ESAIM: COCV 30 (2024) 87
https://doi.org/10.1051/cocv/2024076

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

THE MAXIMUM PRINCIPLE FOR LUMPED-DISTRIBUTED CONTROL SYSTEMS

E. M. Marchini[1,*] and R. B. Vinter[2]

Abstract. This paper concerns the optimal control of lumped-distributed systems, that is control systems comprising interacting infinite and finite dimensional subsystems. An examplar lumped-distributed system is an assembly of rotating components connected by flexible rods. The underlying mathematical model is a controlled semilinear evolution equation, in which nonlinear terms involve a projection of the full state onto a finite dimensional subspace. We derive necessary conditions of optimality in the form of a maximum principle, for a problem formulation which involves pathwise and end-point constraints on the lumped components of the state variable. A key feature of these necessary conditions is that they are expressed in terms of a costate variable taking values in a finite dimensional subspace (the subspace of the state space associated with the lumped variables). By contrast, costate trajectories in earlier-derived necessary conditions for optimal control of evolution equations evolve in the full (infinite dimensional) state space. The computational implications of the reduction techniques introduced in this paper to prove the maximum principle, which permit us to replace the original optimal control problem by one involving a reduced, finite dimensional, state space, will be explored in future work.

Mathematics Subject Classification. 34G25, 35Q93, 46N10, 47J22, 47N70, 49J20, 93C20.

Received March 4, 2024. Accepted September 19, 2024.

1. Introduction

Lumped-distributed control systems are assemblages of interconnected subsystems, some of which have a finite dimensional state space while others infinite dimensional state spaces. Figure 1 illustrates one such control system, which comprises a left and a right inertial mass, with moments of inertia J_0 and J_1 about the axis of rotation, respectively, that are connected by a rod with rotational flexibility, of uniform circular cross-section, modulus of rigidity G and mass per unit length ρ. The control is the exogenous torque applied to the left inertial mass. If the control signal is the field voltage of a DC motor, then the left inertial mass is that of the rotor. The right inertial mass is that of the load. The lumped components are the two inertial masses, for which the states are the angular displacements and velocities. The distributed component is the rod, whose state is the angular strain along the rod and its time rate of change. A related optimal control problem will involve, say, controlling the torque to minimize a weighted sum involving the displacement error of the initial mass, w.r.t. a new set

Keywords and phrases: Optimal control, infinite dimensional systems, state constraints, maximum principle.

[1] Dipartimento di Matematica, Politecnico di Milano Piazza Leonardo da Vinci 32, 20133 Milano, Italy.
[2] Department of Electrical and Electronic Engineering, Imperial College Exhibition Road, London SW7 2AZ, UK.

[*] Corresponding author: elsa.marchini@polimi.it

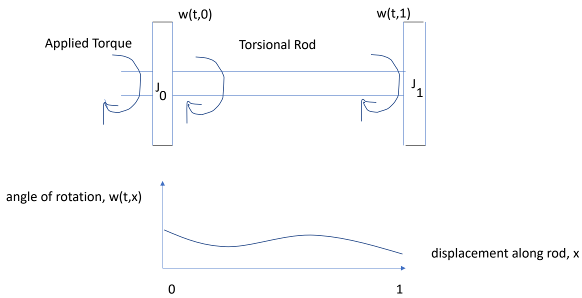

FIGURE 1. Lumped/Distributed Control System.

point, and the energy consumption associated with the exogenous torque. The path-wise state constraint on the inertial masses may correspond to bounds on speed, for safety or validity of the dynamic model.

Take as state variables

$$\begin{aligned} v(y) &: \text{angular velocity at point } y \text{ along the rod } (0 \le y \le 1), \\ s(y) &: \text{torsional strain at point } y \text{ along the rod } (0 \le y \le 1), \\ \theta &: \text{angular velocity of left inertial mass}, \\ \psi &: \text{angular velocity of right inertial mass.} \end{aligned}$$

Define the constants $c^2 := G/\rho$, $d_0 := J_0^{-1}$ and $d_1 := J_1^{-1}$.

The control system governing the time evolution of these state variables (now written $v(t,y)$, $s(t,y)$, $\theta(t)$ and $\psi(t)$) over the time period $[0,T]$, under the action of the control $u(t)$ is taken to be

$$\left.\begin{aligned} \tfrac{\partial v}{\partial y}(t,y) &= c^2 \tfrac{\partial s}{\partial y}(t,y) \\ \tfrac{\partial s}{\partial y}(t,y) &= \tfrac{\partial v}{\partial y}(t,y) \end{aligned}\right\} \quad \text{distributed component,}$$

$$\left.\begin{aligned} \tfrac{d\theta}{dt}(t) &= d_0 s(0,t) + u(t) \\ \tfrac{d\psi}{dt}(t) &= d_1 s(1,t) \end{aligned}\right\} \quad \text{lumped component.}$$

We shall interpret the evolving state of the system $x = (v, s, \phi, \psi)$, governed by these equations, as the mild solution in the Hilbert space $X = L^2(0,1) \times L^2(0,1) \times \mathbb{R} \times \mathbb{R}$ of the controlled evolution equation

$$\dot{x}(t) = \mathcal{A}x(t) + M \circ f(\Lambda \circ x(t), u(t)), \tag{1.1}$$

in which the infinitesimal generator $\mathcal{A}$ is the linear mapping

$$\mathcal{A} \begin{bmatrix} v \\ s \\ \theta \\ \psi \end{bmatrix} = \begin{bmatrix} c^2 \frac{\mathrm{d}s}{\mathrm{d}y}(y) \\ \frac{\mathrm{d}v}{\mathrm{d}y}(y) \\ d_0 s(0) \\ -d_1 s(1) \end{bmatrix}, \tag{1.2}$$

with domain

$$\mathcal{D}(\mathcal{A}) = \big\{ (v, s, \theta, \psi) \in W^{1,2}(0,1) \times W^{1,2}(0,1) \times \mathbb{R} \times \mathbb{R} \,:\, v(0) = \theta \text{ and } v(1) = \psi \big\}. \tag{1.3}$$

The function $f : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$, that takes account of the control input and also of the 'third power' nonlinear damping on the right inertial mass, is

$$f(z_3, z_4, u) = (f_1(u), f_2(z_4)) := (u, -(z_4)^3).$$

The linear mappings $M : \mathbb{R} \times \mathbb{R} \to L^2 \times L^2 \times \mathbb{R} \times \mathbb{R}$ and $\Lambda : L^2 \times L^2 \times \mathbb{R} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}$ are the emersion and projection

$$M(f_1, f_2) := (0, 0, f_1, f_2) \quad \text{and} \quad \Lambda(z_1, \ldots, z_4) := (z_3, z_4).$$

We are precise about the terms 'evolution equation', 'infinitesimal generator', *etc.*, in the next section.

The distinctive feature of this formulation is that the inhomogeneous term '$f(\Lambda \circ z, u)$' in the state equation is affected only by the projection of the state onto a finite dimensional subspace of the state space. If pathwise and end-point constraints are imposed only on the inertial masses and we consider an end-point cost that concerns the behavior only of the inertial masses, then the cost and constraint functions, will be expressed, not in terms of the full state, but in terms, once again, of its projection on a finite dimensional subspace of the state space.

Descriptions of this nature arise in mechanical energy transmission systems, such as wind generators, automobile and aeronautical engineering. They are encountered in communication systems where a transmission line has an active load, and in thermal systems where a distributed thermal channel interacts with heat sinks and sources. Such descriptions are to be found also in the optimal control of hereditary systems, when the dynamic constraint is reformulated as a 'delay-free' evolution equation with an infinite dimensional state space.

Zabczyk [1] has identified an important class of lumped distributed control systems that conform to the framework of this paper. For such systems, boundary controls to a distributed system are applied, not directly, but as the output of a lumped system. (In concrete terms, the overall model includes the dynamics of the control actuators.)

This paper provides necessary conditions of optimality for optimal control problems involving lumped-distributed control systems, having the special structure (1.1). Our analysis, which is based on perturbational techniques due to Clarke, allows for end-point state constraints, pathwise state constraints and terminal costs that depend only on state variable components associated with the lumped components of the overall system.

A distinctive feature of these new necessary conditions is that they are 'reduced', in the sense that they are expressed in terms of a costate trajectory taking values in a finite dimensional space (the state space of the lumped component of the control system). Earlier derived necessary conditions of optimality, for problems in which the dynamic constraint is a semi-linear evolution equation, are expressed in terms of costate variable that evolves in the whole, infinite dimensional, state space (See Fattorini's monograph [2] (see also references therein) including [3] and [4]). The fact that necessary conditions takes this simpler, 'finite dimensional' form, for evolution equations describing lumped linear systems, has not previously been revealed.

In the linear case, the lumped-continuous control system (1.1), with the projected output variable $y = \Lambda(x)$, gives rise to a matrix transfer function, relating input to output, whose entries are irrational functions. Frequency domain techniques have been employed [5], [6] to design controllers for such systems and to solve optimal control problems with a quadratic cost. The reformulation of the optimal control problem as one involving a finite dimensional projection of the state can be regarded as a time domain analogue, which allows nonlinearity and non-quadratic costs, of the irrational transfer function descriptions of the frequency domain literature.

We mention that the reduction procedure, that is central to this paper, can be exploited to construct efficient computational schemes for solving the optimal control problem based, say, on Polak and Mayne's strong variational algorithms [7]. These potential applications will be the subject of future investigation.

Notation: The Euclidean norm of a vector $x \in \mathbb{R}^n$ is written $|x|$. $\mathbb{B}$ indicates the closed unit ball in $\mathbb{R}^n$. The distance function associated with a given set $A \subset \mathbb{R}^n$, is

$$d_A(x) \; := \; \inf\{|x - y| \, : \, y \in A\}, \quad \text{for } x \in \mathbb{R}^n.$$

The convex hull of the set A is written $\text{co}\,A$. Given a multifunction $F(.):\mathbb{R}^n \rightsquigarrow \mathbb{R}^\kappa$, we denote by $\text{Gr}\,F(.)$ the graph of $F(.)$.

$L^1([a,b];\mathbb{R}^n)$ and $L^\infty([a,b];\mathbb{R}^n)$ (abbreviated as L^1 and L^∞) have their usual meanings as spaces of integrable and essentially bounded, measurable n-vector valued functions on $[a,b]$. $\mathcal{C}([a,b];\mathbb{R}^n)$ (abbreviated as $\mathcal{C}$) denotes the space of continuous functions $x:[a,b]\to\mathbb{R}^n$, while $W^{1,1}([a,b];\mathbb{R}^n)$ denotes the space of absolutely continuous functions $x:[a,b]\to\mathbb{R}^n$.

$NBV^+[a,b]$ denotes the space of nondecreasing, real-valued functions μ on $[a,b]$ (which have bounded variation), vanishing at the point a and right continuous on (a,b). The total variation of a function $\mu\in NBV^+[a,b]$ is written $\|\mu\|_{T.V.}$. As is well known, each element $\mu\in NBV^+[a,b]$ defines a Borel measure on $[a,b]$. This associated measure is also denoted μ. The support of the measure μ is written $\text{supp}\{\mu\}$.

We make use of some constructs from nonsmooth analysis, described in detail, for example, in [8, 9]: given a closed set $E\subset\mathbb{R}^n$ and $x\in E$, the *proximal normal cone to E at x* is

$$N_E^P(x) := \{\zeta\in\mathbb{R}^n : \exists\,\epsilon>0 \textit{ and } M>0 \textit{ s.t. } \zeta\cdot(y-x)\le M|x-y|^2 \textit{ for all } y\in E\cap x+\epsilon\mathbb{B}\}.$$

The *limiting normal cone to E at x* is

$$N_E(x) := \big\{\lim_{i\to\infty}\zeta_i : \zeta_i\in N_E^P(x_i) \textit{ and } x_i\in E \textit{ for all } i, \textit{and } x_i\to x\big\}.$$

Given a lower semicontinuous function $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ and a point $x\in\mathbb{R}^n$ such that $f(x)<\infty$, the *proximal subdifferential of f at x* is the set

$$\partial^P f(x) := \left\{ \begin{array}{ll} \zeta\in\mathbb{R}^n : & \exists\,\sigma>0 \text{ and } \epsilon>0 \text{ such that } \forall\, y\in x+\epsilon\mathbb{B} \\ & f(y)-f(x)\ge\ \zeta\cdot(y-x)-\sigma|y-x|^2 \end{array} \right\}.$$

The *limiting subdifferential of f at x* is

$$\partial f(x) := \big\{\lim_{i\to\infty}\zeta_i : \zeta_i\in\partial^P f(x_i), x_i\to x, f(x_i)\to f(x)\big\}.$$

2. Problem formulation

We consider the following optimal control problem, relating to the class of lumped-distributed control systems discussed in the introduction, in which $I:=[0,T]$:

$$(P)\left\{ \begin{array}{l} \text{Minimize } g(\Lambda\circ x(T)) + \int_I L(t,\Lambda\circ x(t),u(t))\mathrm{d}t \\[2ex] \text{over functions } x\in\mathcal{C}(I;X) \text{ and measurable functions } u:I\to\mathbb{R}^m \text{ s.t.} \\[2ex] \dot{x}(t) = \mathcal{A}x(t) + M\circ f(t,\Lambda\circ x(t),u(t)), \quad \text{on } I, \\[2ex] u(t)\in U, \quad \text{for a.e. } t\in I, \\[2ex] h(\Lambda\circ x(t))\le 0, \quad \text{for all } t\in I, \\[2ex] \Lambda\circ x(T)\in C, \\[2ex] x(0)=x_0, \end{array} \right.$$

the data for which comprise: $T > 0$, a C_0 semigroup $\{S(t) \,:\, t \geq 0\}$ on the real Hilbert space X, with infinitesimal generator $\mathcal{A}$, bounded linear mappings $M : \mathbb{R}^\kappa \to X$ and $\Lambda : X \to \mathbb{R}^n$, functions $g : \mathbb{R}^n \to \mathbb{R}$, $f : I \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^\kappa$, $L : I \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}$, a Borel set $U \subset \mathbb{R}^m$, a closed set $C \subset \mathbb{R}^n$ and a point $x_0 \in X$.

The dynamic relation

$$\dot{x}(t) = \mathcal{A}x(t) + M \circ f(t, \Lambda \circ x(t), u(t)), \qquad \text{a.e. } t \in I, \tag{2.1}$$

in (P) is referred to as a controlled evolution equation.

A pair of functions (x, u) is said to be a process for (P), if u is a measurable function such that $u(t) \in U$ a.e. and x is a mild solution of (2.1) (for the specified initial state x_0), in the sense that

$$x(t) = S(t)x_0 + \int_0^t S(t-s)M \circ f(s, \Lambda \circ x(s), u(s))\mathrm{d}s, \qquad \text{for all } t \in I.$$

The process is said to be an admissible process if, also, it satisfies the end-point and pathwise state constraints, $\Lambda \circ x(T) \in C$, and $h(\Lambda \circ x(t)) \leq 0$, for all $t \in I$. The first and second components of (admissible) processes are called (admissible) state trajectories and (admissible) controls respectively.

We say that an admissible process $(\bar{x}, \bar{u})$ is a minimizer for (P) if

$$g(\Lambda \circ \bar{x}(T)) \leq g(\Lambda \circ x(T)) \tag{2.2}$$

for all admissible processes (x, u). We will say that $(\bar{x}, \bar{u})$ is a strong local Λ-minimizer if there exists $\rho > 0$ such that (2.2) is true for all admissible processes (x, u) such that $||\Lambda \circ x - \Lambda \circ \bar{x}||_{\mathcal{C}} \leq \rho$.

3. Necessary conditions

We shall invoke the following hypotheses on the data for problem (P), in which $(\bar{x}, \bar{u})$ is a given process and $\bar{z}(t) := \Lambda \circ \bar{x}(t)$, for each $t \in I$. For some $\epsilon > 0$,

(H1): $(f, L)(., z, .)$ is Lebesgue–Borel measurable for all $z \in \mathbb{R}^n$ and U is a Borel set,

(H2): for each $(t, u) \in I \times U$, $(f, L)(t, ., u)$ is continuously differentiable and there exists $k_f \in L^1(I; \mathbb{R}^+)$ such that,

$$|(f, L)(t, z, u) - (f, L)(t, z', u)| \leq k_f(t)|z - z'|, \text{ for all } z, z' \in \bar{z}(t) + \epsilon\mathbb{B} \text{ and } u \in U, \text{ a.e. } t \in I,$$

(H3): there exists $c_f \geq 0$ such that,

$$|(f, L)(t, z, u)| \leq c_f, \text{ for all } z \in \bar{z}(t) + \epsilon\mathbb{B} \text{ and } u \in U, \text{ a.e. } t \in I,$$

(H4): h is continuously differentiable and there exists $k_h > 0$ such that,

$$|h(z) - h(z')| \leq k_h|z - z'|, \text{ for all } z, z' \in \bar{z}(t) + \epsilon\mathbb{B} \text{ and } t \in I,$$

(H5): g is a continuously differentiable on $(\bar{z}(0), \bar{z}(T)) + \epsilon(\mathbb{B} \times \mathbb{B})$ with Lipschitz constant denoted k_g.

The following theorem provides necessary conditions of optimality, in the form of a maximum principle, for state constrained optimal control problems involving a semilinear evolution equation:

Theorem 3.1. *Let $(\bar{x}, \bar{u})$ be a strong local Λ-minimizer for problem (P). Assume (H1)–(H5).*

Then there exists $\lambda \geq 0$, a continuous function $p : I \to \mathbb{R}^\kappa$ and $\mu \in NBV^+(I)$ satisfying the following conditions, in which

$$q(t) := p(t) + \int_{[t,T]} M^* S^*(s-t)\Lambda^* \nabla h(\Lambda \circ \bar{x}(s))\mu(\mathrm{d}\sigma), \; t \in I:$$

(a): (multiplier non-triviality)

$$(\lambda, p, \mu) \neq (0, 0, 0),$$

(b): (costate equation)

$$p(t) = M^*S^*(T-t)\Lambda^* p(T) + \int_t^T M^*S^*(s-t)\Lambda^*\, \nabla_z\,(q\cdot f - \lambda L)(s,\Lambda\circ\bar{x}(s),\bar{u}(s))\mathrm{d}s, \quad a.e.\ s\in I,$$

(c): (transversality condition)

$$-q(T) = \nabla g(\Lambda\circ\bar{x}(T)) + \xi,\ \text{for some } \xi \in N_C(\Lambda\circ\bar{x}(T)),$$

(d): (Weierstrass condition)

$$\Big(q\cdot f - \lambda L\Big)(t,\Lambda\circ\bar{x}(t),\bar{u}(t)) \;=\; \sup_{u\in U}\Big(q\cdot f-\lambda L\Big)(t,\Lambda\circ\bar{x}(t),u),\ \ a.e.\ t\in I,$$

(e): (complementary slackness)

$$supp\,\{\mu\}\ \subset\ \{t : h(\Lambda\circ\bar{x}(t))=0\}.$$

(In these conditions $S^*(t)$, Λ^* and M^* are the adjoint mappings of $S(t)$, Λ, M respectively. We interpret p and q as row-vector valued functions, gradients (e.g. $\nabla g(z)$) as row vectors and points in the domain of Λ^* and in the range space of M^* as row vectors.)

Comment

For optimal control problems with infinite dimensional state space X, first order necessary conditions are customarily expressed in terms of a costate trajectory that takes values in the (infinite dimension) dual space X^* which, since X is a Hilbert space, can be identified with X. A notable feature of the necessary conditions above is that, owing to the special structure of problem (P), they are expressed in terms of a *reduced* co-state trajectory that takes value in the finite dimensional linear space $\mathbb{R}^\kappa$. This is an important simplification.

In variational analysis, necessary conditions such as those provided by Theorem 3.1 have special significance when, under the stated hypotheses, the dynamic optimization problem (P) is known to have a minimizer. In this case, the search for minimizers can be narrowed to a search over extremals for (P). We can show that (P) does indeed have a minimizer when the hypotheses (H2)–(H5) are satisfied with parameter $\epsilon=\infty$ (and for any bounded function $\bar{z}: I\to\mathbb{R}^n$), when

- there exists an admissible process

and

- the set $\{(f(t,z,u), L(t,z,u)+\alpha\ :\ u\in U \text{ and } \alpha\geq 0\}$ is convex for all $(t,z)\in I\times\mathbb{R}^n$.

The proof, which is omitted, is based on a standard limit taking procedure for minimizing sequences of processes, in which we make the use of the representation of processes provided by Lemma 4.1 and Proposition 4.3 (applied to a state-augmented control system) to establish the existence of a limiting process.

4. Preliminary analysis

The following lemma provides a representation of state trajectories in terms of functions taking values in a finite dimensional space ('reduced state trajectories').

Lemma 4.1. *Suppose that (H1)–(H3) are satisfied (with $\epsilon=+\infty$). Take any measurable function $u: I\to\mathbb{R}^m$ such that $u(t)\in U$, a.e. Then:*

(i): There exists a unique continuous function $z: I\to\mathbb{R}^n$ such that

$$z(t) = \Lambda\circ S(t)x_0 + \int_0^t \Lambda\circ S(t-s)M\circ f(s,z(s),u(s))\mathrm{d}s,\ \ \text{for all } t\in I. \tag{4.1}$$

Furthermore, there exists $R_0>0$ (independent of u) such that $\|z\|_{\mathcal{C}}\leq R_0$.

(ii): equation (2.1) has a unique mild solution x in the class of continuous functions (for the specified control function u) and $z(t) = \Lambda \circ x(t)$, for each $t \in I$, where z is the unique continuous solution to (4.1).

Proof. (i): Take an integer $N > 1$ and a measurable mapping $u : I \to U$. Define the mappings $z \to G_i(z)$, $i = 1, \dots N$, with domain the space $\mathcal{C}([(i-1)(T/N), i(T/N)]; \mathbb{R}^n)$:

$$G_i(z) := \Lambda \circ S(t)x_0 + \int_{(i-1)(T/N)}^{i(T/N)} \Lambda \circ S(t-s)M \circ f(s, z(s), u(s))\mathrm{d}s \text{ for } i = 1, \dots, N.$$

It can be deduced from the dominated convergence theorem and the fact that S is a strongly continuous semigroup that each G_i maps into $\mathcal{C}([(i-1)(T/N), i(T/N)]; \mathbb{R}^n)$. It follows from hypothesis (H2) (with $\epsilon = \infty$) that, for N sufficiently large, each G_i is a contraction. Hence, for each i, there exists a continuous function z_i such that $G_i(z_i) = z_i$. Now let $z : I \to \mathbb{R}^n$ be the concatenation of the functions $z_i : [(i-1)(T/N), i(T/N)] \to \mathbb{R}^n$. z satisfies (4.1).

Both the uniqueness of this solution of (4.1), and also the existence of $R_0 > 0$ (independent of u) such that $\|z\|_{\mathcal{C}} \leq R_0$ follows from the fact that $S(t)$ is bounded in the operator norm uniformly over $t \in [0, T]$, and Gronwall's Lemma.

(ii): We know there exists a continuous solution z to (4.1). Define the continuous function x:

$$x(t) = S(t)x_0 + \int_0^t S(t-s)M \circ f(s, z(s), u(s))\mathrm{d}s, \text{ for all } t \in I.$$

But then, for all $t \in I$,

$$\Lambda \circ x(t) = \Lambda \circ S(t)x_0 + \int_0^t \Lambda \circ S(t-s)M \circ f(s, z(s), u(s))\mathrm{d}s = z(t).$$

It follows that $x(t)$ satisfies

$$x(t) = S(t)x_0 + \int_0^t S(t-s)M \circ f(s, \Lambda \circ x(s), u(s))\mathrm{d}s.$$

We have constructed a mild solution x to (2.1) such that $z(t) = \Lambda \circ x(t)$. The uniqueness of the mild solution follows from Gronwall's lemma. □

This lemma permits us to replace the optimization problem (P) by a 'reduced' optimization problem (R) in which state trajectories x, evolving in the infinite dimensional vector space X are replaced by arcs z ('reduced state trajectories') taking values in the finite dimensional vector space $\mathbb{R}^n$.

$$(R)\begin{cases} \text{Minimize } g(z(T)) + \int_I L(t, z(t), u(t))\mathrm{d}t \\ \text{over functions } z \in \mathcal{C}(I; \mathbb{R}^n) \text{ and measurable functions } u : I \to \mathbb{R}^m \text{ s.t.} \\ z(t) = \Lambda \circ S(t)x_0 + \int_0^t \Lambda \circ S(t-s)M \circ f(s, z(s), u(s))\mathrm{d}s, \quad \text{for all } t \in I, \\ u(t) \in U, \quad \text{for a.e. } t \in I, \\ h(z(t)) \leq 0, \quad \text{for all } t \in I, \\ z(T) \in C, \\ z(0) = \Lambda(x_0). \end{cases}$$

We define, by analogy with the definitions relating to problem (P), processes (comprising pairs (z,u)), reduced state trajectories, control functions, processes for problem (R) and also their admissible versions. We say that an admissible process $(\bar{z},\bar{u})$ is a strong local minimizer for (R) if there exists $\rho\in(0,\infty)\cup\{+\infty\}$ such that $(\bar{z},\bar{u})$ minimizes $g(z(T))$ over all processes such that $\|z-\bar{z}\|_{\mathcal{C}}\le\rho$. (If $\rho=\infty$, we say that $(\bar{z},\bar{u})$ is a minimizer for (R).)

Corollary 4.2. *Suppose that $(\bar{x},\bar{u})$ is a strong local Λ-minimizer for (P). Assume (H1)–(H5). Then, $(\bar{z}:=\Lambda\circ\bar{x},\bar{u})$ is a strong local minimizer for problem (R).*

Proof of Corollary 4.2. Since $(\bar{x},\bar{u})$ is a strong local Λ-minimizer for (P), there exists $\rho>0$ such that $(\bar{x},\bar{u})$ is minimizing w.r.t. admissible processes (x,u) such that $||\Lambda\circ x-\Lambda\circ\bar{x}||_{L^\infty}\le\rho$. We can assume that $\rho\le\epsilon$, where ϵ is as in hypotheses (H2)–(H5). Define the 'truncation' function $\mathrm{tr}_\epsilon:\mathbb{R}^n\to\mathbb{R}^n$:

$$\mathrm{tr}_\epsilon(x):=\begin{cases} x & \text{if } |x|\le\epsilon \\ \epsilon\,\frac{1}{|x|}x & \text{otherwise.}\end{cases}$$

Let $(\tilde{P})$ and $(\tilde{R})$ be the modified versions of (P) and (R) respectively in which g and f are replaced by the functions

$$\tilde{g}(z):=g(\bar{z}(T)+\mathrm{tr}_\epsilon(z-\bar{z}(T)))\ \text{ and }\ (\tilde{f},\tilde{L})(t,z,u):=(f,L)(t,\bar{z}(t)+\mathrm{tr}_\epsilon(z-\bar{z}(t)),u).$$

Let $\bar{z}$ be as defined in the corollary statement. Take any admissible process (z,u) for (R) such that $\|z-\bar{z}\|_{\mathcal{C}}\le\epsilon$. Then $(\bar{z},\bar{u})$ and the data for $(\tilde{R})$ satisfy the hypotheses of (H1)–(H5) (with $\epsilon=\infty$). So, by Lemma 4.1, there exists a process (x,u) for $(\tilde{P})$ such that $z=\Lambda\circ x$. But, because $\|z-\bar{z}\|_{\mathcal{C}}\le\epsilon$, (x,u) is an admissible process also for (P). Then, by strong local Λ-optimality of $(\bar{x},\bar{u})$, $g(z(T))=g(\Lambda\circ x(T))\ge g(\Lambda\circ\bar{x}(T))=g(\bar{z}(T))$. We have shown that $(\bar{z},\bar{u})$ is a strong local minimizer for (R). □

The following 'compactness' proposition will provide information about limits of reduced state and costate trajectories associated with the auxiliary problem (R), under data perturbations:

Proposition 4.3. *Let $\{S(t):X\to X: t\ge0\}$, Λ and M be as in Section 1. Take a multifunction $F:I\times\mathbb{R}^n\rightsquigarrow\mathbb{R}^\kappa$, a sequence of measurable functions $y_i:I\to\mathbb{R}^n$.*

Take also convergent sequences $r_i\to0$ in $L^1(I;\mathbb{R}^n)$ and $\eta_i\to\eta$ strongly in X and a sequence $\{\xi_i\}$ in $L^1(I;\mathbb{R}^\kappa)$. Assume that F has $\mathcal{L}\times\mathcal{B}$ measurable graph, $F(t,.)$ has closed graph for each t and, for some $c_F\in L^1(I;\mathbb{R}^+)$,

$$F(t,y)\subset c_F(t)\mathbb{B}\ \text{for}\ y\in\mathbb{R}^n,\ \text{a.e.}\ t\in I. \tag{4.2}$$

Assume that, for each i,

$$y_i(t)=\Lambda\circ S(t)\eta_i+\int_0^t\Big(\Lambda\circ S(t-s)M\circ\xi_i(s)+r_i(s)\Big)\mathrm{d}s,\ \text{for all}\ t\in I,$$

and

$$\xi_i(t)\in F(t,y_i(t))\ \text{a.e.}\ t\in I. \tag{4.3}$$

Then, along some subsequence (we do not relabel), $y_i(t)\to y(t)$, for any $t\in I$ and some $y\in\mathcal{C}(I;\mathbb{R}^n)$, and there exists $\xi\in L^1(I;\mathbb{R}^\kappa)$ such that

$$y(t)=\Lambda\circ S(t)\eta+\int_0^t\Lambda\circ S(t-s)M\circ\xi(s)\mathrm{d}s,\ \text{for all}\ t\in I$$

and

$$\xi(t) \in co\, F(t, y(t)), \text{ for a.e. } t \in I.$$

If $c_F \in L^\infty(I;\mathbb{R}^+)$, then $y_i \to y$ strongly in $\mathcal{C}(I;\mathbb{R}^n)$.

In the proof of the proposition and in subsequent analysis the constant M_S is taken to be:

$$M_S := \max_{t\in I} \|S(t)\|_{L(X,X)} \tag{4.4}$$

Proof. Since $\{\xi_i\}$ is equi-integrable by (4.2), (4.3), and bounded in $L^1(I;\mathbb{R}^\kappa)$, we have (taking a subsequence and retaining the old notation)

$$\xi_i \rightharpoonup \xi, \qquad \text{weakly in } L^1(I;\mathbb{R}^\kappa),$$

for some $\xi \in L^1(I;\mathbb{R}^\kappa)$. Hence, as $r_i \to 0$ in $L^1(I;\mathbb{R}^n)$,

$$y_i(t) = \Lambda \circ S(t)\eta_i + \int_0^t \Big(\Lambda \circ S(t-s)M \circ \xi_i(s) + r_i(s)\Big)\mathrm{d}s \to$$
$$y(t) = \Lambda \circ S(t)\eta + \int_0^t \Lambda \circ S(t-s)M \circ \xi(s)\mathrm{d}s, \qquad \text{for any } t \in I.$$

Applying a compactness of trajectories theorem, *e.g.* [9], Theorem 2.5.3, we can show that

$$\xi(t) \in \mathrm{co}\, F(t, y(t)).$$

Finally, if $c_F \in L^\infty(I;\mathbb{R}^+)$, then the sequence $\{y_i\}$ is equi-bounded and equi-continuous. Using the fact that, in this case, $\bigcup_i \big\{ \int_0^t S(t-s)M \circ \xi_i(s)\mathrm{d}s\big\}$ is precompact in X, we can show that, for $t < \tau$,

$$\begin{aligned}|y_i(\tau) - y_i(t)| &= \Big| \int_0^\tau \Big(\Lambda \circ S(\tau - s)M \circ \xi_i(s) + r_i(s)\Big)\mathrm{d}s - \int_0^t \Big(\Lambda \circ S(t-s)M \circ \xi_i(s) + r_i(s)\Big)\mathrm{d}s\Big| \\ &\le \int_t^\tau |r_i(s)|\mathrm{d}s + M_S \int_t^\tau c_F(s)\mathrm{d}s + \Big|\Lambda \circ \big(S(\tau - t) - \mathbb{I}\big) \int_0^t S(t-s)M \circ \xi_i(s)\mathrm{d}s\Big|,\end{aligned}$$

where $\mathbb{I}$ is the identity operator. Then, by the Ascoli–Arzelà theorem, $y_i \to y$ strongly in $\mathcal{C}(I;\mathbb{R}^n)$, ending the proof. $\square$

Take a Lipschitz continuous function $d : \mathbb{R}^n \to \mathbb{R}$ with Lipschitz constant k_d. For given $k \in [0,\infty)$ define the quadratic inf-convolution of g:

$$d_k(z) = \inf_{y\in\mathbb{R}^n} \big\{d(y) + k|y - z|^2\big\}. \tag{4.5}$$

The required properties of the quadratic inf-convolution function are assembled in the following lemma (see [8], Sect. 1.5).

Lemma 4.4. *Take $z \in \mathbb{R}^n$ and let $y \in \mathbb{R}^n$ be any vector achieving the infimum in the definition (4.5) of d_k. Let $\eta_k := -2k(y - z)$. Then, for each k,*

(i) d_k *is Lipschitz continuous with Lipschitz constant* k_g*,*
(ii) $d_k(z) \le d(z) \le d_k(z) + \frac{k_d}{k}$*,*
(iii) $d_k(z') - d_k(z) \le \eta_k \cdot (z' - z) + k|z' - z|^2$*, for each* $z' \in \mathbb{R}^n$*,*
(iv) $\eta_k \in \partial^P d(y)$*,*
(v) $|y - z| \le \frac{k_d}{k}$*.*

5. Proof of Theorem 3.1

The proof strategy is, first, to derive necessary conditions of optimality for $(\bar{z} := \Lambda \circ \bar{x}, \bar{u})$ which, from Corollary 4.2, is a strong local minimizer for (R). We subsequently interpret these conditions as necessary conditions for (P). The derivation of necessary conditions for (R) is in several stages.

Step 1: Hypothesis Simplification:

We show, without loss of generality, we can restrict the proof of Theorem 3.1 to the case when

(i): *The hypotheses (H1)–(H5) are satisfied with* $\epsilon = +\infty$
and

(ii): *there is no running cost, i.e.* $L \equiv 0$.

To see that (i) is justified consider the modified version $(\tilde{R})$, involving the truncated data $\tilde{g}$ and $\tilde{f}$. (See Cor. 4.2.) $(\tilde{R})$ satisfies the strengthened form of the hypotheses, with $\epsilon = +\infty$. A similar analysis to that in the proof of Corollary 4.2 tells us that $(\bar{z}, \bar{u})$ is a strong local minimizer for $(\tilde{R})$. We may establish, under the strengthened hypotheses, that $(\bar{z}, \bar{u})$, satisfies first order conditions of optimality, in relation to $(\tilde{R})$. But these necessary conditions are the same as they would be, if we retained the original data (g, f), since (g, f) and $(\tilde{g}, \tilde{f})$ coincide locally. The assertions of the theorem are thereby confirmed. We may assume therefore $\epsilon = +\infty$.

Consider next (ii). Suppose that the theorem has been proved when $L \equiv 0$. Let $(\bar{x}, \bar{u})$ be a strong local Λ-minimizer for (P) (in which L is possibly non-zero). Write $\bar{x}'(t) \equiv \int_0^t L(s, \Lambda \circ \bar{x}(s), u(s))\mathrm{d}s$. Let $(\tilde{P})$ to be a variant of problem (P), in which the infinitesimal generator is $\tilde{\mathcal{A}} := \mathcal{A} \times \{0\}$ on the extended state space $\tilde{X} := X \times \mathbb{R}$ (with domain $\mathcal{D}(\mathcal{A}) \times \mathbb{R}$), the nonlinear term in the dynamics is $\tilde{f}(t, \Lambda(x), x', u) = (f, L)(t, \Lambda(x), u)$, the cost function is $\tilde{g}(z, z') = g(z) + z'$, the running cost is $\tilde{L} \equiv 0$ and the initial state is $(x_0, x'(0) := 0)$. Take also the new linear operators $\tilde{\Lambda} : X \times \mathbb{R} \to \mathbb{R}^n \times \mathbb{R}$ and $\tilde{M} : \mathbb{R}^\kappa \times \mathbb{R} \to X \times \mathbb{R}$ to be

$$\tilde{\Lambda}(x, x') := (\Lambda(x), x') \text{ and } \tilde{M}(y, z') := (M(y), z').$$

The linear operator $\tilde{\mathcal{A}}$ is the generator of a strongly continuous semigroup on $X \times \mathbb{R}$ $\{\tilde{S}(t) : X \times \mathbb{R} \to X \times \mathbb{R} : t \geq 0\}$ which we can explicitly construct from the semigroup $\{S(t) : t \geq 0\}$ as

$$\tilde{S}(t)(x_0, x_0') = (S(t)x_0, x_0') \text{ for } (x_0, x_0') \in X \times \mathbb{R} \text{ and } t \geq 0.$$

We deduce from this representation that, for given u, $((x, x'), u)$ is a process for $(\tilde{P})$ if and only if '(x, u) is a process for (P) and $x'(t) \equiv \int_0^t L(s, x(s), u(s))\mathrm{d}s$'. Furthermore the costs $J(x, u)$ and $J'((x, x'), u)$ coincide, since $J'((x, x'), u) = \tilde{g}(x(T), x'(T)) = g(x(T)) + \int_0^T L(s, x(s), u(s))\mathrm{d}s = J(x, u)$. Noting also that, for any $\rho > 0$ and any essentially bounded function, $(x, x') : I \to X \times \mathbb{R}$, '$||\tilde{\Lambda} \circ ((x, x') - (\bar{x}, \bar{x}'))||_{L^\infty} \leq \rho$' implies '$||\Lambda \circ (x - \bar{x})||_{L^\infty} \leq \rho$', we can conclude that $((\bar{x}, \bar{x}'), \bar{u})$ is a strong local $\tilde{\Lambda}$-minimizer for $(\tilde{P})$. The data for $(\tilde{P})$ satisfies hypotheses (H1)–(H5) (with reference to $(\bar{x}, \bar{x}')$). We are justified then in applying the special case of Theorem 3.1 to problem $(\tilde{P})$, in which there is no running cost, with reference to $((\bar{x}, \bar{x}'), \bar{u})$. We can deduce all the assertions of the theorem for the original problem (P). Note in particular that, since $\tilde{f}(t, (z, z'), u)$ does not depend on the second projected state component z', we deduce from the costate equation and the transversality condition that the costate function for $(\tilde{P})$ has the structure $(p, p'(t) \equiv -\lambda)$, in which λ is the cost multiplier. We are justified then in imposing the extra hypothesis $L \equiv 0$.

Step 2: Necessary Conditions for a Problem with Parameters:

In this step we derive necessary conditions of optimality for the following modified version of (R), whose end-point cost, velocity function and running cost are functions of an additional vector variable $\xi \in \mathbb{R}^\ell$ (called a 'parameter') and for which there is no pathwise state constraint:

$$(R^{(1)})\begin{cases} \text{Minimize } \tilde{g}(z(T),\xi)+\int_0^T \tilde{L}(t,z(t),\xi,u(t))\mathrm{d}t \\ \text{over functions } z\in\mathcal{C}(I;\mathbb{R}^n),\ \xi\in\mathbb{R}^\ell \text{ and measurable functions } u:I\to\mathbb{R}^m \text{ s.t.} \\ z(t)=\Lambda\circ S(t)x_0+\int_0^t \Lambda\circ S(t-s)M\circ\tilde{f}(s,z(s),\xi,u(s))\mathrm{d}s, \quad \text{for all } t\in I, \\ u(t)\in U, \quad \text{for a.e. } t\in I, \\ z(0)=\Lambda(x_0), \\ \|z-\bar{z}\|_{\mathcal{C}}\le\varepsilon' \text{ and } |\xi-\bar{\xi}|\le\varepsilon'. \end{cases}$$

In this problem, $\tilde{f}$, $\tilde{L}$ (replacing f and L) and U assumed to satisfy hypotheses (H1)–(H3) (with $\epsilon=\infty$), when ξ is treated as a additional component of the state variable. However $\tilde{g}$ is required to satisfy merely the modified version of (H5), namely:

(H5)$'$: $\tilde{g}$ is Lipschitz continuous. (Write its Lipschitz constant $k_{\tilde{g}}$.)

(Notice that $\tilde{g}$ is not required to be continuously differentiable.) Here, $(\bar{z},\bar{\xi})$ is some given reduced state/parameter pair and $\varepsilon'>0$. The domain of problem $(R^{(1)})$ comprises triples (x,ξ,u). Processes, admissible processes and minimizers for $(R^{(1)})$, which comprise triples (z,ξ,u), have their obvious meanings.

Proposition 5.1. *Take a minimizer $(\bar{z},\bar{\xi},\bar{u})$ for $(R^{(1)})$. There exists a continuous function $q:I\to\mathbb{R}^\kappa$ and a row vector ζ that satisfy the following conditions:*

$$\begin{aligned} q(t) = -M^*S^*(T-t)\Lambda^*\zeta + \\ \int_t^T M^*S^*(s-t)\Lambda^*\,\nabla_z\,(q\cdot\tilde{f}-\tilde{L})(s,\bar{z}(s),\bar{\xi},\bar{u}(s))\mathrm{d}s \ \ \textit{for all } t\in I, \\ q(t)\cdot\tilde{f}(t,\bar{z}(t),\bar{\xi},\bar{u}(t))-\tilde{L}(t,\bar{z}(t),\,\bar{\xi},\bar{u}(t)) \\ \ge\ q(t)\cdot\tilde{f}(t,\bar{z}(t),\bar{\xi},u))-\tilde{L}(t,\bar{z}(t),\bar{\xi},u), \quad \textit{for all } u\in U, \ \ \textit{a.e. } t\in I, \end{aligned} \tag{5.1}$$

$$(\zeta,\int_I \nabla_\xi\,(q\cdot\tilde{f}-\tilde{L})(s,\bar{z}(s),\bar{\xi},\bar{u}(s))\mathrm{d}s)\ \in\ \partial_\xi\tilde{g}(\bar{x}(T),\bar{\xi}).$$

(Notice that the cost multiplier is unity in these conditions.)

Proof of Proposition 5.1: We can assume that $\tilde{L}\equiv 0$. (See Hypothesis Simplification (ii).)

Consider the following variants of $(R^{(1)})$, $k=1,2\ldots$, with terminal cost the quadratic inf-convolution $\tilde{g}_k$ of $\tilde{g}$ (see (4.5)) in which the state trajectory z and parameter ξ in the dynamic constraint are replaced by control-like variables y and η and the discrepancy is accommodated by an integral penalty term (in which $k_{\tilde{f}}$ is the integrable Lipschitz bound of (H2) for the modified problem),

$$(R_k^{(1)})\begin{cases} \text{Minimize } J_k^{(1)}(z,y,\xi,\eta,u):=\tilde{g}_k(z(T),\xi)+k\int_I k_{\tilde{f}}(t)\big(|z(t)-y(t)|^2+|\xi-\eta(t)|^2\big)\mathrm{d}t, \\ \text{over functions } (z,y,\xi,\eta,u) \text{ such that } y,\eta \text{ and } u \text{ are measurable functions and} \\ z(t)=\Lambda\circ S(t)x_0+\int_0^t\Lambda\circ S(t-s)M\circ\tilde{f}(s,y(s),\eta(s),u(s))\mathrm{d}s, \quad \text{for any } t\in I, \\ u(t)\in U, y(t)\in\mathbb{R}^n \text{ and } \eta(t)\in\mathbb{R}^\ell, \quad \text{a.e. } t\in I, \\ \|z-\bar{z}\|_{\mathcal{C}}+|\xi-\bar{\xi}|\le\tilde{\varepsilon}, \\ \int_I k_{\tilde{f}}(t)(|y(t)|+|\eta(t)|)\mathrm{d}t<\infty, \end{cases}$$

in which $\tilde{\varepsilon}=\min\left\{\frac{\varepsilon'}{2},\frac{\sigma}{2}\right\}$.

Lemma 5.2. *There exists a sequence $\gamma_k \downarrow 0$ such that, for each k,*

$$J_k^{(1)}(\bar{z}, \bar{z}, \bar{\xi}, \bar{\xi}, \bar{u}) \leq \inf(R_k^{(1)}) + \gamma_k,$$

where $\inf(R_k^{(1)})$ *denotes 'infimum cost' of* $(R_k^{(1)})$*, i.e.* $(\bar{z}, \bar{z}, \bar{\xi}, \bar{\xi}, \bar{u})$ *is a* γ_k*-minimizer for* $(R_k^{(1)})$*.*

Proof. Fix k. Take any admissible process (z, y, ξ, η, u) for $(R_k^{(1)})$. According to Lemma 4.1 there exists a continuous function z_k such that

$$z_k(t) = \Lambda \circ S(t)x_0 + \int_0^t \Lambda \circ S(t-s) M \circ \tilde{f}(s, z_k(s), \xi, u(s))\mathrm{d}s.$$

(Observe that z_k is the reduced state trajectory of the process (z_k, ξ, u) for $(R^{(1)})$.) We deduce from (H2) and (4.4) that, for any $t \in I$,

$$\begin{aligned}
|z_k(t) - z(t)| &\leq M_S \int_I |\tilde{f}(s, z_k(s), \xi, u(s)) - \tilde{f}(s, y(s), \eta(s), u(s))|\mathrm{d}s \\
&\leq M_S \int_I k_{\tilde{f}}(s)(|z_k(s) - y(s)| + |\xi - \eta(s)|)\mathrm{d}s \\
&\leq M_S \int_I k_{\tilde{f}}(s)|z_k(s) - z(s)|\mathrm{d}s + M_S \int_I k_{\tilde{f}}(s)(|z(s) - y(s)| + |\xi - \eta(s)|)\mathrm{d}s.
\end{aligned}$$

It now follows from Gronwall's lemma that

$$\|z_k - z\|_{\mathcal{C}} \leq M_S \mathrm{e}^{M_S \int_I k_{\tilde{f}}(t)\mathrm{d}t} \int_I k_{\tilde{f}}(t)(|z(t) - y(t)| + |\xi - \eta(t)|)\mathrm{d}t. \tag{5.2}$$

By Lemma (i) 4.4 and Hölder's inequality, we have

$$\begin{aligned}
J_k(z, y, \xi, \eta, u) - J_k(z_k, z_k, \xi, \xi, u) &= \tilde{g}_k(z(T), \xi) - \tilde{g}_k(z_k(T), \xi) \qquad (5.3)\\
&\quad + k \int_I k_{\tilde{f}}(t)\big(|z(t) - y(t)|^2 + |\xi - \eta(t)|^2\big)\mathrm{d}t \\
&\geq k \int_I k_{\tilde{f}}(t)\big(|z(t) - y(t)|^2 + |\xi - \eta(t)|^2\big)\mathrm{d}t - z(T) - z_k(T)| \\
&\geq k \int_I k_{\tilde{f}}(t)\big(|z(t) - y(t)|^2 + |\xi - \eta(t)|^2\big)\mathrm{d}t - k_{\tilde{g}} \int_I k_{\tilde{f}}(t)(|z(t) - y(t)| + |\xi - \eta(t)|)\mathrm{d}t \\
&\geq k\Big(\int_I k_{\tilde{f}}(t)\mathrm{d}t\Big)^{-1}\Big(\int_I k_{\tilde{f}}(t)|z(t) - y(t)|\mathrm{d}t\Big)^2 - k_{\tilde{g}} \int_I k_{\tilde{f}}(t)|z(t) - y(t)|\mathrm{d}t \\
&\quad + k\Big(\int_I k_{\tilde{f}}(t)\mathrm{d}t\Big)^{-1}\Big(\int_I k_{\tilde{f}}(t)|\xi - \eta(t)|\mathrm{d}t\Big)^2 - k_{\tilde{g}} \int_I k_{\tilde{f}}(t)|\xi - \eta(t)|\mathrm{d}t \\
&\geq 2 \times \min\{k\Big(\int_I k_{\tilde{f}}(t)\mathrm{d}t\Big)^{-1}\alpha^2 - k_{\tilde{g}} : \alpha \in \mathbb{R}\} = C/k,
\end{aligned}$$

in which $C := \frac{1}{2} k_{\tilde{g}}^2 \int_I k_{\tilde{f}}(t)\mathrm{d}t$.

There are two cases to consider. First suppose that

$$M_S \mathrm{e}^{M_S \int_I k_{\tilde{f}}(\tau)\mathrm{d}\tau} \int_I k_{\tilde{f}}(t)(|z(t) - y(t)| + |\xi - \eta(t)|)\mathrm{d}t \leq \varepsilon.$$

Then, by (5.2),

$$\|z_k - \bar{z}\|_{L^\infty} \le \|z_k - z\|_{L^\infty} + \|z - \bar{z}\|_{L^\infty} < 2\varepsilon \le \varepsilon'.$$

It follows that (z_k, ξ, u) is an admissible process for $(R^{(1)})$. Since $(\bar{z}, \bar{\xi}, \bar{u})$ is a minimizer for $(R_k^{(1)})$ and in view of Lemma 4.4 (ii), we have

$$\begin{aligned} J_k^{(1)}(z_k, z_k, \xi, \xi, u) = \tilde{g}_k(z_k(T), \xi) \ge \tilde{g}(z_k(T)), \xi) - k_g\,/k \ge g(\bar{z}(T), \bar{\xi}) - k_g\,/k \\ \ge \tilde{g}_k(\bar{z}(T), \bar{\xi}) - k_g\,/k = J_k^{(1)}(\bar{z}, \bar{z}, \bar{\xi}, \bar{\xi}, \bar{u}) - k_g\,/k. \end{aligned} \tag{5.4}$$

Combining this inequality with (5.3), we obtain

$$J_k^{(1)}(z, y, \xi, \eta, u) - J_k(\bar{z}, \bar{z}, \bar{\xi}, \bar{\xi}, \bar{u}) \ge -(C + k_g)/k.$$

Secondly suppose that

$$M_S \mathrm{e}^{M_S \int_I k_{\tilde{f}}(\tau)\mathrm{d}\tau} \int_I k_{\tilde{f}}(t)(|z(t) - y(t)| + |\xi - \eta(t)|)\mathrm{d}t \ge \varepsilon.$$

Then, from Hölder's inequality, inequality (5.2) and the Lipschitz continuity of $\tilde{g}$, we deduce

$$\begin{aligned} J_k^{(1)}(z, y, \xi, \eta, u) - J_k(\bar{z}, \bar{z}, \bar{\xi}, \bar{\xi}, \bar{u}) \ge k\, \frac{\left(\int_I k_{\tilde{f}}(t)\mathrm{d}t\right)^{-1} \varepsilon^2}{\left(M_S \mathrm{e}^{M_S \int_I k_{\tilde{f}}(\tau)\mathrm{d}\tau}\right)^2} - k_g \epsilon \\ \ge 0, \text{for } k \text{ sufficiently large.} \end{aligned} \tag{5.5}$$

Since (z, y, ξ, η, u) is an arbitrary admissible process, we deduce from (5.4) and (5.5) that $(\bar{z}, \bar{z}, \bar{\xi}, \bar{\xi})$ is a γ_k minimizer for $(R_k^{(1)})$, for k sufficiently large, when we choose

$$\gamma_k := (C + k_g)/k, \text{ for } k \text{ sufficiently large.}$$

□

Consider now the complete metric space

$$\begin{aligned} \mathcal{A} = \big\{ (z, y, \xi, \eta, u) \in \mathcal{C}(I; \mathbb{R}^n) \times L^1_{k_{\tilde{f}}}(I; \mathbb{R}^n) \times \mathbb{R}^\ell \times L^1_{k_{\tilde{f}}}(I; \mathbb{R}^\ell) \times \mathbb{R}^m : \\ u \text{ is measurable and } u(t) \in U \text{ a.e., } (z, y, \xi, \eta, u) \text{ is an admissible process for } (R_k^{(1)}) \big\} \end{aligned}$$

endowed with the metric

$$\begin{aligned} & d_{\mathcal{A}}((z, y, \xi, \eta, u), (z', y', \xi', \eta', u')) \\ & = \int_I k_{\tilde{f}}(t)|y'(t) - y(t)|\mathrm{d}t + \int_I k_{\tilde{f}}(t)|\eta'(t) - \eta(t)|\mathrm{d}t + |\xi' - \xi| + \mathrm{meas}\{t \in I : u'(t) \ne u(t)\}. \end{aligned}$$

(Notice that $d_{\mathcal{A}}$ is metric on processes (z, y, ξ, η, u) for (R_k^1), even though its definition makes no reference to the z component. This is because, for a given (y, ξ, η, u), there is a unique z such that (z, y, ξ, η, u) is a process for $(R_k^{(1)})$.)

$J_k^{(1)}$ is lower semi-continuous on $(\mathcal{A}, d_{\mathcal{A}})$. Noting Lemma 5.2, we deduce from Ekeland's Theorem that, for any k, there exists a minimizer $(z_k, y_k, \xi_k, \eta_k, u_k)$ for the problem

$$(R_k^{(2)}): \quad \min\{J_k^{(2)}(z, y, \xi, \eta, u) \,:\, (z, y, \xi, \eta, u) \in \mathcal{A}\}$$

in which

$$\begin{aligned} J_k^{(2)}(z, y, \xi, \eta, u) &:= J_k^{(1)}(z, y, \xi, \eta, u) + \sqrt{\gamma_k} d_{\mathcal{A}}\big((z, y, \xi, \eta, u), (z_k, y_k, \xi_k, \eta_k, u_k)\big) \\ &= \tilde{g}_k(z(T), \xi) + k \int_I k_{\tilde{f}}(t)\big(|y(t) - z(t)|^2 + |\eta(t) - \xi)|^2\big)\mathrm{d}t \\ &\quad + \sqrt{\gamma_k}\Big(\int_I k_{\tilde{f}}(t)|y_k(t) - y(t)|\mathrm{d}t + \int_I k_{\tilde{f}}(t)|\eta_k(t) - \eta(t)|\mathrm{d}t + |\xi - \xi_k| + \int_I m_k(t, u(t))\mathrm{d}t\Big). \end{aligned}$$

Here,

$$m_k(t, u) = \begin{cases} 0 & \text{if } u = u_k(t) \\ 1 & \text{if } u \neq u_k(t) \end{cases}. \tag{5.6}$$

Furthermore,

$$d_{\mathcal{A}}\big((z_k, y_k, \xi_k, \eta_k, u_k), (\bar{z}, \bar{z}, \bar{\xi}, \bar{\xi}, \bar{u}),\big) \leq \sqrt{\gamma_k}.$$

It follows from property (iii) of the quadratic inf-convolution Lemma 4.4, that $(z_k, y_k, \xi_k, \eta_k, u_k)$ is also a minimizer for the problem:

$$(R_k^{(3)}): \quad \min\{J_k^{(3)}(z, y, \xi, \eta, u) \,:\, (z, y, \xi, \eta, u) \in \mathcal{A}\}.$$

Here,

$$\begin{aligned} J_k^{(3)}(z, y, \xi, \eta, u) &:= (\zeta_k^{(1)}, \zeta_k^{(2)}) \cdot ((z(T), \xi) - (z_k(T), \xi_k)) + k|(z(T), \xi) - (z_k(T), \xi_k)|^2 \\ &\quad + k \int_I k_{\tilde{f}}(t)\big(|y(t) - z(t)|^2 + |\eta(t) - \xi)|^2\big)\mathrm{d}t \\ &\quad + \sqrt{\gamma_k}\Big(\int_I k_{\tilde{f}}(t)|y(t) - y_k(t)|\mathrm{d}t + \int_I k_{\tilde{f}}(t)|\eta(t) - \eta_k(t)|\mathrm{d}t + |\xi - \xi_k| + \int_I m_k(t, u(t))\mathrm{d}t\Big) \end{aligned}$$

and $(\zeta_k^{(1)}, \zeta_k^{(2)})$ is a $2 \times n$ vector such that

$$(\zeta_k^{(1)}, \zeta_k^{(2)}) \in \partial_P \tilde{g}(y_k^{(1)}, y_k^{(2)}), \qquad \text{for some } (y_k^{(1)}, y_k^{(2)}) \in (z_k(T), \xi_k) + k^{-1} k_g \,\mathbb{B}. \tag{5.7}$$

The conditions listed in the following lemma will now be derived from the fact that $(z_k, y_k, \xi_k, \eta_k, u_k)$ is a minimizer for $(R_k^{(3)})$.

Lemma 5.3. *Fix k. Define the functions $q_k : I \to \mathbb{R}^\kappa$ and $r_k : I \to \mathbb{R}^\ell$ to be*

$$q_k(t) = -M^* S^*(T - t)\Lambda^* \zeta_k^{(1)} + 2k \int_t^T M^* S^*(s - t)\Lambda^*\big(k_{\tilde{f}}(s)(y_k(s) - z_k(s)\big)\mathrm{d}s$$

and

$$r_k(t) = -\zeta_k^{(2)} + 2k\int_t^T k_{\tilde{f}}(s)(\eta_k(s) - \xi_k)\mathrm{d}s.$$

Then

$$q_k(t)\cdot\tilde{f}(t, y_k(t), \eta_k(t), u_k(t)) \geq q_k(t)\cdot\tilde{f}(t, y_k(t), \eta_k(t), u) - \sqrt{\gamma_k}, \text{ for all } u \in U, \text{ a.e. } t \in I, \tag{5.8}$$

$$r_k(0) \in \sqrt{\gamma_k}\mathbb{B}, \tag{5.9}$$

$$-2kk_{\tilde{f}}(t)(y_k(t) - z_k(t))^T \in -q_k(t)\cdot\tilde{f}_z(t, y_k(t)\cdot\eta_k(t), u_k(t)) + \sqrt{\gamma_k}\mathbb{B}, \text{ a.e. } t \in I, \tag{5.10}$$

and

$$-2kk_{\tilde{f}}(t)(\eta_k(t) - \xi_k)^T \in -q_k(t)\cdot\tilde{f}_\xi(t, y_k(t), \eta_k(t), u_k(t)) + \sqrt{\gamma_k}\mathbb{B}, \text{ a.e. } t \in I. \tag{5.11}$$

Proof. Take any measurable selection u of U, $y \in L^2_{k_{\tilde{f}}}(I;\mathbb{R}^n)$ and $\eta \in L^2_{k_{\tilde{f}}}(I;\mathbb{R}^\ell)$. Let $\bar{t} \in (0,T)$ be an arbitrary Lebesgue point of both

$$s \mapsto \tilde{f}(s, y(s), \eta(s), u(s)) - \tilde{f}(s, y_k(s), \eta_k(s), u_k(s))$$

and

$$\begin{aligned} s \mapsto & kk_{\tilde{f}}(s)\big(|y(s) - z_k(s)|^2 - |y_k(s) - z_k(s)|^2\big) + kk_{\tilde{f}}(s)\big(|\eta(s) - \xi_k(s)|^2 - |\eta_k(s) - \xi_k(s)|^2\big) \\ & + \sqrt{\gamma_k}k_{\tilde{f}}(s)(|y(s) - y_k(s)| + |\eta(s) - \eta_k(s)|). \end{aligned}$$

Take $\sigma_k \in \mathbb{R}^\ell$. Choose any sequence $\varepsilon_j \downarrow 0$. Define, for each k,

$$u_j(t) = \begin{cases} u(t) & \text{if } t_j \leq t \leq \bar{t} \\ u_k(t) & \text{otherwise,} \end{cases} \qquad y_j(t) = \begin{cases} y(t) & \text{if } t_j \leq t \leq \bar{t} \\ y_k(t) & \text{otherwise,} \end{cases}$$

$$\xi_j = \xi_k + \varepsilon_j\sigma_k, \qquad \eta_j(t) = \begin{cases} \eta(t) & \text{if } t_j \leq t \leq \bar{t} \\ \eta_k(t) & \text{otherwise,} \end{cases}$$

in which $t_j = \bar{t} - \varepsilon_j$. We can arrange, by choosing j sufficiently large, that $t_j > 0$. Now take

$$z_j(t) = \Lambda\circ S(t)x_0 + \int_0^t \Lambda\circ S(t-s)M\circ\tilde{f}(s, y_j(s), \eta_j(s), u_j(s))\mathrm{d}s.$$

and

$$\varsigma_k(t) = \Lambda\circ S(t-\bar{t})M\circ(\tilde{f}(\bar{t}, y(\bar{t}), \eta(\bar{t}), u(\bar{t})) - \tilde{f}(\bar{t}, y_k(\bar{t}), \eta_k(\bar{t}), u_k(\bar{t}))).$$

Then

$$\varepsilon_j^{-1}\Big(z_j - z_k\Big) \to \varsigma_k, \quad \text{uniformly on } [\bar{t}, T]. \tag{5.12}$$

Indeed, writing $\psi(s)=\tilde{f}(s,y(s),\eta(s),u(s))-\tilde{f}(s,y_k(s),\eta_k(s),u_k(s))$ we see that

$$\begin{aligned}
\sup_{t\geq\bar t}\Big|\frac{z_j(t)-z_k(t)}{\varepsilon_j}-\varsigma_k(t)\Big| &= \sup_{t\geq\bar t}\Big|\frac{1}{\varepsilon_j}\int_{t_j}^{\bar t}\Lambda\circ S(t-s)M\circ\psi(s)\mathrm{d}s-\Lambda\circ S(t-\bar t)M\circ\psi(\bar t)\Big|\\
&\leq \sup_{t\geq\bar t}\frac{1}{\varepsilon_j}\int_{t_j}^{\bar t}\big|\Lambda\circ S(t-\bar t)\big(S(\bar t-s)M\circ\psi(s)-M\circ\psi(\bar t)\big)\big|\mathrm{d}s\\
&\leq M_S\frac{1}{\varepsilon_j}\int_{t_j}^{\bar t}|S(\bar t-s)M\circ(\psi(s)-\psi(\bar t))|\mathrm{d}s+M_S\frac{1}{\varepsilon_j}\int_{t_j}^{\bar t}|S(\bar t-s)M\circ\psi(\bar t)-M\circ\psi(\bar t)|\mathrm{d}s\\
&\leq M_S^2\frac{1}{\varepsilon_j}\int_{t_j}^{\bar t}|\psi(s)-\psi(\bar t)|\mathrm{d}s+M_S\frac{1}{\varepsilon_j}\int_{t_j}^{\bar t}|S(\bar t-s)M\circ\psi(\bar t)-M\circ\psi(\bar t)|\mathrm{d}s\to 0, \qquad \text{as } \varepsilon_j\to 0^+.
\end{aligned}$$

Now, by optimality of $(z_k,y_k,\xi_k,\eta_k,u_k)$ and in view of the conditions on y and η, we conclude that

$$\begin{aligned}
0 &\leq \frac{1}{\varepsilon_j}\big((\zeta_k^{(1)},\zeta_k^{(2)})\cdot(z_j(T)-z_k(T),\varepsilon_j\sigma_k)+k(|z_j(T)-z_k(T)|^2+\varepsilon_j^2|\sigma_k|^2\big)\\
&\quad+\frac{1}{\varepsilon_j}\int_I kk_{\tilde f}(t)\big(|y_j(t)-z_j(t)|^2-|y_k(t)-z_k(t)|^2+|\eta_j(t)-\xi_j|^2-|\eta_k(t)-\xi_k)|^2\big)\mathrm{d}t\\
&\quad+\frac{\sqrt{\gamma_k}}{\varepsilon_j}\Big(\int_I k_{\tilde f}(t)|y_j(t)-y_k(t)|\mathrm{d}t+\int_I k_{\tilde f}(t)|\eta_j(t)-\eta_k(t)|\mathrm{d}t+\int_I m_k(t,u_j(t))\Big)\mathrm{d}t+\sqrt{\gamma}_k|\sigma_k|\\
&\leq \frac{1}{\varepsilon_j}\big(\zeta_k^{(1)}\cdot(z_j(T)-z_k(T))+k|z_j(T)-z_k(T)|^2\big)+\zeta_k^{(2)}\cdot\sigma_k+\varepsilon_j|\sigma_k|^2\\
&\quad+\frac{1}{\varepsilon_j}\int_{t_j}^{T}kk_{\tilde f}(t)\big(|y_j(t)-z_j(t)|^2-|y_k(t)-z_k(t)|^2+|\eta_j(t)-\xi_j)|^2-|\eta_k(t)-\xi_k|^2\big)\mathrm{d}t\\
&\quad+\frac{\sqrt{\gamma_k}}{\varepsilon_j}\Big(\int_{t_j}^{\bar t}k_{\tilde f}(t)|y(t)-y_k(t)|\mathrm{d}t+\int_{t_j}^{\bar t}k_{\tilde f}(t)|\eta(t)-\eta_k(t)|\mathrm{d}t\Big)+\sqrt{\gamma_k}+\sqrt{\gamma}_k|\sigma_k|.
\end{aligned} \tag{5.13}$$

But

$$\begin{aligned}
\int_{t_j}^{T}k_{\tilde f}(t)\big(|y_j(t)-z_j(t)|^2-|y_k(t)-z_k(t)|^2\big)\mathrm{d}t &\leq \int_{t_j}^{\bar t}k_{\tilde f}(t)\big(|y(t)-z_k(t)|^2-|y_k(t)-z_k(t)|^2\big)\mathrm{d}t\\
&\quad+\int_{t_j}^{\bar t}k_{\tilde f}(t)\big(|y(t)-z_j(t)|^2-|y(t)-z_k(t)|^2\big)\mathrm{d}t+\int_{\bar t}^{T}k_{\tilde f}(t)\big(|y_k(t)-z_j(t)|^2-|y_k(t)-z_k(t)|^2\big)\mathrm{d}t,
\end{aligned}$$

and

$$\begin{aligned}
\int_{t_j}^{T}k_{\tilde f}(t)\big(|\eta_j(t)-\xi_j|^2-|\eta_k(t)-\xi_k|^2\big)\mathrm{d}t &\leq \int_{t_j}^{\bar t}k_{\tilde f}(t)\big(|\eta(t)-\xi_k|^2-|\eta_k(t)-\xi_k|^2\big)\mathrm{d}t\\
&\quad+\int_{t_j}^{\bar t}k_{\tilde f}(t)\big(|\eta(t)-\xi_j|^2-|\eta(t)-\xi_k|^2\big)\mathrm{d}t+\int_{\bar t}^{T}k_{\tilde f}(t)\big(|\eta_k(t)-\xi_j|^2-|\eta_k(t)-\xi_k|^2\big)\mathrm{d}t.
\end{aligned}$$

Passing to the limit $\varepsilon_j \to 0^+$ with the help of (5.12), we conclude that

$$\begin{aligned}\zeta_k^{(1)} \cdot \varsigma_k(T) + &\int_{\bar t}^T \big(-2kk_{\tilde f}(t)(y_k(t)-z_k(t))\big)\cdot \varsigma_k(t)\mathrm{d}t + \zeta_k^{(2)}\cdot \sigma_k + \int_{\bar t}^T \big(-2kk_{\tilde f}(t)(\eta_k(t)-\xi_k)\big)\cdot\sigma_k \mathrm{d}t\\ &+ kk_{\tilde f}(\bar t)\big(|y(\bar t)-z_k(\bar t)|^2 - |y_k(\bar t)-z_k(\bar t)|^2 + |\eta(\bar t)-\xi_k|^2 - |\eta_k(\bar t)-\xi_k|^2\big)\\ &+ \sqrt{\gamma_k}k_{\tilde f}(\bar t)(|y(\bar t)-y_k(\bar t)| + |\eta(\bar t)-\eta_k(\bar t)|) + \sqrt{\gamma_k} + \sqrt{\gamma}_k|\sigma_k| \ge 0.\end{aligned} \tag{5.14}$$

But $\varsigma_k(t) = S(t-\bar t)\big(\tilde f(\bar t, y(\bar t), \eta(\bar t), u(\bar t)) - \tilde f(\bar t, y_k(\bar t), \eta_k(t), u_k(\bar t))\big)$, for each $t > \bar t$.

From (5.14)

$$\begin{aligned}0 \le\ & \zeta_k^1 \cdot S(T-\bar t)\big(\tilde f(\bar t, y(\bar t), \eta(\bar t), u(\bar t)) - \tilde f(\bar t, y_k(\bar t), \eta_k(\bar t), u_k(\bar t))\big)\\ &+ \int_{\bar t}^T \big(-2kk_{\tilde f}(t)(y_k(t)-z_k(t))\big)\cdot S(t-\bar t)\big(\tilde f(\bar t, y(\bar t), \eta(\bar t), u(\bar t)) - \tilde f(\bar t, y_k(\bar t), \eta_k(\bar t), u_k(\bar t))\big)\mathrm{d}t\\ &- r_k(\bar t)\cdot\sigma_k + \sqrt{\gamma_k}k_{\tilde f}(\bar t)(|y(\bar t)-y_k(\bar t)| + |\eta(\bar t)-\eta_k(\bar t)|) + \sqrt{\gamma_k} + \sqrt{\gamma}_k|\sigma_k|\\ &+ kk_{\tilde f}(\bar t)\big(|y(\bar t)-z_k(\bar t)|^2 - |y_k(\bar t)-z_k(\bar t)|^2 + |\eta(\bar t)-\xi_k|^2 - |\eta_k(\bar t)-\xi_k|^2\big)\\ =\ & \Big(S(T-\bar t)^*\zeta_k^1 + \int_{\bar t}^T S(t-\bar t)^*(-2kk_{\tilde f}(y_k(t)-z_k(t))\mathrm{d}t\Big)\cdot\big(\tilde f(\bar t, y(\bar t), \eta(\bar t), u(\bar t)) - \tilde f(\bar t, y_k(\bar t), \eta_k(t), u_k(\bar t))\big)\\ &+ \langle -r_k(\bar t), \sigma_k\rangle + \sqrt{\gamma_k}k_{\tilde f}(\bar t)(|y(\bar t)-y_k(\bar t)| + |\eta(\bar t)-\eta_k(\bar t)|) + \sqrt{\gamma_k} + \sqrt{\gamma}_k|\sigma_k|\\ &+ kk_{\tilde f}(\bar t)\big(|y(\bar t)-z_k(\bar t)|^2 - |y_k(\bar t)-z_k(\bar t)|^2 + |\eta(\bar t)-\xi_k|^2 - |\eta_k(\bar t)-\xi_k|^2\big).\end{aligned}$$

But then,

$$\begin{aligned}-q_k(\bar t)\cdot&\Big(\tilde f(\bar t, y(\bar t), \eta(\bar t), u(\bar t)) - \tilde f(\bar t, y_k(\bar t), \eta_k(t), u_k(\bar t))\Big) - r_k(\bar t)\cdot\sigma_k\\ &+ kk_{\tilde f}(\bar t)\big(|y(\bar t)-z_k(\bar t)|^2 - |y_k(\bar t)-z_k(\bar t)|^2 + |\eta(\bar t)-\xi_k|^2 - |\eta_k(\bar t)-\xi_k|^2\big)\\ &+ \sqrt{\gamma_k}k_{\tilde f}(\bar t)(|y(\bar t)-y_k(\bar t)| + |\eta(\bar t)-\eta_k(\bar t)|) + \sqrt{\gamma_k} + \sqrt{\gamma}_k|\sigma_k| \ge 0.\end{aligned} \tag{5.15}$$

Fix $y = y_k$, $\eta = \eta_k$ and $\sigma_k = 0$. Noting the above relation is valid for any $\bar t$ in a set of full measure, setting $s = \bar t$ and integrating w.r.t. s over I, we deduce

$$\int_I q_k(s)\cdot\Big(\tilde f(s, y_k(s), \eta_k(s), u(s)) - \tilde f(s, y_k(s), \eta_k(t), u_k(s))\Big)ds \ge -T\times\sqrt{\gamma_k}.$$

Since this is true for an arbitrary selector u, we conclude, with the help of a measurable selection theorem, see [10], that condition (5.8) is satisfied.

Now consider relation (5.15) when the selector u is taken to be u_k. Set $y = y_k$ and $\eta = \eta_k$. Then, using $u = u_k$ in (5.13), for all points s is a set of full measure, $-r_k(s)\sigma_k + \sqrt{\gamma_k}|\sigma_k| \ge 0$. Since r_k is a continuous function, $-r_k(0)\sigma_k + \sqrt{\gamma_k}|\sigma_k| \ge 0$. In view of the fact that σ_k is an arbitrary vector, this implies $|r_k(0)| \le \sqrt{\gamma_k}$. We have shown (5.9).

Fix $\sigma_k = 0$ and $u = u_k$. By considering first $\eta = \eta_k$ and arbitrary y and then considering $y = y_k$ and arbitrary η, we can deduce from (5.15) that, for a.e. $t \in I$,

$$y \mapsto kk_{\tilde f}(t)|y - z_k(t)|^2 - q_k(t)\cdot\tilde f(t, y, \eta_k(t), u_k(t)) + \sqrt{\gamma_k}k_{\tilde f}(t)|y - y_k(t)|$$
$$\text{is minimized at } y = y_k(t),$$

and

$$\eta \mapsto kk_{\tilde{f}}(t)|\eta - \xi_k|^2 - q_k(t) \cdot \tilde{f}(t, y_k(t), \eta(t), u_k(t)) + \sqrt{\gamma_k} k_{\tilde{f}}(t)|\eta - \eta_k(t)| \quad \text{is minimized at } \eta = \eta_k(t).$$

These relations imply (5.10) and (5.11). □

The final stage of the proof of Proposition 5.1 is to pass to the limit in the conditions of Lemma 5.3, as $k \to \infty$. Take any k. Then, by (5.10), (5.11) and (H2),

$$| - 2kk_{\tilde{f}}(s)(y_k(s) - z_k(s))|, | - 2kk_{\tilde{f}}(s)(\eta_k(s) - \xi_k)| \leq |q_k(s)|k_{\tilde{f}}(s) + \sqrt{\gamma_k}, \quad \text{a.e.}$$

We deduce, with the help of Gronwall's Lemma, that $\|q_k\|_{\mathcal{C}} \leq C_1$ and therefore $\|r_k\|_{L^\infty} \leq C_1$, for some positive C_1 that does not depend on k. It follows that, for a.e. s,

$$| - 2kk_{\tilde{f}}(s)(y_k(s) - z_k(s))|, | - 2kk_{\tilde{f}}(s)(\eta_k(s) - \xi_k)| \leq C_1(k_{\tilde{f}}(s) + 1).$$

From the fact that

$$d_{\mathcal{A}}((z_k, y_k, \xi_k, \eta_k, u_k), (\bar{z}, \bar{z}, \bar{\xi}, \bar{\xi}, \bar{u})) \to 0,$$

we deduce that, as $k \to \infty$,

$$\|z_k - \bar{z}\|_{\mathcal{C}} \to 0, \quad \|y_k - \bar{z}\|_{L^1_{k_{\tilde{f}}}} \to 0,$$
$$|\xi_k - \bar{\xi}| \to 0, \quad \|\eta_k - \bar{\xi}\|_{L^1_{k_{\tilde{f}}}} \to 0, \quad \text{meas}\{t \in I : u_k(t) \neq \bar{u}(t)\} \to 0.$$

By extracting subsequences, we can arrange that there exists a set $\mathcal{O} \subset I$, of full Lebesgue measure, such that, for each $t \in \mathcal{O}$,

$$u_k(t) = \bar{u}(t), \text{ for all } k \text{ sufficiently large.} \tag{5.16}$$

We know, by (5.7) that, for each k, there exists $\big(\tilde{y}_k^{(1)}, \tilde{y}_k^{(2)}\big) \in (z_k(T), \xi_k) + \frac{k_g}{k}\mathbb{B}$ such that

$$(\zeta_k^{(1)}, \zeta_k^{(2)}) \in \partial_P \tilde{g}\big(\tilde{y}_k^{(1)}, \tilde{y}_k^{(2)}\big) \qquad \text{and} \qquad \big|\tilde{g}\big(\tilde{y}_k^{(1)}, \tilde{y}_k^{(2)}\big) - \tilde{g}(z_k(T), \xi_k)\big| < \frac{k_g}{k}.$$

Applying Proposition 4.3, extracting subsequences and passing to the limit as $k \to \infty$, we arrive at $(\zeta_k^{(1)}, \zeta_k^{(2)}) \to (\zeta, \zeta_0)$, for some $(\zeta, \zeta_0) \in \partial\tilde{g}(\bar{z}(T), \bar{\xi})$, and

$$-q(t) = M^* S(T-t)^* \Lambda^* \zeta - \int_t^T M^* S(s-t)^* \Lambda^* p(s) \cdot \tilde{f}_z(t, \bar{z}(t), \bar{\xi}, \bar{u}(t)) \mathrm{d}s, \quad \text{for each } t \in I.$$

Furthermore,

$$-r_k(t) \to \zeta_0 - \int_t^T q(s) \cdot \tilde{f}_\xi(s, \bar{z}(s), \bar{\xi}, \bar{u}(s)) \mathrm{d}s, \qquad \text{for all } t \in I.$$

Since $r_k(0) \to 0$, we deduce

$$(\zeta, \int_0^T q(s) \cdot \tilde{f}_\xi(s, \bar{z}(s), \bar{\xi}, \bar{u}(s))\mathrm{d}s) \in \tilde{g}(\bar{z}(T), \bar{\xi}).$$

Now, passing to the limit in (5.8) and noting (5.16), we obtain (5.1).

Step 3: Completion of the Proof:

We return to the proof of Theorem 3.1. Recalling Step 1 of the proof, we are justified in assuming that $L = 0$ and the strengthened versions of (H1)–(H5) are satisfied. Take $\gamma_k \downarrow 0$ and, for each k, define

$$g_k(z, \xi) := \max \big\{ g(z) - g(\bar{z}(T)) + \gamma_k, d_C(z), \xi \big\}.$$

We consider now the family of problems $(R_k^{(4)})$, $k = 1, 2, \ldots$, in which, for each k, we minimize the functional

$$J_k^{(4)}(z, u) = g_k\big(z(T), \max_{t \in I} h(z(t))\big)$$

over admissible processes (z, u) for problem (R). We see that $(\bar{z}, \bar{u})$ is a γ_k minimizer for problem $(R_k^{(4)})$. An analysis based on Ekeland's Theorem assures the existence, for each k sufficiently large, of a minimizer (z_k, u_k) for the problem $(R_k^{(5)})$ of minimizing

$$J_k^{(5)}(z, u) := J_k^{(4)}(z, u) + \sqrt{\gamma_k} \int_0^T m_k(t, u(t))\mathrm{d}t$$

over admissible processes (z, u) for (R) and such that

$$\mathrm{meas}\big\{ t : u_k(t) \neq \bar{u}(t) \big\} \leq \sqrt{\gamma_k}. \tag{5.17}$$

Here m_k is the function (5.6).

With the help of Proposition 4.3 we deduce from the preceding estimate that

$$||z_k - \bar{z}||_{\mathcal{C}} \to 0 \text{ as } k \to \infty.$$

Two possible cases need to be considered:

(A): $\max_{t \in I} h(z_k(t)) > 0$, for a finite number of k,

and

(B): $\max_{t \in I} h(z_k(t)) > 0$, for an infinite number of k.

In case (A) we can arrange, by excluding a finite number of index values, that

$$h(z_k(t)) \leq 0, \quad \text{for all } k.$$

In this case, (z_k, u_k) remains a minimizer of the above problem, when the state constraint is removed. We may apply Proposition 5.1 (in the special case that there is no dependence on the parameter ξ) to obtain conditions satisfied by (z_k, u_k). It follows that, for each k, there exists a continuous function $q_k : I \to \mathbb{R}^\kappa$ such that

$$-q_k(t) = M^* S^*(T - t)\Lambda^* \zeta_k - \int_t^T M^* S^*(s - t)\Lambda^* \, q_k(t) \cdot f_z(t, z_k(t), u_k(t))\mathrm{d}s \text{ for all } t \in I,$$

in which $\zeta_k \in \partial \max\left\{g(z_k(T)) - g(\bar{z}(T)) + \gamma_k, d_C(z_k(T))\right\}$. Furthermore,

$$q_k(t) \cdot f(t, z_k(t), u_k(t)) \geq q_k(t) \cdot f(t, z_k(t), u) - \sqrt{\gamma}_k, \quad \text{for all } u \in U, \text{ a.e. } t \in I.$$

Notice that, by optimality,

$$\max\{g(z_k(T)) - g(\bar{z}(T)) + \gamma_k, d_C(z_k(T))\} > 0.$$

Since '$\sigma \in \partial d_C(y)$ and $d_C(y) > 0$' implies '$|\sigma| = 1$', we conclude from the sum rule [9], Theorem 5.4.1 that, for some $\lambda_k \in [0,1]$ and $\varsigma_k \in \partial d_C(z_k(T))$ such that $|\varsigma_k| = 1$,

$$\zeta_k = \lambda_k \nabla g(z_k(T)) + (1 - \lambda_k)\varsigma_k.$$

Extracting convergent subsequences and passing to the limit as in the previous section, we deduce existence of (p, λ) satisfying conditions of (b) and (d) of Theorem 3.1, with μ set to zero. Finally, to confirm (a) and (c), we note that

$$(-q(T) =)\ \zeta^* = \lambda \nabla g(\bar{z}(T)) + (1 - \lambda)\varsigma,$$

for $\zeta^* = \lim_k \zeta_k$ and some $\varsigma \in \partial d_C(\bar{z}(T))$ such that $|\varsigma| = 1$. Here, $\lambda = \lim_k \lambda_k$. Since $\partial d_C(\bar{z}(T)) \subset N_C(\bar{z}(T))$ we have shown $-p(T) \in \lambda \nabla g(\bar{z}(T)) + N_C(\bar{z}(T))$. This is condition (c). The preceding relation implies that $(\lambda, p) \neq 0$. Indeed if $\lambda = 0$, then $|p(T)| = |\varsigma| = 1$ and so $p \neq 0$. Condition (a) is confirmed.

Consider next Case (B). We can arrange by extracting a subsequence that

$$\max_{t \in I} h(z_k(t)) > 0, \quad \text{for all } k.$$

Notice that there exists $\rho_k \downarrow 0$ such that, for any (z, u) process for (R) and $\xi \in \mathbb{R}$, the inequality

$$d_{\mathcal{E}}((z_k, u_k), (z, u)) + |\max_{t \in I} h(z_k(t)) - \xi| < \rho_k^{\frac{1}{2}}, \tag{5.18}$$

leads to

$$g_k\big(z(T), \max_{t \in I} h(z(t))\big) > 0 \ \text{ and } \ \xi > 0. \tag{5.19}$$

Now, fix k and, for arbitrary $K > 0$, consider the problem (R_k^K) of minimizing

$$J_k^K(z, u, \xi) = g_k(z(T), \xi) + \sqrt{\gamma_k} \int_0^T m_k(t, u(t)) \mathrm{d}t + K \int_0^T \big(\max\{h(z(t)) - \xi, 0\}\big)^2 \mathrm{d}t$$

over triples (z, u, ξ) such that (z, u) is an admissible process for (R) and $\xi \in \mathbb{R}$.

Lemma 5.4. *For any k, there exists K_k such that $\big(z_k, u_k, \max_{t \in I} h(z_k(t))\big)$ is a ρ_k minimizer for problem $(R_k^{K_k})$.*

Proof. Fix k and notice that $\inf(R_k^{K'}) \leq \inf(R_k^K) \leq \inf(R_k^{(5)})$, for $0 \leq K' \leq K$). Assume in contradiction that, for any $K_n \uparrow \infty$, there exists $(\tilde{z}_n, \tilde{u}_n, \tilde{\xi}_n)$ such that

$$J_k^{K_n}(\tilde{z}_n, \tilde{u}_n, \tilde{\xi}_n) < J_k^{(5)}(z_k, u_k) - \rho_k.$$

Notice that

$$K_n \int_0^T \big(\max\{h(\tilde z_n(t)) - \tilde\xi_n, 0\}\big)^2 \mathrm{d}t$$

is bounded, hence

$$\int_0^T \big(\max\{h(\tilde z_n(t)) - \tilde\xi_n, 0\}\big)^2 \mathrm{d}t \to 0,$$

and since $\tilde z_n$ is uniformly bounded, the sequence $\{\xi_n\}$ is bounded too. Applying Proposition 4.3 we can deduce that, along a subsequence, $\tilde z_n \to z \in \mathcal{C}(I;\mathbb{R}^n)$ and $\xi_n \to \xi$. We have

$$\int_0^T \big(\max\{h(\tilde z_n(t)) - \tilde\xi_n, 0\}\big)^2 \mathrm{d}t \to \int_0^T \big(\max\{h(z(t)) - \xi, 0\}\big)^2 \mathrm{d}t = 0,$$

whence $\max\{h(z(t)) - \xi, 0\} = 0$, for a.e. t. We also know, by uniform convergence, that

$$\max\big\{\max_{t\in I} h(\tilde z_n(t)) - \tilde\xi_n, 0\big\} \to \max\big\{\max_{t\in I} h(z(t) - \xi, 0\big\} = 0.$$

It follows that, for n sufficiently large,

$$\begin{aligned} J_k^{(5)}(\tilde z_n, \tilde u_n) &= J_k^{K_n}\big(\tilde z_n, \tilde u_n, \max_{t\in I} h(\tilde z_n(t))\big) \le J_k^{K_n}(\tilde z_n, \tilde u_n, \tilde\xi_n) + \max\big\{\max_{t\in I} h(\tilde z_n(t)) - \tilde\xi_n, 0\big\} \\ &\le J_k^{(5)}(z_k, u_k) - \frac{\rho_k}{2}, \end{aligned}$$

which contradicts the optimality of (z_k, u_k). The proof is complete. □

Take any k. By Lemma 5.4 and Ekeland's theorem, there exists a minimizer $(\hat z_k, \hat u_k, \hat\xi_k)$ for the problem

$$(R_k^{(6)}) : \text{ Minimize } \{J_k^{(6)}(z, u, \xi) \,:\, (z, u) \text{ is a process for } (R_k^{(6)}),\ \|z - \bar z\|_{\mathcal{C}} \le \epsilon \text{ and } \xi \in \mathbb{R}\}$$

in which

$$\begin{aligned} J_k^{(6)}(z, u, \xi) = g_k(z(T), \xi) &+ K_k \int_0^T \big(\max\{h(z(t)) - \xi, 0\}\big)^2 \mathrm{d}t + \sqrt{\gamma_k} \int_0^T m_k(t, u(t)) \mathrm{d}t \\ &+ \sqrt{\rho_k}\Big(\int_0^T \hat m_k(t, u(t)) \mathrm{d}t + \big|\max_{t\in I} h(\hat z_k(t)) - \xi\big|\Big), \end{aligned}$$

where

$$\hat m_k(t, u) = \begin{cases} 0 & \text{if } u = \hat u_k(t) \\ 1 & \text{if } u \ne \hat u_k(t) \end{cases}.$$

Furthermore,

$$\text{meas}\,\{t \in I : u_k(t) \ne \hat u_k\} + \big|\max_{t\in I} h(z_k(t)) - \hat\xi_k\big| \le \sqrt{\rho_k}.$$

From (5.17) we conclude that

$$\text{meas}\,\{t \in I : u_k(t) \neq \bar{u}(t)\} + \leq \sqrt{\rho_k} + \sqrt{\gamma_k}.$$

It follows that $||z_k - \bar{z}||_{\mathcal{C}} \to 0$, as $k \to \infty$.

Problem $(R_k^{(6)})$ is an optimization problem involving a parameter ξ, for which Proposition 5.1 provides necessary conditions. We conclude that, for any k, there exists a continuous function $q_k : I \to \mathbb{R}^\kappa$, $r_k \equiv$ constant (the costate component associated with the added state component) and a vector $(\zeta_k^{(1)}, \zeta_k^{(2)}) \in \mathbb{R}^n \times \mathbb{R}^\ell$ satisfying the following conditions, in which μ_k is Borel measure with Lipschitz continuous distribution function

$$\mu_k([0,t]) := 2K_k \int_{[0,t]} \max\{h(\hat{z}_k(s)) - \hat{\xi}_k, 0\}\mathrm{d}s \;\text{ for } t \in I.$$

$$q_k(t) = -M^* S^*(T-t)\Lambda^* \zeta_k^{(1)} + \int_t^T M^* S^*(s-t)\Lambda^* p(t) \cdot f_z(t, \hat{z}_k(t), \hat{u}_k(t))\mathrm{d}s + \int_t^T M^* S^*(s-t)\Lambda^* \nabla h(\hat{z}_k(s))\mathrm{d}\mu_k(s), \tag{5.20}$$

$$(\zeta_k^{(1)}, \zeta_k^{(2)}) \in \partial g_k(\hat{z}_k(T), \hat{\xi}_k) + \sqrt{\rho}_k \mathbb{B}$$

$$\int_I \mathrm{d}\mu_k(s) \in \zeta_k^{(2)} + \sqrt{\rho_k}\mathbb{B} \tag{5.21}$$

and

$$q_k(t) \cdot f(t, \hat{z}_k(t), \hat{u}_k(t)) \geq q_k(t) \cdot f(t, \hat{x}_k(t), u) - \sqrt{\gamma}_k - \sqrt{\rho}_k, \quad \text{for all } u \in U, \text{ a.e. } t \in I. \tag{5.22}$$

Furthermore,

$$\text{supp}\,\{\mu_k\} \;=\; \big\{t \in I : h(\hat{z}_k(t)) \geq \hat{\xi}_k\big\}. \tag{5.23}$$

Now define the continuous function $p_k : I \to \mathbb{R}^m$

$$p_k(t) := q_k(t) - \int_t^T M^* S^*(s-t)\Lambda^* \nabla h(\hat{z}_k(s))\mathrm{d}\mu_k(s).$$

From (5.18), (5.19),

$$g_k\big(\hat{z}_k(T), \max_{t \in I} h(\hat{z}_k(t))\big) > 0 \;\text{ and }\; \hat{\xi}_k > 0. \tag{5.24}$$

It follows then from (5.23) that

$$\text{supp}\,\mu_k = \big\{t \in I : h(\hat{z}_k(t)) \geq 0\big\}.$$

We know that $(\zeta_k^{(1)}, \zeta_k^{(2)}) \in \partial g_k(\hat{z}_k(T), \hat{\xi}_k) + \sqrt{\rho_k}\mathbb{B}$. It follows from sum rule for limiting subdifferentials that there exist $\lambda_k^i \geq 0$, $i = 1, 2, 3$, with $\sum_{i=1}^3 \lambda_k^i = 1$ such that

$$\zeta_k^{(1)} \in \lambda_k^1 \nabla g(\hat{z}_k(T)) + \lambda_k^2(\partial d_C(z_k(T)) \cap \partial\mathbb{B}) + \sqrt{\rho_k}\mathbb{B} \text{ and } \zeta_k^{(2)} \in [\lambda_k^3 - \sqrt{\rho_k}, \lambda_k^3 + \sqrt{\rho_k}], \tag{5.25}$$

where $\partial\mathbb{B}$, as before, denotes the surface of the unit ball in Euclidean space. (We use the positivity condition (5.24) to justify the '$(\partial d_C(z_k(T)) \cap \partial\mathbb{B})$' term in this relation.)

The sequence $\{\mu_k\}$ of non-decreasing bounded variation functions is bounded in total variation, by (5.21). Now extract subsequences so that $d\mu_k \to d\mu$, in the $\mathcal{C}^*$ topology for some normalized BV function $\mu \in NBV^+(I)$. By Lemma 4.3, we can also arrange that $p_k \to p$ in the $\mathcal{C}$ topology and $q_k(t) \to q(t)$ for all values of $t \in I$ in a set of full Lebesgue measure, including the points 0 and T. Furthermore, $(\lambda_k^1, \lambda_k^2) \to (\lambda, \lambda_2)$ and $(\zeta_k^{(1)}, \zeta_k^{(2)}) \to (\zeta, \lambda_3)$, for suitable elements $p, (\lambda, \lambda_2)$. Now pass to the limit as $k \to \infty$. In consequence of (5.21),

$$1 = \lim_{k\to\infty} (\lambda_k^1 + \lambda_k^2 + \lambda_k^3) = \lambda + \lambda_2 + \|\mu\|_{TV}.$$

Since $-q(T) = \zeta$, relation (5.25) implies the transversality condition

$$-q(T) \in \lambda \nabla g(\hat{z}_k(T)) + \lambda_2 \, \partial d_C(z_k(T))$$

and $-q(T) \geq \lambda_2 - \lambda k_g$. (Here, k_g is a local Lipschitz constant for g near $\bar{z}(T)$.) But then

$$|q(T)| + (1 + k_g)\lambda + \|\mu\|_{TV} \geq 1.$$

Since '$\mu = 0$' implies '$p = q$', this last relation implies $(\lambda, p, \mu) \neq (0, 0, 0)$. We have proved the non-degeneracy condition in the statement of the theorem. The costate equation and the Weierstrass condition result from passing to the limit in (5.20) and (5.22).

Appendix A.

In the introduction, we provided a motivating example of a lumped-distributed control system, modelled as a semi-linear evolution equation (1.1), involving the linear operator $\mathcal{A}$ given by (1.2). In this appendix, we show that $\mathcal{A}$ is the infinitesimal generator of a strongly continuous semigroup.

Proposition A.1. *Take $\mathcal{A}$ to be the linear operator defined by (1.2), in which $c, d_0, d_1 > 0$ are positive constants, with domain given by (1.3). Then $\mathcal{A}$ is the infinitesimal generator of a strongly continuous semigroup.*

Proof. We equip Hilbert space $X = L^2(0,1) \times L^2(0,1) \times \mathbb{R} \times \mathbb{R}$ with the inner product

$$\langle (v', s', \phi', \psi'), (v, s, \phi, \psi) \rangle := c^{-2} \int_0^1 v'(y)v(y)\mathrm{d}y + \int_0^1 s'(y)s(y)\mathrm{d}y + d_0^{-1}\phi'\,\phi + d_1^{-1}\psi'\,\psi,$$

which induces the strong product topology on $L^2 \times \mathbb{R} \times \mathbb{R}$. Note to begin that $\mathcal{D}(\mathcal{A})$ is dense in X. It is easy to check that $\mathcal{A}$ has closed graph. Take any $(v, s, \phi, \psi) \in \mathcal{D}(\mathcal{A})$. Then

$$\langle (v, s, \phi, \psi), \mathcal{A}(v, s, \phi, \psi) \rangle = 0 \tag{A.1}$$

Indeed, using the fact that $\phi = v(0)$ and $\psi = v(1)$, applying an integration by parts and noting the cancellation of constants, we see that

$$\begin{aligned}&\langle (v,s,\phi,\psi), \mathcal{A}(v,s,\phi,\psi)\rangle \\ &\quad = \int_0^1 v(y)\frac{\mathrm{d}s}{\mathrm{d}y}(y)\mathrm{d}y + \int_0^1 s(y)\frac{\mathrm{d}v}{\mathrm{d}y}(y)\mathrm{d}y + s(0)\phi - s(1)\psi \\ &\qquad = 0 + s(1)v(1) - s(0)v(0) + s(0)v(0) - s(1)v(1) \;=\; 0.\end{aligned}$$

In view of (A.1), $\mathcal{A}$ is dissipative. It will follow from the Lumer–Phillips Theorem [11] that $\mathcal{A}$ is the infinitesimal generator of a strongly continuous semigroup on X if, for some real $\lambda_0 > 0$, we can confirm the range condition

$$\mathcal{R}(\mathcal{A} - \lambda_0 I) = X.$$

To verify this condition take, for arbitrary $\lambda_0 > 0$, any $v' \in L^2$, $s' \in L^2$, $a' \in R$ and $b' \in R$. We must establish the existence of $(v, s, v(0), v(1)) \in \mathcal{D}(\mathcal{A})$ such that

$$(\mathcal{A} - \lambda_0 I)(v, s, v(0), v(1)) = (v', s', a', b').$$

This condition can be expressed

$$\begin{aligned}&\frac{\mathrm{d}v}{\mathrm{d}y}(y) = \lambda_0 s(y) + s'(y), \\ &c^2\frac{\mathrm{d}s}{\mathrm{d}y}(y) = \lambda_0 v(y) + v'(y), \\ &d_0 s(0) - \lambda_0 v(0) = a', \\ &-d_1 s(1) - \lambda_0 v(1) = b'.\end{aligned}$$

Write the components of the exponential matrix

$$E = \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22}\end{bmatrix} := \exp\begin{bmatrix} 0 & \lambda_0 \\ c^{-2}\lambda_0 & 0\end{bmatrix} = \begin{bmatrix} \cosh\left(\frac{\lambda_0}{c}\right) & c\sinh\left(\frac{\lambda_0}{c}\right) \\ \frac{1}{c}\sinh\left(\frac{\lambda_0}{c}\right) & \cosh\left(\frac{\lambda_0}{c}\right)\end{bmatrix}.$$

Then by consideration of the system of linear differential equations involved, we can show that the conditions have a solution, for some $(v, s, v(0), v(1)) \in \mathcal{D}(\mathcal{A})$ if the following 4×4 matrix is nonsingular:

$$M := \begin{bmatrix} e_{11} & e_{12} & -1 & 0 \\ e_{21} & e_{22} & 0 & -1 \\ -\lambda_0 & d_0 & 0 & 0 \\ 0 & 0 & -\lambda_0 & -d_1\end{bmatrix}.$$

A simple calculation gives

$$\det M = -\big(\lambda_0^2 + \frac{d_0 d_1}{c}\big)\sinh(\frac{\lambda_0}{c}) \neq 0.$$

for any $\lambda_0 > 0$, since $c, d_0, d_1 \geq 0$. We have confirmed the range property (for any real $\lambda_0 > 0$). □

Acknowledgement

The first author is a member of the Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM) and is supported by the GNAMPA project CUP_E55F22000270001.

References

[1] J. Zabczyk, On decomposition of generators. *SIAM J. Control Optim.* **16** (1978) 523–534.

[2] H.O. Fattorini, Infinite-dimensional Optimization and Control Theory. Cambridge University Press, Cambridge (1999).

[3] H.O. Fattorini, Optimal control problems with state constraints for semilinear distributed-parameter system. *J. Optim. Theory Appl.* **88** (1996) 25–59.

[4] H. Frankowska, E.M. Marchini, and M. Mazzola, Necessary optimality conditions for infinite dimensional state constrained control problems. *J. Differ. Equ.* **264** (2018) 7294–7327.

[5] R.F. Curtain and Y. Zhou, A weighted mixed-sensitivity H-infinity control design for irrational transfer matrices. *IEEE Trans. Automatic Control* **41** (1996) 1312–1321.

[6] H. Ozbay, M.C. Smith and A. Tannenbaum, Mixed sensitivity optimization for a class of unstable infinite dimensional systems. *Linear Algebra Appl.* **178** (1993) 43–83.

[7] E. Polak and D.Q. Mayne, First-order strong variation algorithms for optimal control problems with terminal inequality constraints. *J. Optim. Theory Appl.* **16** (1975) 303–325.

[8] F.H. Clarke, Y.S. Ledyaev, R.J. Stern and P.R. Wolenski, Nonsmooth Analysis and Control Theory. Springer-Verlag, New York (1998).

[9] R.B. Vinter, Optimal Control. Birkhäuser, Boston (2000).

[10] J.-P. Aubin and H. Frankowska, Set-Valued Analysis. Birkhäuser, Boston, Basel, Berlin (1990).

[11] G. Lumer and R.S. Phillips, Dissipative operators in a Banach space. *Pacific J. Math.* **11** (1961) 679–698.