Chattering variations, finitely additive measures, and the nonsmooth maximum principle with state space constraints

Héctor J. Sussmann

***Abstract*— We discuss the proof of a version of the maximum principle with state space constraints for data with very weak regularity properties, using the classical method of packets of needle variations (PNVs), as in Pontryagin's book, but coupling it with a nonclassical theory of multivalued differentials, the so-called "generalized differential quotients" (GDQs). The key technical point of our argument is the use of a different type of PNVs, that we call "chattering PNVs." These variations make it possible to get a conclusion involving finitely additive vector-valued measures of finite total variation. The theory presented here applies to control dynamics without uniqueness of trajectories (so that the flow maps are set-valued) and to differential inclusions (so that the "differentials" of maps are also set-valued).**

***Keywords*— Maximum Principle, state constraints, additive measures**

I. Introduction

Since the work of Milyutin and his collaborators in the 1970s, it has been clear that the correct formulation of the Pontryagin maximum principle with state space constraints must involve finitely additive vector-valued measures of finite total variation. In this paper—which should be regarded as a continuation of [1][1] where the reader will find the definitions of all the technical terms occuring in the statement of our main theorem—we focus on the role of "chattering PNVs" (where "PNV" stands for "packet of needle variations") in the proof of the main result of [1], stated there without proof. Here, we will not repeat the rather long list of definitions of [1], and we will instead refer the reader to the paper itself. We will, however, restate the theorem in full, and will then outline the proof, focusing on a detailed explanation of the main new technical point, namely, how chattering PNVs are defined and used and how they lead to finitely additive measures.

II. Chattering PNVs: an introduction

In a standard PNV, one is given a "reference control" $\eta_* : [a, b] \mapsto U$ (where U is the set of control values of our optimal control problem), and one specifies

(1) a finite sequence $\mathbf{t} = (\bar{t}_1, \ldots, \bar{t}_{\bar{p}})$ of distinct times, such that $a \le \bar{t}_1 < \bar{t}_2 < \cdots < \bar{t}_{\bar{p}} < b$;
(2) a sequence $\vec{\mathbf{u}} = (\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(\bar{p})})$ such that, for each index $j \in \{1, \ldots, \bar{p}\}$, $\mathbf{u}^{(j)}$ belongs to U^{p_j}, that is, $\mathbf{u}^{(j)}$ is a finite sequence $(u_1^j, \ldots, u_{p_j}^j)$ of control values.

If $p = \sum_{j=1}^{\bar{p}} p_j$, then one constructs a p-parameter variation—the PNV associated to $\mathbf{t}, \vec{\mathbf{u}}$—by letting $\eta^{\vec{\varepsilon}}$, for $\vec{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_p) \in \mathbb{R}_+^p$, be the control defined as follows:

a. We group the components ε_k of $\vec{\varepsilon}$ into $\bar{p}$ vectors $\vec{\varepsilon}^{(1)}, \ldots, \vec{\varepsilon}^{(\bar{p})}$ of dimensions $p_1, \ldots, p_{\bar{p}}$, by writing $\vec{\varepsilon}^{(j)} = (\varepsilon_1^j, \ldots, \varepsilon_{p_j}^j)$, where $\varepsilon_i^j = \varepsilon_{p_1 + \ldots + p_{j-1} + i}$.
b. We then assign to each pair (i, j) of indices such that $j \in \{1, \ldots, \bar{p}\}$ and $i \in \{1, \ldots, p_j\}$ the interval $I_i^j(\vec{\varepsilon}) = [t^j + \varepsilon_1^j + \cdots + \varepsilon_{i-1}^j, t^j + \varepsilon_1^j + \cdots + \varepsilon_i^j\,[$, so that, if $|\vec{\varepsilon}| = \varepsilon_1 + \ldots + \varepsilon_p$ is sufficiently small, then the $I_i^j(\vec{\varepsilon})$ are pairwise disjoint subintervals of $[a, b]$, and $I_i^j(\vec{\varepsilon})$ has length ε_i^j for each i, j.
c. Finally, we define $\eta^{\vec{\varepsilon}}$ by letting it be the control obtained from the reference control η_* by substituting the constant control value u_i^j for $\eta_*(t)$ for $t \in I_i^j(\vec{\varepsilon})$.

For a chattering PNV, we will do a similar thing, except that the variation involves in addition a positive-integer parameter N and, for $\vec{\varepsilon} \in \mathbb{R}_{p,+}$ such that $|\vec{\varepsilon}|$ is small enough, the controls $\eta^{\vec{\varepsilon}, N}$ are constructed by

(1) first defining $I^j(\vec{\varepsilon})$ to be, for each j, the interval $[t^j, t^j + |\vec{\varepsilon}^{(j)}|\,[$, whose length is $|\vec{\varepsilon}^{(j)}| = \varepsilon_1^j + \cdots + \varepsilon_{p_j}^j$;
(2) subdividing each $I^j(\vec{\varepsilon})$ into N subintervals $I^{j,\ell}(\vec{\varepsilon})$ of length $\frac{|\vec{\varepsilon}^{(j)}|}{N}$, so that if $\ell = 1, \ldots, N$ then $I^{j,\ell}(\vec{\varepsilon}) = \Big[t^j + (\ell - 1)\frac{|\vec{\varepsilon}^{(j)}|}{N}, t^j + \ell\frac{|\vec{\varepsilon}^{(j)}|}{N}\Big[$;
(3) subdividing each $I^{j,\ell}(\vec{\varepsilon})$ into p_j subintervals of lengths $\frac{\varepsilon_1^j}{N}, \ldots, \frac{\varepsilon_{p_j}^j}{N}$.

Once this is done, $\eta^{\vec{\varepsilon}, N}$ is the control obtained by substituting the constant control value u_i^j for $\eta_*(t)$ whenever $t \in I_i^{j,\ell}(\vec{\varepsilon})$ for some ℓ.

In other words, the classical PNV as well as the chattering one involve the substitution of a control u_i^j for the reference control on a set $S_i^j(\vec{\varepsilon})$ of measure ε_i^j located near $\bar{t}_j$, but in the classical PNV this set is itself an interval, whereas in the chattering PNV $S_i^j(\vec{\varepsilon})$ is the union of N intervals $I_i^{j,\ell}(\vec{\varepsilon})$ of length $\frac{\varepsilon_i^j}{N}$, evenly distributed on the interval $I_j(\vec{\varepsilon})$. The effect of this variation is that the trajectories $\xi^{\vec{\varepsilon}, N}$ corresponding to the controls $\eta^{\vec{\varepsilon}, N}$ can be well approximated, for large N, by trajectories of the convexified contol system. This approximation turns out to play a key role in the argument that leads to finitely additive measures, as we will explain in the last section of the paper.

Research supported in part by NSF Grant DMS01-03901

H. J. Sussmann is with the Department of Mathematics, Rutgers, the State University of New Jersey, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA sussmann@math.rutgers.edu

[1]Also available at the author's current papers Web page, *https://www.math.rutgers.edu/˜sussmann/currentpapers.html* .

III. The maximum principle

As in [1], we consider a *fixed time-interval optimal control problem with state space constraints*, of the form

$$
\begin{array}{ll}
\text{minimize} & \varphi(\xi(b)) + \int_a^b f_0(\xi(t),\eta(t),t)\,dt \\
\text{subject to} & \left\{ \begin{array}{l}
\xi(\cdot) \in W^{1,1}([a,b],X)\,, \\
\dot{\xi}(t) = f(\xi(t),\eta(t),t) \;\text{ a.e.}\,, \\
\xi(a) = \bar{x}_* \;\text{ and }\; \xi(b) \in S\,, \\
g_i(\xi(t),t) \le 0 \;\text{for}\; t\in[a,b],\; i=1,\ldots,m, \\
h_j(\xi(b)) = 0 \;\text{ for }\; j = 1,\ldots,\tilde{m}\,, \\
\eta(\cdot) \in \mathcal{U}\,,
\end{array} \right.
\end{array}
$$

and a *reference trajectory-control pair* (ξ_*,η_*).

We assume that the data 14-tuple

$$\mathcal{D} = (X, m, \tilde{m}, U, a, b, \varphi, f_0, f, \bar{x}_*, \mathbf{g}, \mathbf{h}, S, \mathcal{U})$$

satisfies the following conditions (using "FDNRLS" and "ppd" for "finite-dimensional normed real linear space" and "possibly partially defined," respectively):

- (H1) *X is a FDNRLS, $m \in \mathbb{Z}_+$, $\tilde{m} \in \mathbb{Z}_+$; U is a set, $a, b \in \mathbb{R}$, $a < b$, $\bar{x}_* \in X$ and $S \subseteq X$;*
- (H2) *f_0 is a ppd function from $X \times U \times \mathbb{R}$ to $\mathbb{R}$;*
- (H3) *f is a ppd function from $X \times U \times \mathbb{R}$ to X;*
- (H4) *$\mathbf{g} = (g_1,\ldots,g_m)$ is an m-tuple of ppd functions from $X \times \mathbb{R}$ to $\mathbb{R}$;*
- (H5) *$\mathbf{h} = (h_1,\ldots,h_{\tilde{m}})$ is an $\tilde{m}$-tuple of ppd functions from X to $\mathbb{R}$;*
- (H6) *φ is a ppd function from X to $\mathbb{R}$;*
- (H7) *$\mathcal{U}$ is a set of controllers.*

(A ***controller*** is a ppd function from $\mathbb{R}$ to U whose domain is a nonempty compact interval.)

An ***admissible controller*** is a member of $\mathcal{U}$. If $\alpha, \beta \in \mathbb{R}$ and $\alpha \le \beta$, then we use $W^{1,1}([\alpha,\beta],X)$ to denote the space of all absolutely continuous maps $\xi : [\alpha,\beta] \mapsto X$. A ***trajectory*** for a controller $\eta : [\alpha,\beta] \mapsto U$ is a map $\xi \in W^{1,1}([\alpha,\beta],X)$ such that, for almost every $t \in [\alpha,\beta]$, $(\xi(t),\eta(t),t)$ belongs to $\mathrm{Do}(f)$ and $\dot{\xi}(t) = f(\xi(t),\eta(t),t)$. A ***trajectory-control pair*** (abbr. TCP) is a pair (ξ,η) such that η is a controller and ξ is a trajectory for η. The ***domain*** of a TCP (ξ,η) is the domain of η, which is, by definition, the same as domain of ξ. A TCP (ξ,η) is ***admissible*** if $\eta \in \mathcal{U}$.

A TCP (ξ,η) with domain $[\alpha,\beta]$ is ***cost and constraint admissible*** if (i) (ξ,η) is admissible, (ii) the function $[\alpha,\beta] \ni t \mapsto f_0(\xi(t),\eta(t),t)$ is a. e. defined and measurable, (iii) $\int_\alpha^\beta \min\Big(0, f_0(\xi(t),\eta(t),t)\Big)\,dt > -\infty$, (iv) $\xi(\beta) \in \mathrm{Do}(\varphi)$, and (v) ξ satisfies all our state space constraints, that is (using $\mathrm{Do}(\cdot)$ for "domain of"),

- (CA1) $\xi(\alpha) = \bar{x}_*$ *and* $\xi(\beta) \in S \cap \Big(\cap_{j=1}^{\tilde{m}} \mathrm{Do}(h_j)\Big)$
- (CA2) $(\xi(t),t) \in \mathrm{Do}(g_i)$ *and* $g_i(\xi(t),t) \le 0$ *for all* $t \in [\alpha,\beta]$, *and all* $i \in \{1,\ldots,m\}$,
- (CA3) $h_j(\xi(\beta)) = 0$ *for* $j = 1,\ldots,\tilde{m}$.

We use $ADM(\mathcal{D})$ and $ADM_{[a,b]}(\mathcal{D})$ to denote the sets of (i) all cost and constraint admissible TCPs (ξ,η), and (ii) all $(\xi,\eta) \in ADM(\mathcal{D})$ whose domain is $[a,b]$.

It follows that if $(\xi,\eta) \in ADM_{[a,b]}(\mathcal{D})$ then the number $J(\xi,\eta) = \varphi(\xi(\beta)) + \int_\alpha^\beta f_0(\xi(t),\eta(t),t)\,dt$—called the ***cost*** of (ξ,η)—is well defined and belongs to $]-\infty,+\infty]$.

The hypothesis on the reference TCP (ξ_*,η_*) is that it is a cost-minimizer in $ADM_{[a,b]}(\mathcal{D})$. In other words,

- (H8) $(\xi_*,\eta_*) \in ADM_{[a,b]}(\mathcal{D})$, $J(\xi_*,\eta_*) < +\infty$, *and* $J(\xi_*,\eta_*) \le J(\xi,\eta)$ *for all members* (ξ,η) *of* $ADM_{[a,b]}(\mathcal{D})$.

The "cost-augmented dynamics" $\mathbf{f}$ and the "epi-augmented dynamics" $\check{\mathbf{f}}$ are the set-valued maps from $X \times U \times \mathbb{R}$ to $\mathbb{R} \times X$ such that $\mathrm{Do}(\mathbf{f}) = \mathrm{Do}(\check{\mathbf{f}}) = \mathrm{Do}(f_0) \cap \mathrm{Do}(f)$ and, for $z = (x,u,t) \in X \times U \times \mathbb{R}$,

$$\mathbf{f}(z) = \{(f_0(z), f(z))\} \;\text{ and }\; \check{\mathbf{f}}(z) = [f_0(z), +\infty[\times \{f(z)\}$$

(so $\mathbf{f}$ is actually single-valued).

We will also use the ***constraint indicator maps*** $\chi^{co}_{g_i} : X \times \mathbb{R} \mapsto\!\!\!\!\rightarrow \mathbb{R}$, for $i = 1,\ldots,m$, and the ***epifuncion*** $\check{\varphi} : X \mapsto\!\!\!\!\rightarrow \mathbb{R}$, defined as follows (where "$A : B \mapsto\!\!\!\!\rightarrow C$" stands for "$A$ is a set-valued map from B to C"):

- $\chi^{co}_{g_i}(x,t) = \emptyset$ if $g_i(x,t) \le 0$ or $(x,t) \notin \mathrm{Do}(g_i)$, and $\chi^{co}_f(x,t) = [0,+\infty[$ if $g(x,t) > 0$.
- $\check{\varphi}(x) = \{\varphi(x) + v : v \in \mathbb{R}, v \ge 0\}$ if $x \in \mathrm{Do}(\varphi)$, and $\check{\varphi}(x) = \emptyset$ if $x \notin \mathrm{Do}(\varphi)$.

For $i \in \{1,\ldots,m\}$, we let

$$
\begin{array}{l}
\sigma^{\mathbf{f}}_*(t) = \mathbf{f}(\xi_*(t),\eta_*(t),t) \;\text{ and }\; \sigma^{g_i}_*(t) = 0 \quad \text{if } t \in [a,b], \\
Av_{g_i} = \{(x,t) \in X \times [a,b] : g_i(x,t) > 0\}\,,
\end{array}
$$

(so the Av_{g_i} are the "sets to be avoided"). We then define K_i to be the set of all $t \in [a,b]$ such that $(\xi_*(t),t)$ belongs to the closure of Av_{g_i}. Then K_i is obviously a compact subset of $[a,b]$..

We now make technical hypotheses on $\mathcal{D}$, ξ_*, η_*, and five new objects called $\Lambda^{\mathbf{f}}$, $\Lambda^{\mathbf{g}}$, $\Lambda^{\mathbf{h}}$, Λ^{φ}, and C. To state these hypotheses, we let $\mathcal{U}_{c;[a,b]}$ denote the set of all constant U-valued functions defined on $[a,b]$, and define $\mathcal{U}_{c;[a,b];*} = \mathcal{U}_{c;[a,b]} \cup \{\eta_*\}$. We use $\mathcal{T}^X(\xi_*,\delta)$ to denote the tube $\{(x,t) \in X \times [a,b] : \|x - \xi_*(t)\| \le \delta\}$, and write $f_\eta(x,t) = f(x,\eta(t),t)$, $f_{0,\eta}(x,t) = f_0(x,\eta(t),t)$, and $\mathbf{f}_\eta(x,t) = \mathbf{f}(x,\eta(t),t)$. We use $\mathcal{L}(X)$ to denote the set of all linear maps from X to X.

- (H9) *For each $\eta \in \mathcal{U}_{c;[a,b];*}$. there exist a positive number δ_η such that*
  - (H9.a) *$\mathbf{f}_\eta(x,t)$ is defined whenever (x,t) belongs to $\mathcal{T}^X(\xi_*,\delta_\eta)$,*
  - (H9.b) *the map f_η is co-IBIC[2] on $\mathcal{T}^X(\xi_*,\delta_\eta)$, and the function $\mapsto f_{0,\eta}$ is co-ILBILSC on $\mathcal{T}^X(\xi_*,\delta_\eta)$.*
- (H10) *The number δ_{η_*} can be chosen so that (i) each function g_i is defined on $\mathcal{T}^X(\xi_*,\delta_{\eta_*})$, and (ii) for each $i \in \{1,\ldots,m\}$, $t \in [a,b]$, the set $\{x \in X : g_i(x,t) > 0, \|x - \xi_*(t)\| \le \delta_{\eta_*}\}$ is relatively open in the ball $\{x \in X : \|x - \xi_*(t)\| \le \delta_{\eta_*}\}$,*

[2] "co-IBIC" and "co-ILBILSC" stand for "co-integrably bounded integrally continuous" and "co-integrably lower bounded integrally lower semicontinuous," respectively. These concepts are defined in [1].

(H11) $\Lambda^{\mathbf{f}}$ *is a measurable integrably bounded set-valued map from* $[a,b]$ *to* $X^\dagger \times \mathcal{L}(X)$ *with compact convex values such that*[3] $\Lambda^{\mathbf{f}} \in VG^{L^1,ft}_{GDQ}(\check{\mathbf{f}};[a,b];\xi_*,\sigma^{\mathbf{f}}_*;X\times\mathbb{R})$,

(H12) $\Lambda^{\mathbf{g}}$ *is an* m*-tuple* $(\Lambda^{g_1},\ldots,\Lambda^{g_m})$ *such that, for each* $i\in\{1,\ldots,m\}$, Λ^{g_i} *is an upper semicontinous set-valued map from* $[a,b]$ *to* $X^\dagger$ *with compact convex values, such that* $\Lambda^{g_i}\in VG^{pw,rob}_{GDQ}(\chi^{co}_{g_i};\xi_*,\sigma^{g_i}_*,Av_{g_i})$,

(H13) $\Lambda^{\mathbf{h}}$ *is a generalized differential quotient*[4] *of* $\mathbf{h}$ *at* $(\xi_*(b),\mathbf{h}(\xi_*(b)))$ *in the direction of* X.

(H14) Λ^{φ} *is a generalized differential quotient of the epifunction* $\check{\varphi}$ *at* $(\xi_*(b),\varphi(\xi_*(b)))$ *in the direction of* X,

(H15) C *is a limiting Boltyanskii approximating cone of* S *at* $\xi_*(b)$.

In our last hypothesis, we use the abbreviation ETIVN for "equal-time interval-variational neighborhood," and assume

(H16) *The class* $\mathcal{U}$ *is an ETIVN*[5] *of* η_*.

We are now almost ready to state our version of the maximum principle. All we need is a few preliminary definitions.

First, we define the ***Hamiltonian*** to be the ppd function H_α from $X{\times}U{\times}X^\dagger{\times}\mathbb{R}$ to $\mathbb{R}$ (depending on a real parameter α) given by $H_\alpha(x,u,p,t)=p\cdot f(x,u,t)-\alpha f_0(x,u,t)$.

Next, we use $Int([a,b])$ to denote the set of all real intervals J such that $J\subseteq[a,b]$. (So $J\in Int([a,b])$ if and only if J is a connected subset of $[a,b]$.) We let $PDSeq([a,b])$ denote the set of all finite sequences of pairwise disjoint members of $Int([a,b])$, and write $PDSeq^{\#}([a,b])$ to denote the set of all members $(J_1,\ldots,J_m)$ of $PDSeq([a,b])$ such that $J_1\cup\cdots\cup J_m$ belongs to $\mathrm{Int}([a,b])$. If Y is a FDNRLS, a ***finitely additive*** Y***-valued interval set function on*** $[a,b]$ (or "additive measure on $[a,b]$") is a map $\mu: Int([a,b])\mapsto Y$ having the property that $\mu(J_1\cup\cdots\cup J_m)=\mu(J_1)+\cdots+\mu(J_m)$ whenever $m\in\mathbb{Z}$, $m>0$, and $(J_1,\ldots,J_m)$ belongs to $PDSeq^{\#}([a,b])$. The ***total variation*** of μ is the supremum $\|\mu\|_{tv}$ of the real numbers $\|\mu(J_1)\|+\cdots+\|\mu(J_m)\|$, ranging over all $(J_1,\ldots,J_m)\in PDSeq([a,b])$. We say that μ is of ***of bounded variation*** if $\|\mu\|_{tv}<\infty$. We say that μ is ***nonnegative*** if $Y=\mathbb{R}$ and $\mu(J)\geq 0$ for every $J\in\mathrm{Int}([a,b])$. (Then every nonnegative μ is of bounded variation and satisfies $\|\mu\|_{tv}=\mu([a,b])$.)

If Y is a FDNRLS, we use $Bvadd([a,b],Y)$ to denote the set of all additive Y-valued interval set functions on $[a,b]$ that are of bounded variation, and write $bvadd([a,b],Y)$ to denote the subset of $Bvadd([a,b],Y)$ whose members are the $\mu\in Bvadd([a,b],Y)$ such that $\mu(\{t\})=0$ for every $t\in[a,b]$. We let $bvadd_+([a,b])$ denote the set of all $\mu\in bvadd([a,b],\mathbb{R})$ that are nonnegative.

If $\nu\in bvadd_+([a,b])$ and Y is a FDNRLS, we would like to be able to multiply ν by a bounded Borel measurable map $\gamma:[a,b]\mapsto Y$ and obtain a $\mu\in bvadd([a,b],Y)$ such that, formally, $d\mu=\gamma\cdot d\nu$. It turns out that this is not quite the right thing to do, and that what really can be done is multiply ν by a "bounded Borel measurable pair," as we now explain. If Z is any FDNRLS, then every $\mu\in bvadd([a,b],Z)$ has a unique decomposition $\mu=\mu_{at,-}+\mu_{at,+}+\mu_c$, where μ_c is a continuous (i.e., nonatomic) countably additive Z-valued Borel measure on $[a.b]$, and $\mu_{at,-}$, $\mu_{at,+}$ are, respectively, left-atomic and right-atomic members of $bvadd([a,b],Z)$. (Given $t\in[a,b[$, and $\bar{z}\in Z$, the ***right delta function at*** t ***with value*** $\bar{z}$ is the member $\delta^{\bar{z}}_{+,t}$ of $bvadd([a,b],Z)$ such that $\delta^{\bar{z}}_{+,t}(J)=\bar{z}$ if the interval J contains the set $]t,t+\varepsilon[$ for some positive ε, and $\delta^{\bar{z}}_{+,t}(J)=0$ otherwise[6]. A $\mu\in bvadd([a,b],Z)$ is ***right-atomic*** if it is the sum of a series of right delta functions, converging in the total variation norm. The ***left delta functions*** $\delta^{\bar{z}}_{-,t}$ are defined in a similar way, for $t\in]a,b]$, and then it is clear what is meant by a ***left-atomic*** $\mu\in bvadd([a,b],Z)$.) Let us define a ***bounded measurable pair of*** Y***-valued maps on*** $[a,b]$ to be an ordered pair $\gamma=(\gamma_-,\gamma_+)$ of bounded Borel measurable Y-valued maps on $[a,b]$ such that $\gamma_-(t)=\gamma_+(t)$ for all t in the complement of a finite or countable set. Given such a pair $\gamma=(\gamma_-,\gamma_+)$, and a $\nu\in bvadd_+([a,b])$, the ***product*** $\mu=\gamma\cdot\nu\in bvadd([a,b],Y)$ is given by $\mu=\mu_{at,-}+\mu_{at,+}+\mu_c$, where $\mu_{at,-}$, $\mu_{at,+}$, μ_c are defined, in terms of the canonical decomposition $\nu=\nu_{at,-}+\nu_{at,+}+\nu_c$, by letting μ_c be the product of ν_c with γ_- or[7] γ_+, while $\mu_{at,-}$, $\mu_{at,+}$ are obtained by multplying $\nu_{at,-}$, $\nu_{at,+}$ by γ_- and γ_+, respectively[8].

We are now, finally, ready to state the main result.

Theorem 3.1: Assume that (H1-16) hold, and let $I=\{i\in\{1,\ldots,m\}: K_i\neq\emptyset\}$. Then there exist

1. a covector $\bar{\pi}\in X^\dagger$, a nonnegative real number π_0, and an $\tilde{m}$-tuple $\boldsymbol{\lambda}=(\lambda_1,\ldots,\lambda_{\tilde{m}})$ of real numbers,
2. a measurable map $[a,b]\ni t\mapsto(L_0(t),L(t))\in\Lambda^{\mathbf{f}}(t)$,
3. bounded Borel measurable pairs $\gamma^i=(\gamma^i_-,\gamma^i_+)$ of selections of the set-valued maps Λ^{g_i}, defined on K_i,

[3] The complicated expressions $VG^{L^1,ft}_{GDQ}(\check{\mathbf{f}};[a,b];\xi_*,\sigma^{\mathbf{f}}_*;X\times\mathbb{R})$ and $VG^{pw,rob}_{GDQ}(\chi^{co}_{g_i};\xi_*,\sigma^{g_i}_*,Av_{g_i})$ refer, respectively, to the set of all L^1 fixed-time GDQ variational generators of $\check{f}$ along $(\xi_*,\sigma^{\mathbf{f}}_*)$ in the direction of $X\times\mathbb{R}$ and the set of all pointwise robust GDQ variational generators of $\chi^{co}_{g_i}$ along $(\xi_*,\sigma^{g_i}_*)$ in the direction of Av_{g_i}. These concepts are defined in [1].

[4] Generalized differential quotients (GDQs) and limiting Boltyanski approximating cones are defined in [1].

[5] Again, the concept of ETIVN is defined in [1]. What (H16) means is, essentially, that whenever we consider a packet of needle variations of η_*, the controls corresponding to a parameter $\vec{\varepsilon}$ are admissible as long as $|\vec{\varepsilon}|$ is small enough.

[6] Notice that $\delta^{\bar{z}}_{+,t}([t,t+\varepsilon])=\bar{z}$ for every positive ε; hence $\lim_{j\to\infty}\delta^{\bar{z}}_{+,t}([t,t+2^{-j}])=0$, while on the other hand $\delta^{\bar{z}}_{+,t}(\{t\})=0$. So $\delta^{\bar{z}}_{+,t}$ is not countably additive.

[7] The products $\gamma_-\cdot\nu$ and $\gamma_+\cdot\nu$ are equal, because $\gamma_--\gamma_+$ vanishes outside a countable set, and ν_c is nonatomic.

[8] The reason for considering pairs (γ_-,γ_+) is, of course, that ν may contain both a left atom and a right atom at a point t, and in that case one should allow these two atoms to be multiplied by two *different* vectors $\gamma_-(t)$, $\gamma_+(t)$.

for $i \in I$,

4. a member $L^{\mathbf{h}} = (L^{h_1}, \dots, L^{h_{\tilde{m}}}) \in (X^\dagger)^{\tilde{m}}$ of $\Lambda^{\mathbf{h}}$,
5. a member L^φ of Λ^φ,
6. a family $\{\nu_i\}_{i \in I}$ of nonnegative finitely additive interval set functions $\nu_i \in bvadd_+([a,b])$ such that $\mathrm{support}(\nu_i) \subseteq K_i$ for every $i \in I$,

having the property that, if we define $\mu_i = \gamma_i \cdot \nu_i$, and let $\pi : [a,b] \mapsto X^\dagger$ be the unique solution of the adjoint Cauchy problem

$$\begin{cases} d\pi(t) = (-\pi(t)L(t) + \pi_0 L_0(t))dt + \sum_{i \in I} d\mu_i(t) \\ \pi(b) = \bar{\pi} - \sum_{j=1}^{\tilde{m}} \lambda_j L_j^{\mathbf{h}} - \pi_0 L^\varphi \end{cases} \quad (1)$$

then the following three conditions are satisfied:

I. the ***Hamiltonian maximization condition***:

$$H_{\pi_0}(\xi_*(\bar{t}), \eta_*(\bar{t}), \pi(\bar{t})) \geq H_{\pi_0}(\xi_*(\bar{t}), u, \pi(\bar{t}))$$

whenever $u \in U$, $\bar{t} \in [a,b]$ are such that $(\xi_*(\bar{t}), \bar{t})$ is a point of approximate continuity[9] of both augmented vector fields $(x,t) \mapsto \mathbf{f}(x,u,t)$ and $(x,t) \mapsto \mathbf{f}(x, \eta_*(t), t)$,

II. the ***transversality condition***, which asserts that $-\bar{\pi}$ belongs to $C^\dagger$ (where $C^\dagger$ is the polar cone of C, i.e., $C^\dagger = \{\omega \in X^\dagger : \omega \cdot c \leq 0 \text{ whenever } c \in C\}$),

III. the ***nontriviality condition***

$$\|\bar{\pi}\| + \pi_0 + \sum_{j=1}^{\tilde{m}} |\lambda_j| + \sum_{i \in I} \|\nu_i\| > 0\,.$$

Remark 3.2: The precise interpretation of the adjoint equation (1) is as follows: define $\mu_i^{cd}(t) = -\mu_i([t,b])$, so that μ_i^{cd} is the "cumulative distribution" of μ_i (in the sense that, for example, $\mu_i^{cd}(s) + \mu_i([s,t]) = \mu_i^{cd}(t)$ whenever $a \leq s \leq t \leq b$), normalized so that $\mu_i^{cd}(b) = 0$. Let $\mu^{cd} = \sum_{i \in I} \mu_i^{cd}$. Then π is a solution of (1) if and only if

$$\pi(t) - \mu^{cd}(t) \;=\; \bar{\pi} - \sum_{j=1}^{\tilde{m}} \lambda_j L_j^{\mathbf{h}} - \pi_0 L^\varphi - \pi_0 \int_t^b L_0(s)\,ds + \int_t^b (\pi(s) - \mu^{cd}(s)) \cdot L(s)\,ds + \int_t^b \mu^{cd}(s) \cdot L(s)\,ds$$

for all $t \in [a,b]$. Equivalently, if we write $\pi^*(t) = \pi(t) - \mu^{cd}(t)$, then π^* (which is absolutely continuous) must be a solution of the differential equation $\dot{\pi}^*(t) = -\pi^*(t) \cdot L(t) + \pi_0 L_0(t) - \mu^{cd}(t) \cdot L(t)$, with terminal value $\pi^*(b) = \pi(b) = \bar{\pi} - \sum_{j=1}^{\tilde{m}} \lambda_j L_j^{\mathbf{h}} - \pi_0 L^\varphi$. ■

IV. Outline of the proof of Theorem 3.1

For simplicity, we make the additional assumption[10] that $K_i \neq \emptyset$ for all $i \in \{1, \dots, m\}$, so that the set I occurring in the statement of Theorem 3.1 is $\{1, \dots, m\}$.

Packets of chatttering needle variations. Let $\mathcal{V}$ be the set of all pairs $(u, \bar{t})$ such that $u \in U$, $\bar{t} \in [a,b]$, and $(\xi_*(\bar{t}), \bar{t})$ is a point of approximate continuity of both augmented vector fields $(x,t) \mapsto \mathbf{f}(x,u,t)$ and $(x,t) \mapsto \mathbf{f}(x, \eta_*(t), t)$.

Fix a pair $(\mathbf{V}, \prec)$ such that (i) $\mathbf{V}$ is a nonempty finite subset of $\mathcal{V}$, and (ii) $\prec$ is a total ordering of $\mathbf{V}$ which is time-compatible, in the sense that $(u,t) \prec (u',t')$ whenever $(u,t) \in \mathbf{V}$, $(u',t') \in \mathbf{V}$, and $t < t'$. For $(\mathbf{V}, \prec)$, we define a "packet of chattering needle control variations" as follows. First let p be the cardinality of $\mathbf{V}$, and let $\mathbf{T}(\mathbf{V})$ be the set of all times t that occur in $\mathbf{V}$, so that $t \in \mathbf{T}(\mathbf{V})$ if and only if there exists $u \in U$ such that $(u,t) \in \mathbf{V}$. Let $(\bar{t}_1, \dots, \bar{t}_{\bar{p}})$ be the strictly increasing sequence of members of $\mathbf{T}(\mathbf{V})$. We set $\bar{t}_0 = a$, $\bar{t}_{\bar{p}+1} = b$, so that the $\bar{t}_j$ satisfy $a = \bar{t}_0 \leq \bar{t}_1 < \bar{t}_2 < \cdots < \bar{t}_{\bar{p}} < \bar{t}_{\bar{p}+1} = b$. We let $\bar{r} = \min\{\bar{t}_{j+1} - \bar{t}_j : j = 1, \dots, \bar{p}\}$, so that $\bar{r} \leq \bar{t}_{j+1} - \bar{t}_j$ for $j = 1, \dots, \bar{p}$.

For each index $j \in \{1, \dots, \bar{p}\}$, we use $\mathbf{V}[\bar{t}_j]$ to denote the set of all pairs $(u,t) \in \mathbf{V}$ such that $t = \bar{t}_j$. We then define p_j to be the cardinality of $\mathbf{V}[\bar{t}_j]$ (so $p = \sum_{j=1}^{\bar{p}} p_j$) and let $\Big((u_j^1, \bar{t}_j), (u_j^2, \bar{t}_j), \dots, (u_j^{p_j}, \bar{t}_j)\Big)$ be the $\prec$-ordered sequence of members of $\mathbf{V}[\bar{t}_j]$.

We identify the space $\mathbb{R}_p$ with the Cartesian product $\mathbb{R}_{p_1} \times \mathbb{R}_{p_2} \times \cdots \times \mathbb{R}_{p_{\bar{p}}}$. Hence, if $\vec{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_p) \in \mathbb{R}_p$, we can also write $\vec{\varepsilon} = (\vec{\varepsilon}_{(1)}, \dots, \vec{\varepsilon}_{(\bar{p})})$, where $\vec{\varepsilon}_{(j)} = (\varepsilon_j^1, \dots, \varepsilon_j^{p_j}) \in \mathbb{R}_{p_j}$ for $j = 1, \dots, \bar{p}$, so that $\varepsilon_j^\ell = \varepsilon_{p_1 + \cdots + p_{j-1} + \ell}$ for $j \in \{1, \dots, \bar{p}\}$, $\ell \in \{1, \dots, p_j\}$. We define

$$\mathbb{R}_{p,+} = \{\vec{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_p) \in \mathbb{R}_p : \varepsilon_j \geq 0 \text{ for } j = 1, \dots, p\}$$

and, for each positive real number r, we let $\mathcal{S}_p(r)$ be the simplex $\{\vec{\varepsilon} \in \mathbb{R}_{p,+} : |\vec{\varepsilon}| \leq r\}$, where $|\vec{\varepsilon}| \stackrel{\mathrm{def}}{=} \varepsilon_1 + \cdots + \varepsilon_p$.

For $\vec{\varepsilon} \in \mathcal{S}_p(\bar{r})$, we define intervals $I_j(\vec{\varepsilon})$, for $j \in \{1, \dots, \bar{p}\}$, by letting $I_j(\vec{\varepsilon}) \stackrel{\mathrm{def}}{=} [\bar{t}_j, \bar{t}_j + |\vec{\varepsilon}_{(j)}|\,[$. We remark that the intervals $I_j(\vec{\varepsilon})$, for $j \in \{1, \dots, \bar{p}\}$, are contained in $[a,b]$ and pairwise disjoint, since $|\vec{\varepsilon}_{(j)}| \leq |\vec{\varepsilon}| \leq \bar{r} \leq \bar{t}_{j+1} - \bar{t}_j$. For $N \in \mathbb{N}$ and $k \in \{1, \dots, N\}$, we define times $\bar{t}_j^{N,k}(\vec{\varepsilon}) \in [a,b]$ and subintervals $I_j^{N,k}(\vec{\varepsilon})$ of $[a,b]$ by letting

$$\bar{t}_j^{N,k}(\vec{\varepsilon}) \stackrel{\mathrm{def}}{=} \bar{t}_j + \frac{k|\vec{\varepsilon}_{(j)}|}{N}\,, \quad I_j^{N,k}(\vec{\varepsilon}) \stackrel{\mathrm{def}}{=} \Big[\bar{t}_j^{N,k-1}(\vec{\varepsilon}), \bar{t}_j^{N,k}(\vec{\varepsilon})\Big[\,,$$

so the $I_j^{N,k}(\vec{\varepsilon})$, as k varies from 1 to N, constitute a subdivision of $I_j(\vec{\varepsilon})$ into N equal subintervals of length $\frac{|\vec{\varepsilon}_{(j)}|}{N}$. Finally, for each $j \in \{1, \dots, \bar{p}\}$, $N \in \mathbb{N}$, $k \in \{1, \dots, N\}$, we divide the interval $I_j^{N,k}(\vec{\varepsilon})$ into p_j subintervals $I_j^{N,k,\ell}(\vec{\varepsilon})$ of length $\frac{\varepsilon_j^\ell}{N}$, by letting $\bar{t}_j^{N,k,\ell}(\vec{\varepsilon}) \stackrel{\mathrm{def}}{=} \bar{t}_j^{N,k}(\vec{\varepsilon}) + \frac{1}{N}\sum_{\lambda=1}^{\ell} \varepsilon_j^\lambda$ and $I_j^{N,k,\ell}(\vec{\varepsilon}) \stackrel{\mathrm{def}}{=} \Big[\bar{t}_j^{N,k-1,\ell-1}(\vec{\varepsilon}), \bar{t}_j^{N,k-1,\ell}(\vec{\varepsilon})\Big[$ for $\ell = 0, \dots, p_j$.

For $\vec{\varepsilon} \in \mathcal{S}_p(\bar{r})$, $N \in \mathbb{N}$, we define a control $\eta^{\vec{\varepsilon},N}$, by letting $\eta^{\vec{\varepsilon},N}$ be the U-valued function on $[a,b]$ given by

$$\begin{aligned} \eta^{\vec{\varepsilon},N}(t) &= u_j^\ell \ \text{ if } \ t \in I_j^{N,k,\ell}(\vec{\varepsilon})\,, \\ \eta^{\vec{\varepsilon},N}(t) &= \eta_*(t) \ \text{ if } \ t \in [a,b] \backslash \mathbf{I}(\vec{\varepsilon})\,. \end{aligned}$$

where $\mathbf{I}(\vec{\varepsilon}) \stackrel{\mathrm{def}}{=} \bigcup_{j=1}^{\bar{p}} I_j(\vec{\varepsilon})$.

[9]This concept is defined in [1]

[10]It is a trivial exercise to get rid of this assumption.

It follows from (H16) that we can pick N-dependent numbers $\hat{r}^N \in]0,\bar{r}]$ such that $\eta^{\vec{\varepsilon},N} \in \mathcal{U}$ whenever $|\vec{\varepsilon}| \le \hat{r}^N$. For each $\vec{\varepsilon} \in \mathcal{S}_p(\bar{r})$, let $\mathcal{N}(\vec{\varepsilon})$ be the set of all $N \in \mathbb{N}$ such that $|\vec{\varepsilon}| \le \hat{r}^N$.

Definition 4.1: The set-valued map $\boldsymbol{\eta}^{\mathbf{V},\prec}$ that assigns to each $\vec{\varepsilon} \in \mathcal{S}_p(\bar{r})$ the set $\{\eta^{\vec{\varepsilon},N} : N \in \mathcal{N}(\vec{\varepsilon})\}$ is the ***packet of chattering needle control variations*** corresponding to the pair $(\mathbf{V},\prec)$. ■

Definition 4.2: The N-***chattering parameter-to-trajectory map*** corresponding to the pair $(\mathbf{V},\prec)$ and the positive integer N is the set-valued map $\Xi^{\mathbf{V},\prec,N} : \mathcal{S}_p(\hat{r}^N) \mapsto C^0([a,b],X)$ that assigns to each $\vec{\varepsilon} \in \mathcal{S}_p(\hat{r}^N)$ the set $\Xi^{\mathbf{V},\prec,N}(\vec{\varepsilon})$ of all absolutely continuous maps $\xi : [a,b] \mapsto X$ that satisfy the conditions (i) $\xi(a) = \bar{x}$, and (ii) $\dot{\xi}(t) = f(\xi(t),\eta^{\vec{\varepsilon},N}(t),t)$ for almost all $t \in [a,b]$.

The set-valued map $\Xi^{\mathbf{V},\prec} : \mathcal{S}_p(\bar{r}) \mapsto C^0([a,b],X)$ that assigns to each $\vec{\varepsilon} \in \mathcal{S}_p(\bar{r})$ the set $\bigcup_{N \in \mathcal{N}(\vec{\varepsilon})} \Xi^{\mathbf{V},\prec,N}(\vec{\varepsilon})$ is the ***combined parameter-to-trajectory map*** corresponding to the pair $(\mathbf{V},\prec)$. ■

We now define $\mathcal{X} = \mathbb{R} \times X \times \mathbb{R}_m \times \mathbb{R}_{\tilde{m}}$ and $\tilde{\mathcal{X}} = (\mathbb{R} \cup \{+\infty\}) \times X \times \mathbb{R}_m \times \mathbb{R}_{\tilde{m}}$, and introduce set-valued maps $\mathcal{E}^{\mathbf{V},\prec,N} : \mathcal{S}_p(\hat{r}^N) \mapsto \tilde{\mathcal{X}}$ (called the ***N-chattering augmented endpoint maps***) by letting $\mathcal{E}^{\mathbf{V},\prec,N}(\vec{\varepsilon})$ be, for $\vec{\varepsilon} \in \mathcal{S}_p(\hat{r}^N)$, the set of all 4-tuples $(x_0,x,\vec{w},\vec{z}) \in \tilde{\mathcal{X}}$ such that, for some trajectory $\xi \in \Xi^{\mathbf{V},\prec,N}(\vec{\varepsilon})$, the following conditions hold:

$$x_0 \ \ge\ \varphi(\xi(b)) + \int_a^b f_0(\xi(t),\eta^{\vec{\varepsilon},N}(t),t)\,dt\,, \tag{2}$$

$$x \ =\ \xi(b)\,, \tag{3}$$

$$w_i \ \ge\ 0 \quad \text{if} \quad i \in \{1,\ldots,m\} \ \text{and} \ \sup\{g_i(\xi(t),t) : t \in [a,b]\} > 0\,, \tag{4}$$

$$\vec{z} \ =\ \mathbf{h}(\xi(b))\,, \tag{5}$$

where $\vec{w} = (w_1,\ldots,w_m)$. In addition, we also define the ***combined augmented parameter-to-trajectory map*** corresponding to the pair $(\mathbf{V},\prec)$ to be the set-valued map $\mathcal{E}^{\mathbf{V},\prec} : \mathcal{S}_p(\bar{r}) \mapsto \tilde{\mathcal{X}}$ such that $\mathcal{E}^{\mathbf{V},\prec}(\vec{\varepsilon}) = \bigcup_{N \in \mathcal{N}(\vec{\varepsilon})} \mathcal{E}^{\mathbf{V},\prec,N}(\vec{\varepsilon})$ whenever $\vec{\varepsilon} \in \mathcal{S}_p(\bar{r})$.

The separation property. The crucial property of the set-valued map $\mathcal{E}^{\mathbf{V},\prec}$ is the following separation result. In the statement, $\hat{S}$ is the subset of $\mathcal{X}$ given by

$$\hat{S} = \Big\{(x_0,x,\vec{w},z) : x_0 < x_{0,*}\,,\ x \in S\,,\ \vec{w} < \vec{0}_m\,,\ \vec{z} = \vec{0}_{\tilde{m}}\Big\}\,,$$

where $\vec{0}_\nu$ is, for any ν, the origin of $\mathbb{R}_\nu$, and "$\vec{w} < \vec{0}_m$" means "the inequality $w_j < 0$ holds for $j = 1,\ldots,m$, if $\vec{w} = (w_1,\ldots,w_m)$." In addition, $x_{0,*}$ is the reference cost, so that $x_{0,*} = \varphi(\xi_*(b)) + \int_a^b f_0(\xi_*(t),\eta_*(t),t)$, and $x_{0,*} \in \mathbb{R}$, because of (H8).

The following fact is then a trivial corollary of the optimality of our reference trajectory-control pair.

Proposition 4.3: The image $\mathcal{E}^{\mathbf{V},\prec}\Big(\mathcal{S}_p(\bar{r})\Big)$ does not intersect the set $\hat{S}$. ■

Construction of a GDQ of the augmented endpoint map. The crucial point of the proof of our theorem is to find a GDQ of $\mathcal{E}^{\mathbf{V},\prec}$ at the point $P_{*,p} = \Big(\vec{0}_p,(x_{0,*},\xi_*(b),\vec{0}_m,\vec{0}_{\tilde{m}})\Big) \in \mathbb{R}_p \times \mathcal{X}$ in the direction of the nonnegative orthant $\mathbb{R}_{+,p}$ of $\mathbb{R}_p$. To do this, we first construct a set $\mathcal{G}$ of linear maps from $\mathbb{R}_p$ to $\mathcal{X}$. For each $i \in \{1,\ldots,m\}$, we let $\mathcal{M}_i$ be the set of all measures $\mu \in bvadd([a,b],X^\dagger)$ such that μ is of the form $\gamma \cdot \nu$, for some finitely additive probability measure ν on $[a,b]$ supported by K_i and some bounded Borel measurable pair γ of selections of Λ^{g_i}. For $j = 1,\ldots,\bar{p}$, $\ell = 1,\ldots,p_j$, we let $\bar{x}_j = \xi_*(\bar{t}_j)$, $X_j = (\bar{x}_j,\bar{t}_j)$, $V_j = f^*(X_j)$, $V_{j,0} = f_0^*(X_j)$, $W_j^\ell = f^{u_j^\ell}(X_j)$, $W_{j,0}^\ell = f_0^{u_j^\ell}(X_j)$, $Z_j^\ell = W_j^\ell - V_j$, and $Z_{j,0}^\ell = W_{j,0}^\ell - V_{j,0}$.

We then let $\mathcal{G}_0 = \mathcal{G}_0^{\mathbf{V},\prec}$ be the set of all 5-tuples $\mathcal{Z} = (L_0,L,\boldsymbol{\mu},L^\varphi,L^{\mathbf{h}})$ such that (i) (L_0,L) is a measurable selection of $\Lambda^{\mathbf{f}}$, (ii) $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_m) \in \mathcal{M}$ (where $\mathcal{M} = \mathcal{M}_1 \times \cdots \times \mathcal{M}_m$), (iii) $L^\varphi \in \Lambda^\varphi$, and, finally, (iv) $L^{\mathbf{h}} = (L^{h_1},\ldots,L^{h_{\tilde{m}}}) \in \Lambda^{\mathbf{h}}$.

For any given member $\mathcal{Z} = (L_0,L,\boldsymbol{\mu},L^\varphi,L^{\mathbf{h}})$ of $\mathcal{G}_0$, we construct a linear map $\mathcal{L}^{\mathcal{Z}} : \mathbb{R}_p \mapsto \mathcal{X}$ by defining $\mathcal{L}^{\mathcal{Z}}(\vec{\varepsilon}) = \sum_{j=1}^{\bar{p}} \sum_{\ell=1}^{p_j} \varepsilon_j^\ell Q_j^{\ell,\mathcal{Z}}$ for each $\vec{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_p) = (\vec{\varepsilon}_{(1)},\ldots,\vec{\varepsilon}_{(\bar{p})}) \in \mathbb{R}_p$, where

$$\zeta_j^{\ell,L}(t) = \begin{cases} M_L(t,\bar{t}_j)\cdot Z_j^\ell & \text{if} \quad t > \bar{t}_j\,, \\ 0 & \text{if} \quad t \le \bar{t}_j\,, \end{cases}$$

$$\zeta_{j,0}^{\ell,L,L_0}(t) = \begin{cases} Z_{j,0}^\ell + \int_a^b L_0(s)\zeta_j^{\ell,L}(s)ds & \text{if} \quad t > \bar{t}_j\,, \\ 0 & \text{if} \quad t \le \bar{t}_j\,, \end{cases}$$

$$\begin{aligned}
Q_j^{\ell,\mathcal{Z}} &= (Q_j^{\ell,\mathcal{Z},1},Q_j^{\ell,\mathcal{Z},2},Q_j^{\ell,\mathcal{Z},3},Q_j^{\ell,\mathcal{Z},4})\,,\\
Q_j^{\ell,\mathcal{Z},1} &= L^\varphi\cdot\zeta_j^{\ell,L}(b) + \zeta_{j,0}^{\ell,L,L_0}(b)\,,\\
Q_j^{\ell,\mathcal{Z},2} &= \zeta_j^{\ell,L}(b)\,,\\
Q_j^{\ell,\mathcal{Z},3} &= (Q_j^{\ell,\mathcal{Z},3,1},\ldots,Q_j^{\ell,\mathcal{Z},3,m})\,,\\
Q_j^{\ell,\mathcal{Z},3,i} &= \int_{[a,b]} \langle\zeta_j^{\ell,L},d\mu_i\rangle\,,\\
Q_j^{\ell,\mathcal{Z},4} &= (Q_j^{\ell,\mathcal{Z},4,1},\ldots,Q_j^{\ell,\mathcal{Z},4,\tilde{m}})\,,\\
Q_j^{\ell,\mathcal{Z},4,s} &= L^{h_s}\cdot\zeta_j^{\ell,L}(b)\,,
\end{aligned}$$

for $j = 1,\ldots,\bar{p}$, $\ell = 1,\ldots,p_j$, $i = 1,\ldots,m$, $s = 1,\ldots,\tilde{m}$. (Here $M_L : [a,b]\times[a,b] \mapsto \mathcal{L}(X)$ is the fundamental matrix solution of the linear differential equation $\dot{M} = L\cdot M$, so M_L is characterized by the integral condition $M_L(t,s) = \mathrm{id}_X + \int_s^t L(r)\cdot M_L(r,s)\,dr$.)

We define $\mathcal{G} = \mathcal{G}^{\mathbf{V},\prec}$ to be the set of all linear maps $\mathcal{L}^{\mathcal{Z}}$, for all $\mathcal{Z} \in \mathcal{G}_0^{\mathbf{V},\prec}$.

The following is then the key result.

Lemma 4.4: The set $\mathcal{G}^{\mathbf{V},\prec}$ is a generalized differential quotient of the augmented endpoint map $\mathcal{E}^{\mathbf{V},\prec}$ at $P_{*,p}$ in the direction of $\mathbb{R}_{+,p}$.

Once Lemma 4.4 is proved, the main theorem follows by standard arguments: Proposition 4.3 implies a restricted form of the theorem, in which all the conditions of the conclusion are satisfied, except only for the fact that the inequalities of the Hamiltonian maximization condition only holds for those pairs $(u,t) \in \mathbf{V}$; a compactness argument is then used to pass to the limit and obtain one adjoint covector π and multipliers such that the inequalities hold for all $(u,t) \in \mathcal{V}$.

The proof of Lemma 4.4 is based on a long series of estimates, to be described in detail in a forthcoming paper.

Here, we limit ourselves to pointing out why the "chattering variations" are crucial for the proof, and why ordinary packets of needle variations (corresponding to $N = 1$) do not suffice. This can be understood by means of the following highly simplified example. Suppose $p = 2$ and $\bar{p} = 1$, so we are dealing with a 2-parameter variation at a single time $\bar{t}$, using two control values u^1, u^2. Assume, moreover, that $m = 1$, so there is a single state space constraint $g(x,t) \leq 0$ and a single set $\mathcal{M}_j = \mathcal{M}$. Also, assume global existence and uniquenes of trajectories for all possible controls, so for each $\eta \in \mathcal{U}$ there exists a unique trajectory $\xi^\eta : [a,b] \mapsto X$ with initial condition $\xi^\eta(a) = \bar{x}_*$. Even more strongly, assume that all the vector fields $(x,t) \mapsto f(x,u,t)$ are in fact constant, so $f(x,u,t) = f(u)$, and the reference control η_* is equal to a constant u_*. Finally, assume that g is independent of t and of class C^1.

Let us see what happens when we use a classical packet of needle variations, with $N = 1$. For a vector $(\varepsilon_1, \varepsilon_2)$, write $\xi^{\varepsilon_1,\varepsilon_2} = \xi^{\eta^{\varepsilon_1,\varepsilon_2}}$, and let $\theta(\varepsilon_1, \varepsilon_2)$ be the supremum of the numbers $g(\xi^{\varepsilon_1,\varepsilon_2}(t), t)$, for $t \in [a,b]\}$. Then let Θ be the set-valued map that sends each $(\varepsilon_1, \varepsilon_2)$ to $[0, +\infty[$ if $\theta(\varepsilon_1, \varepsilon_2) > 0$, and to $\mathbb{R}$ if $\theta(\varepsilon_1, \varepsilon_2) \leq 0$. Then Θ is the state space constraint part of the augmented endpoint map $\mathcal{E}$, and to prove Lemma 4.4 in this special situation we have to show that a GDQ of Θ is given by the set A of all linear maps $(\varepsilon_1, \varepsilon_2) \mapsto \langle \varepsilon_1\zeta_1 + \varepsilon_2\zeta_2, \mu\rangle$, for $\mu \in \mathcal{M}$, where ζ_j is the function given by $\zeta_j(t) = 0$ for $t < \bar{t}$, and $\zeta_j(t) = V_j$ for $t \geq \bar{t}$. (Here $V_j = f(u_j) - f(u_*)$, and $\langle \cdot, \cdot \rangle$ is the pairing that sends a curve $\zeta : [a,b] \mapsto X$ with left and right limits at each point and a measure $\mu \in bvadd([a,b], X^\dagger)$ to the integral $\langle \zeta, \mu \rangle \stackrel{\text{def}}{=} \int_{[a,b]} \zeta(t) d\mu(t)$.) To prove this, it suffices to show that, modulo a small error (where "small" means "$o(|\vec{\varepsilon}|)$"), for every sufficiently small $(\varepsilon_1, \varepsilon_2) \in \mathbb{R}_{2,+}$ there exists a member L of A such that $L(\varepsilon_1, \varepsilon_2) \geq 0$ whenever $\theta(\varepsilon_1, \varepsilon_2) > 0$. (Indeed, if we let $B(\varepsilon_1, \varepsilon_2)$ be the set of all $L \in A$ such that $L(\varepsilon_1, \varepsilon_2) \geq 0$ if $\theta(\varepsilon_1, \varepsilon_2) > 0$, and $B(\varepsilon_1, \varepsilon_2) = A$ if $\theta(\varepsilon_1, \varepsilon_2) \leq 0$, then it would follow that B is a map with compact convex nonempty values. Furthermore, B is upper semicontinuous, because if $\{(\varepsilon_1^k, \varepsilon_2^k, L^k)\}_{k \in \mathbb{N}}$ is a sequence in the graph of B such that the $(\varepsilon_1^k, \varepsilon_2^k)$ converge to a limit $(\varepsilon_1^\infty, \varepsilon_2^\infty)$, then the compactness of A enables us to pass to a subsequence and assume that the L^k converge to a limit $L^\infty \in A$. To prove that $(\varepsilon_1^\infty, \varepsilon_2^\infty, L^\infty)$ belongs to the graph of B, we have to show that if $\theta(\varepsilon_1^\infty, \varepsilon_2^\infty) > 0$ then $L^\infty(\varepsilon_1^\infty, \varepsilon_2^\infty) \geq 0$. But if $\theta(\varepsilon_1^\infty, \varepsilon_2^\infty) > 0$ then $g(\xi^{\varepsilon_1^\infty, \varepsilon_2^\infty}(t), t) > 0$ for some t, so $g(\xi^{\varepsilon_1^k, \varepsilon_2^k}(t), t) > 0$ for sufficiently large k, because of (H10). Then, for large k, $\theta(\varepsilon_1^k, \varepsilon_2^k) > 0$, so $L^k(\varepsilon_1^k, \varepsilon_2^k) \geq 0$, and passage to the limit as $k \to \infty$ yields the desired inequality $L^\infty(\varepsilon_1^\infty, \varepsilon_2^\infty) \geq 0$. It then follows that B is a Cellina continuously approximable map, and by construction $L(\varepsilon_1, \varepsilon_2) \in \Theta(\varepsilon_1, \varepsilon_2)$ whenever $L \in B(\varepsilon_1, \varepsilon_2)$, showing that A is indeed a GDQ of Θ.)

So let $(\varepsilon_1, \varepsilon_2)$ be given, and let us prove the existence of L. Naturally, the case when $\theta(\varepsilon_1, \varepsilon_2) \leq 0$ is trivial, since in this situation any $L \in A$ will do. So assume that $\theta(\varepsilon_1, \varepsilon_2) > 0$, i.e., that our curve $\xi^{\varepsilon_1,\varepsilon_2}$ violates the state space constraint at some time τ. Then, if $\omega = \nabla g(\xi_*(\tau))$, the number $\langle \omega, \xi^{\varepsilon_1,\varepsilon_2}(\tau) - \xi_*(\tau)\rangle$ is ≥ 0 (modulo a small error), because $g(\xi^{\varepsilon_1,\varepsilon_2}(\tau)) > 0$ and $g(\xi_*(\tau)) \leq 0$. If we could replace $\xi^{\varepsilon_1,\varepsilon_2}(\tau) - \xi_*(\tau)$ by its "linearization" $\zeta^{\varepsilon_1,\varepsilon_2}(\tau)$ (where $\zeta^{\varepsilon_1,\varepsilon_2}(t) \stackrel{\text{def}}{=} \varepsilon_1\zeta_1(t) + \varepsilon_2\zeta_2(t)$), we would get the inequality $\langle \omega, \zeta^{\varepsilon_1,\varepsilon_2}(\tau)\rangle \geq 0$, i.e., $\langle \zeta^{\varepsilon_1,\varepsilon_2}(\tau), \mu\rangle \geq 0$, where μ is the right delta function at τ with value ω. Since $\mu \in \mathcal{M}$, the linear map L given by $L(\varepsilon_1, \varepsilon_2) = \langle \varepsilon_1\zeta_1 + \varepsilon_2\zeta_2, \mu\rangle$ belongs to A and satisfies $L(\varepsilon_1, \varepsilon_2) \geq 0$, as desired.

The above argument works as long as $\xi^{\varepsilon_1,\varepsilon_2}(\tau) - \xi_*(\tau)$ is well approximated by $\zeta^{\varepsilon_1,\varepsilon_2}(\tau)$. It is easy to see that $\xi^{\varepsilon_1,\varepsilon_2}(t) - \xi_*(t) - \zeta^{\varepsilon_1,\varepsilon_2}(t)$ is $o(|\vec{\varepsilon}|)$ as long as t does not belong to the interval $I(\vec{\varepsilon}) = [\bar{t}, \bar{t} + \varepsilon_1 + \varepsilon_2]$. But the approximation fails when $t \in I(\vec{\varepsilon})$. (For example, $\zeta^{\varepsilon_1,\varepsilon_2}(\bar{t}+) = \varepsilon_1 V_1 + \varepsilon_2 V_2$, but $\xi^{\varepsilon_1,\varepsilon_2}(\bar{t}) - \xi_*(\bar{t}) = 0$.)

It turns out that the argument can be modified so as to make it work even on the bad interval $I(\vec{\varepsilon})$. For this purpose, suppose first that the map $t \mapsto \xi^{\varepsilon_1,\varepsilon_2}(t) - \xi_*(t)$ was actually linear affine on $I(\vec{\varepsilon})$. Then, if $\tau \in I(\vec{\varepsilon})$, we can conclude as before that $\langle \omega, \xi^{\varepsilon_1,\varepsilon_2}(\tau) - \xi_*(\tau)\rangle \geq 0$ modulo a small error. But $\xi^{\varepsilon_1,\varepsilon_2}(\tau) - \xi_*(\tau)$ is a convex combination of $\xi^{\varepsilon_1,\varepsilon_2}(\bar{t}) - \xi_*(\bar{t})$ and $\xi^{\varepsilon_1,\varepsilon_2}(\bar{t} + |\vec{\varepsilon}|) - \xi_*(\bar{t} + |\vec{\varepsilon}|)$, so at least one of the two numbers $\langle \omega, \xi^{\varepsilon_1,\varepsilon_2}(\bar{t}) - \xi_*(\bar{t})\rangle$ and $\langle \omega, \xi^{\varepsilon_1,\varepsilon_2}(\bar{t} + |\vec{\varepsilon}|) - \xi_*(\bar{t} + |\vec{\varepsilon}|)\rangle$ is ≥ 0 (modulo a small error). On the other hand, $\xi^{\varepsilon_1,\varepsilon_2}(t) - \xi_*(t)$, for *both* $t = \bar{t}$ and $t = \bar{t} + |\vec{\varepsilon}|$, is well appproximated by the corresponding vector $\zeta^{\varepsilon_1,\varepsilon_2}(t-)$, so we conclude that $\langle \omega, \xi^{\varepsilon_1,\varepsilon_2}(t) - \xi_*(t)\rangle \geq 0$ for $t = \bar{t}$ or $t = \bar{t} + |\vec{\varepsilon}|$. The argument then proceeds as before.

The purpose of the chattering parameter N is to make $\xi^{\varepsilon_1,\varepsilon_2}(t) - \xi_*(t)$, approximately, a linear affine function of t on $I(\vec{\varepsilon})$. By choosing N large, in an $\vec{\varepsilon}$-dependent way, we achieve the desired approximability by a linear affine function of t up to an $o(|\vec{\varepsilon}|)$ error.

To conclude, we remark that in the above sketch of a proof it is clear that, if $t = \bar{t}$, then one has to use $\langle \omega, \zeta^{\varepsilon_1,\varepsilon_2}(t-)\rangle$, since $\zeta^{\varepsilon_1,\varepsilon_2}(t+)$ is *not* a good approximation to $\xi^{\varepsilon_1,\varepsilon_2}(\bar{t}) - \xi_*(\bar{t})$. (If $t = \bar{t} + |\vec{\varepsilon}|$ then one can just use $\langle \omega, \zeta^{\varepsilon_1,\varepsilon_2}(t)\rangle$, since $\zeta^{\varepsilon_1,\varepsilon_2}$ is continuous at t.) This shows that the distinction between left and right delta functions matters, thus providing a partial explanation for the occurence of finitely additive measures in our setting.

References

[1] Sussmann, H. J., "A very nonsmooth maximum principle with state space constraints." In *Proceedings of the 44rd IEEE 2005 Conference on Decision and Control (Sevilla, Spain, December 12-15, 2005)*, IEEE Publications, New York, 2005.