Local Minimum Principle for an Optimal Control Problem with a Nonregular Mixed Constraint

A.V. Dmitruk*, N.P. Osmolovskii†

Abstract

We consider the simplest optimal control problem with one nonregular mixed constraint $G(x,u) \leqslant 0$, i.e. when the gradient $G_u(x,u)$ can vanish on the surface $G=0$. Using the Dubovitskii–Milyutin theorem on the approximate separation of convex cones, we prove a first order necessary condition for a weak minimum in the form of the so-called "local minimum principle", which is formulated in terms of functions of bounded variation, integrable functions, and Lebesgue-Stieltjes measures, and does not use functionals from $(L^\infty)^*$. Two illustrative examples are given. The work is based on the book by Milyutin [3].

Keywords: normed space, convex cone, dual cone, approximate separation theorem, mixed constraint, phase point, Pontryagin function, Lebesgue-Stieltjes measure, singular measure, costate equation.

1 Introduction

Consider the optimal control problem on a fixed interval of time $[t_0, t_1]$:

$$\mathcal{J}(x,u) := J(x(t_0), x(t_1)) \to \min, \tag{1}$$

$$\dot{x} = f(x,u), \tag{2}$$

$$G(x,u) \leqslant 0, \tag{3}$$

where the functions $J : \mathbb{R}^{2n} \to \mathbb{R}$, $f : \mathbb{R}^{n+m} \to \mathbb{R}^n$, and $G : \mathbb{R}^{n+m} \to \mathbb{R}$ are continuously differentiable. This problem will be called *Problem P*.

Condition (3) is called *mixed state-control constraint* or simply *mixed constraint*. The presence of this constraint determines the main difficulties in obtaining necessary optimality condition for this problem. These difficulties largely disappear if one assumes that the gradient $G_u(x,u)$ does not vanish at the points $(x,u) \in \mathbb{R}^{n+m}$ where $G(x,u) = 0$. In this case we say that the mixed constraint (3) is *regular*. Traditionally, the regularity assumption (properly modified for more general problems) is present in the works on necessary optimality conditions for problems with mixed constraints (see e.g. [7]–[16]). One of the few exceptions is the recent work [17], which will be discussed

*Russian Academy of Sciences, Central Economics and Mathematics Institute, Moscow, Russia; and Lomonosov Moscow State University, Moscow, Russia (optcon@mail.ru).

†Systems Research Institute, Polish Academy of Sciences, Warszawa, Poland (nikolai.osmolovskii@ibspan.waw.pl.).

later. Note that the regularity assumption for mixed constraint does not allow one to consider the *pure state constraint* $g(x) \leqslant 0$ as a special case of the mixed constraint. In this paper, we do not impose any assumptions on the mixed constraint (3), except for the smoothness condition for the function G.

A pair $(x,u) \in \mathbb{R}^{n+m}$ is called *phase point* (of the mixed constraint) if $G(x,u) = 0$ and also $G_u(x,u) = 0$. As mentioned above, it is the presence of such points, which creates the main difficulties in studying the problem and, in addition, gives rise to the main changes even in the formulation of the necessary optimality conditions compared to the regular case.

We consider Problem P for $x \in AC([t_0,t_1],\mathbb{R}^n)$ and $u \in L^\infty([t_0,t_1],\mathbb{R}^m)$, using the notation

$$w = (x,u) \in W \;=\; AC([t_0,t_1],\mathbb{R}^n) \times L^\infty([t_0,t_1],\mathbb{R}^m)$$

and $\xi = (x_0,x_1) = (x(t_0),x(t_1))$. The norm of a pair $w=(x,u)$ is

$$\|w\| = \;\|x\|_{AC} + \|u\|_\infty \;=\; |x(t_0)| + \int_{t_0}^{t_1} |\dot x(t)|\,\mathrm{d}t \,+\, \operatorname*{ess\,sup}_{t\in[t_0,t_1]} |u(t)|.$$

Obviously, the local minimum in this norm is equivalent to the standard *weak minimum*[1]. The goal of this paper is to obtain first-order necessary conditions for a weak minimum in problem (1)–(3) in the form of the so-called *local minimum principle* (LMP)[2].

As is known, an efficient method for obtaining LMP in constrained problems was proposed by Dubovitskii and Milyutin in [1]. The idea was simple (and therefore became very popular): at the minimum point, one should consider the convex cones of first order approximation for the cost and constraints, that should not intersect. Then the separation theorem is applied and the resulting Euler–Lagrange (stationarity) equation is analyzed. This leads to a LMP with multipliers from the spaces dual to the image spaces of the constraints.

However, a difficulty arises in this method: since the image space of the mixed constraints is L^∞, the separating functionals should belong to the conjugate space $(L^\infty)^*$, which has an essentially complex structure. In problems with *regular* mixed constraints, one can prove that the corresponding multipliers are represented by functions from L^1 (see [9, 15]). Unfortunately, this is not possible for problems with *nonregular* mixed constraints.

To overcome this difficulty, Dubovitskii and Milyutin [2] proposed the idea of not exact but *approximate* separation of the cones. For the case of two cones, it looks as follows. Let Y and X be normed spaces with $Y^* = X$, let $H_0,\, H_1 \subset Y$ and $\Omega_0,\, \Omega_1 \subset X$ be nonempty convex cones, Ω_1 open, such that $H_0^* = \overline{\Omega}_0$ and $H_1^* = \overline{\Omega}_1$, where the bar denotes the closure in the strong topology of X and the star denotes the dual (conjugate) cone.

Let $x_1^0 \in \Omega_1$ be a given point. Then the following is true: if $\Omega_0 \cap \Omega_1 = \varnothing$, then for any $\varepsilon > 0$ there exist $h_0 \in H_0$ and $h_1 \in H_1$ such that $\langle x_1^0, h_1\rangle = 1$ and $\|\, h_0 + h_1\| < \varepsilon$. The converse is also true.

[1]By definition, the latter is the minimum in the norm $||x||_C + ||u||_\infty$.

[2]Dubovitskii and Milyutin used the term *local maximum principle* [2]. Both these terms are not completely adequate; nevertheless, following the authors', we use the above term.

A similar result for a finite number of cones allowed Dubovitskii and Milyutin to obtain in [2] the LMP in a problem with a finite number of mixed constraints, given as inclusions to closed sets in $\mathbb{R}^{n+m}$. However, the book [2] is published in Russian in a small number of copies and is very difficult to read.

Many years later, Milyutin presented the same result in the book [3], where he considered a general problem with nonregular mixed constraints. This time these constraints are given by smooth functions, in the form of a finite number of inequalities like (1.3) and equalities $g(x,u) = 0$, assuming that the latter satisfy the full rank condition: $\operatorname{rank} g_u(x,u) = \dim g$ on the surface $g(x,u) = 0$, but without any assumptions on the joint independency of the derivatives $G_u(x,u)$, $g_u(x,u)$. The problem admits also a finite number of endpoint constraints of the form $F(x(t_0), x(t_1)) \leqslant 0$ and $K(x(t_0), x(t_1)) = 0$. Moreover, the smoothness assumption for the inequality constraints, both the endpoint and mixed ones, were essentially weakened to just the convexity of their directional derivatives at the reference point, while the equality constraints were always assumed to be smooth. The problem can also admit a pure control constraint of the inclusion type $u_2(t) \in U(t)$ on a part of control components, where the full control vector is split into two parts: $u = (u_1, u_2)$, and $U(t)$ is a measurable set-valued mapping. In this case, the full rank condition should be considered w.r.t. the first group only: $\operatorname{rank} g_{u_1}(x, u_1, u_2) = \dim g$, as well as the (non)regularity of all the collection of mixed constraints. For this general problem, Milyutin obtained a necessary condition for a weak minimum (the local maximum principle), and further developed it to a necessary condition for a strong minimum (the global maximum principle). A brief account of these results can be found in [18].

Compared to the book [2], the presentation of LMP in [3] is much clearer, with shorter proofs, but still difficult even for Russian-speaking readers. Moreover, these results have never been published in English. All this has led to the fact that the Dubovitskii–Milyutin's general theory of the maximum principle for nonregular mixed constraints, which in our opinion is an outstanding achievement in optimal control, still remains unknown even to specialists.

Because of the difficulties in the study, mentioned above, the nonregular mixed constraints until recently remained outside the scope of specialists' interests in the West. However, now this interest has arisen, as evidenced by the paper [17]. Without analyzing this publication, we will only say that the authors did not achieve the goal that could be set: to get rid of the functionals from $(L^\infty)^*$ in the final result, which are still present in [17], though in integral form. At the same time, the Dubovitskii–Milyutin's LMP is devoid of this drawback.

All this prompted us to write this article. To be as clear as possible in presentation of the specificity caused by the mixed constraints, we chose the simplest possible problem for the first study: it includes the Mayer cost functional, the control system, and just one mixed constraint. (A more general problem will be considered in our future paper.) In many ways, we follow the ideas of the book [3], and yet our presentation differs markedly from that book. We hope that this publication will draw attention of specialists to the ideas and results contained in [3].

The paper is organized as follows. In Section 2, we give definitions of the closure of a measurable set and a measurable function with respect to the measure, proposed by Dubovitskii and Milyutin, and recall some facts about equiintegrable sequences of functions in L^1, which are used in the proof of LMP. Section 3 is devoted to the approximate separation theorem for a finite number of convex cones, which plays a key role in the proof of LMP. We formulate LMP in Section 4 and give two illustrative examples in Section 5. The proof of LMP is given in Section 6.

2 Preliminaries

2.1 The closure with respect to a measure

We start with an important concept introduced by Dubovitskii and Milyutin in [2]. Let $M \subset \mathbb{R}$ be a (Lebesgue) measurable set. The set

$$\operatorname{clm} M \;=\; \{t \in \mathbb{R} : \;\operatorname{mes}(\omega \cap M) > 0 \quad \text{for any open set } \omega \ni t\}$$

is called *closure of M with respect to (the Lebesgue) measure.*

Obviously, $\operatorname{clm} M$ is a closed set. Moreover, $\operatorname{clm} M \subset \overline{M}$, but not the reverse. However, $\operatorname{mes} M \leqslant \operatorname{mes}(\operatorname{clm} M)$ (since almost all $t \in M$ are points of its density), but not the reverse.

In fact, $\operatorname{clm} M$ is the *topological support* of the measure $\mathrm{d}\mu$ that has density $\mathrm{d}\mu/\,\mathrm{d}t = \chi_M(t)$, where χ_M is the characteristic function of the set M.

Now, let $\hat{u} : [t_0, t_1] \to \mathbb{R}^m$ be a measurable function. Consider its graph $\Gamma = \{(t, \hat{u}(t)) : \; t \in [t_0, t_1]\}$ and the projector $\pi : \mathbb{R}^{1+m} \to \mathbb{R}, \quad (t, v) \mapsto t$. The set

$$\operatorname{clm}(\hat{u}) \;=\; \{(t, v) \in \mathbb{R}^{1+m} : \;\operatorname{mes} \pi(O \cap \Gamma) > 0 \quad \text{for any open set } O \ni (t, v)\}$$

is called *the closure of the function $\hat{u}$ with respect to (the Lebesgue) measure*, or in short, *the closure in measure of $\hat{u}$.* (Note that this definition can be applied in fact to any measurable set $\Gamma \subset \mathbb{R}^{1+m}$, and hence, to any measurable set-valued function.)

The following simple properties of $\operatorname{clm}(\hat{u})$ should be noted. By $B_r(u)$ we denote the closed ball in $\mathbb{R}^m$ of radius r centered at u, and by $O_\varepsilon(t, u)$ the open set $\{(t', u') : |t' - t| < \varepsilon, \; |u' - u| < \varepsilon\}$.

Lemma 1 *If $\hat{u} \in L^\infty([t_0, t_1], \mathbb{R}^m)$, then a) $\operatorname{clm}(\hat{u})$ is a compact set in $[t_0, t_1] \times \mathbb{R}^m$, which does not depend on the choice of a particular representative of the function $\hat{u}$, and b) the projector π is surjective on $\operatorname{clm}(\hat{u})$, i.e. $\pi \operatorname{clm}(\hat{u}) = [t_0, t_1]$.*

Proof. Obviously, the set $\operatorname{clm}(\hat{u})$ is closed and bounded, which proves the first assertion. To prove the second one, suppose the contrary, i.e. that $\exists\, t_* \notin \pi \operatorname{clm}(\hat{u})$. Set $r = \|\hat{u}\|_\infty$. Then, for any $u \in B_r(0)$ there is an $\varepsilon > 0$ such that $\operatorname{mes} \pi(O_\varepsilon(t_*, u) \cap \Gamma) = 0$. Since $B_r(0)$ is compact, there exists a finite number of $\varepsilon_i > 0$ and $u_i \in B_r(0)$, $i = 1, \ldots, k$ such that the union $\mathcal{C} := \bigcup_i O_{\varepsilon_i}(t_*, u_i)$ gives $\operatorname{mes} \pi(\mathcal{C} \cap \Gamma) = 0$. Define $\varepsilon_* = \min \varepsilon_i$ and $\omega = (t_* - \varepsilon_*, \, t_* + \varepsilon_*)$. Obviously, the set $Z = \omega \times B_r(0)$ is contained in $\mathcal{C}$, whence $\operatorname{mes} \pi(Z \cap \Gamma) = 0$. But the latter means that $|\hat{u}(t)| > r$ for a.a. $t \in \omega$, and then $\|\hat{u}\|_\infty > r$, a contradiction. $\square$

In fact, passing from $\hat{u}$ to $\operatorname{clm}(\hat{u})$, we obtain a set-valued mapping

$$\operatorname{clm}(\hat{u})(\cdot) : \;\; t \mapsto \; \{v \;:\; (t, v) \in \operatorname{clm}(\hat{u})\,\}$$

such that $\hat u(t) \in \mathrm{clm}\,(\hat u)(t)$ for almost all $t \in [t_0, t_1]$. Clearly, this mapping is upper semicontinuous.

Another way to define the closure in measure is $\mathrm{clm}\,(\hat u) := \bigcap_{u\sim\hat u} \overline{\mathrm{Graph}\,(u)}$, where the equivalence $u \sim \hat u$ means that $u(t) = \hat u(t)$ for almost all $t \in [t_0, t_1]$. One can easily show that

$$\mathrm{clm}\,(\hat u) \;=\; \bigcap_{\mathrm{mes}\ E = t_1 - t_0} \overline{\mathrm{Graph}\,(\hat u|_E)}\,,$$

where this time the intersection is taken over all measurable subsets $E \subset [t_0, t_1]$ of full measure, and by definition $\mathrm{Graph}\,(\hat u|_E) = \{(t,u) \in \mathbb{R}^{1+m}:\ t \in E,\ u = \hat u(t)\}$.

Note also that, if $\hat x(t)$ is a continuous function, then

$$\mathrm{clm}\,(\hat x, \hat u)(t) \;=\; (\hat x(t), \mathrm{clm}\,(\hat u)(t)) \quad \text{for all} \quad t \in [t_0, t_1]. \tag{4}$$

2.2 Uniformly integrable families of functions

A family $\mathcal{F}$ of functions from $L^1([t_0, t_1], \mathbb{R}^m)$ is called *uniformly integrable* (or *equi-integrable*) if for any $\varepsilon > 0$ there is a $\delta > 0$ such that for any measurable set $E \subset [t_0, t_1]$ of $\mathrm{mes}\ E < \delta$ we have $\int_E |\,\lambda(t)|\,\mathrm{d}t < \varepsilon, \quad \forall\, \lambda \in \mathcal{F}$.

Obviously, this is equivalent to the fact that the functions of $\mathcal{F}$ possess a common modulus of integrability, i.e. a function $\nu: \mathbb{R}_+ \to \mathbb{R}_+$ such that $\nu(\delta) \to 0$ as $\delta \to 0+$, and for any measurable set $E \subset [t_0, t_1]$ we have

$$\int_E |\,\lambda(t)|\,\mathrm{d}t \;\leqslant\; \nu(\mathrm{mes}\ E), \qquad \forall\, \lambda \in \mathcal{F}.$$

Since the functions of $L^1([t_0, t_1], \mathbb{R}^m)$ generate absolutely continuous vector-valued measures on $[t_0, t_1]$, a uniformly integrable family of functions generates a uniformly absolutely continuous family of vector-valued measures.

We will use these concepts in the case when $\mathcal{F}$ is a sequence of functions $\lambda^k \in L^1([t_0, t_1], \mathbb{R}^m)$, $k = 1, 2, \ldots$. By the Dunford–Pettis theorem [4, 5], any uniformly integrable sequence $\lambda^k \in L^1$ contains an L^∞-weakly convergent subsequence λ^{k_s}. The latter means that there exists a function $\lambda \in L^1$ such that, for any $u \in L^\infty$ we have

$$\int_{t_0}^{t_1} \langle \lambda^{k_s}, u\rangle\,\mathrm{d}t \;\to\; \int_{t_0}^{t_1} \langle \lambda, u\rangle\,\mathrm{d}t \qquad (s \to \infty).$$

We write in this case $\lambda^{k_s} \xrightarrow{w} \lambda \quad (s \to \infty)$.

In general, it is impossible to extract a weakly convergent sequence from an arbitrary bounded set of functions in L^1, since this set can be not uniformly integrable. Nevertheless, the following important fact holds true (see, e.g. [6] and references therein)[3].

Lemma 2 *(The biting lemma.) Let a sequence* $\lambda^k \in L^1([t_0, t_1], \mathbb{R})$ *be bounded, i.e.* $\|\lambda^k\|_1 \leqslant$ *const for all* $k = 1, 2, \ldots$. *Then there exists a sequence of measurable sets* $A^k \subset [t_0, t_1]$ *such that* $\mathrm{mes}\ A^k \to (t_1 - t_0)$ *and the sequence* $\lambda^k_A := \lambda^k \chi_{A^k}$ *is uniformly integrable, hence it contains a weakly convergent subsequence.*

[3] Dubovitskii and Milyutin, being not aware of these works, proved this fact independently in [2, 3].

2.3 Functions of bounded variation and charges

Denote by $\mathbb{R}^{n*}$ the space of row vectors of the dimension n, and by $BV([t_0,t_1],\mathbb{R}^{n*})$ the space of functions $p:[t_0,t_1]\to\mathbb{R}^{n*}$ of bounded variation for which the values $p(t_0-0)$ and $p(t_1+0)$ are also defined. By definition, *the jump* of p at a point $t\in[t_0,t_1]$ is $[p](t):=p(t+0)-p(t-0)$. We define the Radon measure (or charge) $\mathrm{d}p$, which corresponds to the function p, by the following condition: if $[a,b]\subset[t_0,t_1]$, then $\int_{[a,b]}\mathrm{d}p = p(b+0)-p(a-0)$. Note that we always prefer to denote Radon measures on $[t_0,t_1]$ by $\mathrm{d}p$, rather than p or $p(\mathrm{d}t)$, as is customary. This makes it possible to distinguish measures from the functions of bounded variation that define them, without introducing new notation.

If $l:C([t_0,t_1],\mathbb{R}^n)\to\mathbb{R}$ is a linear continuous functional, then by the Riesz theorem, there exists a function $p\in BV([t_0,t_1],\mathbb{R}^{n*})$ such that

$$\langle l,x\rangle=\int_{t_0}^{t_1}x(t)\,\mathrm{d}p\qquad\forall\,x\in C([t_0,t_1],\mathbb{R}^n),\tag{5}$$

but this function is not unique[4]. It is unique under the additional requirement that the function p vanishes at t_0 (or at t_1) and is one-way continuous, for example continuous from the left. For the definiteness, we will assume that the functions $p\in BV$ are left-continuous, i.e., $p(t-0)=p(t)$ for all $t\in[t_0,t_1]$, and $p(t_0)=0$. If p belongs to BV, we write $\mathrm{d}p\in C^*$, keeping in mind the relations (5).

If the function $p\in BV([t_0,t_1],\mathbb{R}^{n*})$ is absolutely continuous, then the measure $\mathrm{d}p$ and the functional l defined by (5) are also called *absolutely continuous*.

We say that $t_*\in[t_0,t_1]$ is a *point of continuity* of a measure $\mathrm{d}p$ if $[p](t_*)=0$, i.e., if the measure $\mathrm{d}p$ has no atom at this point. Recall that the measure $\mathrm{d}p$ can have atoms in at most countably many points. Therefore, the set of continuity points of $\mathrm{d}p$ is dense in $[t_0,t_1]$. A point $t^*\in[t_0,t_1]$ where the measure $\mathrm{d}p$ has an atom, i.e., $[p](t_*)>0$, is often called a *jump point* of the measure.

As usual, we say that a sequence of measures $\mathrm{d}p^k$ weakly* converges to a measure $\mathrm{d}p\in C^*$ (i.e. $C-$converges in C^*) if

$$\int_{t_0}^{t_1}x(t)\,\mathrm{d}p^k\;\to\;\int_{t_0}^{t_1}x(t)\,\mathrm{d}p\quad\text{ as }\quad k\to\infty$$

for all $x\in C([t_0,t_1],\mathbb{R}^n)$, and we write in this case $\mathrm{d}p^k\xrightarrow{*}\mathrm{d}p$.

Let λ^k be a sequence of functions in L^1. Consider the corresponding sequence of absolutely continuous measures $\mathrm{d}p^k:=\lambda^k\,\mathrm{d}t$. Assume that $\mathrm{d}p^k$ is weakly* convergent to some measure $\mathrm{d}p\in C^*$, that is $\lambda^k\,\mathrm{d}t\xrightarrow{*}\mathrm{d}p$. Denote by $\Theta\subset[t_0,t_1]$ the set of all continuity points of the measure $\mathrm{d}p$. Then for any $\tau_0,\tau_1\in\Theta$ with $\tau_0<\tau_1$, we have

$$\int_{[\tau_0,\tau_1]}\lambda^k(t)\,\mathrm{d}t\;\to\;\int_{[\tau_0,\tau_1]}\mathrm{d}p\quad(k\to\infty).$$

[4] Strictly speaking, we should write $\mathrm{d}p\,x$, but it is more convenient to write $x\,\mathrm{d}p$.

3 An approximate separation theorem

Let X and Y be normed spaces, such that $X = Y^*$. Let $\Omega \subset X$ be a nonempty convex cone and $\overline{\Omega}$ its closure. We say that a cone $H \subset Y$ is *thick on the cone* Ω (or is *predual to* Ω) if $H^* = \overline{\Omega}$. Here H^* denotes the conjugate cone of H, consisting of all linear continuous functionals that are nonnegative on H (in other words, H^* is the polar cone of $(-H)$). We will need the following properties of these cones.

Lemma 3 *Let $x^0 \in \operatorname{int}\Omega$. Then the set $Sec\, H = \{h : \langle x^0, h\rangle = 1\}$ is bounded, and its conical hull is $H \setminus \{0\}$.*

Proof. Suppose $\exists\, h_k \in Sec\, H$ with $\|h_k\| = r_k \to \infty$. Setting $\widetilde{h}_k = h_k/r_k$ we have $\|\widetilde{h}_k\| = 1$ and $\langle x^0, \widetilde{h}_k\rangle \to 0$. Let $x^0 + B_\delta \subset \Omega$ for some $\delta > 0$, where B_δ is the closed ball in X of radius δ, centered at zero. Then $\langle x^0 + B_\delta, \widetilde{h}_k\rangle \geqslant 0$, whence $\langle B_\delta, \widetilde{h}_k\rangle \geqslant -o(1)$. But here the infimum of the left hand side equals $-\delta$, a contradiction.

Thus, $Sec\, H$ is bounded. To prove the second assertion, take any nonzero $h \in H$. Since $H^* = \overline{\Omega}$, we have $\langle \Omega, h\rangle \geqslant 0$, and since $x^0 \in \operatorname{int}\Omega$, we have $\alpha := \langle x^0, h\rangle > 0$, whence $h/\alpha \in \mathrm{Sec}\, H$, q.e.d. □

Now, let be given two convex cones $H_0\,,\, H_1 \subset Y$ and two convex cones $\Omega_0\,,\, \Omega_1 \subset X$, such that $H_0^* = \overline{\Omega}_0$ and $H_1^* = \overline{\Omega}_1\,,$ where again $X = Y^*$. The following theorem is an approximate analog of the Hahn–Banach separation theorem for the case of two convex cones, in which the separating functionals are taken not from the dual but from the predual space.

Theorem 1 *Let Ω_1 be open and $x_1^0 \in \Omega_1$. Then $\Omega_0 \cap \Omega_1 = \varnothing \iff \forall\, \varepsilon > 0$ $\exists\, (h_0, h_1) \in H_0 \times H_1$ such that $\langle x_1^0, h_1\rangle = 1$ and $\|h_0 + h_1\| < \varepsilon$.*

Proof. ($\Longleftarrow$) Suppose, on the contrary, that $\exists\, \hat{x} \in \Omega_0 \cap \Omega_1$. Without loss of generality assume that $\|x_1^0\| = \|\hat{x}\| = 1$ and $B_r(\hat{x}) \subset \Omega_1$ for some $r > 0$, where $B_r(\hat{x})$ is the closed ball in X of radius r centered at $\hat{x}$. Set $\varepsilon = r/2$ and take any pair (h_0, h_1) with the above properties. They imply $\|h_1\| \geqslant 1$. Set $y = h_0 + h_1$. Then $\|y\| \leqslant \varepsilon$ and $h_0 + h_1 - y = 0$, whence

$$\langle \hat{x}, h_0\rangle + \langle \hat{x},\, h_1 - y\rangle = 0. \tag{6}$$

The first summand here is nonnegative. Now, the inequality $\langle \hat{x} - B_r, h_1\rangle \geqslant 0$ implies that $\langle \hat{x}, h_1\rangle \geqslant \sup\langle B_r, h_1\rangle \geqslant r$, and since $|\langle \hat{x}, y\rangle| \leqslant \varepsilon$, the second summand in (6) can be estimated as $\langle \hat{x},\, h_1 - y\rangle \geqslant r - \varepsilon = r/2 > 0$, so the sum in (6) cannot be zero, a contradiction.

($\Longrightarrow$) We have to show that $\inf \|H_0 + Sec\, H_1\| = 0$. Suppose the contrary: this $\inf > r > 0$, i.e. the distance from the set $Sec\, H_1$ to the cone $-H_0$ is greater than r. Therefore, $(-H_0) \cap (Sec\, H_1 + B_r) = \varnothing$. Then, by the classical separation theorem, $\exists\, \hat{x} \in X$, $\|\hat{x}\| = 1$, such that $\langle \hat{x}, -H_0\rangle \leqslant 0$ and $\langle \hat{x},\, Sec\, H_1 + B_r\rangle \geqslant 0$. The first relation implies $\hat{x} \in H_0^* = \overline{\Omega}_0\,,$ and the second one $\langle \hat{x},\, Sec\, H_1\rangle \geqslant \sup\langle \hat{x}, B_r\rangle = r$, i.e. $\forall\, h_1 \in Sec\, H_1$ we have $\langle \hat{x}, h_1\rangle \geqslant r$. By Lemma 3 $\|Sec\, H_1\| \leqslant d$ for some $d > 0$. Take any positive $\varepsilon < r/d$. Then it follows that for any $h_1 \in Sec\, H_1$

$$\langle \hat{x} + B_\varepsilon\,, h_1\rangle \;\geqslant\; r - \sup\,\langle B_\varepsilon\,, h_1\rangle \;\geqslant\; r - \varepsilon d > 0,$$

and since the conical hull of $Sec\, H_1$ is $H_1 \setminus \{0\}$, we get $\hat{x} + B_\varepsilon \subset H_1^* = \overline{\Omega}_1$, so $\hat{x} \in \Omega_1$. Thus, $\hat{x} \in \overline{\Omega}_0 \cap \Omega_1$, and since Ω_1 is open, there exists an element $x' \in \Omega_0 \cap \Omega_1$, a contradiction. $\square$

Note that in the proof of implication $\Longrightarrow$, instead of separating the cones Ω_0 and Ω_1 by an element of X^*, we use the classical theorem to separate the cone H_0 and an extension of the cone H_1 by an element of X.

The general case. Now, let be given a finite number of convex cones $\Omega_0, \Omega_1, \ldots, \Omega_m$ in X, among which the last m are open, and convex cones $H_0, H_1, \ldots, H_m$ in Y such that $H_i^* = \overline{\Omega}_i$ for all $i = 0, 1, \ldots, m$ (i.e. each H_i is thick on Ω_i). As before, $X = Y^*$. Let be also given elements $x_i^0 \in \Omega_i$, $i = 1, \ldots, m$, of the open cones. The following theorem is an approximate analog of the Dubovitskii–Milyutin "multi-separation" theorem for convex cones[5] (see [1, Theorem 2.1]).

Theorem 2 $\Omega_0 \cap \Omega_1 \cap \ldots \cap \Omega_m = \emptyset \iff \forall \varepsilon > 0 \;\; \exists\, h_i \in H_i$, $i = 0, 1, \ldots, m$, *such that*

$$\sum_{i=1}^{m} \langle\, x_i^0, h_i \rangle = 1 \qquad \text{and} \qquad \|\, h_0 + \sum_{i=1}^{m} h_i \| < \varepsilon. \tag{7}$$

The first of these conditions can be regarded as a *normalization condition*, while the second one is an *approximate Euler–Lagrange equation.* Note that the cone Ω_0 does not appear in the first condition, it appears only in the second one.

Proof. ($\Longleftarrow$) Without loss of generality assume that $\sum_1^m \|x_i^0\| = 1$. We have to show that all Ω_i do not intersect. Suppose, on the contrary, $\exists\, \hat{x} \in \bigcap_{i=0}^{n} \Omega_i$. Without loss of generality assume that $\|\hat{x}\| = 1$, and let $r > 0$ be such that $B_r(\hat{x}) \subset \Omega_i$ for all $i \geqslant 1$.

Set $\varepsilon = r/2$ and take any collection $(h_0, h_1, \ldots, h_m)$ of elements in $H_0, H_1, \ldots, H_m$, respectively, satisfying (7). The first of these conditions together with the relation $\sum_1^m \|x_i^0\| = 1$ imply that $\max\limits_{1 \leqslant i \leqslant m} \|h_i\| \geqslant 1$. Let, for definiteness, $\|h_m\| \geqslant 1$.

Set $y = \sum_0^m h_i$. Then $\|y\| \leqslant \varepsilon$ and $\sum_0^{m-1} h_i + h_m - y = 0$, whence

$$\sum_{i=0}^{m-1} \langle\, \hat{x}, h_i \rangle \;+\; \langle \hat{x},\, h_m - y \rangle \;=\; 0. \tag{8}$$

The first m summands here are nonnegative. Now, the inequality $\langle \hat{x} - B_r, h_m \rangle \geqslant 0$ implies that $\langle \hat{x}, h_m \rangle \geqslant \sup \langle B_r, h_m \rangle \geqslant r$, and since $|\langle \hat{x}, y \rangle| \leqslant \varepsilon$, the last summand in (8) can be estimated as $\langle \hat{x},\, h_m - y \rangle \geqslant r - \varepsilon = r/2 > 0$, so the left hand side in (8) cannot be zero, a contradiction.

($\Longrightarrow$) We prove by induction. Suppose the theorem holds for all $m' < m$ open cones and consider the case of m open cones.

If $\Omega_0 \cap \Omega_m = \emptyset$, then by Theorem 1 $\forall \varepsilon > 0 \;\; \exists\, h_0 \in H_0$ and $h_m \in H_m$ such that $\langle x_m^0, h_m \rangle = 1$ and $\|h_0 + h_m\| < \varepsilon$. Choosing all h_i for $i = 1, \ldots, m-1$ to be arbitrary sufficiently small elements of H_i, we get $\sum_1^m \langle x_i^0, h_i \rangle \geqslant 1$ and $\|h_0 + \sum_1^m h_i\| < 2\varepsilon$. Obviously, this implies the statement of Theorem.

[5] In [1], the condition for separating the cones was called the Euler–Lagrange equation.

Now, suppose that $W_0 := \Omega_0 \cap \Omega_m \neq \emptyset$. Set $K_0 = H_0 + H_m$ and notice that in this case $K_0^* = \overline{\Omega}_0 \cap \overline{\Omega}_m = \overline{\Omega_0 \cap \Omega_m} = \overline{W}_0$, that is K_0 is thick on W_0. (The second equality holds because both Ω_0 and Ω_m are convex and the last one is open.) Consider the cones $K_0, H_1, \ldots, H_{m-1} \subset Y$ and the corresponding cones $W_0, \Omega_1, \ldots, \Omega_{m-1} \subset X$, where the last collection does not intersect. The dual cones to the first ones are equal to the closure of the last ones, so we have the situation of Theorem 2 for $m-1$ open cones. By the premise of induction, $\forall \varepsilon > 0 \;\; \exists k_0, h_1, \ldots, h_{m-1}$ from the cones $K_0, H_1, \ldots, H_{m-1}$, respectively, such that

$$\sum_1^{m-1} \langle x_i^0, h_i \rangle = 1 \quad \text{and} \quad \| k_0 + \sum_1^{m-1} h_i \| < \varepsilon.$$

Setting here $k_0 = h_0 + h_m$ with some $h_0 \in H_0$ and $h_m \in H_m$, we obtain $\sum_1^{m-1} \langle x_i^0, h_i \rangle + \langle x_m^0, h_m \rangle \geqslant 1$ and still $\|h_0 + \sum_1^{m-1} h_i + h_m\| < \varepsilon$. Multiplying the obtained collection by some $\lambda \leqslant 1$ we get the required. $\square$

4 Local minimum principle

Consider the set

$$\mathcal{N}(G) \;:=\; \{(x,u) \in \mathbb{R}^{n+m} : \quad G(x,u) = 0, \quad G_u(x,u) = 0\}.$$

Clearly, $\mathcal{N}(G)$ is closed. It is called the *set of phase points*. We assume that this set is nonempty (otherwise the mixed constraint is regular).

Define the following set-valued mapping $(x,u) \in \mathbb{R}^{n+m} \rightrightarrows S(x,u) \subset \mathbb{R}^{n*}$:

(i) if $(x,u) \in \mathcal{N}(G)$, then $S(x,u) = \{G_x(x,u)\}$,

(ii) if $(x,u) \notin \mathcal{N}(G)$, then $S(x,u) = \emptyset$.

Thus, $S(x,u)$ is a singleton $\{G_x(x,u)\}$ or an emptyset.

For any nonempty set $M \subset \mathbb{R}^{n+m}$ we define $S(M) = \bigcup_{(x,u)\in M} S(x,u)$.

Let $\hat{w} = (\hat{x}, \hat{u}) \in W$ be a given admissible process in Problem P investigated for optimality. Denote for short $\hat{\xi} = (\hat{x}(t_0), \hat{x}(t_1))$. Let us formulate the conditions of local minimum principle for the process $\hat{w}$.

Recall that for the function $\hat{u}$ we introduced (in Sec. 2) the set-valued mapping $\operatorname{clm}(\hat{u})(t) = \{u \in \mathbb{R}^m : (t,u) \in \operatorname{clm}(\hat{u})\}$, and recall also that $\big(\hat{x}(t), \operatorname{clm}(\hat{u})(t)\big) = \operatorname{clm}(\hat{w})(t)$ for all $t \in [t_0, t_1]$. Define a set

$$\mathcal{D} \;:=\; \{t \in [t_0, t_1] : \quad \operatorname{clm}(\hat{w})(t) \,\cap\, \mathcal{N}(G) \neq \emptyset\}. \tag{9}$$

Since the set $\operatorname{clm}(\hat{w})$ is compact and $\mathcal{N}(G)$ is closed, $\mathcal{D}$ is a closed (possibly empty) subset in $[t_0, t_1]$. Denote by $\chi_{\mathcal{D}}$ its characteristic function.

For any $t \in \mathcal{D}$ consider the set $\operatorname{conv} S(\operatorname{clm}(\hat{w})(t)) = \operatorname{conv} S\big(\hat{x}(t), \operatorname{clm}(\hat{u})(t)\big)$, where conv stands for the convex hull. We call it the *set of possible directions of jumps* of the adjoint variable at the point t.

For any nonempty set $M \subset \mathbb{R}^{n+m}$ we define $G_x(M) = \bigcup_{(x,u)\in M} G_x(x,u)$. It follows from the definitions that for any $t \in \mathcal{D}$ we have

$$S(\mathrm{clm}\,(\hat{w})(t)) = G_x\big(\mathrm{clm}\,(\hat{w})(t) \cap \mathcal{N}(G)\big) \neq \emptyset.$$

Now, define *the Pontryagin function* $H(x,u,p) = p\,f(x,u)$, where $p \in \mathbb{R}^{n*}$ is a costate (adjoint) row-vector.

The conditions of *local minimum principle* (LMP) at the point $\hat{w}$ are as follows: there exist multipliers

$$\hat{\alpha}_0 \in \mathbb{R}, \quad \hat{p} \in BV([t_0,t_1],\mathbb{R}^{n*}), \quad \hat{\lambda} \in L^1([t_0,t_1],\mathbb{R}), \quad \mathrm{d}\hat{\eta} \in C^*([t_0,t_1],\mathbb{R}), \tag{10}$$

such that

$$\hat{\alpha}_0 \geqslant 0, \quad \hat{\lambda} \geqslant 0, \quad \hat{\lambda}\,G(\hat{w}) = 0, \quad \mathrm{d}\hat{\eta} \geqslant 0, \quad \mathrm{d}\hat{\eta}\cdot\chi_{\mathcal{D}} = \mathrm{d}\hat{\eta}, \tag{11}$$

$$\hat{\alpha}_0 + \|\hat{\lambda}\|_1 + \int_{[t_0,t_1]} \mathrm{d}\hat{\eta} > 0, \tag{12}$$

and a $\mathrm{d}\hat{\eta}$-measurable essentially bounded function $\hat{s} : [t_0,t_1] \to \mathbb{R}^{n*}$ such that

$$\hat{s}(t) \in \mathrm{conv}\, S\big(\hat{x}(t), \mathrm{clm}\,(\hat{u})(t)\big) \quad \text{for almost all } t \text{ in } d\hat{\eta} - \text{measure}, \tag{13}$$

moreover, the following adjoint equation in terms of measures:

$$-\,\mathrm{d}\hat{p} = H_x(\hat{w},\hat{p})\,\mathrm{d}t + \hat{\lambda}\,G_x(\hat{w})\,\mathrm{d}t + \hat{s}\,\mathrm{d}\hat{\eta}, \tag{14}$$

and the transversality conditions:

$$-\,\hat{p}(t_0-) = \hat{\alpha}_0 J_{x_0}(\hat{\xi}), \qquad \hat{p}(t_1+) = \hat{\alpha}_0 J_{x_1}(\hat{\xi}) \tag{15}$$

are fulfilled, and finally, the stationarity condition with respect to the control is satisfied:

$$H_u(\hat{w},\hat{p}) + \hat{\lambda}\,G_u(\hat{w}) = 0. \tag{16}$$

The last equation means that $H_u(\hat{w}(t),\hat{p}(t)) + \hat{\lambda}(t)\,G_u(\hat{w}(t)) = 0,\quad$ a.e. in $[t_0,t_1]$, where "a.e." means "almost everywhere with respect to the Lebesgue measure".

Condition (13) means that there exists a set $\mathcal{R} \subset \mathcal{D}$ of full $\mathrm{d}\hat{\eta}$-measure (i.e., $\int_{\mathcal{R}} \mathrm{d}\hat{\eta} = \int_{[t_0,t_1]} \mathrm{d}\hat{\eta}$) such that the inclusion in (13) holds for all $t \in \mathcal{R}$.

The values $\hat{s}(t)$ for $t \notin \mathcal{D}$ are of no importance.

Note that conditions (10)–(16) differ from that for problems with *regular* mixed constraints only by the presence of the term $\hat{s}\,\mathrm{d}\hat{\eta}$ in the adjoint equation (14). If $\mathcal{D} = \emptyset$, this term vanishes.

The adjoint equation can be understood in the integral form: for almost all t

$$\hat{p}(t) = \hat{p}(t_0 - 0) + \int_{t_0}^{t} \big(H_x(\hat{w},\hat{p}) + \hat{\lambda}\,G_x(\hat{w})\big)\,\mathrm{d}\tau + \int_{t_0-0}^{t+0} \hat{s}(\tau)\,\mathrm{d}\hat{\eta}(\tau).$$

(The last integral is taken over the interval $[t_0,t]$ including its endpoints.)

Remark. It is convenient to introduce the so-called *augmented Pontryagin function* $\overline{H}(x,u,p,\lambda) = p\,f(x,u) + \lambda\,G(x,u)$, whence the costate equation (14) and the stationarity condition in the control (16) take the following shorter form, respectively:

$$-\,\mathrm{d}\hat p \;=\; \overline{H}_x(\hat w,\hat p,\hat\lambda)\,\mathrm{d}t \;+\; \hat s\,\mathrm{d}\hat\eta, \tag{17}$$

$$\overline{H}_u(\hat w,\hat p,\hat\lambda) \;=\; 0. \tag{18}$$

Theorem 3 *If $\hat w$ is a weak local minimum in Problem P, then it satisfies the local minimum principle (10)–(16).*

Some particular cases.

1. Let t_* be an isolated point in $\mathcal{D}$ and the function $\hat u$ be continuous at t_*. Then $\operatorname{clm}(\hat u)(t_*) = \hat u(t_*)$ and $s(t_*) = G_x(\hat w(t_*))$, so the measure $\mathrm{d}\hat\eta$ can have an atom at this point: $\mathrm{d}\hat\eta(\{t_*\}) > 0$, and the costate variable have the jump $[p](t_*) = -G_x(\hat w(t_*))\,\mathrm{d}\hat\eta(\{t_*\})$.

If the control $\hat u$ has a discontinuity of the first kind at t_*, then $(\operatorname{clm}\hat u)(t_*)$ consists of two points: $\hat u(t_*-0)$ and $\hat u(t_*+0)$. If both the corresponding points $(\hat x(t_*),\hat u(t_*-0))$ and $(\hat x(t_*),\hat u(t_*+0))$ belong to $\mathcal{N}(G)$, then $s(t_*) = s_0\,G_x(\hat x(t_*),\hat u(t_*-0)) + s_1\,G_x(\hat x(t_*),\hat u(t_*+0))$, where $s_0 \geqslant 0$, $s_1 \geqslant 0$, $s_0+s_1 = 1$, and the costate variable has the jump $[p](t_*) = -s(t_*)\,\mathrm{d}\hat\eta(\{t_*\})$.

2. Consider the case when the function G does not depend on u, i.e. $G(x,u) = g(x)$. Then the mixed constraint (3) reduces to *a pure state constraint* $g(x) \leqslant 0$. In this case $\mathcal{N}(g) = \{(x,u):\; g(x) = 0\}$, i.e., each point on the boundary of the state constraint[6] is a phase point, the set $\mathcal{D} = \{t:\, g(\hat x(t)) = 0\}$ consists of contact points, and by setting $\mathcal{R} = \mathcal{D}$, we get $s(t) = g'(\hat x(t))$ at any point $t\in\mathcal{D}$.

Consequently, the formulation of LMP in this case is as follows: there exist multipliers $\hat\alpha_0 \in \mathbb{R}$, $\hat p \in BV([t_0,t_1],\mathbb{R}^{n*})$, $\hat\lambda \in L^1([t_0,t_1],\mathbb{R})$, and $\mathrm{d}\hat\eta \in C^*([t_0,t_1],\mathbb{R})$ such that

$$\begin{aligned}
&\hat\alpha_0 \geqslant 0, \quad \hat\lambda \geqslant 0, \quad \hat\lambda\, g(\hat x) = 0, \quad \mathrm{d}\hat\eta \geqslant 0, \quad g(\hat x)\,\mathrm{d}\hat\eta = 0,\\
&\hat\alpha_0 \;+\; \|\hat\lambda\|_1 \;+ \textstyle\int_{[t_0,t_1]} \mathrm{d}\hat\eta \;>\; 0,\\
&-\,\mathrm{d}\hat p \;=\; H_x(\hat w,\hat p)\,\mathrm{d}t + g'(\hat x)\big(\hat\lambda\,\mathrm{d}t + \mathrm{d}\hat\eta\,\big),\\
&-\hat p(t_0-) = \hat\alpha_0 J_{x_0}(\hat\xi), \quad \hat p(t_1+) = \hat\alpha_0 J_{x_1}(\hat\xi),\\
&H_u(\hat w,\hat p) \;=\; 0,
\end{aligned}$$

where $H(x,u,p) = pf(x,u)$. Setting $\hat\lambda\,\mathrm{d}t + \mathrm{d}\hat\eta =: \mathrm{d}\hat\mu$, we get

$$\begin{aligned}
&\hat\alpha_0 \geqslant 0, \quad \mathrm{d}\hat\mu \in C^*([t_0,t_1],\mathbb{R}), \quad \mathrm{d}\hat\mu \geqslant 0, \quad g(\hat x)\,\mathrm{d}\hat\mu = 0, \quad \hat\alpha_0 + \textstyle\int_{[t_0,t_1]} \mathrm{d}\hat\mu > 0,\\
&-\,\mathrm{d}\hat p \;=\; H_x(\hat w,\hat p)\,\mathrm{d}t \;+\; g'(\hat x)\,\mathrm{d}\hat\mu.
\end{aligned}$$

The transversality conditions and the condition $H_u(\hat w,\hat p) = 0$ do not change. Thus we obtain the well-known conditions of LMP for the problem with a state constraint.

[6] To be precise, this is indeed the boundary if $g'(x) \neq 0$ on it.

5 Examples

Example 1: the measure $\mathrm{d}\hat{\eta}$ has atoms

Let $[t_0, t_1]$ be a fixed interval, $t_0 < t_1$, $x \in \mathbb{R}$, $u \in \mathbb{R}$. Consider the problem

$$J := x(t_0)\, x(t_1) \to \min, \qquad \dot{x} = u, \qquad G := \frac{1}{2}\, u^2 - x + 1 \leqslant 0.$$

Conditions $G = 0$, $G_u = 0$ select here the only phase point $(x, u) = (1, 0)$. Since $(x-1) \geqslant \frac{1}{2}\, u^2$, we always have $x \geqslant 1$, hence $\inf J \geqslant 1$. Then, the process $\hat{x}(t) \equiv 1$, $\hat{u}(t) \equiv 0$ is a solution to the problem. Therefore, $\mathcal{D} = [t_0, t_1]$.

Further, we have (removing the hats over the multipliers): $H = p\, u$, $\overline{H} = p\, u + \lambda\, G$, $\overline{H}_u = p + \lambda\, u$, $\overline{H}_x = -\lambda$, $s = G_x = -1$. The condition $\overline{H}_u = 0$ gives $p + \lambda \hat{u} = 0$, whence $p(t) = 0$ for all $t \in (t_0, t_1)$, and therefore $p(t_0+) = p(t_1-) = 0$. The transversality conditions give $p(t_0-) = -\alpha_0\, x(t_1) = -\alpha_0$, $p(t_1+) = \alpha_0\, x(t_0) = \alpha_0$, so the jumps of p at the endpoints are: $[p](t_0) := p(t_0+) - p(t_0-) = \alpha_0$ and $[p](t_1) := p(t_1+) - p(t_1-) = \alpha_0$.

The adjoint equation $-\,\mathrm{d}p = \overline{H}_x\, \mathrm{d}t + s\, \mathrm{d}\eta$ reduces to $\mathrm{d}p = \lambda\, \mathrm{d}t + \mathrm{d}\eta$, $\lambda \geqslant 0$, $\mathrm{d}\eta \geqslant 0$. Since $p(t) = 0$ for $t \in (t_0, t_1)$, we have $\lambda = 0$ and $\mathrm{d}p = \mathrm{d}\eta$.

If $\alpha_0 = 0$, then $\mathrm{d}p = \mathrm{d}\eta = 0$, which contradicts the nontriviality condition (12). Therefore, we can set $\alpha_0 = 1$. Then the measure $\mathrm{d}p$ is the sum of δ-functions at t_0 and t_1, respectively, and the same is true for $\mathrm{d}\eta$.

Example 2: the measure $\mathrm{d}\hat{\eta}$ is absolutely continuous

Fix any $T > 0$ and consider the problem on the interval $[-T, T]$:

$$\begin{aligned} &\dot{y} = x, \quad \dot{x} = u, \qquad G(y, x, u) = \frac{1}{2}\, u^2 - x \leqslant 0, \\ &J \,=\, y(T) - y(-T) \,-\, \frac{m}{2}\Big(x(-T) + x(T)\Big) \,\to\, \min, \end{aligned} \tag{19}$$

where $m \in (0, T)$ is a given number.

Here $y(T) - y(-T) = \int_{-T}^{T} x\, dt$, so the variable y is in fact inessential. The set $\mathcal{N}(G) = \{(x, u) :\ G = 0,\ G_u = 0\}$ consists of the only point $(x, u) = (0, 0)$, and since $(G_y, G_x) = (0, -1)$, the direction of possible jumps of the costate vector $p = (p_y, p_x)$ is $s = (s_y, s_x) = (0, -1)$, where the subscripts indicate coordinates, not partial derivatives.

Set $b = T - m$, and consider the following process:

$\hat{x}(t) = \hat{u}(t) = 0$ on the interval $[-b, b]$,

$\hat{x}(t) = \frac{1}{2}\, (t-b)^2$ and $\hat{u}(t) = t - b$ on $[b, T]$,

$\hat{x}(t) = \frac{1}{2}\, (t+b)^2$ and $\hat{u}(t) = t + b$ on $[-T, -b]$.

Obviously, this process is admissible. Let us show that it is globally optimal in the problem. To do this, choose any value $h \in [0, T]$, fix the endpoints $x(-T) = x(T) = \frac{1}{2}\, h^2$, and find the minimum of $\int_{-T}^{T} x\, dt$ under the given mixed constraint $G \leqslant 0$, i.e.

$|u| \leqslant \sqrt{2\,x}$. Clearly, this minimum is attained at the lowest possible curve, i.e. the one satisfying $\dot x = \sqrt{2\,x}$ on $[0,T]$, $x(T) = h$, and symmetrically on $[-T,0]$. Therefore,

$$\begin{aligned} x(t) &= 0 \quad \text{on the interval } [-T+h,\, T-h], \\ x(t) &= \tfrac{1}{2}\,(t-T+h)^2 \quad \text{on } [T-h,\, T], \\ x(t) &= \tfrac{1}{2}\,(-t+T-h)^2 \quad \text{on } [-T,\, -T+h]. \end{aligned} \tag{20}$$

Then $J(h) = \frac{1}{3}h^3 - \frac{m}{2}\,h^2$, and we have to find the minimum of this function over $h \in [0,T]$. The equation $J'(h) = h^2 - mh = 0$ has the only positive solution $h = m$, and since $J'(h) < 0$ for $h < m$, and $J'(h) > 0$ for $h > m$, we conclude that $J(h)$ has a global minimum over $h \in [0,T]$ at $h = m$. Clearly, no $h > T$ can give a smaller cost value, so $h = m$ provides the global minimum of $J(h)$, and the corresponding curve (20), coinciding with $\hat x(t)$, provides the global minimum in problem (19).

Let us check the LMP for this curve. According to Theorem 3, there exist $\alpha_0 \geqslant 0$, $\lambda \in L^1$, $\lambda(t) \geqslant 0$ a.e. on $[-T,T]$, a measure $\mathrm{d}\eta \in C^*$ supported on $\mathcal{D} = [-b,b]$, the function $s = (s_y, s_x) = (0,-1)$ a.e. in $[-b,b]$ w.r.t. $\mathrm{d}\eta$, and the function $p = (p_y, p_x) \in BV$, such that this collection is nontrivial:

$$\alpha_0 \,+\, \|\lambda\|_1 \,+ \int_{[t_0,t_1]} \mathrm{d}\eta \,>0, \tag{21}$$

generates the augmented Pontryagin function $\overline{H} \,=\, p_y x \,+\, p_x u \,+\, \lambda\,(\frac{1}{2}\,u^2 - x)$, and satisfies the conditions (14)–(16).

The condition $\overline{H}_u = 0$ gives $p_x + \lambda u = 0$. Since $\overline{H}_y = 0$ and $s_y = G_y = 0$, the adjoint equation $-\,\mathrm{d}p_y = \overline{H}_y\,\mathrm{d}t + s_y\,\mathrm{d}\eta$ reduces to $\mathrm{d}p_y = 0$, whence $p_y = \text{const}$. The transversality conditions for p_y are: $p_y(-T) = -\alpha_0 J_{y(-T)} = \alpha_0$ and $p_y(T) = \alpha_0 J_{y(T)} = \alpha_0$. Consequently, $p_y = \alpha_0$.

Since $H_x = p_y = \alpha_0$ and $s_x = G_x = -1$ (a.e. in $[-b,b]$ w.r.t. $\mathrm{d}\eta$), the adjoint equation $-\,\mathrm{d}p_x = \overline{H}_x\,\mathrm{d}t + s_x\,\mathrm{d}\eta$ has the form

$$-\,\mathrm{d}p_x \,=\, \alpha_0\,\mathrm{d}t - \lambda\,\mathrm{d}t - \mathrm{d}\eta. \tag{22}$$

The transversality conditions for p_x are:

$$p_x(-T) = -\alpha_0 J_{x(-T)} = \alpha_0\, m/2\,, \qquad p_x(T) = \alpha_0 J_{x(T)} = -\alpha_0\, m/2\,. \tag{23}$$

If $\alpha_0 = 0$, then $p_y \equiv 0$, and (22) with (23) reduce to $p_x = \lambda\,\mathrm{d}t + \mathrm{d}\eta \geqslant 0$ and $p_x(-T) = p_x(T) = 0$, whence $\lambda = 0$ and $\mathrm{d}\eta = 0$, which contradicts the nontriviality (21). Thus, we can set $\alpha_0 = 1$, and also $p_y = 1$. Then

$$\mathrm{d}p_x \,=\, (\lambda - 1)\;\mathrm{d}t + \mathrm{d}\eta, \qquad p_x(-T) = m/2\,, \quad p_x(T) = -m/2\,.$$

Since $u = 0$ on $D = [-b,b]$ and $p_x = -\lambda u$, we get $p_x = 0$ and $\lambda\,\mathrm{d}t + \mathrm{d}\eta = \mathrm{d}t$ there. So, $\mathrm{d}\eta$ is absolutely continuous on D and is not unique: both λ and $\dot\eta$ are just nonnegative and bounded by the relation $\lambda(t) + \dot\eta(t) = 1$.

Consider the interval $(b,T]$. We have there $\mathrm{d}\eta = 0$,

$$x \,=\, (t-b)^2/2, \quad u = \dot x \,=\, t-b, \quad p_x \,=\, -\lambda\, u \,=\, -\lambda(t-b), \quad \dot p_x = \lambda - 1,$$

whence $\dot\lambda(t-b) = 1-2\lambda$. Setting $\sigma = \lambda - 1/2$ and $\tau = t-b$, we get $\dot\sigma\,\tau = -2\sigma$, which easily gives $\sigma = c/\tau^2$, and so, $\lambda = \frac{1}{2} + c/\tau^2$ with some constant c. Then $c=0$ and $\lambda = 1/2$ (otherwise $\lambda \notin L^1$), whence $p_x = -\frac{1}{2}(t-b) < 0$ and $p_x(b+0) = 0$, so the jumps $[p_x](b) = [\eta](b) = 0$.

The symmetric picture is on the interval $[-T,-b)$. Here $\lambda = 1/2$, $p_x = -\frac{1}{2}(t+b) > 0$ and $p_x(-b-0) = 0$, so the jumps $[p_x](-b) = [\eta](-b) = 0$.

6 Proof of LMP

In this section we prove Theorem 3. We will assume that

$$\operatorname*{ess\,sup}_{t\in[t_0,t_1]} G(\hat w(t)) \,=\, 0, \tag{24}$$

otherwise the mixed constraint is redundant for the weak minimality of the process $\hat w$. For any $\delta > 0$, define a set

$$M_\delta \,=\, \{t \in [t_0,t_1]: \;\; G(\hat w(t)) \geqslant -\delta\}.$$

In view of (24), $\operatorname{mes} M_\delta > 0$ for all $\delta > 0$.

6.1 Application of approximate separation theorem

Let us consider as independent variables in Problem P the pair $(x_0,u) \in \mathbb{R}^n \times L^\infty$, while the state $x(t)$ is determined by the latter as the solution to equation (2) with the initial condition $x(t_0) = x_0$, so that $x = x(x_0,u)$ is a nonlinear operator of (x_0,u), which maps $\mathbb{R}^n \times L^\infty$ to the space C. The Problem P has then the form

$$J(x_0,\, x(x_0,u)(t_1)) \to \min, \qquad G(x(x_0,u)(t),\, u(t)) \leqslant 0. \tag{25}$$

Note that the weak minimality of the pair $\hat w = (\hat x,\hat u)$ in Problem P is equivalent to the local minimality of the pair $(\hat x(t_0),\hat u)$ in Problem (25).

1. Let $\hat w = (\hat x,\hat u)$ be a reference process. Consider the equation in variations for the control system (2):

$$\dot{\bar x} \,=\, f_x(\hat w)\,\bar x \,+\, f_u(\hat w)\,\bar u, \qquad \bar x(t_0) = \bar x_0\,, \tag{26}$$

and define the corresponding linear operator

$$A:\; (\bar x_0,\bar u) \in \mathbb{R}^n \times L^\infty([t_0,t_1],\mathbb{R}^m) \;\to\; \bar x \in C([t_0,t_1],\mathbb{R}^n),$$

where $\bar x$ is the solution to (26) for the given pair $(\bar x_0,\bar u)$.

Recall the following well-known fact, which relates to the nonlinear operator $(x_0,u) \to x$ defined by the original equation (2) with $x(t_0) = x_0$. This operator maps a neighborhood of the point $(\hat x(t_0),\hat u) \in \mathbb{R}^n \times L^\infty([t_0,t_1],\mathbb{R}^m)$ to the space $C([t_0,t_1],\mathbb{R}^m)$ endowed with its standard norm $\|x\|_C = \max_t |x(t)|$.

Lemma 4 *The operator A is the Frechet derivative at $(\hat{x}(t_0), \hat{u})$ of the nonlinear operator $(x_0, u) \to x$. Hence, for any solution $\bar{w} = (\bar{x}, \bar{u})$ to (26), there is a correction $\widetilde{x}_\varepsilon$ parametrized by $\varepsilon > 0$ with $\widetilde{x}_\varepsilon(t_0) = 0$ and $\|\widetilde{x}_\varepsilon\|_C = o(\varepsilon)$ as $\varepsilon \to 0+$, such that the pair $w_\varepsilon = (\hat{x} + \varepsilon\bar{x} + \widetilde{x}_\varepsilon,\, \hat{u} + \varepsilon\bar{u})$ satisfies (2) with the initial condition $\hat{x}_0 + \varepsilon\bar{x}_0$.*

2. Introduce a Banach space $\mathcal{Y} = \mathbb{R}^{n*} \times L^1([t_0, t_1], \mathbb{R}^{m*}) \times \mathbb{R}$ with elements $y = (c_0, v, r)$, and its dual space $\mathcal{X} = \mathcal{Y}^* = \mathbb{R}^n \times L^\infty([t_0, t_1], \mathbb{R}^m) \times \mathbb{R}$ with elements $\varkappa = (x_0, u, q)$. The pairing between these spaces is given by

$$\langle y, \varkappa \rangle \;=\; c_0\, x_0 \,+ \int_{t_0}^{t_1} v(t)\, u(t)\; \mathrm{d}t \,+\, rq.$$

To prove Theorem 3, we follow the Dubovitskii–Milyutin approach. First of all, we define, in the space $\mathcal{X}$, the following cones of first order approximations of the cost and constraint. For any $\delta > 0$ we set

$$\begin{aligned}
\Omega_0 \;&=\; \{\varkappa \in \mathcal{X}: \;\; J_{x_0}(\hat{\xi})\, x_0 \,+ J_{x_1}(\hat{\xi})\, x(t_1) \,+ q < 0, \;\; \text{where } \; x = A(x_0, u)\,\},\\
\Omega_\delta \;&=\; \{\varkappa \in \mathcal{X}: \;\; \operatorname*{ess\,sup}_{t \in M_\delta} \Big(G_x(\hat{w})\, x + \,G_u(\hat{w})\, u \Big) \,+ q \,< 0, \;\; \text{where } \; x = A(x_0, u)\,\},\\
\Omega \;&=\; \{\varkappa \in \mathcal{X}: \;\; q > 0\}.
\end{aligned}$$

Obviously, all these cones are convex, open, and nonempty (since the first two contain the triple $(0, 0, -1)$, and the last one contains $(0, 0, 1)$).

Remark. In what follows, our aim will be to separate these cones by elements of the predual space $\mathcal{Y}$. The variable q and the third cone Ω are introduced because without q the second cone Ω_δ can be empty, which prevents application of the separation theorem. To avoid the analysis of this case that can be tedious, we, following [3], introduce the additional variable $q > 0$. The price for this trick is negligible in the case of present simplest problem P. In a more general problem, it would be a bit more essential, but still acceptable.

The first step in the Dubovitskii–Milyutin approach is to show that the cones of first order approximations do not intersect.

Lemma 5 *If $\hat{w}$ is a weak minimum in problem P, then for any $\delta > 0$*

$$\Omega_0 \,\cap\, \Omega_\delta \,\cap\, \Omega \;=\; \emptyset. \tag{27}$$

Proof. Suppose, on the contrary, there exist a $\delta > 0$ and a

$$\bar{\varkappa} \;=\; (\bar{x}_0, \bar{u}, \bar{q}) \;\in\; \Omega_0 \cap \Omega_\delta \cap \Omega.$$

Set $\bar{x} = A(\bar{x}_0, \bar{u})$, $\bar{w} = (\bar{x}, \bar{u})$, and take the curve $w_\varepsilon = (\hat{x} + \varepsilon\bar{x} + \widetilde{x}_\varepsilon,\, \hat{u} + \varepsilon\bar{u})$ from Lemma 4. Since $G(\hat{w}) \leqslant 0$ a.e. on $[t_0, t_1]$, and $G'(\hat{w})\bar{w} = G'_x(\hat{w})\,\bar{x} + G'_u(\hat{w})\,\bar{u} < -\bar{q} < 0$ on M_δ, we have for sufficiently small $\varepsilon > 0$:

$$G(w_\varepsilon) = \; G(\hat{w}) + \varepsilon G'(\hat{w})\bar{w} + o(\varepsilon) \;<\; -\varepsilon\bar{q} + o(\varepsilon) < 0 \quad \text{a.e. on } M_\delta\,.$$

For a.a. $t \notin M_\delta$, we have $G(\hat{w}) \leqslant -\delta$, whence we obviously obtain $G(w_\varepsilon) \leqslant -\delta/2 < 0$ for small $\varepsilon > 0$, so the pair w_ε satisfy the mixed constraint of problem (25).

Now, consider the reference endpoints $\hat{\xi} = (\hat{x}_0, \hat{x}_1)$ and set $\bar{x}_1 = \bar{x}(t_1)$, $\bar{\xi} = (\bar{x}_0, \bar{x}_1)$. Since $\bar{\varkappa} \in \Omega_0$ and $\bar{q} > 0$, we have $J'(\hat{\xi})\bar{\xi} = J_{x_0}(\hat{\xi})\bar{x}_0 + J_{x_1}(\hat{\xi})\bar{x}_1 < -\bar{q} < 0$, and then, for sufficiently small $\varepsilon > 0$

$$\mathcal{J}(w_\varepsilon) = J(\hat{x}_0 + \varepsilon\bar{x}_0,\, \hat{x}_1 + \varepsilon\bar{x}_1 + \widetilde{x}_\varepsilon(t_1)) =$$
$$= J(\hat{\xi}) + \varepsilon J'(\hat{\xi})\,\bar{\xi} + J'_{x_1}(\hat{\xi})\,\widetilde{x}_\varepsilon(t_1) + o(\varepsilon) < J(\hat{\xi}) - \varepsilon\,\bar{q} + o(\varepsilon) < \mathcal{J}(\hat{w}),$$

which contradicts the weak minimality at $\hat{w}$. The lemma is proved. □

3. Next, we define cones H_0, H_δ, H in $\mathcal{Y}$ that are thick on Ω_0, Ω_δ, Ω, respectively. Let us start with the cone Ω_0. Consider a functional $l : \mathbb{R}^n \times L^\infty \to \mathbb{R}$ such that

$$l(x_0, u) := J_{x_1}(\hat{\xi})\, x(t_1), \quad \text{where} \quad x = A(x_0, u).$$

As is known, one can give its explicit dependence of (x_0, u). To this end, introduce the usual adjoint function $p_0 \in AC$ determined by the adjoint equation to (26):

$$-\dot{p}_0 = p_0 f_x(\hat{w}) \quad \text{with} \quad p_0(t_1) = J_{x_1}(\hat{\xi}).$$

Then, obviously $\frac{d}{dt}(p_0\, x) = p_0 f_u(\hat{w})\, u$, whence integrating we get

$$J_{x_1}(\hat{\xi})\, x(t_1) = p_0(t_0)\, x_0 + \int_{t_0}^{t_1} p_0\, f_u(\hat{w})\, u\, \mathrm{d}t \qquad \forall\, (x_0, u) \in \mathbb{R}^n \times L^\infty.$$

Consequently,

$$J_{x_0}(\hat{\xi})x_0 + J_{x_1}(\hat{\xi})\, x(t_1) + q = \Big(J_{x_0}(\hat{\xi}) + p_0(t_0)\Big)x_0 + \int_{t_0}^{t_1} p_0 f_u(\hat{w})\, u\, \mathrm{d}t + q \tag{28}$$

for all $\varkappa = (x_0, u, q) \in \mathcal{X}$. Define a triple

$$\hat{y}_0 = (J_{x_0}(\hat{\xi}) + p_0(t_0),\; p_0 f_u(\hat{w}),\; 1) \in \mathcal{Y}.$$

In view of (28), Ω_0 is an open half-space: $\Omega_0 = \{\varkappa \in \mathcal{X} : \langle \hat{y}_0, \varkappa\rangle < 0\}$, and its closure is $\overline{\Omega}_0 = \{\varkappa \in \mathcal{X} : \langle \hat{y}_0, \varkappa\rangle \leqslant 0\}$. Setting $H_0 = \{-\alpha_0\, \hat{y}_0 : \alpha_0 \geqslant 0\}$, we obtain $H_0^* = \overline{\Omega}_0$, that is H_0 is thick on Ω_0.

4. Consider the cone Ω_δ. First, we claim that

$$\overline{\Omega}_\delta = \{\varkappa \in \mathcal{X} : \operatorname*{ess\,sup}_{t \in M_\delta} \Big(G_x(\hat{w})\, x + G_u(\hat{w})\, u\Big) + q \leqslant 0, \text{ where } x = A(x_0, u)\}.$$

Indeed, for any such $\varkappa$, taking a smaller $q' < q$ we get a point $\varkappa' \in \Omega_\delta$, q.e.d.

Define a cone H_δ consisting of all functionals $y_\delta = (c_0, v, r) \in \mathcal{Y}$ that for all $\varkappa = (x_0, u, q) \in \mathcal{X}$ act as follows:

$$\langle y_\delta, \varkappa\rangle = -\int_{t_0}^{t_1} \lambda(t)\Big(G_x(\hat{w})\, x + G_u(\hat{w})\, u + q\Big)\, dt, \quad \text{where} \quad x = A(x_0, u),$$

and $\lambda \in L^1$ is an arbitrary nonnegative function concentrated on M_δ, that is $\lambda \geqslant 0$ and $\lambda\chi_{M_\delta} = \lambda$, where χ_{M_δ} is the characteristic function of the set M_δ.

Lemma 6 $H_\delta^* = \overline{\Omega}_\delta$.

Proof. If $y_\delta \in H_\delta$ and $\varkappa \in \overline{\Omega}_\delta$, then obviously $\langle y_\delta, \varkappa\rangle \geqslant 0$, whence $\varkappa \in H_\delta^*$. Therefore, $\overline{\Omega}_\delta \subset H_\delta^*$. Let us prove the converse inclusion $\overline{\Omega}_\delta \supset H_\delta^*$.

Indeed, take any $\bar\varkappa \in H_\delta^*$, that is $\langle y_\delta, \bar\varkappa\rangle \geqslant 0$ for all $y_\delta \in H_\delta$. This means that

$$\langle y_\delta, \bar\varkappa\rangle \;=\; -\int_{t_0}^{t_1} \lambda(t)\Big(G_x(\hat w)\,\bar x \;+\; G_u(\hat w)\,\bar u \;+\bar q\Big)\,dt \;\geqslant\; 0$$

for all nonnegative functions $\lambda \in L^1$ concentrated on M_δ. This obviously implies $G_x(\hat w)\,\bar x \;+\; G_u(\hat w)\,\bar u \;+\bar q \;\leqslant 0$ a.e. on M_δ, that is $\bar\varkappa \in \overline{\Omega}_\delta$. Thus, $H_\delta^* \subset \overline{\Omega}_\delta$, q.e.d. $\square$

5. Take any $y_\delta \in H_\delta$ and the corresponding function $\lambda \in L^1$. Represent it in the canonical form $y_\delta = (c_0, v, r)$. In fact, we only have to find a representation of the term $\int \lambda(t)\, G_x(\hat w)\, x\, dt$. To this aim, define a function $p_\delta \in AC$ from the equation

$$-\dot p_\delta \;=\; p_\delta f_x(\hat w) \;+\; \lambda\, G_x(\hat w), \quad p_\delta(t_1) = 0.$$

Since $\dot x = f_x(\hat w)\,x + f_u(\hat w)\,u$, we have $\frac{\mathrm{d}}{\mathrm{d}t}(p_\delta\, x) \;=\; -\lambda\, G_x(\hat w)\,x \;+\; p_\delta f_u(\hat w)\,u$, whence

$$\int_{t_0}^{t_1} \lambda\, G_x(\hat w)\,x\;\mathrm{d}t \;=\; p_\delta(t_0)\,x_0 \;+\int_{t_0}^{t_1} p_\delta f_u(\hat w)\,u\;\mathrm{d}t.$$

Then, for any $\varkappa = (x_0, u, q) \in \mathcal{X}$ we have

$\langle y_\delta, \varkappa\rangle \;=\; -p_\delta(t_0)\,x_0 \;- \displaystyle\int_{t_0}^{t_1} \Big(\big(p_\delta f_u(\hat w) + \lambda G_u(\hat w)\big)\,u \;+\lambda q\Big)\;\mathrm{d}t,$ and so

$$y_\delta = \;-\left(p_\delta(t_0),\;\; p_\delta f_u(\hat w) + \lambda\, G_u(\hat w),\;\; \int_{t_0}^{t_1} \lambda\,\mathrm{d}t\right).$$

6. Finally, consider the cone Ω. Set $\hat y \;=\; (0,0,1) \;\in\; \mathcal{Y}$ and $H \;=\; \{\alpha\hat y \;:\; \alpha \;\geqslant\; 0\}$. Then $H^* = \overline{\Omega}$, that is H is thick on Ω.

7. Set $\varkappa^0 = (0,0,1) \in \mathcal{X}$ (here $x_0 = 0$, $u = 0$, $q = 1$). Obviously,

$$-\varkappa^0 \in \Omega_0 \cap \Omega_\delta\,, \qquad \varkappa^0 \in \Omega.$$

Now, we apply Theorem 2 to condition (27). According to this theorem, for any $\delta > 0$ and any $\varepsilon > 0$ there exist functionals

$$y_0 \in H_0, \quad y_\delta \in H_\delta, \quad y \in H \tag{29}$$

such that

$$\langle y_0,\, -\varkappa^0\rangle \;+\; \langle y_\delta,\, -\varkappa^0\rangle \;=\; 1, \tag{30}$$

$$\|y_0 \;+\; y_\delta \;+\; y\| \;< \varepsilon. \tag{31}$$

(Here we choose the cone Ω to be excluded from the normalization condition (30).)

Analysis of these conditions will lead to the local minimum principle.

6.2 Analysis of conditions (29)–(31)

According to the definitions of H_0, H_δ, and H, conditions (29) mean that

$$\begin{aligned}
y_0 &= -\alpha_0\big(p_0(t_0) + J_{x_0}(\hat{\xi}),\ p_0 f_u(\hat{w}),\ 1\big), \quad \alpha_0 \geqslant 0,\\
y_\delta &= -\big(p_\delta(t_0),\ p_\delta f_u(\hat{w}) + \lambda\, G_u(\hat{w}),\ \textstyle\int_{t_0}^{t_1} \lambda\, \mathrm{d}t\big), \quad \lambda \geqslant 0, \quad \lambda\chi_\delta = \lambda,\\
y &= \alpha\,(0,\ 0,\ 1), \quad \alpha \geqslant 0.
\end{aligned}$$

Condition (30) gives

$$\alpha_0 + \int_{t_0}^{t_1} \lambda\, \mathrm{d}t \;=\; 1. \tag{32}$$

In view of this relation, we get

$$-(y_0 + y_\delta + y) \;=$$
$$= \Big(\alpha_0 p_0(t_0) + p_\delta(t_0) + \alpha_0 J_{x_0}(\hat{\xi}),\ \ (\alpha_0 p_0 + p_\delta) f_u(\hat{w}) + \lambda G_u(\hat{w}),\ \ 1-\alpha\Big).$$

Set $p = \alpha_0 p_0 + p_\delta$. Then

$$-\,\dot{p} \;=\; p f_x(\hat{w}) + \lambda G_x(\hat{w}), \qquad p(t_1) = \alpha_0 J_{x_1}(\hat{\xi}), \tag{33}$$

and $-(y_0 + y_\delta + y) \;=\; \Big(p(t_0) + \alpha_0 J_{x_0}(\hat{\xi}),\ p f_u(\hat{w}) + \lambda G_u(\hat{w}),\ 1-\alpha\Big).$
Condition (31) implies

$$|\, p(t_0) + \alpha_0 J_{x_0}(\hat{\xi})| \;+\; \|\, p f_u(\hat{w}) + \lambda G_u(\hat{w})\|_1 \;+\; |1-\alpha| \;<\; \varepsilon. \tag{34}$$

Recall that such α_0, α, λ, p exist for all $\delta > 0$ and $\varepsilon > 0$.

Thus, there exist two countable sequences $\{(\alpha_0^k, \alpha^k, \lambda^k, p^k)\}_{k=1}^\infty$ and $\{\delta^k\}_{k=1}^\infty$, where $\delta^k \to 0+$,

$$\alpha_0^k \in \mathbb{R}, \quad \alpha^k \in \mathbb{R}, \quad \lambda^k \in L^1, \quad p^k \in AC, \tag{35}$$

$$\alpha_0^k \geqslant 0, \quad \alpha^k \geqslant 0, \quad \lambda^k \geqslant 0, \quad \lambda^k \chi_{M_{\delta^k}} = \lambda^k, \tag{36}$$

$$-\delta^k \leqslant G(\hat{w}(t)) \leqslant 0 \quad \text{a.e. on } M_{\delta^k} \tag{37}$$

(the latter follows from the definition of M_δ), such that

$$\alpha^k \to 1, \quad \alpha_0^k + \|\lambda^k\|_1 \;=\; 1, \tag{38}$$

$$-\dot{p}^k \;=\; p^k f_x(\hat{w}) \;+\; \lambda^k G_x(\hat{w}), \tag{39}$$

$$p^k(t_0) + \alpha_0^k J_{x_0}(\hat{\xi}) \to 0, \qquad p^k(t_1) = \alpha_0^k J_{x_1}(\hat{\xi}), \tag{40}$$

$$\|\, p^k f_u(\hat{w}) \;+\; \lambda^k G_u(\hat{w})\|_1 \to 0. \tag{41}$$

Hereinafter, we do not write the condition $k \to \infty$. Note also that superscript k is always used to denote the number of a member in the sequence and never used to denote the degree.

Without loss of generality we assume that $\alpha_0^k \to \hat{\alpha}_0 \geqslant 0$. Then

$$\hat{\alpha}_0 \;+\; \|\lambda^k\|_1 \to 1. \tag{42}$$

Moreover, conditions (40) imply

$$p^k(t_0) \to -\hat{\alpha}_0 J_{x_0}(\hat{\xi}), \qquad p^k(t_1) \to \hat{\alpha}_0 J_{x_1}(\hat{\xi}). \tag{43}$$

It follows that the sequences $p^k(t_0)$ and $p^k(t_1)$ are bounded, and in view of (42) the norms $\|\lambda^k\|_1$ are also bounded. Therefore, by (39) and the Gronwall's inequality, the norms $\|p^k\|_\infty$ are uniformly bounded as well.

Now, we rewrite the adjoint equation (39) in the form of measures:

$$-\,\mathrm{d}p^k \;=\; p^k f_x(\hat{w})\,\mathrm{d}t \;+\; \lambda^k G_x(\hat{w})\,\mathrm{d}t. \tag{44}$$

Define a measure

$$\mathrm{d}\mu^k \;:=\; \lambda^k\, G_x(\hat{w})\,\mathrm{d}t. \tag{45}$$

Equation (44) then takes the form

$$-\,\mathrm{d}p^k \;=\; p^k f_x(\hat{w})\,\mathrm{d}t \;+\; \mathrm{d}\mu^k. \tag{46}$$

Clearly, the sequence $\|\,\mathrm{d}\mu^k\|$ is bounded. Without loss of generality we assume that $\mathrm{d}\mu^k$ weakly* converges to some measure $\mathrm{d}\hat{\mu} \in C^*$ (i.e. $C-$converges in C^*), and denote this as

$$\mathrm{d}\mu^k \;\xrightarrow{*}\; \mathrm{d}\hat{\mu}. \tag{47}$$

Conditions (47), (46), and (43) imply that there is a function $\hat{p} \in BV$ such that at every point $t \in [t_0, t_1]$ of continuity of the limiting measure $\mathrm{d}\hat{\mu}$ (hence almost everywhere) we have $p^k(t) \to \hat{p}(t)$, and moreover,

$$-\,\mathrm{d}\hat{p} \;=\; \hat{p}\, f_x(\hat{w})\,\mathrm{d}t \;+\; \mathrm{d}\hat{\mu}, \tag{48}$$

$$-\hat{p}(t_0-) \;=\; \hat{\alpha}_0 J_{x_0}(\hat{\xi}), \qquad \hat{p}(t_1+) \;=\; \hat{\alpha}_0 J_{x_1}(\hat{\xi}). \tag{49}$$

Since the sequence $\|p^k\|_\infty$ is bounded, we also have

$$\|p^k - \hat{p}\|_1 \;\to\; 0. \tag{50}$$

Now, our aim is to find a more detailed representation of the measure $d\hat{\mu}$.

6.3 Representation of the absolutely continuous part of $d\hat{\mu}$

1. Since the sequence $\|\lambda^k\|_1$ is bounded, then, according to Lemma 2, there exists a sequence of measurable sets $A^k \subset [t_0, t_1]$ such that $\operatorname{mes} A^k \to (t_1 - t_0)$, and the sequence $\lambda^k_A := \lambda^k \chi_{A^k}$ is uniformly integrable, hence it contains a weakly convergent (with respect to L^∞) subsequence. Without loss of generality we assume that the sequences λ^k_A itself weakly converges to some function $\hat{\lambda} \in L^1$:

$$\lambda^k_A \;\xrightarrow{w}\; \hat{\lambda}. \tag{51}$$

Since $\lambda^k_A \geqslant 0$ for all k, we have $\hat{\lambda}(t) \geqslant 0$ a.e. in $[t_0, t_1]$ and

$$\|\lambda^k_A\| \;\to\; \|\hat{\lambda}\|. \tag{52}$$

Further, conditions $\lambda^k \chi_{M_{\delta^k}} = \lambda^k \geqslant 0$ (see (36)), and $\lambda^k_A = \lambda^k \chi_{A^k}$ imply in view of (37) that $|\lambda^k_A(t)\, G(\hat w(t))| \leqslant \lambda^k_A(t)\, \delta^k$ a.e. on $[t_0, t_1]$, hence $\|\lambda^k_A\, G(\hat w)\| \to 0$. The more so, $\lambda^k_A\, G(\hat w) \to 0$ weakly in L^1.

On the other hand, (51) implies $\lambda^k_A\, G(\hat w) \xrightarrow{w} \hat\lambda\, G(\hat w)$, whence

$$\hat\lambda(t)\, G(\hat w(t)) \,=\, 0 \quad \text{a.e. in } [t_0, t_1], \tag{53}$$

i.e., the complementary slackness condition in (11) holds true.

2. Consider more thoroughly condition (41). Define a function $p^k_A := p^k \chi_{A^k} \in L^\infty$. Since the set $B^k := [t_0, t_1] \setminus A^k$ has mes $B^k \to 0$, and the sequence $\|p^k\|_\infty$ is bounded, we get $\|p^k_A - p^k\|_1 = \|\, p^k \chi_{B^k}\|_1 \to 0$, which in view of (50) yields

$$\|p^k_A - \hat p\|_1 \,\to\, 0. \tag{54}$$

Condition (41) means that $p^k f_u(\hat w) + \lambda^k G_u(\hat w) \,=\, z^k$, where $\|\, z^k\|_1 \to 0$. Multiplying it by χ_{A^k}, we get $p^k_A\, f_u(\hat w) \,+\, \lambda^k_A\, G_u(\hat w) \,=\, z^k_A \,:=\, z^k \chi_{A^k}\,, \quad \|\, z^k_A\|_1 \,\to 0$. This and condition (54) imply $\|\, \hat p\, f_u(\hat w) + \lambda^k_A\, G_u(\hat w)\|_1 \,\to\, 0$. Finally, since $\lambda^k_A \xrightarrow{w} \hat\lambda$, we obtain $\hat p\, f_u(\hat w) \,+\, \hat\lambda\, G_u(\hat w) \,=\, 0$, i.e. condition (16) of LMP holds true.

3. Now, introduce the sequence $\lambda^k_B \,:=\, \lambda^k \chi_{B^k} \,\in\, L^1$. Obviously, $\lambda^k_B \geqslant 0$ and $\lambda^k_A \,+\, \lambda^k_B \,=\, \lambda^k$. Note that both λ^k_A and λ^k_B are supported on M_{δ^k}, since they are restrictions of λ^k supported on M_{δ^k}. Therefore, if we narrow the set B^k to the set $M_{\delta^k} \cap B^k$, the function λ^k_B would not change. So, we will assume that $B^k \subset M_{\delta^k}$, that is $B^k = M_{\delta^k} \setminus A^k$. Setting

$$\mathrm{d}\mu^k_A \,=\, \lambda^k_A\, G_x(\hat w)\, \mathrm{d}t, \qquad \mathrm{d}\mu^k_B \,=\, \lambda^k_B\, G_x(\hat w)\, \mathrm{d}t,$$

we obtain two sequences of measures $\mathrm{d}\mu^k_A$ and $\mathrm{d}\mu^k_B$ in C^*. Since by (45) $\mathrm{d}\mu^k = \lambda^k\, G_x(\hat w)\, \mathrm{d}t$, we have $\mathrm{d}\mu^k_A + \mathrm{d}\mu^k_B = \mathrm{d}\mu^k$. Since $\lambda^k_A \xrightarrow{w} \hat\lambda$, we have

$$\mathrm{d}\mu^k_A \xrightarrow{*} \mathrm{d}\hat\mu_A \,:=\, \hat\lambda\, G_x(\hat w)\, \mathrm{d}t. \tag{55}$$

Since $\mathrm{d}\mu^k \xrightarrow{*} \mathrm{d}\hat\mu$ and $\mathrm{d}\mu^k_A \xrightarrow{*} \mathrm{d}\hat\mu_A$, there exists a measure $\mathrm{d}\hat\mu_B \in C^*$ such that

$$\mathrm{d}\mu^k_B \xrightarrow{*} \mathrm{d}\hat\mu_B, \qquad \mathrm{d}\hat\mu_A + \mathrm{d}\hat\mu_B \,=\, \mathrm{d}\hat\mu.$$

Now we aim to specify the measure $\mathrm{d}\hat\mu_B$, and this is the main part of our study.

6.4 Representation of the singular part of $\mathrm{d}\hat\mu$

1°. We have $\lambda^k_B := \lambda^k \chi_{B^k}$, where $B^k = M_{\delta^k} \setminus A^k$, mes $B^k \to 0$, and

$$\mathrm{d}\mu^k_B \,=\, \lambda^k_B\, G_x(\hat w)\, \mathrm{d}t \xrightarrow{*} \mathrm{d}\hat\mu_B\,. \tag{56}$$

Since the sequence of norms $\|\lambda^k\|_1$ is bounded, the sequence of measures $\lambda^k \chi_{B^k}\, \mathrm{d}t$ in C^* is also bounded. Therefore, without loss of generality we assume that there is a measure $\mathrm{d}\hat\eta \in C^*$ such that $\mathrm{d}\hat\eta \geqslant 0$ and

$$\lambda^k_B\, \mathrm{d}t \,=\, \lambda^k\, \chi_{B^k}\, \mathrm{d}t \xrightarrow{*} \mathrm{d}\hat\eta. \tag{57}$$

Since $\|\lambda^k\|_1 = \|\lambda^k_A\|_1 + \|\lambda^k_B\|_1\,,$ conditions (42) and (52) imply

$$\hat\alpha_0 \,+\, \|\hat\lambda\|_1 \,+\, \|\lambda^k_B\|_1 \,\to 1. \tag{58}$$

Moreover, since $\lambda^k_B \geqslant 0,$ relation (57) yields $\|\lambda^k_B\|_1 \,\to \|\,\mathrm{d}\eta\|,$ whence

$$\hat\alpha_0 \,+\, \|\hat\lambda\|_1 \,+\, \|\,\mathrm{d}\eta\| \;=\; 1, \tag{59}$$

which is equivalent to the nontriviality condition (12).

There are two possible cases: $\|\,\mathrm{d}\eta\| = 0$ and $\|\,\mathrm{d}\eta\| > 0.$ In the first, trivial case, $\|\lambda^k_B\|_1 \,\to 0,$ the more so $\|\,\mathrm{d}\mu^k_B\| \to 0,$ then $\mathrm{d}\hat\mu_B = 0,$ i.e. the singular part of $\mathrm{d}\hat\mu$ does not appear in the LMP. Setting here $\hat s = 0$ and $\mathcal{R} = \emptyset,$ we obtain the costate equation (14) with properties (13) that are trivially satisfied.

2°. Consider now the main case, where

$$\|\,\mathrm{d}\eta\| \;=\; \lim_k \|\lambda^k_B\|_1 \;=:\; r_B > 0. \tag{60}$$

Here we will slightly narrow the sets B^k in order to obtain more properties of λ^k_B. To do this, we need the following

Lemma 7 *Let be given two sequences of functions $a_n \geqslant 0$ and $b_n \geqslant 0$ in $L^1([t_0,t_1],\mathbb{R})$, and a sequence of measurable sets $B_n \subset [t_0,t_1]$ of* mes $B_n \to 0$ *such that*

$$\int_{B_n} a_n(t)\,\mathrm{d}t \,\to 1, \qquad \int_{B_n} b_n(t)\,\mathrm{d}t \,\to 0.$$

Then there is a sequence of measurable sets $E_n \subset B_n$ such that

$$a_n(t) > 0 \;\; a.e. \;\; on \;\; E_n\,, \qquad \int_{E_n} a_n(t)\,\mathrm{d}t \to 1, \qquad and \qquad \operatorname*{ess\,sup}_{t\in E_n} \frac{b_n(t)}{a_n(t)} \;\to\; 0.$$

Proof. Narrowing if necessary the sets B_n, we assume that $a_n(t) > 0$ a.e. on B_n. Take any sequence $\omega_n \to 0+$ such that $\displaystyle\int_{B_n} b_n(t)\,\mathrm{d}t \;=\; o(\omega_n),$ and define a sequence of sets $E_n = \{t \in B_n : \;\; b_n(t) \leqslant \omega_n\, a_n(t)\}.$ Then

$$\int_{B_n\setminus E_n} a_n(t)\,\mathrm{d}t \;\leqslant\; \frac{1}{\omega_n}\int_{B_n\setminus E_n} b_n(t)\,\mathrm{d}t \;\leqslant\; \frac{1}{\omega_n}\int_{B_n} b_n(t)\,\mathrm{d}t \;\to\; 0,$$

which gives the required properties. □

3°. Consider the L^1- functions

$$\sigma^k := \, p^k f_u(\hat w) + \lambda^k G_u(\hat w). \tag{61}$$

According to (41), $\|\,\sigma^k\|_1 \to 0.$ Then also $\displaystyle\int_{B^k} (1 + |\,\sigma^k|)\,\mathrm{d}t \,\to 0.$ By Lemma 7, there exists a sequence of measurable sets $E^k \subset B^k$ such that $\lambda^k_B(t) > \,0$ a.e. on E^k,

$$\int_{E^k} \lambda^k_B\,\mathrm{d}t \,\to\, r_B > 0, \qquad \text{and} \qquad \omega^k := \operatorname*{ess\,sup}_{E^k} \frac{1 + |\sigma^k|}{\lambda^k_B} \,\to\, 0. \tag{62}$$

The first relation here means that $\displaystyle\int_{B^k\setminus E^k} \lambda_B^k\,\mathrm{d}t \to 0,$ therefore (56) and (57) imply

$$\lambda_B^k\, G_x(\hat w)\,\chi_{E^k}\,\mathrm{d}t \ \stackrel{*}{\rightharpoonup}\ \mathrm{d}\hat\mu_B\,, \qquad \lambda_B^k\,\chi_{E^k}\,\mathrm{d}t \ \stackrel{*}{\rightharpoonup}\ \mathrm{d}\hat\eta.$$

Set $\lambda_E^k := \lambda_B^k\chi_{E^k} = \lambda^k\chi_{E^k}$. Then $\lambda^k(t) = \lambda_E^k(t)$ a.e. on E^k,

$$\lambda_E^k\,\mathrm{d}t \ \stackrel{*}{\rightharpoonup}\ \mathrm{d}\hat\eta, \qquad \lambda_E^k\, G_x(\hat w)\,\mathrm{d}t \ \stackrel{*}{\rightharpoonup}\ \mathrm{d}\hat\mu_B\,, \tag{63}$$

so the "narrowed" sequence λ_E^k has the same limit properties as the original λ_B^k does.

Since $E^k \subset M_{\delta^k}$, relations (37) imply

$$-\,\delta^k \leqslant G(\hat w(t)) \leqslant 0 \quad \text{ a.e. on } \quad E^k. \tag{64}$$

Moreover, in view of definition (61), for all k

$$G_u(\hat w) \;=\; \frac{\sigma^k - p^k f_u(\hat w)}{\lambda^k} \qquad \text{a.e. on} \quad E^k.$$

The second condition in (62) and the boundedness of the sequence $\|p^k\|_\infty$ imply

$$\varepsilon^k := \operatorname*{ess\,sup}_{E^k} |\, G_u(\hat w)| \;=\; \operatorname*{ess\,sup}_{E^k} \frac{\big|\sigma^k - p^k f_u(\hat w)\big|}{\lambda^k} \;\to\; 0,$$

whence

$$|\, G_u(\hat w(t))| \;\leqslant\; \varepsilon^k \quad \text{ a.e. on } \quad E^k. \tag{65}$$

4°. We will need the following constructions. Recall that we introduced the set of phase points $\mathcal{N}(G) := \{(x,u) \in \mathbb{R}^{n+m}:\ G(x,u) = 0,\ \ G_u(x,u) = 0\}$ and assumed that this set is nonempty.

Now, for any $\delta > 0$ and $\varepsilon > 0$, introduce its extension up to δ, ε:

$$\mathcal{N}_{\delta,\varepsilon}(G) \;:=\; \{(x,u) \in \mathbb{R}^{n+m}\;:\; -\delta \leqslant G(x,u) \leqslant 0, \quad |\, G_u(x,u)| \leqslant \varepsilon\,\}.$$

Obviously it is closed, and $\mathcal{N}(G) \;=\; \bigcap_{\delta>0,\ \varepsilon>0} \mathcal{N}_{\delta,\varepsilon}(G)$.

By analogy with the mapping $S(x,u)$, for any $\delta > 0$ and $\varepsilon > 0$, define a set-valued mapping

$$(x,u) \in \mathbb{R}^{n+m} \;\rightrightarrows\; S_{\delta,\varepsilon}(x,u) \,\subset \mathbb{R}^{n*}:$$

(i) if $(x,u) \in \mathcal{N}_{\delta,\varepsilon}(G)$, then $S_{\delta,\varepsilon}(x,u) \;=\; \{G_x(x,u)\}$,

(ii) if $(x,u) \notin \mathcal{N}_{\delta,\varepsilon}(G)$, then $S_{\delta,\varepsilon}(x,u) \;=\; \varnothing$.

Obviously, this mapping is compact-valued, upper semicontinuous, and

$$\bigcap_{\delta>0,\ \varepsilon>0} S_{\delta,\varepsilon}(x,u) \;=\; S(x,u) \qquad \text{for all } (x,u). \tag{66}$$

For any nonempty set $M \subset \mathbb{R}^{n+m}$, define

$$S_{\delta,\varepsilon}(M) := \bigcup_{(x,u)\in M} S_{\delta,\varepsilon}(x,u).$$

and $S_{\delta,\varepsilon}(\emptyset) = \emptyset$. Clearly, if M is compact, the set $S_{\delta,\varepsilon}(M)$ is compact as well. Note that for any $M \subset \mathbb{R}^{n+m}$ we have $S_{\delta,\varepsilon}(M) = S_{\delta,\varepsilon}(M \cap \mathcal{N}_{\delta,\varepsilon}(G))$.

Now, consider the reference process $\hat{w}(t) = (\hat{x}(t), \hat{u}(t))$. We will assume that the corresponding set $\mathcal{D}$, defined in (9), is nonempty, i.e. there exists a point $t_* \in [t_0, t_1]$ such that $\operatorname{clm}(\hat{w})(t_*) \cap \mathcal{N}(G) \neq \emptyset$.

Since the set $\operatorname{clm}(\hat{u})$ is compact, the set $S_{\delta,\varepsilon}(\hat{x}(t), \operatorname{clm}(\hat{u})(t))$ is also compact for any t and upper semicontinuous in t.

For any points $\tau_0 < \tau_1$ in $[t_0, t_1]$, define a set

$$Q_{\delta,\varepsilon}[\tau_0, \tau_1] := \bigcup_{t\in[\tau_0,\tau_1]} S_{\delta,\varepsilon}(\hat{x}(t), \operatorname{clm}(\hat{u})(t)). \tag{67}$$

By the above argument, the right hand side here is a compact set. Moreover, relation (66) implies that

$$Q[\tau_0, \tau_1] := \bigcap_{\delta>0,\ \varepsilon>0} Q_{\delta,\varepsilon}[\tau_0, \tau_1] = \bigcup_{\tau\in[\tau_0,\tau_1]} S(\hat{x}(\tau), \operatorname{clm}(\hat{u})(\tau)), \tag{68}$$

and this set is also compact.

Finally, for any t_*, if $\tau_0 \to t_* - 0$ and $\tau_1 \to t_* + 0$, then obviously

$$Q[\tau_0, \tau_1] \to S(\hat{x}(t_*), \operatorname{clm}(\hat{u})(t_*)). \tag{69}$$

in the Hausdorf sense.

5°. Now we can describe the relationship between the measures $\mathrm{d}\hat{\mu}_B$ and $\mathrm{d}\hat{\eta}$. Recall that $\|\,\mathrm{d}\hat{\eta}\| > 0$ by (60).

Lemma 8 *The measure $\mathrm{d}\hat{\mu}_B$ admits a representation*

$$\mathrm{d}\hat{\mu}_B = \hat{s}(t)\,\mathrm{d}\hat{\eta} \tag{70}$$

with some $\mathrm{d}\hat{\eta}$-measurable essentially bounded function $\hat{s} : [t_0, t_1] \to \mathbb{R}^{n*}$, and there is a set $\mathcal{R} \subset \mathcal{D}$ of full $\mathrm{d}\hat{\eta}$-measure (i.e., $\int_{\mathcal{R}} \mathrm{d}\hat{\eta} = \int_{[t_0,t_1]} \mathrm{d}\hat{\eta}$) such that

$$\hat{s}(t) \in conv\, S\big(\hat{x}(t), \operatorname{clm}(\hat{u})(t)\big) \quad \textit{for all} \quad t \in \mathcal{R}. \tag{71}$$

Proof. a) In view of (63), $|\,\mathrm{d}\hat{\mu}_B| \leqslant M\,\mathrm{d}\hat{\eta}$, where $M = \|G_x(\hat{w})\|_\infty$. Hence, the measure $\mathrm{d}\hat{\mu}_B$ is absolutely continuous with respect to $\mathrm{d}\hat{\eta}$, and therefore, by the Radon-Nikodym theorem it admits representation (70), where $\hat{s}(t)$ is a $\mathrm{d}\hat{\eta}$-measurable function taking values in $\mathbb{R}^{n*}$ and satisfying $|\hat{s}(t)| \leqslant M$ a.e. in $\mathrm{d}\hat{\eta}$.

b) Let us prove inclusion (71) with some $\mathcal{R} \subset \mathcal{D}$ of full $\mathrm{d}\hat{\eta}$-measure. Fix a point $t_* \in [t_0, t_1]$ with the following properties:

(i) if $t_* \in [\tau_0, \tau_1] \subset [t_0, t_1]$, $\tau_0 < \tau_1$, then $\int_{[\tau_0,\tau_1]} \mathrm{d}\hat\eta > 0$,

(ii) if $t_* \in [\tau_0, \tau_1] \subset [t_0, t_1]$, $\tau_0 < \tau_1$, $\tau_0 \to t_*$, $\tau_1 \to t_*$, then[7]

$$\hat s_{[\tau_0,\tau_1]} \;:=\; \frac{\int_{[\tau_0,\tau_1]} \mathrm{d}\hat\mu_B}{\int_{[\tau_0,\tau_1]} \mathrm{d}\hat\eta} \;\to\; \hat s(t_*), \tag{72}$$

As is known, the set of such points t_* has a full $\mathrm{d}\hat\eta$-measure in $[t_0, t_1]$ (since it includes the Lebesgue points of the function $\hat s$ with respect to the measure $\mathrm{d}\hat\eta$). Denote this set by $\mathcal{R}$. We have $\int_{\mathcal{R}} \mathrm{d}\hat\eta \;=\; \int_{[t_0,t_1]} \mathrm{d}\hat\eta$.

Take any $[\tau_0,\, \tau_1]$ containing t_*. Then $\int_{[\tau_0,\tau_1]} \lambda_E^k\, \mathrm{d}t > 0$ for all sufficiently large k. (Otherwise $\lambda_E^k = 0$ a.e. in $[\tau_0,\, \tau_1]$ for a subsequence $k \to \infty$, which implies that also $\mathrm{d}\hat\eta = 0$ in $[\tau_0,\, \tau_1]$, a contradiction with the choice of t_*).

Therefore, we can define a row-vector

$$s^k_{[\tau_0,\tau_1]} \;:=\; \frac{\int_{[\tau_0,\tau_1]} \lambda_E^k\, G_x(\hat w)\; \mathrm{d}t}{\int_{[\tau_0,\tau_1]} \lambda_E^k\; \mathrm{d}t}\;. \tag{73}$$

Let Θ be the set of continuity of the measures $\mathrm{d}\hat\mu_B$ and $\mathrm{d}\hat\eta$, i.e., the set of all those t, which are not atoms neither of $\mathrm{d}\hat\mu_B$ nor of $\mathrm{d}\hat\eta$. Note that Θ is dense in $[t_0, t_1]$. If $\tau_0,\, \tau_1 \in \Theta$, then

$$s^k_{[\tau_0,\tau_1]} \;\to\; \hat s_{[\tau_0,\tau_1]} \quad \text{as} \quad k \to \infty, \tag{74}$$

since both the numerator and denominator tend to the corresponding limits.

Notice that the right hand side of (73) is a "convex combination" of the vectors $G_x(\hat w(t))$, in its continuous version.

c) In view of (64) and (65), we have $\hat w(t) \in \mathcal{N}_{\delta^k\varepsilon^k}(G)$ a.e. on E^k, and so

$$G_x(\hat w(t)) \;\in\; S_{\delta^k\varepsilon^k}(\hat x(t), \hat u(t)) \quad \text{a.e. on} \quad E^k.$$

Since $\hat u(t) \in \mathrm{clm}\,(\hat u)(t)$ a.e. on $[t_0, t_1]$, we get

$$G_x(\hat w(t)) \;\in\; S_{\delta^k\varepsilon^k}(\hat x(t),\, \mathrm{clm}\,(\hat u)(t)) \quad \text{a.e. on} \quad E^k. \tag{75}$$

Recall that $t_* \in \mathcal{R}$. Take any $[\tau_0,\, \tau_1]$ containing t_*. The last inclusion and definition (67) imply that for almost all $t \in E^k \cap [\tau_0,\, \tau_1]$

$$G_x(\hat w(t)) \;\in\; \bigcup_{\tau\in[\tau_0,\tau_1]} S_{\delta^k\varepsilon^k}(\hat x(\tau),\, \mathrm{clm}\,(\hat u)(\tau)) \;=\; Q_{\delta^k\varepsilon^k}[\tau_0, \tau_1]\,,$$

and the more so, for almost all $t \in E^k \cap [\tau_0,\, \tau_1]$

$$G_x(\hat w(t)) \;\in\; \mathrm{conv}\; Q_{\delta^k\varepsilon^k}[\tau_0, \tau_1].$$

[7] If $\tau_0 = t_*$ we do not need to tend $\tau_0 \to t_*$, so only tend $\tau_1 \to t_*$. The same concerns τ_1.

By the Caratheodory theorem, the right hand side here is a convex compact set. Then, the definition (73) gives (since λ_E^k is supported on E^k) :

$$s^k_{[\tau_0,\tau_1]} \in \operatorname{conv} Q_{\delta^k \varepsilon^k}[\tau_0, \tau_1]\,.$$

Now, assume that $\tau_0,\, \tau_1 \in \Theta, \quad \tau_0 < t_* < \tau_1$. Taking the limit as $k \to \infty$ in view of (74) and (68), we get

$$\hat{s}_{[\tau_0,\tau_1]} \in \operatorname{conv} Q(x,u)[\tau_0, \tau_1]\,.$$

Finally, taking the limit as $\tau_0 \to t_*\,, \quad \tau_1 \to t_*$ along Θ in view of (72) and (69), we obtain

$$\hat{s}(t_*) \in \operatorname{conv} S(\hat{x}(t_*),\, \operatorname{clm}(\hat{u})(t_*)). \tag{76}$$

Thus, the set $S(\hat{x}(t_*),\, \operatorname{clm}(\hat{u})(t_*))$ is nonempty, which by the definition (9) means that $t_* \in \mathcal{D}$. Since the point $t_* \in \mathcal{R}$ is arbitrary, it follows that $\mathcal{R} \subset \mathcal{D}$. Consequently, inclusion (71) holds.

d) For $t \in [t_0, t_1] \setminus \mathcal{R}$ we can redefine (if necessary) $\hat{s}(t)$ by zero, without violating the conditions of LMP. The lemma is proved. $\square$

Thus, in view of this lemma and (55), the adjoint equation (48) takes the form

$$-\, d\hat{p} \;=\; \hat{p} f_x(\hat{w}) + d\hat{\mu}_A + d\hat{\mu}_B \;=\; \hat{p} f_x(\hat{w}) + \hat{\lambda}\, G_x(\hat{w})\, dt + \hat{s}\, d\hat{\eta}, \tag{77}$$

i.e. condition (14) holds true.

Thus, all conditions (10)–(12) of the local minimum principle are satisfied. Theorem 3 is completely proved.

References

[1] A.Ya. Dubovitskii, A.A. Milyutin, *Extremum problems in the presence of restrictions,* USSR Comput. Math, and Math. Phys. v. 5 (1965), p. 1–80 (translated from Zh. Vychislit. Mat. i Mat. Fiz., 5 (1965), no. 3, 395–453, in Russian).

[2] A.Ya. Dubovitskii and A.A. Milyutin. *Necessary Conditions of a Weak Minimum in the General Optimal Control Problem*, Nauka, Moscow, 1971 (in Russian).

[3] A.A Milyutin. *Maximum Principle in the General Problem of Optimal Control*, Fizmatlit, Moscow, 2001 (in Russian).

[4] N. Dunford and J. Schwartz. *Linear Operators, Part 1: General Theory*, Wiley-Interscience, 1968, N.-Y., London.

[5] R.E. Edwards. *Functional Analysis*, Holt, Rienhart and Winston, New York, et al., 1965.

[6] M. Saadoune, M. Valadier. *Extraction of a good subsequence from a bounded sequence of integrable functions,* J. of Convex Analysis, v. 2 (1995), no. 1/2, 345-357.

[7] K. Makowsky, L.W. Neustadt, *Optimal control problems with mixed control-phase variable constraints,* SIAM J. on Control and Optimization, 1974, v. 12, no. 2, p. 184-228.

[8] A.A. Milyutin, *Maximum principle for the regular systems,* The Necessary Condition in Optimal Control (A.P. Afanas'ev, V.V. Dikusar, A.A. Milyutin, S.A. Chukanov; ed. A.A. Milyutin), Nauka, Moscow, 1990 (in Russian), Ch. 5, pp. 132–157.

[9] A.V. Dmitruk, *Maximum principle for a general optimal control problem with state and regular mixed constraints,* Comput. Math. Modeling, vol. 4 , pp. 364–377, 1993 (translated from Optimal'nost' Upravlyaemyh Dinamicheckih Sistem, M., Nauka, v. 14, p. 26–42, 1990, in Russian).

[10] U. Ledzewicz, *On abnormal optimal control problems with mixed equality and inequality constraints,* J. of Math. Analysis and Appl., v. 173, no. 1, pp. 18–42, 1993.

[11] R.F. Hartl, S.P. Sethi, R.G. Vickson. *A survey of the maximum principles for optimal control problems with state constraints,* SIAM Review, v. 37, no. 2, pp. 181–218, 1995.

[12] A.A. Milyutin, N.P. Osmolovskii. *Calculus of Variations and Optimal Control*, American Mathematical Society, Providence, Rhode Island, v.180, 1998, Part 1, Chapter 2 and Chapter 5.

[13] J.F. Bonnans, A. Hermant. *Second-order analysis for optimal control problems with pure state constraints and mixed control-state constraints*, Ann. Inst. H. Poincare, Nonlinear Analysis, v. 26, no. 2, 561–598, 2009.

[14] F.H. Clarke and MdR de Pinho, *Optimal control problems with mixed constraints,* SIAM J. Control Optim., vol. 48, no. 7, pp. 4500–4524, 2010.

[15] A.V. Dmitruk, N.P. Osmolovskii, *Necessary Conditions for a weak minimum in optimal control problems with integral equations subject to state and mixed constraints*, SIAM J. Control Optim., vol. 52, no. 6, pp. 3437–3462, 2014.

[16] A. Boccia, MdR de Pinho and R. Vinter, *Optimal control problems with mixed and pure state constraints,* SIAM J. Control Optim., vol. 54, no. 6, pp. 3061–3083, 2016.

[17] J.A. Becerril and M.D.R. de Pinho. *Optimal control with nonregular mixed constraints: an optimization approach,* SIAM J. Control Optim., vol. 59, no. 3, pp. 2093–2120, 2021.

[18] A.V. Dmitruk, *On the development of Pontryagin's Maximum Principle in the works of A. Ya. Dubovitskii and A. A. Milyutin,* Control and Cybernetics, 2009, v. 38, no. 4A, pp. 923–957.