset$.

The set $\mathcal{F}$ is obviously convex. If (a) holds, then we can separate S and $\mathcal{F}$ by the hyperplane in $\mathbb{R}^{m+1}$ generated by the vector $0 \neq \lambda \in \mathbb{R}^{m+1}$, i.e., $\langle \lambda, y \rangle \geq \langle \lambda, z \rangle$ for all $y \in S$ and $z \in \mathcal{F}$. It easily follows from the structure of $\mathcal{F}$ that $\lambda_0 \geq 0$ and also that $\langle \lambda, y \rangle \geq 0$ for all $y \in S$ as a consequence of the separation. Consider now the Lagrangian ϕ as in (21) constructed with the multipliers $\lambda = (\lambda_0, \ldots, \lambda_m)$ only for the cost function φ_0 and the equality constraints φ_i from (42). Then proceeding similarly to the proof of Theorem 5.1 in the case of the inequality constraints in Sect. 6, we find the adjoint trajectories $p(t) \in D$ and $q(t) \in \widetilde{D}$ satisfying (24) and (25) and such that the maximum condition (26) holds along $\widehat{p}(t) = p(t) + q(t)$. This verifies Theorem 5.1 for the equality constraints in case (a).

It remains to justify the theorem in case (b). In this case we have by Lemma 7.1 that $0 \in \text{bd}\,(\Omega)$. Hence there is a hyperplane generated by $0 \neq c \in \mathbb{R}^m$, which supports Ω at 0. Defining finally the vector $\lambda := (0, c) \in \mathbb{R}^{m+1}$ gives us $\langle \lambda, y \rangle \geq 0$ for all $y \in S$. Now we can proceed similarly to the derivation above and thus complete the proof of the theorem. □

Remark 7.1 (*Explicit dependence of F on pointwise control values*) Consider the case when the operator $F(x, u)$ explicitly depends on the control values $u(t) \in U$; in particular, via the representation $F(x, u)(t) = f(x(t), u(t))$ discussed in Sect. 5 and surely encompassed the ODE setting. In this case the maximum condition (26) can be understood in the standard sense as the maximization of the Hamilton–Pontryagin function $\mathcal{H}(\overline{x}(t), \phi'(\overline{x})(t), \widehat{p}(t), u)$ over the control value set U for a.e. $t \in I$. To proceed in more detail, we need to use appropriate facts from measurable set-valued mappings assuming that U is a *Souslin* subset (i.e., a continuous image of a Borel set) of a Banach spaces. Arguing by contradiction, suppose that the maximum condition (26) does not hold along the optimal pair $(\bar{u}, \bar{x})$ and then find a measurable set $E \subset I$ with $\text{mes}(E) \neq 0$ such that $V(t) \neq \emptyset$ for all $t \in E$, where the set-valued mapping $V\colon E \rightrightarrows U$ is defined by

$$V(t) := \left\{ u \in U \,\middle|\, \mathcal{H}(\overline{x}(t), \phi'(\overline{x})(t), \widehat{p}(t), u) \right\} < \sup_{v \in U} \left\{ \mathcal{H}(\overline{x}(t), \phi'(\overline{x})(t), \widehat{p}(t), v) \right\}.$$

It follows from the theory of measurable set-valued mappings with taking into account that U is Souslin (see, e.g., [10, Chapter 6] and the references therein) that the mapping $V(\cdot)$ is measurable and admits a measurable single-valued *selection* $v(t) \in V(t)$ on E. Then we get (38) and continue as in the proof of Theorem 5.1 given above.

8 Conclusions

This paper demonstrates the possibility of deriving necessary optimality conditions of the Maximum Principle type for a general class of optimal control problems governed by nonlinear self-adjoint operator equations with state constraints in infinite-dimensional spaces. Among important and challenging problems of the future research, we mention the following:

(i) To clarify the validity of the *technical assumption* (H5) for particular classes of preudodifferential operator equations as well as functional differential, partial differential, integral and other types of equations besides those mentioned above.
(ii) To explore the possibility of extending the results obtained here to the case of *nondifferentiable* initial data in optimal control problems.
(iii) To replace the one-dimensional interval $I \subset \mathbb{R}$ in the optimal control problem (P) formulated in Sect. 5 by a *multidimensional* region $\Omega \subset \mathbb{R}^n$. It seems that we can proceed in the later case with constructing "needle-type" variations of controls and thus make it possible to incorporate the general results of nonlinear operator theory discussed in Sects. 1–4 to the scheme of deriving necessary optimality conditions developed in Sects. 5–7.

Acknowledgments The authors thank the King Fahd University of Petroleum and Minerals for excellent facilities provided to support scientific research. We are indebted to two anonymous referees for their careful reading the original manuscript and providing many suggestions and remarks, which allowed us to significantly improve the original presentation.

References

1. Polak, E.: Optimization: Algorithms and Consistent Approximations. Springer, New York (1997)
2. Pontryagin, L.C., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: The Mathematical Theory of Optimal Processes. Wiley, New York (1962)
3. Neimark, M.A.: Linear Differential Operators: Part II. Frederick Ungar, New York (1968)
4. Warga, J.: Optimal Control of Differential and Functional Equations. Academic Press, New York (1972)
5. Brézis, H.: Operateurs Maximauz Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert. North-Holland, Amsterdam (1973)
6. Lasiecka, I., Triggiani, R.: Control Theory for Partial Differential Equations. Cambridge University Press, Cambridge (2000)
7. Li, X., Yong, J.: Optimal Control Theory for Infinite-Dimensional Systems. Birkhäuser, Boston (1995)
8. Shubin, M.A.: Preudodifferential Operators and Spectral Theory, 2nd edn. Springer, New York (2001)
9. Giannessi, F.: Constrained Optimization and Image Space Analysis. Springer, Berlin (2005)
10. Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation, II: Applications. Springer, Berlin (2006)
11. El-Gebeily, M.A., Mordukhovich, B.S., Alshahrani, M.M.: Optimal control of self-adjoint nonlinear operator equations in Hilbert spaces. Appl. Anal. **93**, 210–222 (2014)
12. Alshahrani, M.M., El-Gebeily, M.A., Mordukhovich, B.S.: Maximum principle in optimal control of singular ordinary differential equations in Hilbert spaces. Dyn. Syst. Appl. **21**, 219–234 (2012)
13. McShane, E.J.: Necessary conditions for generalized curve problems in the calculus of variations. Duke Math. J. **7**, 1–27 (1940)