The Hamiltonian Inclusion for Non-Convex Velocity Sets

Richard B. Vinter*

June 4, 2014

Abstract

Since Clarke’s 1973 proof of the Hamiltonian inclusion for optimal control problems with convex velocity sets, there has been speculation (and, more recently, speculation relating to a stronger, partially convexified version of the Hamiltonian inclusion) as to whether these necessary conditions are valid in the absence of the convexity hypothesis. The issue was in part resolved by Clarke himself when, in 2005, he showed that $L^\infty$ local minimizers satisfy the Hamiltonian inclusion. In this paper it is shown, by counter-example, that the Hamiltonian inclusion (and so also the stronger partially convexified Hamiltonian inclusion) are not in general valid for non-convex velocity sets, when the local minimizer in question is merely a $W^{1,1}$ local minimizer, not a $L^\infty$ local minimizer. The counter-example demonstrates that the need to consider $L^\infty$ local minimizers, not $W^{1,1}$ local minimizers, in the proof of the Hamiltonian inclusion for nonconvex velocity sets is fundamental, not just a technical restriction imposed by currently available proof techniques. The paper also establishes the validity of the partially convexified Hamiltonian inclusion for $W^{1,1}$ local minimizers, under a normality assumption, thereby correcting earlier assertions in the literature.

Keywords: Optimal Control, Differential Inclusions, Hamiltonian Inclusion.
1 Introduction

One of the most fruitful areas of study in optimal control theory since the 1970’s has been the derivation of first order necessary conditions of optimality for problems in which the dynamic constraint takes the form of a differential inclusion. A common formulation for such problems is:

\[
(P) \begin{cases} 
\text{Minimize } g(x(0), x(1)) \\
\text{over absolutely continuous functions } x(\cdot) : [0, 1] \to \mathbb{R}^n \text{ satisfying} \\
\dot{x}(t) \in F(t, x(t)) \text{ for a.e. } t \in [0, 1] \\
(x(0), x(1)) \in C,
\end{cases}
\]

in which \(g(\cdot, \cdot) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}\) is a given function, \(F(\cdot, \cdot) : [0, 1] \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n\) is a given multifunction taking values closed, non-empty subsets of \(\mathbb{R}^n\) and \(C \subset \mathbb{R}^n \times \mathbb{R}^n\) is a given closed set.

We shall refer to absolutely continuous functions \(x(\cdot) : [0, 1] \to \mathbb{R}^n\) that satisfy \(\dot{x}(t) \in F(t, x(t))\), a.e. as \(F\)-trajectories. An \(F\)-trajectory that satisfies the constraints of the optimal control problem is called a feasible \(F\)-trajectory. Two notions of ‘local minimizer’ will be considered:

A feasible \(F\)-trajectory \(\bar{x}(\cdot)\) is a

(a): \(L^\infty\) local minimizer if, for some \(\delta > 0\),

\[ g(x(0), x(1)) \geq g(\bar{x}(0), \bar{x}(1)) \tag{1} \]

for all feasible \(F\)-trajectories satisfying

\[ ||x(\cdot) - \bar{x}(\cdot)||_{L^\infty} \leq \delta, \]

(b): \(W^{1,1}\) local minimizer if, for some \(\delta > 0\), (1) is satisfied for all feasible \(F\)-trajectories satisfying

\[ ||x(\cdot) - \bar{x}(\cdot)||_{W^{1,1}} \leq \delta. \]

Here the \(W^{1,1}\) norm of \(x(\cdot)\) is

\[ ||x(\cdot)||_{W^{1,1}} = |x(0)| + \int_0^1 |\dot{x}(t)| dt. \]

Initial impetus to this research was given by Clarke’s discovery, in 1973, of the Hamiltonian Inclusion condition [2], so called because, in this condition, the minimizer \(\bar{x}(\cdot)\) and costate trajectory \(p(\cdot)\) are required to satisfy the differential inclusion

\[ (-\dot{p}(t), \dot{x}(t)) \in \text{co} \partial_{x,p}H(t, x(t), p(t)), \text{ a.e.}, \tag{2} \]

where \(H(\cdot, \cdot, \cdot) : [0, 1] \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}\) is the Hamiltonian function:

\[ H(t, x, p) = \max_{v \in F(t, x)} p \cdot v. \tag{3} \]

\(\partial_{x,p}H(t, \bar{x}(t), p(t))\) denotes the limiting subdifferential of \(H(t, \cdot, \cdot)\) at \((\bar{x}(t), p(t))\), defined below. The Hamiltonian inclusion can be interpreted as a generalization, to allow for a dynamic
constraint in the form of a differential inclusion and nonsmooth data, of classical necessary conditions from the calculus of variations often referred to as Hamilton’s system of equations.

Numerous researchers, including Clarke, Rockafellar, Ioffe, Loewen, Mordukhovich, Vinter and Zheng subsequently worked on refining the Hamiltonian inclusion condition, and also the derivation of necessary conditions for local minimizers, that combine generalizations of the classical Euler Lagrange and Weierstrass conditions. [3], [6], [5], [7], [8], [12] is an incomplete list of relevant references. Ignoring, for simplicity, developments to permit dynamic constraints in the form of unbounded differential inclusions, pathwise state constraints, etc., we can summarize the state of play, circa 1997, as follows. (Information on the history of these, and preceding, results can be found, for example, in the introductions to [5], [6] and in [11, notes to Chap. 3].)

We refer to the following hypotheses, in which \( \bar{x}(\cdot) \) is the \( F \)-trajectory of interest, \( c(\cdot) \) and \( k(\cdot) \) are integrable functions, and \( \epsilon > 0 \) is a given number:

(H1): \( F(\cdot,\cdot) \) is \( \mathcal{L} \times \mathcal{B} \) measurable, where \( \mathcal{L} \) and \( \mathcal{B} \) denote the \( \sigma \)-algebras of Lebesgue subsets of \([0,1]\) and Borel subsets of \( \mathbb{R}^n \) respectively. \( g(\cdot,\cdot) \) is locally Lipschitz continuous and \( C \) is closed,

(H2): \( F(t,x) \subset c(t)\mathbb{B} \),

(H3): \( F(t,x) \subset F(t,x') + k(t)|x-x'|\mathbb{B} \)

for all \( x,x' \in \bar{x}(t) + \epsilon\mathbb{B} \), a.e. \( t \in [0,1] \).

**Theorem 1.1 (Euler Lagrange Inclusion)**

Let \( \bar{x}(\cdot) \) be a \( W^{1,1} \) local minimizer for \((P)\). Assume that (H1)-(H3) are satisfied. Then there exist \( p(\cdot) \in W^{1,1}([0,1];\mathbb{R}^n) \) and, \( \lambda \geq 0 \) such that

(i): \( (p(\cdot),\lambda) \neq (0,0) \),

(ii): \( (p(0),-p(1)) \in \lambda \partial g(\bar{x}(0),\bar{x}(1)) + N_C(\bar{x}(0),\bar{x}(1)) \),

(iii): \( \hat{p}(t) \in \text{co} \{ q \mid (q,p(t)) \in N_{\text{Gr}(F(\cdot,\cdot))}(\bar{x}(t),\dot{\bar{x}}(t)) \} \) a.e.,

(iv): \( p(t) \cdot \dot{\bar{x}}(t) = H(t,\bar{x}(t),p(t)) \) a.e.

The notation \( \partial g(y) \) and \( N_A(y) \) employed in these relations is made precise at the end of this section; these symbols refer to the limiting subdifferential of the function \( g(\cdot) \) and the limiting normal cone of the set \( A \) at a point \( y \) respectively. \( \text{Gr}(\cdot) \) denotes the graph of a multifunction.

If \( F(t,x) \) is convex then, under the hypotheses (H1)-(H3), it is known [9] (see also [1]) that

\[
\text{co} \{ q \mid (q,p(t)) \in N_{\text{Gr}(F(\cdot,\cdot))}(\bar{x}(t),\dot{\bar{x}}(t)) \} \subset \text{co} \{ -q \mid (q,\bar{x}(t)) \in \partial_x p H(t,\bar{x}(t),p(t)) \} .
\]

It follows that the Euler Lagrange inclusion condition implies
Theorem 1.2 (Partially Convexified Hamiltonian Inclusion)

Let $\bar{x}(.)$ be a $W^{1,1}$ local minimizer for $(P)$. Assume that (H1)-(H3) are satisfied. Assume also that

$$(C): F(t,x) \text{ is convex for all } (t,x) \in [0,1] \times \mathbb{R}^n.$$ 

Then there exists $p(.) \in W^{1,1}([0,1];\mathbb{R}^n)$ and $\lambda \geq 0$, such that

(i): $(p(.),\lambda) \neq (0,0),$

(ii): $(p(0),-p(1)) \in \lambda \partial g(\bar{x}(0),\bar{x}(1)) + N_C(\bar{x}(0),\bar{x}(1)) ,$

(iii): $(-\dot{p}(t),\dot{\bar{x}}(t)) \in \text{co} \{ q | (q,\bar{x}(t)) \in \partial_{x,p}H(t,\bar{x}(t),p(t)) \}$ a.e.

It is unnecessary to append the Weierstrass condition (Thm 1.1, (iv)) to the above conditions because it is implied by the Hamiltonian inclusion (2) (and so also by the stronger, partially convexified, condition above).

In the sequel, we shall often summarize the statement ‘there exists non-zero $(\lambda,p(.))$ satisfying the Euler Lagrange inclusion (iii), the transversality (ii) and the Weierstrass conditions (iv) in Thm. 1.1’, simply, as the ‘Euler Lagrange inclusion’. Likewise reference to the ‘(partially convexified) Hamiltonian inclusion’ is intended to convey the transversality condition (ii), in addition to the (partially convexified) Hamiltonian inclusion in Thm. 1.2 or thm. 1.3.

There has been speculation, dating back to the derivation of the original Hamiltonian inclusion for convex $F(.,.)$’s, as to whether the Hamiltonian inclusion, or indeed the more refined, partially convexified Hamiltonian inclusion, continue to be satisfied when $F(.,.)$ is no longer convex valued. In this more general setting, neither the Hamiltonian inclusion nor the partially convexified Hamiltonian follow in general from the Euler Lagrange inclusion, and so the Hamiltonian inclusion (and its partially convexified counterpart) would appear to be distinct necessary conditions.

Clarke took an important step towards settling the issue when, in his 2005 monograph on necessary conditions [3], he proved:

Theorem 1.3 (Hamiltonian Inclusion for Non-convex $F(.,.)$)

Let $\bar{x}(.)$ be a $L^\infty$ local minimizer for $(P)$. Assume that (H1)-(H3) are satisfied. Then there exists $p(.) \in W^{1,1}([0,1];\mathbb{R}^n)$ and $\lambda \geq 0$, such that

(i): $(p(.),\lambda) \neq 0, $

(ii): $(p(0),-p(1)) \in \lambda \partial g(\bar{x}(0),\bar{x}(1)) + N_C(\bar{x}(0),\bar{x}(1)) .$

(iii): $(-\dot{p}(t),\dot{\bar{x}}(t)) \in \text{co} \partial_{x,p}H(t,\bar{x}(t),p(t))$ a.e.

and, in so doing, pioneering the use of Stegall’s theorem in the derivation of first order optimality conditions by perturbational methods, in place of earlier mainstays, such as the variational principles of Ekeland or Borwein and Preiss.
We draw attention to the fact that this last theorem requires \( \bar{x}(.) \) to be a \( L^\infty \) local minimizer, not merely a \( W^{1,1} \) local minimizer, as in the two preceding theorem statements.

Knowledge of validity of the Hamiltonian inclusion for nonconvex \( F(.,.) \)'s and \( L^\infty \) local minimizers still leaves some unanswered questions.

- Is the partially convexified Hamiltonian inclusion valid for nonconvex \( F(.,.) \)'s and \( L^\infty \) local minimizers?
- Is the Hamiltonian inclusion (or its partially convexified counterpart) for nonconvex \( F(.,.) \)'s still valid when the requirement in Thm. 3 that \( \bar{x}(.) \) is a \( L^\infty \) local minimizer is replaced by the weaker requirement that it is merely a \( W^{1,1} \) minimizer?

The purpose of this paper is provide some answers. Concerning the first question, we shall show, by counter-example, that the Hamiltonian inclusion (and so certainly also the stronger partially convexified Hamiltonian inclusion) are not in general valid for non-convex \( F(.,.) \), when the local minimizer in question is merely a \( W^{1,1} \) local minimizer. The counter-example therefore demonstrates that the consideration of \( L^\infty \) local minimizers, in place of \( W^{1,1} \) local minimizers, in the proof of the Hamiltonian inclusion is essential, and not just a technical requirement arising from the analytical techniques earlier employed in the literature.

It has been earlier claimed [5] that \( W^{1,1} \) local minimizers satisfy the partially convexified Hamiltonian inclusion under an additional normality hypothesis. In the final section of the paper we show that this assertion is valid, for a modified notion of 'normality' and when the local minimizer in question is a \( L^\infty \) local minimizer. The counter-example demonstrates that, without these changes, the original claim is incorrect.

**Notation:** In Euclidean space, the length of a vector \( x \) is denoted by \( |x| \), and the closed unit ball \( \{ x \mid |x| \leq 1 \} \) by \( B \). \( d_D(x) \) denotes the Euclidean distance of the point \( x \) from a given closed set \( D \), namely \( \min \{|x - x'| \mid x' \in D \} \).

Some standard constructs from nonsmooth analysis are employed in this paper. Background material may be found in a number of texts, examples of which include [3], [4], [10] and [11].

Take a closed set \( D \subset \mathbb{R}^k \) and a point \( \bar{x} \in D \). The limiting normal cone \( N_D(\bar{x}) \) of \( D \) at \( \bar{x} \) is defined to be

\[
N_D(\bar{x}) := \{ p \mid \exists x_i \xrightarrow[D \to \bar{x}]{} \bar{x}, p_i \xrightarrow{x \to x_i} p \text{ s.t. } \limsup_{x \to x_i} \frac{p_i \cdot (x - x_i)}{|x - x_i|} \leq 0 \text{ for each } i \},
\]

in which \( x_i \xrightarrow{D \to \bar{x}} \bar{x} \) is notation conveying the information '\( x_i \to \bar{x} \)' and '\( x_i \in D \) for all \( i \)'.

Take a lower semicontinuous function \( f(\cdot) : \mathbb{R}^k \to \mathbb{R} \) and a point \( \bar{x} \in \mathbb{R}^k \). The limiting subdifferential of \( f(\cdot) \) at \( \bar{x} \) is

\[
\partial f(\bar{x}) = \{ \xi \mid \exists \xi_i \to \xi \text{ and } x_i \to \bar{x} \text{ s.t. } \limsup_{x \to x_i} \frac{\xi_i \cdot (x - x_i) - f(x) + f(x_i)}{|x - x_i|} \leq 0 \text{ for each } i \}.
\]
2 An Example

Let $\theta(.)$ be the function $\theta(r) = \min\{r, 0\}$. Consider the optimization problem:

$$\begin{align*}
(E) \quad & \text{Minimize } \frac{1}{2} x_1(1) - x_2(1) \\
& \text{over absolutely continuous functions } x(.) = (x_1(.), x_2(.), x_3(.)) \text{ satisfying }
\begin{bmatrix}
\dot{x}_1(t) \\
\dot{x}_2(t) \\
\dot{x}_3(t)
\end{bmatrix} 
\in \begin{bmatrix}
0 & x_1(t) \\
1 & 0
\end{bmatrix} \cup \begin{bmatrix}
0 & x_1(t) \\
-1 & 0
\end{bmatrix} \cup \begin{bmatrix}
0 & \theta(x_1)
\end{bmatrix}
\end{align*}$$

$$(x_1(0), x_2(0), x_3(0)) \in [0, \infty) \times \{0\} \times \{0\},$$
$$(x_1(1), x_2(1), x_3(1)) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}.$$ 

This will be recognized as a special case of $(P)$ in which $F(t, x) = F(x)$ is

$$F(x) = \begin{bmatrix}
0 & x_1 \\
1 & 0
\end{bmatrix} \cup \begin{bmatrix}
0 & x_1 \\
-1 & 0
\end{bmatrix} \cup \begin{bmatrix}
0 & \theta(x_1)
\end{bmatrix}$$

and

$$g(x^0, x^1) = \frac{1}{2} x_1^1 - x_2^1, \quad C = ([0, \infty) \times \{0\} \times \{0\}) \times \mathbb{R}^3.$$ 

The second two coordinates of the multifunction $F(x)$ are illustrated in Fig. 1, for positive, zero and negative values of $x_1$. 

![Figure 1:](image-url)
Salient features of this example are summarized in the following proposition\(^1\):

**Proposition 2.1** The data of problem \((E)\) satisfy hypotheses \((H1)-(H3)\) with reference to the feasible \(F\)-trajectory \((\bar{x}(.) \equiv (0, 0, 0))\), and

(a): \((\bar{x}(.) \equiv (0, 0, 0))\) is a \(W^{1,1}\) local minimizer,

(b): \((\bar{x}(.) \equiv (0, 0, 0))\) is not a \(L^\infty\) local minimizer,

(c): There does not exist non-zero \((\lambda, p(\cdot))\) such that the Hamiltonian inclusion condition \(\text{(iii)}\) and the transversality condition \(\text{(ii)}\) in Thm. 1.3 are both satisfied, with reference to \(\bar{x}(\cdot)\). Neither therefore is it possible to find non-zero \((\lambda, p(\cdot))\) satisfying both the (stronger) partially convexified Hamiltonian inclusion \(\text{(iii)}\) and the transversality condition \(\text{(ii)}\) in Thm. 1.2, with reference to \(\bar{x}(\cdot)\).

\[\text{Figure 2: Switching Function}\]

**Proof**

(a): Take any \(\epsilon \in (0, 1/2)\) and any feasible \(F\)-trajectory \((x_1(\cdot), x_2(\cdot), x_3(\cdot))\) such that

\[||x(\cdot)||_{W^{1,1}} \leq \epsilon.\]

Then \(x_1(\cdot)\) is constant (write the constant value \(x_1\)) and, from the left endpoint constraint, \(x_1 \geq 0\). Since \(x_2(0) = 0\), we deduce from \(\dot{x}(t) \in F(t, x(t))\) that

\[x_2(1) = \int_I x_1 dt + \int_{[0,1]\setminus I} 0 dt = x_1 \times \text{meas}\{I\}\]

and

\[|\dot{x}_3(t)| = \begin{cases} 1 & \text{a.e. } t \in I \\ 0 & \text{a.e. } t \in [0, 1]\setminus I. \end{cases}\]

Here, \(I := \{t \in [0, 1]: \dot{x}_3(t) \neq 0\}\).

It follows that

\[||\dot{x}_3(\cdot)||_{L^1} = \text{meas}\{I\}.\]

\[^1\]A reviewer helpfully points out that the assertions of Prop. 2.1 remain valid when the function \(\theta(\cdot)\) involved in the construction of the example is allowed to be any Lipschitz continuous function such that \(\theta(0) = 0, \theta(r) \leq r\) for \(r \geq 0\) and \(\theta(r) \leq \frac{1}{4}\) for \(r < 0\) and the success of the example depends on the non-differentiability, and failure of Clarke regularity [4, p. 89], of any such function \(\theta(\cdot)\) at \(r = 0\).
Since \( x_3(0) = 0 \), we know \( ||\dot{x}_3(.)||_{L^1} = ||x_3(.)||_{W^{1,1}} \). But then
\[
\text{meas}\{I\} = ||x_3(.)||_{W^{1,1}} \leq ||x(.)||_{W^{1,1}} < \epsilon,
\]
and so \( x_2(1) \leq \epsilon x_1 \). We conclude that
\[
g(x(0), x(1)) = \frac{1}{2} x_1 - x_2(1) \geq x_1(1/2 - \epsilon) \geq 0.
\]
We have shown that \( x(.) \) has cost not lower than that of \( \bar{x}(.) \).

(b): Take any \( \epsilon > 0 \) and write \( \alpha = \epsilon/\sqrt{3} \). Let \( sw(.) \) be the switching function illustrated in Fig. 2
\[
sw(t) = \begin{cases} 
+1 & \text{for } t \in [2k - 2, 2k - 1) \\
-1 & \text{for } t \in [2k - 1, 2k), \quad \text{for } k = 1, 2, \ldots
\end{cases}
\]
Now let \( x(.) = (x_1(.), x_2(.), x_3(.)) \) be the feasible \( F \)-trajectory defined by the relations
\[
\begin{align*}
\{ \dot{x}_1(t), \dot{x}_2(t), \dot{x}_3(t) \} &= (0, \alpha, sw(M\epsilon t)) \quad \text{for } t \in [0, 1], \\
\{ x_1(0), x_2(0), x_3(0) \} &= (\alpha, 0, 0),
\end{align*}
\]
in which \( M\epsilon \) is an even positive integer such that
\[
M\epsilon \geq \frac{\sqrt{3}}{\epsilon}.
\]
We see immediately that \( (x_1(1), x_2(1)) = (\alpha, \alpha) \). A simple calculation (‘integrate the switching function’) yields
\[
|x_3(t)| \leq \epsilon/\sqrt{3} \text{ for all } t.
\]
Since \( |x_1(t)| \) and \( |x_2(t)| \) are bounded by \( \alpha = \epsilon/\sqrt{3} \),
\[
||x(.)||_{L^\infty} \leq (3 \times \left( \epsilon/\sqrt{3} \right)^2)^{\frac{1}{2}} = \epsilon.
\]
Also,
\[
g(x(0), x(1)) = \frac{1}{2} x_1(1) - x_2(1) = \left( \frac{1}{2} - 1 \right) \alpha < 0.
\]
On the other hand, \( F \)-trajectory \( \bar{x}(.) \), which is feasible, has cost \( g(\bar{x}(0), \bar{x}(1)) = 0 \). It follows that there exists a feasible \( F \)-trajectory arbitrarily close to \( \bar{x}(.) \equiv 0 \) w.r.t. the \( L^\infty \) norm, with lower cost, i.e. \( \bar{x}(.) \) is not an \( L^\infty \) local minimizer. The \( F \)-trajectory \( x(.) \) will be recognized as a ‘chattering’ approximation to the feasible co-\( F \) trajectory with constant velocity \((0, \alpha, 0)\).

(c): Suppose that, contrary to the assertion, \( \bar{x}(.) \) does satisfy the Hamiltonian inclusion. Then there exist a costate trajectory \( p(.) = (p_1(.), p_2(.), p_3(.)) \) and \( \lambda \geq 0 \) satisfying the conditions listed in Thm. 1.3. In view of hypotheses (H1)-(H3), which are satisfied by the data for \( (E) \), and since the right endpoints of feasible \( F \)-trajectories are free, we must have \( \lambda > 0 \). By scaling \((\lambda, p(.)\)), we can arrange that \( \lambda = 1 \). The transversality conditions give:
\[
p_1(0) \leq 0, \quad p_1(1) = -\frac{1}{2}, \quad p_2(1) = 1 \quad \text{and} \quad p_3(1) = 0.
\]
The Hamiltonian is
\[
H(t, x, p) = \max \{ (p_2 x_1 + |p_3|), p_2 \times \theta(x_1) \}.
\]
The Hamiltonian inclusion yields \( \dot{p}_2(.) \equiv 0 \), whence \( p_2(.) \equiv 1 \). But then, for each \( t \in [0, 1] \), we have on a neighbourhood of \((\bar{x}(t), p(t))\),
\[
H(t, x, p) = \max \{ (p_2 x_1 + |p_3|), (p_2 \times \min \{x_1, 0\}) \} = p_2 x_1 + |p_3|.
\]
It follows that
\[
\text{co } \partial_{x,p} H(t, \bar{x}(t), p(t)) = \{ (p_2(t), 0, 0), (0, 0, \alpha) \in \mathbb{R}^6 : \alpha \in [-1, +1] \}.
\]
Using this information, we deduce from the Hamiltonian inclusion that \( \dot{p}_1(.) = -p_2(.) \equiv -1 \). Since \( p_1(1) = -\frac{1}{2} \), it follows that \( p_1(0) = +\frac{1}{2} \). But this contradicts \( p_1(0) \leq 0 \). We have confirmed that the Hamiltonian inclusion conditions are not satisfied.

**Exercise.** Because \( \bar{x}(.) \) is a \( W^{1,1} \) local minimizer, it must satisfy the Euler Lagrange inclusion (iii), transversality condition (ii) and Weierstrass condition (iv) in Thm. 1.1. With reference to the Euler Lagrange inclusion, the costate trajectory components \((p_1(.), p_2(.), p_1(.))\) satisfy similar conditions to those associated with the Hamiltonian inclusion, with this crucial difference: Now \( \dot{p}_1(t) \in [-p_2(t), 0] \), yielding merely the information that \( p_1(0) \geq -\frac{1}{2} \). This condition is too weak to establish a contradiction with the transversality condition. We leave the checking of these facts as an exercise.

### 3 Discussion

(i): The example of the previous section establishes that, under hypotheses (H1)-(H3), the Hamiltonian inclusion and transversality condition are not in general both satisfied, with reference to \( W^{1,1} \) local minimizers, if \( F(\cdot) \) is not convex valued. Since the partially convexified Hamiltonian inclusion implies the Hamiltonian inclusion, neither are the partially convexified Hamiltonian inclusion and transversality condition in general both satisfied, with reference to \( W^{1,1} \) local minimizers, if \( F(\cdot) \) is not convex valued. The interplay of the (partially convexified) Hamiltonian inclusion condition and the transversality condition is crucial here. It is possible to find non-zero \((\lambda, p(\cdot))\) satisfying the Hamiltonian inclusion condition alone, but not also the transversality condition, as will be clear from the constructions in the proof of Prop. 2.1.

(ii): The example \((E)\) is not inconsistent with Thm. 1.3, which asserts the validity of the Hamiltonian inclusion for \( L^\infty \) local minimizers, because the \( F\)-trajectory \( \bar{x}(\cdot) \) in example \((E)\) is only a \( W^{1,1} \) local minimizer, not a \( L^\infty \) local minimizer. Rather, it highlights a crucial fact about Thm. 1.3, namely it is an optimality condition only for \( L^\infty \) local minimizers, not for the larger class of \( W^{1,1} \) local minimizers.

(iii): While the partially convexified Hamiltonian inclusion is not in general valid for problem \((P)\), with reference to \( W^{1,1} \) local minimizers, when \( F(\cdot) \) is not convex valued, it might reasonably be conjectured that a version of this necessary condition is valid for ‘finite Lagrangian’ problems:

\[
(P_{\text{finite}}) \quad \text{Minimize } \left\{ \int_0^1 L(t, x(t), \dot{x}(t))dt + l(x(0), x(1)) \right\}
\]
in which \( l(\cdot, \cdot) \) is a locally Lipschitz continuous function and \( L(\cdot, \cdot, \cdot) \) is a finite valued function satisfying hypotheses such as those listed in (6), Section 4 or (11), Section 7.3) or similar, but more stringent, hypotheses. However we can construct from example \((E)\) a finite Lagrangian example that disproves the conjecture. Indeed, because \((E)\) is a free right endpoint problem, \( \bar{x}(\cdot) \) is also a \( W^{1,1}_\cdot \) local minimizer for the finite Lagrangian problem

\[
g(x(0), x(1)) + K \left( d_C(x(0), x(1)) + \int_0^1 d_F(x(t))(\dot{x}(t))dt \right)
\]

in which \( g(\cdot, \cdot), C \) and \( F(\cdot) \) have the identifications of \((E)\) and \( K \) is a suitably large constant. But the appropriate partially convexified Hamiltonian (and transversality) condition for this problem are the same as for \((E)\). So we have arrived at a finite Lagrangian problem for which the convexified Hamiltonian inclusion is not satisfied, with reference to a \( W^{1,1}_\cdot \) local minimizer.

\( \text{(iv)} \): Another insight provided by this example is that the Hamiltonian inclusion can provide information about local minimizers (in the case when \( F(t, x) \) is not convex) unavailable from the Euler Lagrange inclusion. For this example, the Hamiltonian inclusion tells us that \( \bar{x}(\cdot) \) cannot be a \( L^\infty_\cdot \) local minimizer, and so is definitely not a global minimizer; an analysis of the example based on the Euler Lagrange inclusion is inconclusive in this regard.

\( \text{(v)} \): If we append to the hypotheses (H1)-(H3) the extra condition that \( F(\cdot, \cdot) \) is everywhere convex valued, then the partially convexified Hamiltonian inclusion is valid. Can we relax this extra condition and insist merely that \( F(\cdot, \cdot) \) is convex valued along the \( W^{1,1}_\cdot \) minimizer \( \bar{x}(\cdot) \), i.e. that \( F(t, \bar{x}(t)) \) is convex for all \( t \in [0, 1] \)? The answer is no, we cannot; a counter-example is provided by a variant of example \((E)\), obtained by suitably enlarging the right side of the differential inclusion \( \dot{x} \in F(t, x) \) to include line segments, as indicated in Fig. 3. Notice that \( F(t, \bar{x}(\cdot) \equiv (0, 0, 0)) = \{0\} \times \{0\} \times [-1, +1] \) is a convex set for all \( t \in [0, 1] \). Crucially, however, \( F(t, x) \) is nonconvex for any \( x = (x_1, x_2, x_3) \) such that \( x_1 > 0 \).

### 4 The Partially Convexified Hamiltonian Inclusion Under a Normality Hypothesis

The preceding analysis invites the question: for optimal control problems with non-convex velocity sets, do \( L^\infty_\cdot \) local minimizers satisfy the partially convexified Hamiltonian inclusion? While answering this remains an intriguing open problem, it can at least be shown that the partially convexified Hamiltonian inclusion is valid under an additional ‘normality’ hypothesis. This is the substance of the following theorem.

**Definition** A feasible \( F \)-trajectory \( \bar{x}(\cdot) \) is said to be a normal extremal w.r.t. the Hamiltonian inclusion if there exist no non-zero \( p(\cdot) \in W^{1,1}_\cdot \) such that

\[
(i): (p(0), -p(1)) \in N_C(\bar{x}(0), \bar{x}(1)) \,.
\]

\[
(ii): (-\dot{p}(t), \dot{x}(t)) \in \text{co } \partial_{x,p} H(t, \bar{x}(t), p(t)) \text{ a.e.}
\]
Theorem 4.1 (Partially Convexified Hamiltonian Inclusion Under a Normality Hypothesis)

Let \( \bar{x}(.) \) be an \( L^\infty \) local minimizer for \((P)\). Assume that hypotheseses (H1)-(H3) are satisfied. Assume further that \( \bar{x}(.) \) is a normal extremal w.r.t. the Hamiltonian inclusion. Then \( \bar{x}(.) \) satisfies the partially convexified Hamiltonian condition in normal form, that is: there exists \( p(.) \in W^{1,1}([0,1];\mathbb{R}^n) \) such that

(i): \( (p(0),-p(1)) \in \partial g(\bar{x}(0),\bar{x}(1)) + N_C(\bar{x}(0),\bar{x}(1)) \).

(ii): \(-\dot{p}(t) \in \text{co} \{ q \mid (q,\bar{x}(t)) \in \partial_{x,p} H(t,\bar{x}(t),p(t)) \} \text{ a.e.}\)

Remarks

(a): Typically, there are two steps involved in seeking a solution to an optimal control problem by means of a multiplier rule, of which the partially convexified Hamiltonian inclusion is an example. First, we check that the multiplier rule cannot be satisfied by any feasible element in a manner such that the cost multiplier is zero, i.e. that extremals are normal. Second, we seek a feasible element that satisfies the optimality condition in normal form, in the knowledge that such an element exists. The above theorem tells us that this methodology can be applied using the partially convexified Hamiltonian inclusion but with this crucial difference: in the first step, we must establish normality w.r.t. the Hamiltonian inclusion condition, not w.r.t. the partially convexified Hamiltonian condition.

(b): The idea of establishing that, for non-convex \( F(.,.) \)'s, a local minimizer satisfies the partially convexified Hamiltonian inclusion under a normality hypothesis related to another kind of necessary condition, is due to Ioffe. Indeed it is asserted in ([5], Thm. 7) that \( W^{1,1} \) local minimizers satisfy the partially convexified Hamiltonian inclusion, under a normality...
hypothesis related to the Euler Lagrange inclusion. But this assertion is, in general, incor-
rect, as is revealed by the example of Section 2. This optimal control problem is normal
(in the sense of the Euler Lagrange inclusion), with reference to the specified feasible $F$
trajectory, because it is a free right endpoint problem. But, as we have shown, the partially
convexified Hamiltonian condition is not satisfied. An explanation for the discrepancy is
as follows: the proof of ([5], Thm. 7) depends on ([6], Prop. 9), whose proof is based
on the assumption that, for finite Lagrangian problems, $W^{1,1}$ local minimizers satisfy the
partially convexified Hamiltonian inclusion. This assumption is not in general justified, as
we remarked in discussion item (iii) in Section 3.

The above theorem may be viewed then as a correct statement of these earlier assertions,
in which the local minimizer in question is taken to be an $L^\infty$ local minimizer rather than
a $W^{1,1}$ local minimizer, and normality is defined in relation to the Hamiltonian inclusion,
not the Euler Lagrange inclusion.

Proof. We may assume, without loss of generality that the functions $c(.)$ and $k(.)$ in (H2) and
(H3) are essentially bounded. (This is because, if not, we can introduce a change of time variable
which renders them essentially bounded; we then prove the stated assertions for the transformed
problem, which imply the same assertions for the original problem, c.f. [3, Section 2.4].) By
reducing the size of $\epsilon > 0$ in (H1) and (H3) we can arrange that the $L^\infty$
local minimizer $\bar{x}(.)$ minimizes $g(x(0), x(1))$ over all feasible $F$-trajectories satisfying
$||x(.) - \bar{x}(.)|| \leq \epsilon$. Now define

$$ S := \{ x(.) \in W^{1,1} | \dot{x}(t) \in F(t, x(t)) \text{ a.e. and } ||x(.) - \bar{x}(.)||_{L^\infty} \leq \epsilon/2 \}. $$

For arbitrary $K > 0$ define $J_K(.) : W^{1,2} \to R$

$$ J_K(x(.)) := (1/K)g(x(0), x(1)) + d_C(x(0), x(1)) + \int_0^1 |x(t) - \bar{x}(t)|^2 dt. $$

Here $W^{1,2}$ denotes the Hilbert space of absolutely continuous functions with square integrable
derivatives, with the inner product $<x(.), x'(.)> = x(0) \cdot x'(0) + \int_0^1 \dot{x}(t) \cdot \dot{x}'(t) dt$. There are two alternatives to consider:

(a): There exists $K > 0$ such that

$$ J_K(\bar{x}(.)) = \min \{ J_K(x(.)) | x(.) \in S \}, $$

and

(b): There exist sequences $K_i \uparrow \infty$, $\delta_i \downarrow 0$ such that, for some sequence $\{y_i(.)\}$ in $S$, we have

$$ J_{K_i}(y_i(.)) < \min J_{K_i}(\bar{x}(.)) - \delta_i \text{ for all } i. \quad (4) $$

Consider first (a). In this case, $\bar{x}(.)$ is an $L^\infty$ local minimizer for

$$ \left\{ \begin{array}{l}
\text{Minimize } (1/K)g(x(0), x(1)) + d_C(x(0), x(1)) + \int_0^1 |x(t) - \bar{x}(t)|^2 dt \\
\text{over absolutely continuous functions } x(.) : [0, 1] \to \mathbb{R}^n \text{ satisfying} \\
\dot{x}(t) \in F(t, x(t)) \text{ for a.e. } t \in [0, 1].
\end{array} \right. $$

12
Because there are no endpoint constraints, we can conclude from the Relaxation Theorem [11, Thm. 2.7.3] that \( \bar{x}(.) \) is also an \( L^\infty \) local minimizer for the related relaxed optimal control problem, in which \( c \circ F(t, x) \) replaces \( F(t, x) \). So \( \bar{x}(.) \) satisfies the partially convexified Hamiltonian inclusion conditions in normal form for the relaxed problem. (To see this, we absorb the integral cost term into the state, by state augmentation, and apply Thm. 1.2; because there is no right endpoint constraint, these necessary conditions apply in normal form.) The presence of the integral cost term has no effect on the conditions, because \( x \to |x - \bar{x}(t)|^2 \) is continuously differentiable with vanishing derivative at \( x = \bar{x}(t) \). Notice that the transversality condition is:

\[
(p(0), -p(1)) \in \partial(g + KdC)(\bar{x}(0), \bar{x}(1)) \subset \partial g(\bar{x}(0), \bar{x}(1)) + KdC(\bar{x}(0), \bar{x}(1)) \\
\subset \partial g(\bar{x}(0), \bar{x}(1)) + N_C(\bar{x}(0), \bar{x}(1)).
\]

This is the transversality condition in normal form. Since the Hamiltonian, and so also the partially convexified Hamiltonian inclusion condition, are the same for the original and convexified optimal control problems, we have proved the assertions of the theorem in case (a).

Consider next case (b). We shall complete the proof by showing that this case is incompatible with the normality hypothesis, and therefore cannot arise.

Take sequences \( K_i \uparrow \infty \), \( \delta_i \downarrow 0 \) and \( \{y_i\} \) in \( S \) having the properties associated with case (b). Since \( c(\cdot) \) is essentially bounded, elements \( x(\cdot) \) in \( S \), together with their derivatives \( \dot{x}(\cdot) \), are uniformly essentially bounded. We can therefore find constants \( k_1, k_2, k_3 > 0 \) such that, for any \( x(\cdot) \in S \),

\[
|g(x(0), x(1))| \leq k_1, \quad ||x(\cdot) - \bar{x}(\cdot)||_{L^\infty} \leq k_2 \quad \text{and} \quad ||x(\cdot)||_{L^\infty} \leq k_3||x(\cdot)||_{L^2}.
\]

(The last inequality is valid because, on subsets of \( L^\infty \) in which elements \( x(\cdot) \) and their velocities \( \dot{x}(\cdot) \) are uniformly essentially bounded, the \( L^\infty \) and the \( L^2 \) norms are equivalent.)

For each \( i \), apply Stegall’s Theorem [4, p. 43] to the proper, lower semicontinuous function \( \Phi(x) : W^{1,2} \to \mathbb{R} \cup \{+\infty\} \)

\[
\Phi(x(\cdot)) = \begin{cases} 
J_{K_i}(x(\cdot)) & \text{if } x(\cdot) \in S \\
+\infty & \text{otherwise}
\end{cases}
\]

defined on the Hilbert space \( \mathbb{R}^n \times L^2 \). We thereby obtain \( (d_i, a_i(\cdot)) \in \mathbb{R}^n \times L^2 \) and \( x_i(\cdot) \in S \), such that \( |d_i|^2 + ||a_i(\cdot)||_{L^2}^2 \) is arbitrarily small, and \( x_i(\cdot) \) minimizes

\[
x(\cdot) \to J_{K_i}(x(\cdot)) + d_i \cdot x(0) + \int_0^1 a_i(t) \cdot \dot{x}(t) dt
\]

over \( x(\cdot) \in S \). We can arrange that \( (d_i, a_i(\cdot)) \to (0, 0) \) in \( \mathbb{R}^n \times L^2 \) as \( i \to \infty \). Noting that the \( L^1 \) norm is weaker than the \( L^2 \) norm, we can also arrange that

\[
(k_2 + ||c(\cdot)||_{L^\infty}) \times \max\{|d_i|, ||a_i||_{L^1}\} < \delta_i/4 \quad \text{for each } i,
\]

for each \( i \). From (4) we know

\[
(1/K_i)g(y_i(0), y_i(1)) + d_C(y_i(0), y_i(1)) + \int_0^1 |y_i(t) - \bar{x}(t)|^2 dt \\
\leq (1/K_i)g(\bar{x}(0), \bar{x}(1)) - \delta_i
\]
Taking account of (5) and (6), and also the optimizing property of \(x_i(.)\), we can show that

\[
(1/K_i)g(x_i(0), x_i(1)) + d_C(x_i(0), x_i(1)) + \int_0^1 |x_i(t) - \bar{x}(t)|^2 dt
\]

\[
\leq (1/K_i)g(y_i(0), y_i(1)) + d_C(y_i(0), y_i(1)) + \int_0^1 |y_i(t) - \bar{x}(t)|^2 dt + d_i \cdot (y_i(0) - x_i(0)) + \int_0^1 (\dot{y}_i(t) - \dot{x}_i(t)) \cdot a(t) dt
\]

\[
\leq (1/K_i)g(\bar{x}(0), \bar{x}(1)) - \delta_i + d_i \cdot (y_i(0) - x_i(0)) + \int_0^1 (\dot{y}_i(t) - \dot{x}_i(t)) \cdot a(t) dt
\]

\[
\leq (1/K_i)g(\bar{x}(0), \bar{x}(1)) - \delta_i + 2(k_2 + ||c(.)||_{L^\infty}) \times \max\{|d_i|, ||a_i||_{L^1}\}
\]

\[
\leq (1/K_i)g(\bar{x}(0), \bar{x}(1)) - (\delta_i/2) .
\]  

(7)

This inequality implies that

\[ ||x_i(.) - \bar{x}(.)||_{L^\infty} \leq k_3 \left( \int_0^1 |x_i(t) - \bar{x}(t)|^2 dt \right)^{\frac{1}{2}} \leq k_3 \times \left( \frac{2k_1}{K_i} \right)^{\frac{1}{2}} . \]

It follows that \[ ||x_i(.) - \bar{x}(.)||_{L^\infty} \to 0 \text{ as } i \to \infty . \]

Inequality (7) also implies that \(g(x_i(0), x_i(1)) < g(\bar{x}(0), \bar{x}(1))\). But then \(d_C(x_i(0), x_i(1)) > 0 , \)

for \(i\) sufficiently large since, otherwise, \(\bar{x}(.)\) cannot be a \(L^\infty\) local minimizer for \((P)\).

Let us now investigate the properties of \(x_i(.)\), in the limit as \(i \to \infty\). We have established that \(x_i(.) \to \bar{x}(.)\) in \(L^\infty\) and \((d_i, a_i(.)) \to (0, 0)\) in \(\mathbb{R}^n \times L^2\). By extracting subsequences (we do not re-label), we can arrange that \(a_i(.) \to 0\) a.e. Write

\[ z_i(t) = \int_0^t (a_i(s) \cdot \dot{x}_i(s) + |x_i(s) - \bar{x}(s)|^2) ds . \]

The minimizing properties of the sequence \(\{x_i(.)\}\) can be expressed thus: for each \(i\), \((x_i(.), z_i(.)\)

is a \(L^\infty\) local minimizer for the problem

\[
\left\{ \begin{array}{l}
\text{Minimize } (1/K_i)g(x(0), x(1)) + z(1) + d_C(x(0), x(1)) + d_i \cdot x(0)
\end{array} \right. 
\]

over absolutely continuous functions \((x(.), z(.)) : [0, 1] \to \mathbb{R}^{n+1}\) satisfying

\[
(\dot{x}(t), \dot{z}(t)) \in \{(e, a_i(t) \cdot e) | e \in F(t, x(t))\} + \{(0, |x(t) - \bar{x}(t)|^2)\} \text{ a.e. } ,
\]

\[
z(0) = 0 .
\]

Since this is a free right endpoint problem, \((x_i(.), z_i(.))\) satisfies the Hamiltonian inclusion condition of Thm. 1.3, in normal form. In terms of the costate vector block elements, denoted \((p, -r)\), the Hamiltonian is

\[
\tilde{H}(t, x, p, r) = \max_{e \in F(t, x)} \{(p - ra_i(t)) \cdot e\} - r|x - \bar{x}(t)|^2
\]

\[
= H(t, x, p - ra_i(t)) - r|x - \bar{x}(t)|^2 .
\]  

(9)
Take any \( i \). Let \(( p_i(\cdot), r_i(\cdot) )\) be the costate trajectory. Since \( \tilde{H}(t, \cdot, \cdot, \cdot, \cdot) \) does not depend on \( z, \dot{z}(t) = 0 \). But \( r_i(1) = 1 \), by the transversality condition. We conclude that \( r_i(\cdot) \equiv 1 \). The Hamiltonian inclusion tells us that
\[
(-\dot{p}_i(t), \dot{x}_i(t), \dot{z}_i(t)) \in \partial_{x,p,r} \tilde{H}(t, x_i(t), p_i(t), r_i(t)) \text{ a.e.} \quad (10)
\]
An analysis of subgradients of \( \tilde{H}(t, \cdot, \cdot, \cdot, \cdot) \) at \(( x_i(t), p_i(t), r_i(t) )\), characterized as limits of neighboring proximal subgradients, permits us to deduce from (10) that
\[
(-\dot{p}_i(t), \dot{x}_i(t)) \in \partial_{x,p} H(t, x_i(t), p_i(t) - a_i(t)) + \{(0,0)\} \times 2| x_i(t) - \bar{x}(t)| \mathbb{B} \text{ a.e.} \quad (11)
\]
The transversality condition also yields the information
\[
(p_i(0), -p_i(1)) \in (1/K_i) \partial g(x_i(0), x_i(1)) + |d_i| \mathbb{B} \times \{0\} + \{\eta_i\}, \quad (12)
\]
in which \( \eta_i \in \partial d_C(x_i(0), x_i(1)) \). In consequence of (8) (see, e.g., [11, Lemma 2.5.3]), we know that \( |\eta_i| = 1 \). So we can write
\[
\eta_i \in \partial d_C(x_i(0), x_i(1)) \cap \{\eta \in \mathbb{R}^n \times \mathbb{R}^n \mid |\eta| = 1\} . \quad (13)
\]
Since, by (11) and (13), \( p_i(\cdot) \) and its derivatives \( \dot{p}_i(\cdot) \) are uniformly bounded, we can arrange, by subsequence extraction, that
\[
p_i(\cdot) \to p(\cdot) \text{ uniformly, and } \dot{p}_i(\cdot) \to \dot{p}(\cdot) \text{ weakly in } L^1 ,
\]
for some absolutely continuous function \( p(\cdot) \).

A standard convergence analysis (see, e.g., [11, proof of Thm. 2.5.3]), which makes use of the fact that \( a_i(\cdot) \to 0 \) a.e. and \( d_i \to 0 \), yields, in the limit,
\[
(-\dot{p}(t), \dot{x}(t)) \in \partial_{x,p} H(t, \bar{x}(t), p(t)) \text{ a.e.}
\]
From (12) and (13) we obtain
\[
(p(0), -p(1)) \in N_C(\bar{x}(0), \bar{x}(1)) \quad \text{and} \quad |(p(0), -p(1))| = 1 .
\]
But this is not possible under the normality hypothesis. We have confirmed that case (b) is never encountered.

Acknowledgments. This work was supported by the European Union under the 7th Framework Programme FP7-PEOPLE-2010-ITN Grant agreement number 264735-SADCO and by EPSRC under the grant EP/G066477/1.

References


