ANNALES DE L'I. H. P., SECTION C

S. ARTSTEIN

A variational convergence that yields chattering systems

Annales de l'I. H. P., section C, tome S6 (1989), p. 49-71

<http://www.numdam.org/item?id=AIHPC_1989__S6__49_0>

A VARIATIONAL CONVERGENCE THAT YIELDS CHATTERING SYSTEMS

Z. ARTSTEIN
Department of Theoretical Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israël

Abstract

Chattering variational problems are introduced and employed as limits of a variational convergence mode which is compact and preserves the optimal values. The phenomenon arises in systems with highly oscillatory coefficients. Robustness of, and approximations to, optimal controls in such cases are examined.

1. Introduction

We study a problem of variational convergence, namely the convergence of the data in variational problems. The convergence we seek is such that the values of the individual problems and the optimal solutions vary continuously with the data. Such considerations are basic in the study of approximations to and sensitivity of variational problems, and arise when the proper limit of problems with refined data has to be determined. An excellent account of variational convergence is the monograph by Attouch [4].

The systems we deal with in this paper are of the optimal control type and have highly oscillatory coefficients. The common convergences, e.g. weak-L_1, or strong L_1, are not suitable for such systems We propose a graph-type convergence which works, but in turn yields a new type of variational problems as possible limits. We call these by the suggestive name, *chattering equations.* The structure of the chattering problems enables a study of approximations and sensitivity, and induces a proper notion of feedback.

The paper is organized as follows: The framework of our analysis is displayed in the next section together with the goals and the motivation. The technical conditions and the technical setting are given in Section 3. An informal discussion is presented in Section 4 where the heuristic of the new concept is explained. This discussion is done with reference to the simpler case of linear quadratic problems; in fact we offer a full description of this case but with informal, and at points ad hoc, proofs. The general nonlinear setting is analyzed in Sections 5 and 6; the former introduces the chattering equations and the latter describes the relevant topologies and proves the continuity. The sensitivity and approximation questions are addressed in the closing section.

2. Framework and Goals

We deal with a family of minimization problems, each of the form

$$P(f,g) \qquad \begin{aligned} &\text{minimize} \int_a^b Q(x(t),u(t))dt \\ &\text{subject to } \frac{dx}{dt} = f(x,t) + g(u,t) \\ &x(a) = x_0 \end{aligned}$$

Here $x \in R^n$ and $u \in R^m$. In this paper we consider the time interval $[a,b]$, the cost function $Q(x,u)$ and the initial condition x_0 fixed. The functions $f(\cdot,\cdot)$ and $g(\cdot,\cdot)$ which determine the constraint may be different for different problems, and therefore the problem is denoted $P(f,g)$. We assume that f and g belong to prescribed ensembles $\mathcal{F}$ and $\mathcal{G}$. (The exact structure of $\mathcal{F}$ and $\mathcal{G}$ and the other specifications, e.g. of $Q(\cdot,\cdot)$, are given in the next section.)

Our technique applies to a more general setting, for instance to a constraint of the form $\dot{x} = f(x,u,t)$ or Q depending on t. We restrict the discussion to the separable case $P(f,g)$ for the sake of clarity. In fact, some of the ideas are demonstrated best with the aid of the simpler case of the linear quadratic minimization, as follows:

$$LQ(A,B) \qquad \begin{aligned} &\text{minimize} \int_a^b (|x(t)|^2 + |u(t)|^2)dt \\ &\text{subject to } \frac{dx}{dt} = A(t)x + B(t)u \\ &x(a) = x_0 \end{aligned}$$

Here we assume that the matrix valued coefficients $A(t)$ and $B(t)$ belong to prescribed ensembles $\mathcal{A}$ and $\mathcal{B}$ of matrix valued functions. The discussion in Section 4 refers therefore to the LQ case.

Here is a main goal of the analysis.

Goal 1. Find a topology on $\mathcal{F}$ and a topology on $\mathcal{G}$ such that

(i) both topologies are compact, and

(ii) the infimal value of the problem $P(f,g)$, denoted by $val(f,g)$, depends continuously on f and g.

When the two topologies desired by the first goal are determined, we can set the second goal.

Goal 2. Discover the continuity and sensitivity properties of solutions, or approximate solutions, $u(t)$ of $P(f,g)$ with respect to the data f and g.

We elaborate on the statement of the second goal, but before that we disclose briefly the motivation behind the analysis. (Many examples of similar variational convergence problems can be found in Attouch [4]. Convergence of control problems was studied by Buttazzo and Dal Maso [8], and Buttazzo [6], [7].)

The variety of possible data may arise in two ways: either from uncertainty, due to errors of estimations and fluctuations, or we may have a sequence of problems with increasingly refined structure, and we want to deduce from the behavior of the limit information about the refined approximations. In particular, the proper notion of limit has to be defined. Condition (ii) is a natural link between the sequence and its limit in this situation. Condition (i) enables us to extract converging subsequences at least from, say, a sequence of refined structures, and to determine their limits. Another application for compactness is the existence of uniform bounds. Suppose we know that $val(f,g)$ is finite for all (f,g); then (i) and (ii) together imply that there is a uniform bound for all costs $val(f,g)$.

Note that the two requirements set in Goal 1 oppose each other. The continuity may require many open sets, but too many open sets may harm compactness. In addition to the two formal conditions we wish the topologies to be not too abstract so that information on the possible solutions can be deduced along the lines of the second goal.

The second goal is concerned with the sensitivity of the system to perturbations, either in the parameters or in the controls that are used. Queries that we want answers to are: Suppose $u_0(t)$ is a solution of $P(f_0, g_0)$; if f and g are close to f_0, g_0, is $u_0(t)$ an approximate solution of $P(f, g)$? What happens if both the data f_0, g_0 and the control $u_0(t)$ are perturbed and what is then the proper notion of a small perturbation of $u_0(t)$? Do the optimal solutions, or near optimal solutions, of $P(f, g)$ vary continuously with (f, g)?

3. The Technical Setting

The norm of the vector x in R^n is denoted by $|x|$; the controls u are in R^m. The derivative dx/dt is also denoted by $\dot{x}$. If it matters, vectors are thought of as column vectors; the transpose of x or A are denoted by x^T and A^T.

We wish to include highly oscillatory data; other than that, we are willing (in this paper) to impose restrictive conditions. Accordingly, we specify $Q, \mathcal{F}$ and $\mathcal{G}$ as follows:

The function $Q(x, u)$ is assumed continuous in (x, u) and coarse, namely $Q(x, u) \geq \gamma_0(|x|^2 + |u|^2)$ for some constant $\gamma_0 > 0$.

The family $\mathcal{F}$ is determined by prescribed constants $\lambda_1 > 0$ and κ_1. The family consists of all the functions $f(x, t) : R^n \times [a, b] \longrightarrow R^n$, which are continuous in x, measurable in t and such that $|f(x, t)| \leq \lambda_1(|x| + 1)$ and $|f(x, t) - f(y, t)| \leq \kappa_1|x - y|$.

The family $\mathcal{G}$ is determined by constants $\lambda_2 > 0$ and κ_2. It consists of all the functions

$g(u,t): R^m \times [a,b] \longrightarrow R^n$, continuous in u, measurable in t and such that $|g(u,t)| \leq \lambda_2(|u|+1)$ and $|g(u,t) - g(v,t)| \leq \kappa_2|u-v|$.

Notice that we do not impose conditions that guarantee the existence of an optimal solution. The value $val(f,g)$ is therefore only an infimum, and the approximation and sensitivity queries raised in the previous section may refer to approximate solutions.

In the particular case of the LQ problems, the previous conditions translate as follows. The ensembles $\mathcal{A}$ and $\mathcal{B}$ consist of all the $n \times n$, respectively $n \times m$, matrix valued functions bounded by κ_1, respectively κ_2.

The admissible controls for a pair (f,g) in $\mathcal{F} \times \mathcal{G}$ are measurable functions $u(t)$: $[a,b] \longrightarrow R^n$, which are integrable over $[a,b]$. The conditions on $\mathcal{F}$ and $\mathcal{G}$ imply that for a given admissible control $u(t)$ the solution $x(t)$ of the constraint differential equation $\dot{x} = f(x,t) + g(u(t),t)$, $x(a) = x_0$, is unique. Hence the cost

$$cost(u(t),f,g) = \int_a^b Q(x(t),u(t))dt$$

is well defined; the cost is finite if $u(t)$ is bounded.

4. Discussion

We display in this section the difficulties that arise when trying to achieve Goal 1 and the arguments that lead to the solution we offer. To make our considerations transparent we limit the discussion to the Linear Quadratic problem and the ensembles $\mathcal{A}$ and $\mathcal{B}$. The rather simple structure of the LQ problems enables us to provide proofs based on known formulas. These are not available for the general nonlinear case which is treated in the next section.

It is possible to fulfill the requirements set in Goal 1 for the ensemble $\mathcal{A}$ by employing a standard convergence. Consider the weak-L_1 topology on $\mathcal{A}$. It is compact since elements in $\mathcal{A}$ are bounded by κ_1. There are explicit formulas for the minimal value $val(A, B)$ of $LQ(A, B)$. For instance

$$val(A, B) = x_0^T K x_0 \tag{4.1}$$

where K is positive definite and given as $K = K(a)$ with $K(t)$ solving the Ricatti equation

$$\dot{K} = -KA(t) - A(t)^T K + KB(t)B(t)^T K - I \ , \quad K(b) = 0 \tag{4.2}$$

see Athans and Falb [3, page 761]. A standard result on continuous dependence of solutions on parameters (we quote a more general one in Section 6) implies that $K(a)$ depends continuously on variations in $A(t)$ with respect to the weak-L_1 topology; hence the latter is the desired topology.

The situation with $\mathcal{B}$ is more involved. The weak-L_1 topology is not suitable since $val(A, B)$ is then not a continuous function of B. This is reflected in the term $B(t)B(t)^T$ in (4.2). As a concrete counterexample let $A(t) = 0$ and $B_k(t) = sin\ kt$, both scalars. Then $B_k(t)$ converge weakly to $B_0(t) = 0$. The Ricatti equation for the limit is $\dot{K} = 1$, while for each k the Ricatti equation is $\dot{K} = K^2 sin^2 kt - 1$. The limit of the latter (since $sin^2 kt$ converge weakly to $\frac{1}{2}$) is the equation $\dot{K} = \frac{1}{2}K^2 - 1$, which governs the limit of the values. Hence continuity fails.

A common topology which yields the continuity is the strong-L_1 topology, but it is not compact. It seems difficult to bridge the gap; even more so since the topology we seek has to be revealing. Thus, for instance, the oscillations of, say, $sin\ kt$ in the example should not be washed out in the limit; this since the optimal controls clearly follow these oscillations. We therefore introduce the following idea: We allow limits of sequences in

$\mathcal{B}$ to be outside $\mathcal{B}$; in fact, we allow these limits not to be matrix valued functions at all; then we may need to modify the definition of the LQ problem in order to accommodate these limits. We show in the sequel how this completion idea can be carried out. We call the terms added to $\mathcal{B}$ *chattering systems.*

But before displaying the completion we want to remind the reader that the idea of completion and compactification with unordinary items is not new in the calculus of variations and optimal control. The generalized curves of L.C. Young (see [13] and references therein) is one celebrated example. The relaxed controls introduced by J. Warga (see [12] and references therein) is another important example. Both, however, deal with the space of solutions, either trajectories or controls, and not with the space of equations. A completion in the space of differential equations was performed by Kurzweil [10]. Sequences of optimal control problems for which an extra term appears in the limit were described by Buttazzo [7].

Recall that we want to maintain the effect of oscillations in the limit. A way of modelling instantaneous oscillations is to allow as the limit of $B_k(t)$, functions, say $\beta(t)$ with values being probability distributions over the space of $n \times m$ matrices. The case $B(t)$ is then the particular case of a measure concentrated at $\{B(t)\}$. Keeping in mind that the controls $u_k(t)$ may respond to the rapid oscillations of the coefficients, we wish to allow the controls at the limit equation to respond to the instantaneous changes. A way to model this is to let the control u at the time t be a function of the matrix B; we write it as $u(t, B)$. The contribution to the dynamics is the weighted average of $Bu(t, B)$ with respect to the probability measure $\beta(t)$. Similarly, the contribution to the cost in the LQ problem is the weighted average of $|u(t, B)|^2$. The chattering LQ problem is then as follows.

$$LQ(A,\beta)\qquad \begin{aligned} &\text{minimize} \int_a^b (|x(t)|^2 + \int_M |u(t,B)|^2 d_B\beta(t)dt \\ &\text{subject to } \dot{x} = A(t)x + \int_M Bu(t,B)d_B\beta(t) \\ &x(a) = x_0 \end{aligned}$$

where M denotes the space of $n \times m$ matrices. The measure valued mapping $\beta(t)$ has values supported in the compact subset $K_2 = \{B \in M : |B| \leq \kappa_2\}$, see Section 3 for the definition of κ_2. We denote the space of probability measures on K_2 by $Prob(K_2)$. Then $\beta : [a,b] \longrightarrow Prob(K_2)$, and we assume that β is measurable, when on $Prob(K_2)$ we consider, say, weak convergence of measures (see Billingsley [5]). The ensemble of all such chattering coefficients $\beta(t)$ is denoted by $\mathcal{P}$. We still have to specify the admissible controls. We may choose them measurable in (t,B) or continuous in B. (The particular case of $LQ(A,B)$ is realized as an $LQ(A,\beta)$ with a control $u(t,B)$ continuous in B.) For definiteness we decide to have $u(t,B)$ Borel measurable in (t,B).

The goals set for the collections $\mathcal{A}$ and $\mathcal{B}$ relate now to $\mathcal{A}$ and $\mathcal{P}$. We define the desired topology on $\mathcal{P}$. To this end we identify an element $\beta(\cdot)$ in $\mathcal{P}$ with the measure, say $\boldsymbol{\beta}$, on $[a,b] \times K_2$, obtained by integrating $\beta(\cdot)$ with respect to the Lebesgue measure, namely, if $D \subset [a,b] \times K_2$ is Borel and D_t denotes its t-section, then $\boldsymbol{\beta}(D) = \int_a^b \beta(t)(D)dt$. On the space of these $(b-a)$-times probability measures we adopt the weak convergence of measures which is metrizable and compact since $[a,b] \times K_2$ is compact, see Billingsley [5]. This is the convergence we want, as stated in the next proposition.

Note that the restriction of the weak convergence on P to the ensemble of ordinary coefficients $\mathcal{B}$ is identical with the L_1 norm convergence.

Proposition 4.1. The weak-L_1 topology on $\mathcal{A}$ and the weak convergence of measures on $\mathcal{P}$ are compact, and $val(A,\beta)$ is continuous on $\mathcal{A} \times \mathcal{D}$.

Compactness of the two topologies is a known property. The continuity will follow from the general results in the sequel, but a relatively simple proof can be crafted along the lines of the derivations of the Ricatti equation (4.2). It is not hard to show, but we do not do it here, that $val(A, \beta)$ is equal to $x_0^T K x_0$ with $K = K(a)$ where $K(t)$ satisfies the Ricatti equation

$$\dot{K} = -KA(t) - A(t)^T K + K(\int_M BB^T d_B\beta(t))K - I \ , \quad K(b) = 0 \ . \tag{4.3}$$

Standard observations imply now that $K(a)$ depends continuously on the respective topologies.

The LQ problems have a nice form of a feedback solution. For $LQ(A, B)$ the optimal solution is a function of t and $x(t)$ given by

$$u(t, x) = -B(t)^T K(t)x(t) \tag{4.4}$$

with K satisfying (4.2), see Athans and Falb [3, page 763]. It is possible to get a similar formula for the chattering case. Then $u = u(t, B, x)$

$$u(t, B, x) = -B^T K(t)x(t) \tag{4.5}$$

with $K(t)$ satisfying (4.3). Notice that in the chattering formulation the optimal control is continuous in all variables.

Example 4.2. We apply our solution to the example mentioned before, namely $A(t) = 0$ and $B_k(t) = sin\ kt$, both scalars. The range of the coefficients B_k is $K_2 = [-1, 1]$. It is not difficult to compute $\beta(t)$; indeed, $\beta(t) = \beta_0$ is constant and $\beta_0([\sigma, \tau]) =$

$\pi^{-1}(\arcsin \tau - \arcsin \sigma)$. The limit problem is then

$$\text{minimize} \quad \int_a^b (|x(t)|^2 + \int_{-1}^1 u(t,\sigma)^2(1-\sigma^2)^{-\frac{1}{2}}\,d\sigma)dt$$

$$\text{subject to } \dot{x} = \int_{-1}^1 \sigma u(t,\sigma)(1-\sigma^2)^{-\frac{1}{2}}\,d\sigma$$

$$x(0) = x_0 \ .$$

If we employ the Ricatti equation (4.3) in this case we get

$$\dot{K} = \frac{\pi}{2}K^2 - 1 \ , \quad K(b) = 0$$

and the optimal solution of the continuous problem is, in a feedback form,

$$u(t,\sigma,x) = -\sigma K(t)x$$

and $K(t)$ can be found explicitly from the previous differential equation; the simple calculation shows that

$$K(t) = (\frac{2}{\pi})^{\frac{1}{2}}(1 - exp((2\pi)^{\frac{1}{2}}(t-b))(1 + exp((2\pi)^{\frac{1}{2}}(t-b))^{-1} \ .$$

5. Chattering Systems

In this section we introduce the chattering variational problems in the nonlinear case. The basic idea is along the lines of the discussion in the previous section, with the necessary modifications.

Let G be the collection of continuous functions $g(u):\ R^m \longrightarrow R^n$ satisfying $|g(u)| \leq \lambda_2(|u|+1)$ and $|g(u) - g(v)| \leq \kappa_2|u-v|$. The constants λ_2 and κ_2 are taken from the description of $\mathcal{G}$, see Section 2. We consider G as a metric space with the uniform convergence on bounded sets; a possible distance is

$$\|g_1 - g_2\| = \sum_{N=1}^{\infty} \frac{1}{N} 2^{-N} max\{|g_1(u) - g_2(u)| : \ |u| \leq N\} .$$

We need the following for a reference.

Lemma 5.1. The space G with the uniform convergence on bounded sets is compact.

Proof. Follows easily from the boundedness on bounded sets and the uniform Lipschitz condition.

We denote by $Prob(G)$ the family of probability measures on G. Let $\mathcal{P}$ denote the ensemble of measurable mappings $\zeta(t) : \ [a,b] \longrightarrow Prob(G)$, when the latter is endowed with the metric structure of weak convergence of measures, see Billingsley [5].

An admissible control is a function $u(t,g) : \ [a,b] \times G \longrightarrow R^m$, which is Borel measurable in (t,g).

The chattering variational problem is determined by a pair $f \in \mathcal{F}$ and $\zeta \in \mathcal{P}$ as follows:

$$P(f,\zeta) \qquad \begin{aligned} &\text{minimize} \quad \int_a^b (\int_G Q(x(t), u(t,g)) d_g \zeta(t)) dt \\ &\text{subject to } \dot{x} = f(x,t) + \int_G g(u(t,g)) d_g \zeta(t) \\ &x(a) = x_0 \ . \end{aligned}$$

The following result is needed to assure that the constraint differential equation is well defined.

Lemma 5.2. Let $u(t,g)$ be an admissible control. Then $g(u(t,g)) : \ [a,b] \times G \longrightarrow R^n$ is a measurable function.

<u>**Proof.**</u> $g(u(t,g))$ is the composition of the continuous function $(g,v) \longrightarrow g(v)$, from $G \times R^m$ into R^n, with the measurable mapping $(t,g) \longrightarrow (g,u(t,g))$, from $[a,b] \times G$ into $G \times R^m$.

Like the ordinary case we denote by $val(f,\zeta)$ the infimal value of $P(f,\zeta)$, and by $cost(u(\cdot,\cdot),f,\zeta)$ the cost of applying the control u with the data (f,ζ). It is clear that the ordinary problem $P(f,g)$ can be viewed as a particular case of $P(f,\zeta)$ with $\zeta(t)$ concentrated on $g(\cdot,t)$. The ordinary admissible control $u(t)$ can be formally extended to an admissible control of the chattering case by letting $u(t,g) = u(t)$. (Specifying such an extension is needed when we consider chattering perturbations of ordinary problems.)

Once the ensemble $\mathcal{G}$ is extended to the ensemble $\mathcal{P}$, the two goals set in Section 2 may refer now to data in $\mathcal{F} \times \mathcal{P}$. The analysis of these is done in the next two sections.

6. Compactness and Continuity

In this section we introduce the topologies for $\mathcal{P}$ and $\mathcal{F}$, and verify the first goal, namely the compactness of $\mathcal{P}$ and $\mathcal{F}$ and the continuity of $val(f,\zeta)$.

The definition of the topology on $\mathcal{P}$ follows the outline of the preceding section. We identify an element $\zeta(t): [a,b] \longrightarrow Prob(G)$ with the measure ς on $[a,b] \times G$ obtained by integrating $\zeta(t)$, namely $\varsigma(D) = \int_a^b \zeta(t)(D_t)dt$, where D_t is the t-section of D. Thus ς is a multiple of a probability measure by $b-a$. Convergence of sequences ς_k is taken as the weak convergence of measures (see Billingsley [5]).

A simple computation shows that if $\zeta_0(t)$ represents nonchattering coefficients, say $\zeta_0(t)$ is supported at the singleton $\{g_0(\cdot,t)\}$, and $\zeta_k(t)$ converge to $\zeta_0(t)$, then

$$\int_a^b \int_G \|g_0(\cdot,t) - g\| d_g \zeta_k(t) dt \quad \text{converge to zero} .$$

In particular, the restriction of our convergence to the data in $\mathcal{G}$ coincides with the L_1 convergence in the sense that $g_k \longrightarrow g_0$ is equivalent to $\int_a^b \|g_k(\cdot,t) - g_0(\cdot,t)\|dt \longrightarrow 0$.

Proposition 6.1. The convergence in $\mathcal{P}$ is metrizable and compact.

Proof. Compactness and metrizability of the weak convergence follow from the compactness of $[a,b] \times G$ (see Lemma 5.1) and the Prohorov Theorem (see Billingsley [5, pages 37, 240]). Therefore, the only point to verify is that the limit, say, ς_0, of a sequence ς_k in $\mathcal{P}$ is still in $\mathcal{P}$, namely ς_0 can be disintegrated with respect to the Lebesgue measure on $[a,b]$ and the values of the resulting mapping $\zeta_0(t)$ are probability measures on G. But this follows easily from the observation that $\varsigma_k([c,d] \times G) = d - c$ for all k and all $a \leq c \leq d \leq b$. This completes the proof.

The definition of the topology on $\mathcal{F}$ is a weak-L_1 type (compare with the topology on $\mathcal{A}$ in Section 4) localized at each x. It is a standard topology in continuous dependence considerations of ordinary differential equations; it goes back to Gikhman [9]. The definition is as follows. The sequence f_k converges to f_0 if for every $x \in R^n$ and every $t \in [a,b]$ the sequence $\int_a^t f_k(x,s)ds$ converges to $\int_a^t f_0(x,s)ds$.

Proposition 6.2. The convergence on $\mathcal{F}$ is metrizable and compact.

Proof. See Proposition 2.4 and Theorem 2.4 in Artstein [1].

Recall the following notions. Let h be a measurable function from a measure space (T,μ) into a metric space Y. The distribution of h, denoted by Dh, is the measure on Y given by $Dh(C) = \mu(h^{-}(C))$. A sequence h_k converges in distribution to h_0 if Dh_k converges to Dh_0 with respect to weak convergence of measures. The latter is metrizable in our case and we use $dist(\eta_1, \eta_2)$ to denote the Prohorov distance between the measures η_1 and η_2; the Prohorov distance is equivalent to weak convergence of measures, see Billingsley [5].

The following will be used several times.

<u>**Lemma 6.3.**</u> Let $h(\tau) : T \longrightarrow Y$ be measurable and let μ be a measure on T. Let $\ell : Y \longrightarrow R^n$ be measurable. Then $\int_Y \ell(y) d(Dh) = \int_T \ell(h(\tau)) d\mu$.

<u>**Proof.**</u> Standard.

It will be convenient to use shorter notations for some of the integrals that define the chattering variational problem. We use the convention that functionals with the same index (double index occasionally) relate to each other.

Let $u_k(t, g)$ be an admissible control applied to the chattering problem $P(f_k, \zeta_k)$. We denote by $\gamma_k(t)$ the resulting forcing term in the constraint differential equation, namely

$$\gamma_k(t) = \int_G g(u_k(t, g)) d_g \zeta_k(t) \ .$$

The solution then of the constraint differential equation is denoted by

$$x_k(t) \ .$$

We also denote

$$c_k(t) = \int_G Q(x_k(t), u_k(t, g)) d_g \zeta_k(t) \ ,$$

namely $c_k(t)$ is the integrand of the cost functional when $u_k(t, g)$ is used.

For the control $u_k(t, g)$ we define $h_k : \ [a, b] \times G \longrightarrow [a, b] \times G \times R^m$ by

$$h_k(t, g) = (t, g, u_k(t, g)) \ .$$

We agree to compute Dh_k with respect to ζ_k, and for the sequence $u_{0,k}$ that we have we compute $Dh_{0,k}$ with respect to ζ_0.

Proposition 6.5. Let (f_k, ζ_k) be a sequence in $\mathcal{F} \times \mathcal{P}$ and let $u_k(t,g)$ and $u_{0,k}(t,g)$ be sequences of admissible controls, uniformly bounded. Then

(a) If $dist(Dh_k, Dh_{0,k})$ converge to 0 as $k \longrightarrow 0$, then $\gamma_k(t) - \gamma_{0,k}(t)$ converge to 0 in the weak-L_1 topology.

(b) If in addition $f_k \longrightarrow f_0$ in $\mathcal{F}$, then $x_k(t) - x_{0,k}(t)$ converge uniformly to 0, and

(c) The difference of costs, $cost(u_k, f_k, \zeta_k) - cost(u_{0,k}, f_0, \zeta_0)$ converge to 0 as $k \longrightarrow \infty$.

Proof. (a) Let $[c,d]$ be a subinterval of $[a,b]$ and denote by r_k and $r_{0,k}$ the restrictions of h_k and $h_{0,k}$ to $[c,d] \times G$. Since the first two coordinates of h are the identical maps, it follows from the structure of ς and the condition on Dh_k that $dist(Dr_k, Dr_{0,k})$ also converges to 0.

Define $\ell(t,g,v) = g(v)$, then $\ell : \; [a,b] \times G \times R^m \longrightarrow R^n$ is continuous. Since all the measures have a common compact support, by boundedness of the controls, it follows (Billingsley [5, page 113]) that $\int \ell d(Dr_k) - \int \ell d(Dr_{0,k})$ converge to zero. The composition of ℓ with r_k and $r_{0,k}$ yields the functions $g(u_k(t,g))$ and $g(u_{0,k}(t,g))$. If we now use Lemma 6.3 to rewrite $\int \ell d(Dr_k)$ and $\int \ell d(Dr_{0,k})$ as integrals with respect to ς_k and ς_0, the convergence (using the notation $\gamma(t)$) translates to

$$\int_c^d \gamma_k(t)dt - \int_c^d \gamma_{0,k}(t)dt \quad \text{converge to zero.}$$

Since all $\gamma_k, \gamma_{0,k}$ are bounded and $[c,d]$ is arbitrary, the latter convergence implies the desired weak convergence.

(b) By the preceding paragraph we have that the distance in $\mathcal{F}$ between $f_k(x,t) + \gamma_k(t)$ and $f_0(x,t) + \gamma_{0,k}(t)$ converges to zero. Continuous dependence of solutions with respect to data in $\mathcal{F}$ is a known result, see e.g. Artstein [1, Theorem 3.1].

(c) Define $q(t,g,v) = Q(y_0(t),v)$ where $y_0(t)$ is a continuous function. Then q is continuous, hence $\int q d(Dh_k) - \int q d(Dh_{0,k})$ converge to zero. The composition of q with h_k and $h_{0,k}$ yields the mappings $Q(y_0(t), u_k(t,g))$ and $Q(y_0(t), u_{0,k}(t,g))$. If we use Lemma 6.3 and write the latter convergence with respect to ζ_k and ζ_0 we get

$$\int_a^b \int_G Q(y_0(t), u_k(t,g)) d_g\zeta_k(t)dt - \int_a^b \int_G Q(y_0(t), u_{0,k}(t,g)) d_g\zeta_0(t)dt \longrightarrow 0 \ .$$

If we only could replace $y_0(t)$ in the first term of the convergence by $x_k(t)$ and $y_0(t)$ in the second term by $x_{0,k}(t)$, without harming the convergence, we would be done, since it would be the differences of costs that converge to 0. But the replacements are allowed, since $y_0(t)$ can be chosen a common limit point of $x_k(t)$ and $x_{0,k}(t)$ (by part (b) and the obvious uniform continuity) and then the continuity of $Q(\cdot,\cdot)$ provides the necessary estimates that guarantee the convergence. This completes the proof.

Remark. The previous result holds of course when the sequence $u_{0,k}$ is replaced by one control function u_0. Then $\gamma_k(t)$ converges weakly to $\gamma_0(t)$, the trajectories $x_k(t)$ converge to $x_0(t)$ and the costs $cost(u_k, f_k, \zeta_k)$ converge to $cost(u_0, f_0, \zeta_0)$. The reason we go to the trouble of considering a sequence of controls $u_{0,k}$ applied to (f_0, ζ_0) is the lack of compactness in the space of admissible controls (although the distributions $Dh_{0,k}$, and the functions $\gamma_{0,k}(t)$ and $x_{0,k}(t)$ converge). It is the same difficulty that implies the possible lack of optimal solutions. (We could introduce relaxed controls but this is beyond the scope of this paper.)

We need one more lemma; it guarantees the fulfillment of the conditions of the last proposition. We continue to use the same conventions.

Proposition 6.6. Suppose ζ_k converge to ζ_0 in $\mathcal{P}$. If $u_0(t,g)$ is a bounded admissible control then there exists a sequence of uniformly bounded admissible controls

$u_k(t,g)$ such that Dh_k converges to Dh_0. Conversely, if $u_k(t,g)$ is a uniformly bounded sequence of admissible controls, then there exists a sequence of controls $u_{0,k}(t,g)$ such that $dist(Dh_k, Dh_{0,k})$ converge to zero.

Proof. Let C be a compact set in R^m which contains all the possible values of the controls $u_k(t,g)$, $k = 0,1,\ldots$, in the two parts of the proposition. It exists by the uniform boundedness. Define $H(t,g) = \{(t,g,v) : v \in C\}$. Then H is a set-valued map. Consider the sequence of set-valued maps H_k obtained when H is considered a map of the measure space $([a,b] \times G, \varsigma_k)$. Clearly, since ς_k converge weakly to ς_0, the maps H_k converge in distribution to H_0 (where convergence in distribution of set-valued maps is defined in a natural way, see e.g. Artstein [2]). By Artstein [2, Theorem 6.3] the closure of $\{Dh : h$ a selection of $H_n\}$ converges in the Hausdorff metric to the closure of $\{Dh : h$ a selection of $H_0\}$, where closure and Hausdorff distance are taken with respect to the $dist(\cdot,\cdot)$ metric. A selection h_k of H_k and selections $h_{0,k}$ of H_0 correspond to admissible controls u_k and $u_{0,k}$ respectively. Therefore, the convergence of the closures of selections implies our result. This completes the proof.

Theorem 6.7. The given topologies in $\mathcal{F}$ and $\mathcal{P}$ are compact, and $val(f,\zeta)$ is continuous (namely Goal 1 is achieved).

Proof. Compactness was verified in Propositions 6.1 and 6.2. To prove the continuity note first that the optimal or near optimal controls for the problems $P(f,\zeta)$ are all bounded by the same uniform bound, say r. This follows from the coarsivity of $Q(x,u)$ and the uniform local boundedness of $f \in \mathcal{F}$ and $g \in G$. Let now (f_k,ζ_k) converge to (f_0,ζ_0) in $\mathcal{F} \times \mathcal{P}$ and let $u_0(t,g)$ be an admissible control for $P(f_0,\zeta_0)$. By Proposition 6.6 there are uniformly bounded admissible controls $u_k(t,g)$ for $P(f_k,\zeta_k)$ for which Dh_k converge to Dh_0. By part (c) of Proposition 6.5, $cost(u_k,f_k,\zeta_k)$ converge to $cost(u_0,f_0,\zeta_0)$. Since u_0

is arbitrary it follows that $limsup\ val(f_k, \zeta_k) \leq val(f_0, \zeta_0)$. Let now $u_k(t, g)$ be admissible controls for $P(f_k, \zeta_k)$, bounded by the aforementioned constants. By Proposition 6.6 there are uniformly bounded admissible controls $u_{0,k}(t, g)$, all for $P(f_0, \zeta_0)$, for which $dist(Dh_k, Dh_{0,k})$ converge to zero. By part (c) of Proposition 6.5 the difference between $cost(u_k, f_k, \zeta_k)$ and $cost(u_{0,k}, f_0, \zeta_0)$ tends to zero. Since u_k are arbitrary within the bound r it follows that $liminf\ val(f_k, \zeta_k) \geq val(f_0, \zeta_0)$. The two inequalities together constitute the desired continuity.

7. Robustness and Approximations

We provide in this section some answers to the queries induced by the second goal of Section 2. We pursue the analysis within the framework of the chattering problems $P(f, \zeta)$. The results of course make sense also for the ordinary problems, and the translation to this case is easy.

Our first result is concerned with the robustness to changes in the data.

Proposition 7.1. Let $u_0(t, g)$ be a bounded admissible control which is continuous in the variable g. Let (f_k, ζ_k) converge to (f_0, ζ_0) in $\mathcal{F} \times \mathcal{P}$. Then $cost(u_0, f_k, \zeta_k)$ converge to $cost(u_0, f_0, \zeta_0)$.

Proof. We write $u_k(t, g)$ for $u_0(t, g)$ when u_0 is applied in $P(f_k, \zeta_k)$. If we can only show that Dh_k converge to Dh_0 (see Proposition 6.5), then by part (c) of Proposition 6.5 the convergence follows. The convergence of Dh_k to Dh_0 is trivial if $u_0(t, g)$ is continuous in both variables. If $u_0(\cdot, g)$ is only measurable, then by the Scorza Dragoni ([11]) extension of the Lusin theorem, for every $\varepsilon > 0$ there exists a set $T \subset [a, b]$ with Lebesgue measure less than ε, and a function $v_0(t, g)$ which is bounded by r, continuous and $v_0(t, g) = u_0(t, g)$

if $t \notin T$. For v_0 we then have that desired convergence. But since $\varsigma_k(T \times G) \leq \varepsilon$ for all k, and since ε is arbitrarily small, the convergence holds for u_0 as well. This completes the proof.

The continuity of $u_0(t,g)$ in the variable g cannot be dropped from the conditions of the previous result. Indeed, $u_0(t,\cdot)$ discontinuous in sensitive to even a uniform small perturbation in $g(\cdot,t)$ in the ordinary case. The continuity assumption, however, does not seem to be severe. For instance, we showed in Section 4 (see (4.5)) that the optimal solutions of the LQ problems are continuous in (t,B).

We consider now perturbations in both the data and the controls. A basic question is then, in what sense to measure perturbations in the controls such that a small perturbation results in a small deviation of the cost? We offer two answers. One when the perturbation should result in a small error regardless of the perturbation in the data, and another when compatible perturbations are considered.

Proposition 7.2. Let $u_k(t,g)$ be a uniformly bounded sequence of admissible controls, each continuous in the g-variable and such that $\int_a^b \|u_k(t,\cdot) - u_0(t,\cdot)\| dt$ converge to 0 (with $\|\cdot\|$ being the sup norm). Let (f_k,ζ_k) converge to (f_0,ζ_0) in $\mathcal{F} \times \mathcal{P}$. Then $cost(u_k,f_k,\zeta_k)$ converge to $cost(u_0,f_0,\zeta_0)$.

Proof. By the Egorov theorem a subsequence of $u_k(t,\cdot)$ converges uniformly to $u_0(t,\cdot)$ in the sup norm, on sets $([a,b]\backslash T) \times G$ with T of arbitrarily small measure. Then (with the argument we used in the preceding result) Dh_k converges to Dh_0, and by part (c) of Proposition 6.5, the proof is complete.

Proposition 7.3. Let (f_k,ζ_k) converge to (f_0,ζ_0) in $\mathcal{F} \times \mathcal{P}$. Let $u_k(t,g)$ be perturbations of the bounded $u_0(t,g)$ as follows: $u_k(t,g) = u_0(t,g) + p_k(t,g)$ with p_k uniformly bounded and the distributions Dp_k computed with respect to ς_k, converging to a measure supported at $\{0\}$. If also $u_0(t,g)$ is continuous in the g-variable, then $cost(u_k, f_k, \zeta_k)$ converge to $cost(u_0, f_0, \zeta_0)$.

Proof. It is straightforward to show that the distance between Dh_k and Dh_0, both computed with respect to ζ_k, tends to zero. We already proved in the proof of Proposition 7.1 that the distributions of h_0 when computed with respect to ζ_k converges to Dh_0 computed with respect to ζ_0. Together with part (c) of Proposition 6.5, the proof is complete.

Our final result is concerned with the convergence of optimal controls. Again, since we did not post conditions guaranteeing existence, uniqueness or lower closure for optimal solutions, the statements involve approximating sequence. It is straightforward to restate the result referring to optimal controls when the aforementioned properties hold.

We continue with the convention that the distribution Dh_k (with $h(t,g) = (t, g, u(t,g))$) is computed with respect to ζ_k and $Dh_{0,k}$ is computed with respect to ζ_0. Notice that the statement $dist(Dh_k, Dh_{0,k}) \longrightarrow 0$ carries considerable more information here about the controls, than ordinary convergence in distribution implies; this is due to the first two coordinates of h.

Theorem 7.4. Let (f_k, ζ_k) converge to (f_0, ζ_0) in $\mathcal{F} \times \mathcal{P}$. Let $u_k(t,g)$ be uniformly bounded admissible controls such that $cost(u_k, f_k, \zeta_k) - val(f_k, \zeta_k)$ converge to zero. Then there is a sequence $u_{0,k}(t,g)$ of uniformly bounded admissible controls such that $cost(u_{0,k}, f_0, \zeta_0) - val(f_0, \zeta_0)$ tends to zero, and such that $dist(Dh_k, Dh_{0,k})$ tends to zero as $k \longrightarrow \infty$.

Proof. The existence of $u_{0,k}$ with the property $dist(Dh_k, Dh_{0,k}) \longrightarrow 0$ was proved in Proposition 6.6. Part (c) of Proposition 6.5 implies then that
$cost(u_k, f_k, \zeta_k) - cost(u_{0,k}, f_0, \zeta_0)$ tends to zero. The continuity of $val(f, \zeta)$, established

in Theorem 6.7, implies then that indeed $cost(u_{0,k}, f_0, \zeta_0)$ converge to $val(f_0, \zeta_0)$. This completes the proof.

References

[1] Z. Artstein, Topological dynamics of an ordinary differential equation. *J. Differential Equations* 23, 1977, 216-223.

[2] Z. Artstein, Distributions of random sets and random selections. *Israel J. Math.* 46, 1983, 313-324.

[3] M. Athans and P.L. Falb, *Optimal Control.* McGraw Hill, New York, 1966.

[4] H. Attouch, *Variational Convergence for Functions and Operators.* Pitman, Applicable Mathematics Series, Boston, 1984.

[5] P. Billingsley, *Convergence of Probability Measures.* John Wiley, New York, 1968.

[6] G. Buttazzo, Some relaxation problems in optimal control theory. *J. Math. Anal. Appl.* 125, 1987, 272-287.

[7] G. Buttazzo, Relaxation and Γ-limits in optimal control theory. Lecture in the Congrês Franco-Quebecois D'Analyse Non Linéaire Appliquée, Perpignan, Juin 1987.

[8] G. Buttazzo and G. Dal Maso, Γ-convergence and optimal control problems. *J. Optimiz. Theory Appl.* 38, 1982, 385-407.

[9] J.I. Gikhman, On a theorem of N.N. Bogolyubov. *Ukranain Math. J.* 4 (2), 1952, 215-218.

[10] J. Kurzweil, Generalized ordinary differential equations. *Czech. Math. J.* 8, 1958, 360-388.

[11] G. Scorza Dragoni, Una theorema sulla funzioni continue rispetto ad una i misurable rispetto at ultra variable. *Rend. Sem. Mat. Univ. Padova* 17, 1948, 102-106.

[12] J. Warga, *Optimal Control of Differential and Functional Equations.* Academic Press, New York 1972.

[13] L.C. Young, *Calculus of Variations and Optimal Control Theory.* W.B. Saunders, Philadelphia 1969.