THE EULER AND WEIERSTRASS CONDITIONS FOR NONSMOOTH VARIATIONAL PROBLEMS[1]

(to appear in Calculus of Variations and Partial Differential Equations)

A. D. Ioffe[2] and R. T. Rockafellar[3]

Abstract. Necessary conditions are developed for a general problem in the calculus of variations in which the Lagrangian function, although finite, need not be Lipschitz continuous or convex in the velocity argument. For the first time in such a broadly nonsmooth, nonconvex setting, a full subgradient version of Euler's equation is derived for an arc that furnishes a local minimum in the classical weak sense, and the Weierstrass inequality is shown to accompany it when the arc gives a local minimum in the strong sense. The results are achieved through new techniques in nonsmooth analysis.

---

[1]**AMS (1991) subject classification** 49K05, 49J52, 58C20

This research was supported in part by funds from the U.S.-Israel Science Foundation under grant 90–00455, and also by the Fund for the Promotion of Research at the Technion under grant 100–954 and by the U.S. National Science Foundation under grant DMS-9200303.

[2]Dept. of Mathematics, Technion, Haifa 34608 Israel; e–mail: ioffe@techunix.technion.ac.il

[3]Depts. of Mathematics and Applied Mathematics, University of Washington, Seattle 98195 USA e–mail: rtr@washington.math.edu

1 Introduction

Let W_p^1 denote the Banach space consisting of all R^n-valued absolutely continuous functions $x(\cdot)$ on $[0,1]$ with $\dot{x}(\cdot) \in L_p$. (The norm can be taken as $\|x(\cdot)\|_p^1 = |x(0)| + \|\dot{x}(\cdot)\|_p$, where $|\cdot|$ is the Euclidean norm of a vector in R^n.) The *generalized problem of Bolza*, which will be the focus of our attention in this paper, concerns the minimization over W_1^1 of a functional

$$J(x(\cdot)) = l(x(0), x(1)) + \int_0^1 L(t, x(t), \dot{x}(t))dt, \tag{1}$$

where the functions l and L, in contrast to the traditional setting for the calculus of variations, need not be differentiable or even continuous, and for some purposes may be extended-real-valued. These features account for the ability of such a seemingly simple problem statement to cover the many complications built into the classical problem of Bolza, including constraints of all kinds on $(x(0), x(1))$ and $(x(t), \dot{x}(t))$: such constraints can be represented through penalties, finite or infinite. The generalized problem covers not only classical models but problems in optimal control as well.

In the inherited terminology of the subject, an arc $x_*(\cdot) \in W_1^1$ furnishes a *strong* local minimum of J if $J(x_*(\cdot)) \leq J(x(\cdot))$ for all $x(\cdot)$ in a set of the form

$$\{\, x(\cdot) \in W_1^1 : |x(t) - x_*(t)| \leq \epsilon \text{ for all } t \in [0,1] \,\}$$

for some $\epsilon > 0$, whereas it gives a *weak* local minimum if the comparison merely holds on the smaller kind of set

$$\{\, x(\cdot) \in W_1^1 : |x(t) - x_*(t)| \leq \epsilon \text{ and } |\dot{x}(t) - \dot{x}_*(t)| \leq \epsilon \text{ for a.e. } t \in [0,1] \,\}.$$

For an arc furnishing a weak minimum, the classical first-order necessary condition is a transversality condition plus the Euler-Lagrange equation. In the case of a strong minimum, the Euler-Lagrange equation is supplemented by the Weierstrass condition. In both cases the Euler-Lagrange equation can be written instead as the Hamiltonian equation, as long as L has positive-definite a second-derivative matrix with respect to the velocity argument. But these conditions rely on differential calculus for their derivation and even their formulation. Attempts at extending them to the general problem of Bolza must confront possible nonsmoothness and discontinuity in l and L, and this calls for a different level of analysis.

While much has already been accomplished in this direction, as will be reviewed presently, most derivations of an extended Euler condition for functions L of general character have dealt only with the case of a strong minimum, yet they have not at the same time provided the anticipated necessity of the Weierstrass condition, except in situations where convexity assumptions make it virtually automatic. We aim to address this discrepancy in two ways. First, we establish the necessity of a state-of-the-art Euler condition in the case of a weak minimum for problems with integrands L much more general than those successfully handled in this respect so far. Second, we show that in the case of a strong minimum involving such integrands, the Weierstrass condition must indeed hold as well. By the Weierstrass condition we mean the inequality

$$L(t, x_*(t), y) \geq L(t, x_*(t), \dot{x}_*(t)) + \langle p(t),\, y - \dot{x}_*(t)\rangle \quad \text{for all } y, \text{ a.e. in } t \tag{2}$$

holding for an adjoint arc $p(\cdot) \in W_1^1$ that is paired with $x_*(\cdot)$ through the Euler condition, whatever it may be. The classical Weierstrass condition can also be viewed from another angle as relating to a lower semicontinuity property of the functional (1), but that mode of thinking leads instead to relaxation theory—the study of the relation between the problem with L and the same problem with $\bar{L}$, the integrand obtained by convexifying $L(t, x, y)$ in the y argument—which is not our topic here. It may be observed however that if (2) does hold, then $L(t, x_*(t), \dot{x}_*(t)) = \bar{L}(t, x_*(t), \dot{x}_*(t))$ a.e., so an important implication for relaxation theory is immediate.

Bolza problems with nondifferentiable, possibly extended-real-valued data elements l and L were first studied in papers of Rockafellar [25, 26, 27], in the fully convex case, where $l(x, y)$ and $L(t, x, y)$ are convex functions of $(x, y) \in R^n \times R^n$. In that case subgradients in the sense of convex analysis can be utilized in place of gradients for the purpose of expressing extensions of Euler's equation and other conditions. Euler's equation relative to an optimal arc x_* comes out as asserting the existence of an adjoint arc p such that

$$(\dot{p}(t), p(t)) \in \partial L(t, x_*(t), \dot{x}_*(t)) \quad \text{a.e. in } t, \tag{3}$$

where the set on the right consists of the subgradients of the convex function $L(t, \cdot, \cdot)$ at $(x_*(t), \dot{x}_*(t))$. Due to full convexity, this condition is always sufficient for a global minimum and, under a constraint qualification (with respect to the constraints implicit in the problem formulation with ∞), it is also necessary. Indeed, no distinction is needed between a strong or weak minimum, and the Weierstrass condition is automatic. Also, the Euler condition can always be recast equivalently as a subgradient type of Hamiltonian equation.

The 1973 dissertation of Clarke [3] greatly broadened the scope of this approach by introducing a robust concept of subgradient for *nonconvex* functions. Convexity was replaced to an important degree by assumptions of Lipschitz continuity on L or its associated epigraphical mapping. Clarke showed in [4] that a generalized subgradient version of Euler's equation,

$$(\dot{p}(t), p(t)) \in \bar{\partial} L(t, x_*(t), \dot{x}_*(t)) \quad \text{a.e. in } t, \tag{4}$$

where $\bar{\partial}$ signifies subgradients in his broader sense, must hold as a necessary condition for a weak minimum in problems without full convexity, as long as certain assumptions are satisfied, which include $L(t, \cdot, \cdot)$ being Lipschitz continuous around $(x_*(t), \dot{x}(t))$. In [6] he obtained a subgradient Hamiltonian condition for a strong minimum, but without the intermediary of an Euler condition. In [9] (see also the book [10], p. 187) he did state, for a strong minimum, the Weierstrass condition (2) in combination with an Euler condition, but the latter in the form

$$\dot{p}(t) \in \bar{\partial}_x L(t, x_*(t), \dot{x}_*(t)) \text{ and } p(t) \in \bar{\partial}_y L(t, x_*(t), \dot{x}_*(t)) \quad \text{a.e. in } t. \tag{5}$$

Although (4) and (5) would be identical if one were dealing with gradients, they can be quite different for subgradients. Under the assumption of local Lipschitz continuity made by Clarke in his presentation of this result, (5) is typically weaker than (4), as he well recognized in [10], since the closure of the graph of the mapping $(x, y) \mapsto \bar{\partial}_x L(t, x, y) \times \bar{\partial}_y L(t, x, y)$ includes the graph of the mapping $(x, y) \mapsto \bar{\partial} L(t, x, y)$. (In special circumstances, (5) can provide some information beyond (4), but this is exceptional; an example will be seen at the end of this paper.) Time independence of L was also assumed by Clarke in this connection, but he based the arguments on his work with the maximum principle in [7], which would also support an approach under less serious restrictions.

Since these contributions, Loewen and Rockafellar [20, 21] have used Clarke's Hamiltonian condition to derive a sharper version of the Euler condition than (4) as necessary for a strong minimum. This version, described below, involves less convexification in the formation of subgradient sets. It is the Euler condition we adopt here. Loewen and Rockafellar did not require L to be finite and locally Lipschitz continuous, but they imposed convexity of $L(t, x, y)$ in y. For this reason, they were easily able to get the Weierstrass condition too, on the side; see also Clarke [12] for a subsequent contribution under such assumptions, and Ioffe [17] and Rockafellar [28, 29] for subgradient connections between Lagrangians and Hamiltonians more generally in the case of partially convex integrands.

The move toward application of a less-convexified subgradient set received much impetus from work of Mordukhovich [22, 23]. Recently also, Mordukhovich succeeded in demonstrating in [24] that a sharpened subgradient version of the Euler condition is necessary for an "intermediate" minimum with certain integrands $L(t, x, y)$ that are not convex in y, an intermediate minimum being a minimum relative to a neighborhood in the W^1_p norm for some $p < \infty$. This did not entail the Weierstrass condition at the same time, however.

As a final observation in this vein, we note that for integrands not depending on x the Weierstrass condition in combination with the Euler condition is an easy outcome of the application of convex analysis to the description of solutions of optimal control problems exhibiting linearity with respect to state variables; cf. Ioffe and Tikhomirov [19]. (We are indebted to Tikhomirov for the observation that in the Euler–type condition (5) with separately taken subdifferentials, one of them, namely with respect to y, can be the smaller non–convex subdifferential while the other must still be the generalized gradient of Clarke.)

These results have left the status of the Euler condition unclear with regard to its necessity for a weak minimum in general problems with non-Lipschitz L, and more importantly its pairing with the Weierstrass condition in the case of a strong minimum.

We do not exclude the possibility that the recent arguments of Loewen-Rockafellar and Mordukhovich might be adjustable to cover the case of the weak minimum but this is not obvious and would require effort. Maybe the same can be said about Clarke's landmark derivation of the Euler condition in [6] based on calmness as a constraint qualification. Here, though, we apply different tactics in attempting to resolve the issues. As far as the Weierstrass condition is concerned, it would be too much to hope that it would hold for every adjoint arc $p(\cdot)$ appearing in the Euler condition, but one may naturally ask whether it must hold for at least one such arc.

We are not able at the current stage of our technique to take on the full generality of extended-real-valued L, but at least we can handle integrands that can well be discontinuous. While integrands with ∞ have already yielded to analysis of Euler possibilities for a strong minimum, only Lipschitz continuous integrands, or ones with convexity in y, have previously been covered by work in the directions in which we are headed here. The endpoint function l poses fewer hurdles. We do allow it to be extended-real-valued, and our results therefore apply to problems with constraints on $x(0)$ and $x(1)$ as represented through infinite penalties.

To formulate our main result precisely, we must begin by clarifying the kind of subgradients that are involved. These need only be described in a setting of Hilbert spaces, inasmuch as that includes not only R^n but the subspace W_2^1 of W_1^1 in which our methodology allows us to work through truncation.

For any Hilbert space X and function $f : X \to \bar{R}$ (where $\bar{R}$ denotes the extended real line $[-\infty, \infty]$), a *proximal subgradient* of f at a point x with $f(x)$ finite is an element $x^* \in X^*$ such that there exist $\epsilon > 0$ and $k > 0$ for which

$$f(x+u) - f(x) - \langle x^*, u\rangle \geq -k\|u\|^2, \text{ if } \|u\| < \epsilon.$$

The set of all proximal subgradients of f at x will be denoted by $\partial_p f(x)$. The *limiting proximal subdifferential* of f at x, when f is l.s.c. at x, is the set

$$\partial f(x) = \limsup_{\substack{u \to x \\ f(u) \to f(x)}} \partial_p f(u)$$

(with the limsup taken with respect to weak (sequential) convergence of the proximal subgradients).

This is the same as the subdifferential adopted as the vehicle of basic nonsmooth analysis in R^n by Mordukhovich (see [22] and its references) and in infinite-dimensional spaces developed as the *approximate* subdifferential by Ioffe [16], and a considerable body of theory is therefore available to us when utilizing it. It is smaller than the Clarke subdifferential, which would correspond to its convex hull in a certain extended sense.

Our choice of the proximal subdifferential as the starting point for our efforts is explained by a specific advantage that it has over competitors like the Dini or Fréchet subdifferentials in handling integral functionals. If $I(u(\cdot)) = \int_0^1 g(t, u(t))dt$ on L_2 and

$v(\cdot) \in \partial_p I(u(\cdot))$, then $v(t) \in \partial_p g(t, u(t))$ for almost every t (the subgradients being taken in the u argument only). Effective treatment of integral functionals is of course essential to our aims.

In minimizing the functional J over the space W_1^1 we adopt the following assumptions on the functions l and L relative to the candidate arc $x_*(\cdot)$ later to be involved in the necessary condition:

(A_1) $l(x,y)$ is l.s.c., possibly with ∞ but not $-\infty$, and $l(x_*(0), x_*(1))$ finite;

(A_2) $L(t,x,y)$ is an everywhere finite and lower semicontinuous function of (x,y) for almost every $t \in [0,1]$, and it is measurable with respect to t in the sense that the set valued mapping $t \mapsto \operatorname{epi} L(t,\cdot,\cdot)$ from $[0,1]$ into $R^n \times R^n$ is measurable;

(A_3) for any $N > 0$ there are an $\epsilon > 0$ and $k(t) \in L_1,\ c(t) \in L_1$ such that

$$|L(t,x,y) - L(t,x',y)| \le k(t)|x - x'| \quad \text{and} \quad |L(t,x,y)| \le c(t)$$

when $|y - \dot{x}_*(t)| \le N,\ |x - x_*(t)| \le \epsilon,\ |x' - x_*(t)| \le \epsilon$.

Note that convexity of $L(t,x,y)$ with respect to y is *not* assumed. Furthermore, L need not be locally Lipschitz continuous with respect to (x,y), but merely in x for fixed y, and even that only around the points $(x_*(t), \dot{x}_*(t))$. As a function of y alone, $L(t,x,y)$ could well have discontinuities, provided that lower semicontinuity, at least, is maintained.

Theorem 1 *Suppose that $x_*(\cdot)$ furnishes a local minimum of $J(x(\cdot))$ in the norm of W_1^1 (as holds in particular if $x_*(\cdot)$ furnishes a classical strong minimum), and that (A_1)–(A_3) hold. Then there is an arc $p(\cdot)$ in W_1^1 such that following three relations are satisfied:*

(a) Euler condition: $\dot{p}(t) \in \operatorname{conv}\{w : (w, p(t)) \in \partial L(t, x_*(t), \dot{x}_*(t))\}$ a.e. in t;

(b) Weierstrass condition: $L(t, x_*(t), y) \ge L(t, x_*(t), \dot{x}_*(t)) + \langle p(t),\ y - \dot{x}_*(t)\rangle$ for all y, a.e. in t;

(c) transversality condition: $(p(0), -p(1)) \in \partial l(x_*(0), x_*(1))$.

Moreover, the Euler condition (a) and transversality condition (c) hold even if $x_*(\cdot)$ merely furnishes a classical weak minimum.

In the Euler condition (a), the subgradients are those of the limiting proximal subdifferential, also called *approximate* (e.g. [14, 15]) or *generalized derivatives* (e.g. [22]) of $L(t,x,y)$ with respect to (x,y). (Mordukhovich's result in [24], in contrast, appeals to a larger subdifferential generated by taking proximal subgradient limits in (t,x,y) rather than just in (x,y).)

Our technique for proving Theorem 1 rests on the "fuzzy subgradient calculus" that was first developed for Dini and Fréchet subdifferentials by Ioffe in [14, 15, 16]. Here we require an analogous result for *proximal* subgradients in Hilbert spaces. The modified argument that we use to obtain this result (involving the uniform lower semicontinuity condition (ULC) below) can be applied in fact to a wide range of subdifferentials, not just proximal ones; see Borwein and Ioffe [1].

We also draw on various results about convexification, some new. In particular we show how to calculate from the subgradients of L those of the corresponding relaxed Lagrangian. In dealing with the relaxed problem we use a standard-looking theorem which, however, is not completely covered by the other results we are aware of, e.g. in contributions of Warga [30], Ioffe and Tikhomirov [18], Ekeland and Temam [13], and Clarke [5]. Among these, Clarke's relaxation result comes closest to ours, but he assumes that $L(t,\cdot,\cdot)$ is Lipschitz whereas we do not.

We employ the strategy of first establishing the Euler condition under a slightly different set of assumptions. These assumptions are satisfied in particular for the relaxed problem corresponding under the normalization $x_*(t) \equiv 0$ to the truncated ϵ-integrand

$$L_{N\epsilon} = \begin{cases} L(t,x,y) + \epsilon|y|^2 & \text{if } |y| \le N, \\ \infty & \text{if } |y| > N, \end{cases}$$

which we consider on W_2^1 instead of W_1^1. We use our convexification formula to get both the Euler condition and the Weierstrass condition for $L_{N\epsilon}$ in this setting. By passing to the limit as $N \to \infty$ and $\epsilon \to 0$ we then obtain our general result in W_1^1.

The crucial difficulty that has to be overcome in proving the Euler condition is that unless the sequence of adjoint arc derivatives $\dot{p}^\nu(\cdot)$ we generate converges *pointwise* to $\dot{p}(\cdot)$, the condition ultimately obtained in terms of $x_*(\cdot)$ and $p(\cdot)$ can at best be of Clarke type—with full subgradient convexification. We achieve pointwise convergence by introducing $p^\nu(\cdot)$ carefully as a proximal subgradient of the restriction of J to the Hilbert space W_2^1, and then resort to methodology of infinite-dimensional nonsmooth analysis.

This approach has two advantages over the limit approach of Mordukhovich. It does not require continuity of $L(t,x,y)$ in t, and it leads directly to the necessity of the Euler condition for a weak minimum instead of just an "intermediate" kind of minimum. But it requires very delicate analysis as $N \to \infty$ and $\epsilon \to 0$ and therefore, unlike Mordukhovich's, has little claim to providing support for the construction of numerical approximations.

The last section of the paper provides examples which help to clarify the relationship between our necessary conditions and the earlier ones obtained by Clarke.

2 "Fuzzy Calculus" of Proximal Subgradients

We shall begin our discussion by estimating the proximal subgradients of a sum of functions by means of proximal subgradients of the summands. For functions on R^n, the necessary estimate is provided by the "fuzzy calculus" developed in [15]. We refer to [16, 1] for similar results involving other (than proximal) subdifferentials in infinite dimensional spaces.

Theorem 2 *Let X be a Hilbert space and $f_1,\dots, f_k$ be (extended-real-valued) functions defined and lower semicontinuous in a neighborhood of $\bar{x}$ and finite at $\bar{x}$. Assume that the following uniform lower semicontinuity property holds on the diagonal:*

(ULC) *there is a $\delta > 0$ such that for any k sequences $\{x_{ir}\}$, $i = 1,\dots,k$, $r = 1, 2,\dots$ belonging to the δ–ball around $\bar{x}$ and such that $\|x_{ir} - x_{jr}\| \to 0$ as $r \to 0$, there is a sequence $\{u_r\}$ of elements of the ball such that $\|x_{ir} - u_r\| \to 0$ and*

$$\liminf_{r\mapsto\infty} \sum_i (f_i(x_{ir}) - f_i(u_r)) \geq 0.$$

Then for any $x^* \in \partial_p(\sum_i f_i)(\bar{x})$ and any $\epsilon > 0$ there are $u_i, {u_i}^*$, $i = 1,\dots,k$ such that

$$|f_i(u_i) - f_i(\bar{x})| \leq \epsilon, \quad \|u_i - \bar{x}\| < \epsilon, \quad {u_i}^* \in \partial_p f_i(u_i), \quad \|\sum {u_i}^* - x^*\| < \epsilon.$$

Proof. The structure of the proof is very similar to that of the finite-dimensional proof of [15]. First we observe that we can assume without loss of generality that $\bar{x} = 0, f_i(0) = 0$ and $x^* = 0$. (If not, we replace $f_i(x)$ by $f_i(\bar{x} + x) - f_i(\bar{x}) - k^{-1}\langle x^*, x\rangle$.)

By definition the condition $0 \in \partial_p(\sum_i f_i)(0)$ means that there are $N > 0$, $\delta > 0$ such that

$$\sum f_i(x) \geq -N\|x\|^2 \quad \text{when} \quad \|x\| \leq \delta.$$

We can assume δ so small that it does not exceed the δ in (ULC), and $f_i(x) \geq -1$ if $\|x\| \leq \delta$.

Consider the function

$$\phi_r(x_1,\dots,x_k) = \sum_i f_i(x_i) + N\sum_i \|x_i\|^2 + r\sum_{i,j} \|x_i - x_j\|^2.$$

This gives

$$0 = \phi_r(0,\dots,0) \geq \alpha_r = \inf\{\phi_r(x_1,\dots,x_k) : \|x_i\| \leq \delta\} \geq -k.$$

Take x_{ir} to satisfy

$$\phi_r(x_{1r},\dots,x_{kr}) \leq \alpha_r + 1/r.$$

Then

$$-k + r\sum_{i,j} \|x_{ir} - x_{jr}\|^2 \leq \phi_r(x_{1r},\dots,x_{kr}) \leq 1/r,$$

so that $\|x_{ir} - x_{jr}\|^2 \leq (k + 1)/r \to 0$.

By (ULC) there are u_r such that $\|x_{ir} - u_r\| \mapsto 0$ and

$$\sum_i f_i(x_{ir}) \geq \sum_i f_i(u_r) + \mathrm{o}(1).$$

It follows that

$$\begin{aligned} 0 \leq \sum_i f_i(u_r) + kN\|u_r\|^2 &\leq \sum_i f_i(x_{ir}) + N\sum_i \|x_{ir}\|^2 + \mathrm{o}(1) \\ &\leq \phi_r(x_{1r}, \ldots, x_{kr}) + \mathrm{o}(1) \leq 1/r + \mathrm{o}(1) = \mathrm{o}(1), \end{aligned}$$

hence $\sum_i f_i(u_r) + kN\|u_r\|^2 \to 0$. But $\sum_i f_i(u_r) + N\|u_r\|^2 \geq 0$, so u_r must go to zero as well as all x_{ir}. The above inequality now implies that

$$0 \leq \sum_i \liminf_{r\to\infty} f_i(x_{ir}) \leq \limsup_{r\to\infty} \sum_i f_i(x_{ir}) = 0$$

which means that for each i we have $f_i(x_{ir}) \to 0$ or, equivalently, $|f_i(x_{ir}) - f_i(0)| \to 0$.

Take now a sufficiently small σ, say, $\sigma < \delta/2$ and an $r = r(\sigma)$ such that $\|x_{ir}\| < \sigma,\ i = 1, \ldots, k$ and $r^{-1}(\sigma) < \sigma^3$. By the smooth variational principle of Borwein–Preiss [2], there are quadratic functions

$$\Delta_i(x) = \|x\|^2 - \langle a_i, x\rangle + \beta_i$$

with $\|a_i\| \leq 2\delta$ and points $u_i, i = 1, \ldots, k$ such that $\|u_i - x_{ir}\| < \sigma$ and the function

$$g(x_1, \ldots, x_k) = \phi_r(x_1, \ldots, x_k) + \sigma \sum_i \Delta_i(x_i)$$

attains at $(u_1, \ldots, u_k)$ a minimum on the set of $(x_1, \ldots, x_k)$ satisfying $\|x_i\| < \delta$ and

$$g(u_1, \ldots, u_k) \leq g(x_{1r}, \ldots, x_{kr}).$$

As $\|u_i\| \leq 2\sigma < \delta$, this means that the function

$$\sum_i f_i(x_i) + N\sum \|x_i\|^2 + r\sum_{i,j} \|x_i - x_j\|^2 + \sigma(\sum_i(\|x_i\|^2 - \langle a_i, x_i\rangle)$$

attains an unconditional local minimum at $(u_1, \ldots, u_k)$. Setting $x_i = u_i + h_i$, we get from this fact that

$$\begin{aligned} &\sum_i \Big[f_i(u_i + h_i) - f_i(u_i) + (N + \sigma)(\|h_i\|^2 + 2\langle u_i, h_i\rangle) - \sigma\langle a_i, h_i\rangle\Big] \\ &\qquad + r\sum_{i,,j} \Big(\|h_i\|^2 + \|h_j\|^2 + 2\langle u_i - u_j | h_i - h_j\rangle - 2\langle h_i | h_j\rangle\Big) \geq 0 \end{aligned}$$

for all sufficiently small h_i.

Setting consecutively for $i = 1, \dots, k$, $h_j = \delta_{ij} h$, $u_i{}^* = -2(N+\sigma)u_i + \sigma a_i + 2r\sum_j (u_i - u_j)$ and $M = N + \sigma$, we deduce from the last inequality that

$$f_i(u_i + h) - f_i(u_i) \ge \langle u_i{}^*, h\rangle - M\|h\|^2, \ \forall i = 1, \dots, k,$$

which means that

$$u_i{}^* \in \partial_p f_i(u_i).$$

On the other hand,

$$\begin{aligned} \sum_i u_i{}^* &= -2(N+\sigma)\sum_i u_i + \sigma \sum_i a_i \\ &\qquad +2r\sum_{ij}(u_i - u_j) \\ &= -2(N+\sigma)\sum_i u_i + \sigma\sum_i a_i, \end{aligned}$$

that is

$$\|\sum_i u_i{}^*\| \le (2N+\sigma)\sum_i \|u_i\| + \sigma\sum_i \|a_i\| \le 2k(2N+\sigma)\sigma + k\sigma \cdot 2\sigma.$$

It remains, given $\epsilon > 0$, to take a $\sigma < \epsilon/2$ so small that the quantity at the right-hand part of the last inequality is smaller than ϵ and $f_i(x) \ge f_i(0) - \epsilon$ if $\|x\| \le \sigma$.

3 Subgradients Under Convexification

Let $f : R^m \times R^n \to \bar{R}$ be a lower semicontinuous function such that its convex hull with respect to the second variable

$$\bar{f}(z, y) = \mathrm{conv}_y f(z, y)$$

is proper (everywhere greater than $-\infty$ and such that for any z there is a y with $f(z, y) < \infty$). What can be said about interrelation between $\partial\bar{f}$ and ∂f? This is the question we are going to answer in this section.

The main tool in answering this question is a parametric representation for $\bar{f}$ based on Carathéodory's theorem for the convex hull operation:

$$\bar{f}(z, y) = \inf_{\substack{\lambda_0 \ge 0, \dots, \lambda_n \ge 0 \\ \sum_i y_i = y, \sum_i \lambda_i = 1}} \sum_{i=0}^{n} \Phi(z, y_i, \lambda_i),$$

where

$$\Phi(z, y, \lambda) = \begin{cases} \lambda f(z, y/\lambda), & \text{if } \lambda > 0 \\ 0, & \text{if } \lambda = 0 \text{ and } y = 0 \\ +\infty, & \text{otherwise.} \end{cases}$$

This can also been written in terms of the parameter element

$$u = (y_0, \dots, y_n, \lambda_0, \dots, \lambda_n) \in (R^n)^{n+1} \times R^{n+1}$$

as

$$\bar{f}(z, y) = \inf_u F(z, y, u),$$

where

$$F(z, y, u) = \sum_{i=0}^{n} \Phi(z, y_i, \lambda_i) + \delta\Big(u \Big| \sum_{i=0}^{n} y_i = y,\ \sum_{i=0}^{n} \lambda_i = 1\Big).$$

Here $\delta(u|C)$ stands for the indicator function of C (i.e., the one equal to zero on C and $+\infty$ outside of C).

Theorem 3 *Assume that the following two conditions are satisfied:*

(B) *For each $(\bar{z}, \bar{y}) \in R^m \times R^n$ and each $\alpha \in R$, there is an $\epsilon > 0$ such that the set*

$$\{(z, y, u) : |z - \bar{z}| \leq \epsilon,\ \ |y - \bar{y}| \leq \epsilon,\ \ F(z, y, u) \leq \alpha\}$$

is compact.

(C) *The set* $\mathrm{dom} f(z, \cdot) = \{y : f(z, y) < \infty\}$ *does not depend on z, and for every y of this set $f(\cdot, y)$ is locally Lipschitzian.*

Consider any $(\bar{z}, \bar{y})$ with $f(\bar{z}, \bar{y})$ finite. Then for any $(\bar{w}, \bar{v}) \in \partial \bar{f}(\bar{z}, \bar{y})$ there is an element $\bar{u} = (\bar{y}_0, \dots, \bar{y}_n, \bar{\lambda}_0, \dots, \bar{\lambda}_n)$ satisfying $\bar{f}(\bar{z}, \bar{y}) = F(\bar{z}, \bar{y}, \bar{u})$, along with vectors $w_i, i = 0, \dots, n,$ such that

$$\sum_i \bar{\lambda}_i w_i = \bar{w} \tag{6}$$

and such that, for each i with $\bar{\lambda}_i > 0$, one has for $\hat{y}_i = \bar{y}_i / \bar{\lambda}_i$ that

$$(w_i, \bar{v}) \in \partial f(\bar{z}, \hat{y}_i) \quad \forall i. \tag{7}$$

Corollary 1 *Suppose under the assumption given in the theorem that $\bar{y}$ is "exposed" for f at $\bar{z}$, in the sense that the minimum defining $\bar{f}(\bar{z}, \bar{y})$ is attained only by vectors $\bar{u} = (\bar{y}_0, \dots, \bar{y}_n, \bar{\lambda}_0, \dots, \bar{\lambda}_n)$ in which for every i with $\bar{\lambda}_i > 0$ the vector $\hat{y}_i = \bar{y}_i / \bar{\lambda}_i$ agrees with $\bar{y}$. (In other words from*

$$\sum_i \bar{\lambda}_i \hat{y}_i = \bar{y}, \quad \bar{\lambda}_i \geq 0, \quad \sum_i \bar{\lambda}_i = 1, \quad \sum_i \bar{\lambda}_i f(\bar{z}, \hat{y}_i) = \bar{f}(\bar{z}, \bar{y}),$$

it follows that $\hat{y}_i = \bar{y}$ for every i with $\bar{\lambda}_i > 0$.)

Then for $(\bar{w}, \bar{v}) \in \partial \bar{f}(\bar{z}, \bar{y})$, one has

$$\bar{w} \in \mathrm{conv}\ \{w : \ (w, \bar{v}) \in \partial f(\bar{z}, \bar{y})\}$$

and

$$f(\bar{z}, y) \geq f(\bar{z}, \bar{y}) + \langle \bar{v}, y - \bar{y} \rangle, \quad \forall y.$$

Before proving the theorem, we discuss in some more detail the condition (B) introduced in the statement of the theorem.

Proposition 1 *Condition (B) is implied by*

$$(B_1) \qquad \liminf_{\substack{z \to \bar{z} \\ y \to \bar{y} \\ \lambda \searrow 0}} \lambda f(z, y/\lambda) = \infty \quad \text{when } \bar{y} \neq 0.$$

Proof. In combination with f being l.s.c., (B_1) means that Φ is l.s.c. It follows that F is l.s.c. The next claim is that the set

$$\{(z, y, \lambda) : z \in Z,\ \lambda \in [0,1],\ \Phi(z, y, \lambda) \leq \alpha\}$$

is bounded for any bounded set $Z \subset R^m$ and any $\alpha \in R$. If this were not true, we could find sequences $\{z^\nu\} \subset Z$, $\{\lambda^\nu\} \subset [0,1]$ and an unbounded sequence $\{y^\nu\}$ such that $\Phi(z^\nu, y^\nu, \lambda^\nu) \leq \alpha$. Without loss of generality it can be supposed that z^ν converges to some $\bar{z}$, that $\lambda^\nu > 0$ and that $0 < |y^\nu| \to \infty$.

The condition $\Phi(z^\nu, y^\nu, \lambda^\nu) \leq \alpha$ can then be written as

$$\alpha \geq \lambda^\nu f(z^\nu, y^\nu/\lambda^\nu) = \hat{\lambda}^\nu f(z^\nu, \hat{y}^\nu/\hat{\lambda}^\nu)|y^\nu|$$

with $\hat{y}^\nu = y^\nu/|y^\nu|$, $\hat{\lambda}^\nu = \lambda^\nu/|y^\nu| \to 0$. We can suppose that $\hat{y}^\nu \to \hat{y}$, where $|\hat{y}| = 1$. Then

$$\liminf_{\nu\to\infty} \hat{\lambda}^\nu f(z^\nu, \hat{y}_i^\nu/\hat{\lambda}^\nu)| \leq \lim_{\nu\to\infty} (\alpha/|y^\nu|) = 0$$

in contradiction with (B_1).

From this property of Φ we conclude in particular the existence for any bounded $Z \subset R_n$ of $\beta \in R$ such that

$$\Phi(z, y, \lambda) \geq \beta, \text{ when } z \in Z, \lambda \in [0,1].$$

Returning to the formula for $F(z, y, u)$, we apply this with $Z = \{z : |z - \bar{z}| \leq \epsilon\}$ for some $\bar{z}$ and $\epsilon > 0$. We get for $u = (y_0, \ldots, y_n, \lambda_0, \ldots, \lambda_n)$ that

$$\left.\begin{array}{l} |z - \bar{z}| \leq \epsilon \\ |y - \bar{y}| \leq \epsilon \\ F(z, y, u) \leq \alpha \end{array}\right\} \Longrightarrow \Phi(z, y_i, \lambda_i) \leq \alpha - n\beta \text{ and } \lambda_i \in [0,1],\ \forall i = 0, \ldots, n.$$

Since the set of (z, y_i, λ_i) satisfying

$$|z - \bar{z}| \leq \epsilon,\ \lambda_i \in [0,1],\ \Phi(z, y_i, \lambda_i) \leq \gamma$$

is bounded for every $\gamma \in R$, the boundedness property required by (B) does hold.

Proposition 2 *Condition* (B_1) *is equivalent to* $f(z,y)$ *being coercive in* y *locally uniformly in* z *in the sense that*

(B_2) *for any* $\bar{z}$ *and* $\epsilon > 0$ *there is a nondecreasing function* $\theta : [0,\infty) \to \bar{R}$ *with* $\theta(0)$ *finite,*
$\theta(s)/s \to \infty$ *as* $s \to \infty$, *such that*

$$f(z,y) \geq \theta(|y|), \text{ when } |z - \bar{z}| \leq \epsilon.$$

Proof. If (B_2) holds, then (when $\bar{y} \neq 0$)

$$\liminf_{\substack{z \to \bar{z} \\ y \to \bar{y} \\ \lambda \searrow 0}} \lambda f(z, y/\lambda) \geq \liminf_{\substack{y \to \bar{y} \\ \lambda \searrow 0}} \lambda\theta(|y|/\lambda) = |\bar{y}| \liminf_{\substack{y \to \bar{y} \\ \lambda \searrow 0}} \frac{\theta(|y|/\lambda)}{|y|/\lambda} = \infty.$$

Thus (B_1) holds. Conversely, under (B_1) we can get a function θ with the property demanded by (B_2) by taking

$$\theta(s) = \min_{\substack{|z-\bar{z}| \leq \epsilon \\ |y| \geq s \\ \lambda \in [0,1]}} \Phi(z,y,\lambda).$$

This is easily seen from the properties of Φ revealed in Proposition 1.

Proposition 3 *Denote by*

$$h(z,v) = \sup_y \{\langle y,v\rangle - f(z,y)\}$$

the "Hamiltonian" associated with f. *Then* (B_1) *is equivalent to*

(B_3) $\qquad h(z,v)$ *is u.s.c. and everywhere* $< \infty$.

Proof. Sufficiency: if (B_3) holds, one can take θ to be the conjugate of the function

$$\psi(t) = \max_{\substack{|z-\bar{z}| \leq \epsilon \\ |v| \leq t}} h(z,v)$$

to get a function satisfying (B_2).

Necessity: we can think of h as defined by parametric optimization:

$$-h(z,v) = \min_y G(z,v,y) \text{ with } G(z,v,y) = f(z,y) - \langle y,v\rangle.$$

From (B_2) and the lower semicontinuity of f we see that G has the basic properties guaranteeing that $-h$ is l.s.c., as required.

Proof of the theorem. By definition there exist sequences

$(w^\nu, v^\nu) \to (\bar w, \bar v),\ (z^\nu, y^\nu) \to (\bar z, \bar y)$ with $\bar f(z^\nu, y^\nu) \to \bar f(\bar z, \bar y)$, and $(w^\nu, v^\nu) \in \partial_p \bar f(z^\nu, y^\nu)$.

In particular, this implies through the convexity of $\bar f(z, y)$ in y that

$$f(z^\nu, y) \geq \bar f(z^\nu, y) \geq \bar f(z^\nu, y^\nu) + \langle v^\nu, y - y^\nu \rangle,\ \forall y. \tag{8}$$

Because (B) holds, the set

$$U(z, y) = \{u : F(z, y, u) = \bar f(z, y)\}$$

is nonempty when $\bar f(z, y) < \infty$, and the set-valued mapping $U : (z, y) \mapsto U(z, y)$ carries any set of the form $\{(z, y) :\ |z - \bar z| \leq \epsilon,\ |y - \bar y| \leq \epsilon,\ \bar f(z, y) \leq \alpha\}$ for a sufficiently small ϵ into a bounded set of vectors u. Also,

$$\limsup_{\substack{(z', y') \to (z, y) \\ \bar f(z', y') \to \bar f(z, y) < \infty}} U(z', y') \ = \ U(z, y). \tag{9}$$

For any (z, y) of a neighborhood of (z^ν, y^ν) and any u we have (for some $u \in U(z^\nu, y^\nu)$)

$$\begin{aligned} F(z, y, u) \geq \bar f(z, y) &\geq \bar f(z^\nu, y^\nu) + \langle w^\nu, z - z^\nu \rangle + \langle v^\nu, y - y^\nu \rangle + \mathrm{O}(|y - y^\nu|^2 + |z - z^\nu|^2) \\ &\geq F(z^\nu, y^\nu, u^\nu) + \langle w^\nu, z - z^\nu \rangle + \langle v^\nu, y - y^\nu \rangle + \mathrm{O}(|y - y^\nu|^2 + |z - z^\nu|^2), \end{aligned} \tag{10}$$

or in other words,

$$\begin{array}{rcl} \sum_{i=0}^n \Phi(z, y_i, \lambda_i) & \geq & \sum_{i=0}^n \Phi(z^\nu, y_i^\nu, \lambda_i^\nu) \\ & + & \langle w^\nu, z - z^\nu \rangle + \langle v^\nu, \sum_{i=0}^n y_i - \sum_{i=0}^n y_i^\nu \rangle \\ & + & \mathrm{O}(|z - z^\nu|^2 + \sum_{i=0}^n |y_i - y_i^\nu|^2), \end{array}$$

when $\lambda_i \geq 0,\ \sum_i \lambda_i = 1;\ \sum_i y_i = y$.

Consider the functions

$$\phi_i^\nu(z, y) = \Phi(z, y, \lambda_i^\nu).$$

Then the inequality above means that

$$(w^\nu, v^\nu, \ldots, v^\nu) \in \partial_p (\sum_i \phi_i^\nu)(z^\nu, y_0^\nu, \ldots, y_n^\nu). \tag{11}$$

Applying Theorem 2 (with the account of the fact that ϕ_i^ν does not depend on y_j with $j \neq i$), we find, for any $\nu = 1, 2, \ldots,$ vectors $(\hat z_i^\nu, \hat y_i^\nu, \hat w_i^\nu, \hat v_i^\nu)$ such that for all i with $\lambda_i^\nu > 0$

$$\left. \begin{array}{l} |\hat z_i^\nu - z^\nu| \leq 1/\nu,\ |\hat y_i^\nu - y^\nu| \leq 1/\nu, \\ |\hat v_i^\nu - v^\nu| \leq 1/\nu,\ |\sum_i \hat w_i^\nu - w^\nu| \leq 1/\nu, \\ \text{and } \ (\hat w_i^\nu, \hat v_i^\nu) \in \partial_p \phi_i^\nu(\hat z_i^\nu, \hat y_i^\nu). \end{array} \right\} \tag{12}$$

This means that for any i with $\lambda_i^\nu > 0$

$$\begin{aligned} f(z, y_i/\lambda_i^\nu) &\geq f(\hat{z}_i^\nu, \hat{y}_i^\nu/\lambda_i^\nu) \\ &+ \langle (\hat{w}_i^\nu/\lambda_i^\nu) | z - \hat{z}_i^\nu \rangle + \mathrm{O}(|z - \hat{z}_i^\nu|^2) \\ &+ \langle \hat{v}_i^\nu, (y_i - y_i^\nu(\cdot))/\lambda_i^\nu \rangle + \mathrm{O}(|y_i - y_i^\nu(\cdot)|^2). \end{aligned} \tag{13}$$

from which we conclude that

$$((\hat{w}_i^\nu/\lambda_i^\nu), \hat{v}_i^\nu) \in \partial_p f(\hat{z}_i^\nu, \hat{y}_i^\nu/\lambda_i^\nu).$$

From (9) and the local boundedness of the mapping U mentioned earlier, we can suppose that the vectors $u^\nu = ({y^\nu}_0, \ldots, {y^\nu}_n, {\lambda^\nu}_0, \ldots, {\lambda^\nu}_n)$ converge to some $\bar{u} = (\bar{y}_0, \ldots, \bar{y}_n, \bar{\lambda}_0, \ldots, \bar{\lambda}_n) \in U(\bar{z}, \bar{y})$. We also observe that by (C) and (12)(13) the vectors $\hat{w}_i^\nu/\lambda_i^\nu$ are uniformly bounded as well as $\hat{w}_i^\nu$. So we may assume that every $\hat{w}_i^\nu$ converges to a certain p_i with $p_i = 0$ if $\bar{\lambda}_i = 0$. Therefore, as $\hat{z}_i^\nu \to \bar{z}$, $\hat{y}_i^\nu \to \bar{y}_i$ and $\hat{v}_i^\nu \to \bar{v}$ by (13), we get, setting $w_i = p_i/\bar{\lambda}_i$, that for every i with $\bar{\lambda}_i > 0$, both $(w_i, \bar{v}) \in \partial f(\bar{z}, \bar{y}_i/\bar{\lambda}_i)$ and $\sum_i \bar{\lambda}_i w_i = \sum_i p_i = \bar{w}$. This completes the proof of Theorem 3.

4 An Auxiliary Problem - Relaxation

As the first step of the proof of Theorem 1 we shall consider the same functional $J(\cdot)$ but under somewhat different set of assumptions. Namely, instead of (A_2) and (A_3) we assume the following:

(A_2') $L(t,x,y)$ is l.s.c (extended-real-valued) as a function of (x,y), the set-valued map $t \mapsto \mathrm{epi}L(t,\cdot,\cdot)$ is measurable, and the sets

$$\mathrm{dom}L(t,x,\cdot) = \{y : L(t,x,y) < \infty\} = R(t)$$

do not depend on x and are bounded by a square integrable function $r(t)$, that is to say, $|x| \leq r(t)$ if $x \in R(t)$;

(A_3') $|L(t,x,y) - L(t,x',y)| \leq \omega(t, |x - x'|)$ and $|L(t,x,y)| \leq c(t)$,

for all x, x' of a certain ϵ-ball ($\epsilon > 0$) around $x_*(t)$ and $y \in R(t)$, where $\omega(t,\delta)$ is a nonnegative Carathéodory function for almost every t converging monotonously to zero as $\delta \to 0$ and $c(t)$ is summable.

The last condition says that the dependence of L on x is continuous uniformly in $y \in R(t)$ for almost every t. This is definitely the case when $L(t,\cdot,\cdot)$ is continuous (as $R(t)$ is bounded) or if L is obtained from a function satisfying (A_1) - (A_3) by changing its value to $+\infty$ outside of $R(t)$. We also observe that implicit in (A_3') is that $\omega(t,\delta) \leq c(t)$ so that $\int_0^1 \omega(t,\delta)dt \to 0$ as $\delta \to 0$.

For any $\epsilon > 0$ we denote by $\bar{L}_\epsilon(t,x,y)$ the convex hull of $L(t,x,y) + \epsilon|y - \dot{x}_*(t)|^2$ with respect to the third argument, and by $\bar{J}_\epsilon(\cdot)$ the functional obtained from $J(\cdot)$ by replacing $L(t,x,y)$ by $\bar{L}_\epsilon(t,x,y)$ under the integral. We write $\bar{L}$ and $\bar{J}$ for $\bar{L}_\epsilon$ and $\bar{J}_\epsilon$ when $\epsilon = 0$.

Theorem 4 *We posit* (A_1), (A_2') *and* (A_3'), *and assume that* $x_*(\cdot)$*is a local minimum of* $J(\cdot)$ *in* W_1^1. *Then* $x_*(\cdot)$*is a local minimum of* $\bar{J}_\epsilon(\cdot)$ *in* W_1^1, *and* $\bar{J}_\epsilon(x_*(\cdot)) = J(x_*(\cdot))$ *for any sufficiently small* $\epsilon > 0$

This is a kind of relaxation theorem which differs from standard results of that type (cf. [18], [13]) through the fact that we consider a local minimum in the W_1^1-topology rather than a global minimum, or a local minimum in the topology of uniform convergence.

Proof We claim first that

$$\bar{L}(t, x_*(t), \dot{x}_*(t)) = L(t, x_*(t), \dot{x}_*(t)) \text{ a.e..} \tag{14}$$

Supposing the contrary, we deduce that there is a $\gamma > 0$ such that for t in a set of positive measure the set $U_\gamma(t)$ of vectors $u = (y_0, \dots, y_n, \lambda_0, \dots, \lambda_n) \in (R^n)^{n+1} \times R^{n+1}$ with

$$\lambda_i \geq 0, \ \sum_i \lambda_i = 1, \quad \sum_i \lambda_i y_i = \dot{x}_*(t), \quad \sum_i \lambda_i L(t, x_*(t), y_i) \leq L(t, x_*(t), \dot{x}_*(t)) - \gamma, \tag{15}$$

is nonempty. On the other hand, as $R(t)$ is uniformly bounded, we can find a $\delta > 0$ such that $J(x(\cdot)) \geq J(x_*(\cdot))$ for any $x(\cdot)$ with $x(0) = x_*(0)$ and $\dot{x}(t) = \dot{x}_*(t)$ on a set of measure not smaller than $1 - \delta$.

Choose a set Δ of positive measure, less than δ such that $U_\gamma \neq \emptyset$ for all $t \in \Delta$, and let

$$\bar{u}(t) = (\bar{y}_0(t), \dots, \bar{y}_n(t), \bar{\lambda}_0(t), \dots, \bar{\lambda}_n(t))$$

be a measurable selection of $U_\gamma(t)$ on Δ. We can extend $\bar{u}(t)$ to the rest of $[0, 1]$ by $\bar{y}_i(t) = \dot{x}_*(t), \bar{\lambda}_i(t) = (n+1)^{-1}$. Then, of course, $\sum_i \bar{\lambda}_i(t)\bar{y}_i(t) = \dot{x}_*(t)$ for almost all t. Set

$$\begin{aligned}
\bar{\Lambda} &= \{(\lambda_0(\cdot), \dots, \lambda_n(\cdot)) : \ \lambda_i(t) \geq 0, \ i = 0, \dots, n, \ \sum_i \lambda_i(t) = 1 \text{ a.e.}\}; \\
\Lambda &= \{(\lambda_0(\cdot), \dots, \lambda_n(\cdot)) : \ \lambda_i(t) \in \{0, 1\}, \ i = 0, \dots, n, \ \sum_i \lambda_i(t) = 1 \text{ a.e.}\}
\end{aligned}$$

Then, obviously, $(\bar{\lambda}_0(t), \dots, \bar{\lambda}_n(t)) \in \bar{\Lambda}$ for almost all t.

By Lyapunov's theorem on vector measures, for any finite set of vector functions

$$h^s(t) = (h_0^s(t), \dots, h_n^s(t)) \in L_1, \ s = 1, \dots, m,$$

the images of Λ and $\bar{\Lambda}$ under the map

$$(\lambda_0, \dots, \lambda_n) \mapsto \Big(\int_0^1 \sum_i \lambda_i(t) h_i^1(t) dt, \dots, \int_0^1 \sum_i \lambda_i(t) h_i^m(t) dt \Big)$$

coincide (and are convex compact). Hence there is a sequence $\{(\lambda_0^\nu(\cdot),\ldots,\lambda_n^\nu(\cdot))\}$ of elements of Λ such that

$$\int_0^1 \sum_i \lambda_i^\nu(t)\bar{y}_i(t)dt = \int_0^1 \sum_i \bar{\lambda}_i(t)\bar{y}_i(t)dt = \int_0^1 \dot{x}_*(t)dt = x_*(1) - x_*(0)$$

and

$$\int_0^1 \sum_i \lambda_i^\nu(t)h_i(t)dt \to \int_0^1 \sum_i \bar{\lambda}_i(t)h_i(t)dt,\ \forall\ (h_0(\cdot),\ldots,h_n) \in L_1. \tag{16}$$

Set

$$x^\nu(t) = x_*(0) + \int_0^t \sum_i \lambda_i^\nu(\tau)\bar{y}_i(\tau)d\tau.$$

Then (16) along with the boundedness of $R(t)$ imply that $x^\nu(\cdot) \to x_*(\cdot)$ uniformly, and by (A_3') (which justifies the application of the Lebesgue majorized convergence theorem)

$$\begin{aligned}
\lim_{\nu\to\infty} J(x^\nu(\cdot)) &= \lim_{\nu\to\infty}(l(x^\nu(0),x^\nu(1)) + \int_0^1 L(t,x^\nu(t),\dot{x}^\nu(t))dt) \\
&= l(x_*(0),x_*(1)) + \lim_{\nu\to\infty}\int_0^1 L(t,x_*(t),\dot{x}^\nu(t))dt \\
&= l(x_*(0),x_*(1)) + \lim_{\nu\to\infty}\int_0^1 \sum_i \lambda_i^\nu(t)L(t,x_*(t),\bar{y}_i(t))dt \\
&= l(x_*(0),x_*(1)) + \int_0^1 \sum_i \bar{\lambda}_i(t)L(t,x_*(t),\bar{y}_i(t))dt. \qquad (17)
\end{aligned}$$

By (15) it follows that

$$\lim_{\nu\to\infty}(J(x^\nu(\cdot) - J(x_*(\cdot)) = \int_\Delta \Big(\sum_i \bar{\lambda}_i(t)L(t,x_*(t),\bar{y}_i(t)) - L(t,x_*(t),\dot{x}_*(t))\Big)dt < 0.$$

On the other hand, $\dot{x}^\nu(t) = \dot{x}_*(t)$ on the complement of Δ which is a set of measure greater than $1-\delta$ and $x^\nu(0) = x_*(0),\ x^\nu(1) = x_*(1)$ which implies that $J(x^\nu) \geq J(x_*)$ according to the choice of δ. This is a contradiction proving the claim.

We next note the following simple fact.

Lemma 1 *Let f be an l.s.c. function on $R^n \times R^n$ satisfying (B), and let for given $(\bar{z},\bar{y})$ we have $f(\bar{z},\bar{y}) = \bar{f}(\bar{z},\bar{y})$, $\bar{f}$ being the convex hull of f with respect to y. Take an $\epsilon > 0$ and set $f_\epsilon(z,y) = f(z,y) + \epsilon|y-\bar{y}|^2$. Then $\bar{y}$ is exposed for f_ϵ at $\bar{z}$ in the sense of Corollary 1. Moreover, if a sequence of vectors $u^\nu = (y_0^\nu,\ldots,y_n^\nu,\lambda_0^\nu,\ldots,\lambda_n^\nu))$ is such that ($\lambda_i{}^\nu \geq 0,\ \sum_i \lambda_i{}^\nu = 1$) and*

$$\sum_i \lambda_i{}^\nu y_i^\nu \to \bar{y}, \quad \sum_i \lambda_i{}^\nu f_\epsilon(\bar{z},y_i^\nu) \to f_\epsilon(\bar{z},\bar{y}) = f(\bar{z},\bar{y}),$$

then $\sum_i \lambda_i{}^\nu(y_i^\nu - \bar{y})^2 \to 0$.

Proof We may assume that $\sum_i {\lambda_i}^\nu (y_i^\nu - \bar{y})^2$ converges to a certain a. Then

$$\begin{aligned} f(\bar{z},\bar{y}) &= \lim_{\nu\to\infty} \sum_i {\lambda_i}^\nu f_\epsilon(\bar{z}, y_i^\nu) \\ &= \liminf_{\nu\to\infty} \sum_i {\lambda_i}^\nu (f(\bar{z}, y_i^\nu) + \epsilon(y_i^\nu - \bar{y})^2) \\ &\geq \epsilon a + \liminf_{\nu\to\infty} \sum_i {\lambda_i}^\nu f(\bar{z}, y_i^\nu) \\ &\geq \epsilon a + \bar{f}(\bar{z},\bar{y}) = \epsilon a + f(\bar{z},\bar{y}), \end{aligned}$$

whence $a = 0$, which proves the lemma.

Now we can conclude the proof. Assume the theorem is not valid. Then for an $\epsilon > 0$ we can find a sequence of $x^\nu(\cdot)$ converging to $x_*(\cdot)$ in $W^{1,1}$, and satisfying for some positive a^ν

$$\bar{J}_\epsilon(x^\nu(\cdot)) + a^\nu < \bar{J}(x_*(\cdot)) = \bar{J}_\epsilon(x_*(\cdot)). \tag{18}$$

Then there are $\bar{y}_i^\nu(t),\ \bar{\lambda}_i^\nu(t),\ i = 1, ..., n$, such that $\bar{\lambda}^\nu = (\bar{\lambda}_0^\nu, \ldots, \bar{\lambda}_1^\nu) \in \bar{\Lambda}$ and

$$\sum_i \bar{\lambda}_i^\nu(t)\bar{y}_i^\nu(t) = \dot{x}^\nu(t);\ \sum_i \bar{\lambda}_i^\nu(t) L_\epsilon(t, x^\nu(t), \bar{y}_i^\nu(t)) = \bar{L}_\epsilon(t, x^\nu(t), \dot{x}^\nu(t)) \text{ a.e.}$$

As $L_\epsilon(t,\cdot,\cdot)$ satisfies (B) (since $L(t,\cdot,\cdot)$ does) the function $\bar{L}_\epsilon(t,\cdot,\cdot)$ is l.s.c. for all t, so that

$$\liminf_{\nu\to\infty} \sum_i \bar{\lambda}_i^\nu(t) L_\epsilon(t, x^\nu(t), \bar{y}_i^\nu(t)) \geq \bar{L}_\epsilon(t, x_*(t), \dot{x}_*(t)) = L(t, x_*(t), \dot{x}_*(t)) \text{ a.e.}$$

Together with (A_3') this implies that

$$\liminf_{\nu\to\infty} \sum_i \bar{\lambda}_i^\nu(t) L_\epsilon(t, x_*(t), \bar{y}_i^\nu(t)) \geq \bar{L}_\epsilon(t, x_*(t), \dot{x}_*(t)).$$

It follows from Lemma 1 that

$$\sum_i \bar{\lambda}_i^\nu(t)(\bar{y}_i^\nu(t) - \dot{x}_*(t))^2 \to 0 \text{ a.e.} \tag{19}$$

Applying again the theorem of Lyapunov, we find a vector function $(\lambda_0^\nu(t), \ldots, \lambda_n^\nu(t))$ with values in Λ such that

- for the function
$$z^\nu(t) = x^\nu(0) + \int_0^t \sum_i \lambda_i^\nu(\tau)\bar{y}_i^\nu(\tau)d\tau \tag{20}$$
the equality $z^\nu(t) = x^\nu(t)$ holds for $t = 1$ and (for finitely but) sufficiently many other points t to make sure (in view of boundedness of $R(t)$) that $z^\nu(t)$ and $x^\nu(t)$ are close enough to guarantee (owing to (A_3')) that
$$|L(t, x^\nu(t), \bar{y}_i^\nu(t)) - L(t, z^\nu(t), \bar{y}_i^\nu(t))| \leq a^\nu/2, \text{ a.e.}; \tag{21}$$

- for all $i = 0, \dots, n$, and all ν

$$\int_0^1 \sum_i \lambda_i^\nu(t)(\bar{y}_i^\nu(t) - \dot{x}^\nu(t))^2 dt = \int_0^1 \sum_i \bar{\lambda}_i^\nu(t)(\bar{y}_i^\nu(t) - \dot{x}^\nu(t))^2 dt, \tag{22}$$

  and

- for every ν

$$\int_0^1 \sum_i \bar{\lambda}_i^\nu(t) L(t, x^\nu(t), \bar{y}_i^\nu(t)) = \int_0^1 \sum_i \lambda_i^\nu(t) L(t, x^\nu(t), \bar{y}_i^\nu(t)). \tag{23}$$

We observe now that, as every $\lambda_i^\nu(t)$ may assume only values 0 and 1, and their sum is identically one, for any function $\phi(y)$ we have $\sum_i \lambda_i^\nu(t)\phi(\bar{y}_i^\nu(t)) = \phi(\dot{z}^\nu(t))$. Therefore (19), (20) and (22) mean that $\|z^\nu(\cdot) - x_*(\cdot)\|_1^1 \to 0$. For the same reason, we conclude from (23) and (21) that $J(z^\nu(\cdot)) \le \bar{J}(x^\nu(\cdot)) + a^\nu/2$ which gives together with (19) that $J(z^\nu(\cdot)) \le J(x_*(\cdot)) - a^\nu/2$ — a contradiction. This concludes the proof of the theorem.

5 The Auxiliary Problem—A Necessary Condition

We continue to consider the same problem as in the preceding section assuming in addition the following

(A_4) $L(t, \cdot, \cdot)$ is Lipschitz continuous around every point (x, y) with $|x - x_*(t)| \le \epsilon$, $y \in \mathrm{int} R(t)$.

Theorem 5 *We posit (A_1), (A_2'), (A_3') with $\omega(t, \delta) = k(t) \cdot \delta$, $k(t)$ being a summable function, and (A_4) and assume that $J(x_*(\cdot))$ is finite and $x_*(\cdot)$ is a local minimum of $J(x(\cdot))$ in W_1^1. If the measure of the set*

$$\{t : \dot{x}_*(t) \in \mathrm{int} R(t)\},$$

is positive, then there is an absolutely continuous function $p(t)$ with values in R^n such that the Euler condition and the transversality conditions of Theorem 1 are satisfied.

The proof of the theorem consists of several steps.

Step 1 – Reformulation. It follows from (A_2'), because $J(x_*(\cdot))$ is finite, that no loss of generality will result from the assumption that $x_*(t) \equiv 0$ (otherwise we replace x by $x_*(t) + x$). Consider the function $f(t, x, y)$ coinciding with $L(t, x, y)$ if $|x| \le \epsilon$ and equal to $+\infty$ otherwise. Then zero is also a local minimum of the functional

$$I(x(\cdot)) = l(x(0), x(1)) + \int_0^1 L(t, x(t), \dot{x}(t)) dt.$$

Let us further define the following three functionals on $R^n \times L_1 \times L_1$:

$$\begin{aligned} I_0(a, z(\cdot), y(\cdot)) &= l(a, a + \textstyle\int_0^1 y(t)dt); \\ I_1(a, z(\cdot), y(\cdot)) &= \textstyle\int_0^1 L(t, z(t), y(t))dt; \\ I_2(a, z(\cdot), y(\cdot)) &= \textstyle\int_0^1 k(t)|z(t) - a - \int_0^t y(\tau)d\tau|dt. \end{aligned}$$

In terms of these we get from (A_3') that

$$I(x(\cdot)) \le I_0(x(0), z(\cdot), \dot{x}(\cdot)) + I_1(x(0), z(\cdot), \dot{x}(\cdot)) + I_2(x(0), z(\cdot), \dot{x}(\cdot)),$$

from which we conclude that

$$a_* = x_*(0) = 0, \quad z_*(t) = x_*(t) \equiv 0, \quad y(t) = \dot{x}_*(t) \equiv 0,$$

give a local minimum to $\sum_0^2 I_i(a, z(\cdot), y(\cdot))$ on $R^n \times L_1 \times L_1$ and, consequently, also on $R^n \times L_2 \times L_2$.

Step 2 – Verification of (ULC).

Lemma 2 *The functionals $I_i(a, z(\cdot), y(\cdot))$ satisfy the (ULC) condition of Theorem 2.*

Proof. As follows from (A_4'),

$$x_*(1) - x_*(0) \in \operatorname{int} \int_0^1 R(t)dt.$$

Choose $\delta > 0$ so small that $2\delta B \subset \int_0^1 R(t)dt$, where B is the unit ball in R^n. Let the sequences of $(a_i^\nu, z_i^\nu(\cdot), y_i^\nu(\cdot)) \in R^n \times L_2 \times L_2$ be such that

$$|a_i^\nu| + \|z_i^\nu(\cdot)\|_2 + \|y_i^\nu(\cdot)\|_2 \le \delta, \quad \sum_0^2 I_i(a_i^\nu, z_i^\nu(\cdot), y_i^\nu(\cdot)) < \infty$$

for all ν and all $i = 0, 1, 2$.

$$|a_i^\nu - a_j^\nu| + \|z_i^\nu(\cdot) - z_j^\nu(\cdot)\|_2 + \|y_i^\nu(\cdot) - y_j^\nu(\cdot)\|_2 \to 0.$$

We have to find a sequence of $(a^\nu, z^\nu(\cdot), y^\nu(\cdot))$ with the property that

$$|a_i^\nu - a^\nu| + \|z_i^\nu(\cdot) - z^\nu(\cdot)\|_2 + \|y_i^\nu(\cdot) - y^\nu(\cdot)\|_2 \to 0$$

and

$$\liminf_{\nu \to \infty} \sum_{i=0}^2 [I_i(a_i^\nu, z_i^\nu(\cdot), y_i^\nu(\cdot)) - I_i(a^\nu, z^\nu(\cdot), y^\nu(\cdot))] \ge 0.$$

To this end we set

$$\xi^\nu = \int_0^1 (y_1^\nu(t) - y_0^\nu(t))dt \to 0$$

and choose, for any unit vector $e \in R^n$, a measurable selection $u_e(t)$ of $R(t)$ such that

$$\int_0^1 u_e(t)dt = \int_0^1 y_0^\nu(t)dt + \delta e.$$

Let further $v^\nu(t)$ be the $u_e(t)$ corresponding to $e = -\xi^\nu/|\xi^\nu|$ when $\xi^\nu \neq 0$. Then

$$\int_0^1 y_0^\nu(t)dt = (1 + |\xi^\nu|/\delta)^{-1}(\int_0^1 y_1^\nu(t)dt + (|\xi^\nu|/\delta)\int_0^1 v^\nu dt).$$

It easily follows from the Lyapunov vector measure theorem that there is a subset Δ^ν of measure $|\xi^\nu|/(\delta + |\xi^\nu|)$ such that

$$\int_0^1 [\chi_{\Delta^\nu}(t)v^\nu(t) + (1 - \chi_{\Delta^\nu})y_1^\nu(t)]dt = \int_0^1 y_0^\nu(t)dt,$$

where χ_Δ is the characteristic function of Δ. (Just consider the function

$$\Delta \mapsto \Big(\int_0^1 [\chi_\Delta(t)v^\nu(t) + (1 - \chi_\Delta)y_1^\nu(t)]dt, \int_0^1 \chi_\Delta dt\Big)$$

from the collection of Lebesgue measurable subsets of $[0, 1]$ into R^{n+1} and observe that the theorem of Lyapunov can be applied as $v^\nu(t)$ and $y_1^\nu(t)$ are bounded as selections of $R(t)$.)

Now we set $a^\nu = a_0^\nu$, $z^\nu = z_1^\nu(t)$ and define $y^\nu(\cdot)$ to be equal to $y_1^\nu(t)$if $\xi^\nu = 0$ and to

$$y^\nu(\cdot) = (1 - \chi_{\Delta^\nu})y_1^\nu(t) + \chi_{\Delta^\nu}v^\nu(t)$$

if $\xi^\nu \neq 0$. Then

$$I_0(a_0^\nu, z^\nu(\cdot), y^\nu(\cdot)) = I_0(a_0^\nu, z_0^\nu(\cdot), y_0^\nu(\cdot));$$

$$|I_1(a^\nu, z^\nu(\cdot), y^\nu(\cdot)) - I_1(a_1^\nu, z_1^\nu(\cdot), y_1^\nu(\cdot))| \leq 2\int_{\Delta^\nu} c(t)dt \to 0$$

(by (A_3') and as $\operatorname{mes}\Delta^\nu \to 0$) and (setting $k = \int_0^1 k(t)dt$)

$$|I_2(a_2^\nu, z^\nu(\cdot), y^\nu(\cdot)) - I_2(a_2^\nu, z_2^\nu(\cdot), y_2^\nu(\cdot))|$$
$$\leq \int_0^1 k(t)|z_1^\nu(t) - z_2^\nu(t)|dt + k\int_0^1 |y_1^\nu(t) - y_2^\nu(t)|dt + 2k\int_{\Delta^\nu} c(t)dt \longrightarrow 0.$$

This completes the proof of the lemma.

Step 3– Application of the "fuzzy calculus". It follows from the conclusion of Step 1 that

$$0 \in \partial_p(I_0 + I_1 + I_2)(0,0,0),$$

(where p stands for the proximal subgradient (in $R^n \times L_2 \times L_2$)). By Theorem 2, for any $i = 0,1,2$ there are two sequences of triples $(a_i^\nu, z_i^\nu(\cdot), y_i^\nu(\cdot)))$ and $(b_i^\nu, w_i^\nu(\cdot), v_i^\nu(\cdot))$ in $R^n \times L_2 \times L_2$ such that

$$|I_i(a_i^\nu, z_i^\nu(\cdot), y_i^\nu(\cdot)) - I_i(0,0,0)| < 1/\nu, \quad |a_i^\nu| < 1/\nu, \quad \|z_i^\nu(\cdot)\| < 1/\nu, \quad \|y_i^\nu(\cdot)\| < 1/\nu;$$
$$|\sum b_i^\nu| < 1/\nu, \quad \|\sum w_i^\nu(\cdot)\| < 1/\nu, \quad \|\sum v_i^\nu(\cdot)\| < 1/\nu;$$
$$(b_i^\nu, w_i^\nu(\cdot), v_i^\nu(\cdot)) \in \partial_p I_i(a_i^\nu, z_i^\nu(\cdot), y_i^\nu(\cdot))$$

(all norms are, of course, in L_2). The latter means that there are r^ν such that

$$\begin{aligned} I_i(a_i^\nu + \alpha, z_i^\nu(\cdot) + \xi(\cdot), y_i^\nu(\cdot) + \eta(\cdot)) - I_i(a_i^\nu, z_i^\nu(\cdot), y_i^\nu(\cdot)) \\ \geq \langle b_i^\nu, \alpha\rangle + \int_0^1 [\langle w_i^\nu(t), \xi(t)\rangle + \langle v_i^\nu(t), \eta(t)\rangle] dt \\ - r^\nu(|\alpha|^2 + \|\xi(\cdot)\|^2 + \|\eta(\cdot)\|^2) \end{aligned}$$

for all $\alpha, \xi(\cdot), \eta(\cdot)$ sufficiently close to zero.

Step 4 – Analysis. What does the above inequality mean for each of the three functionals? For $i = 0$ it means that

$$\begin{aligned} l(a_0^\nu + \alpha, a_0^\nu + \alpha + \int_0^1 (y_0^\nu(t) + \eta(t)) dt) - l(a_0^\nu, a_0^\nu + \int_0^1 y_0^\nu(t) dt) \\ \geq \langle b_0^\nu, \alpha\rangle + \int_0^1 [\langle w_0^\nu(t), \xi(t)\rangle + \langle v_0^\nu(t), \eta(t)\rangle] dt - r^\nu(|\alpha|^2 + \|\xi(\cdot)\|^2 + \|\eta(\cdot)\|^2). \end{aligned}$$

Setting $\alpha = 0$ and $\eta(\cdot) = 0$ in this inequality, we conclude that $w_0^\nu(t) \equiv 0$. Setting $c_0^\nu = a_0^\nu + \int_0^1 y_0^\nu(t) dt$, $d^\nu = \int_0^1 v_0^\nu(t) dt$, and taking $\eta(t) \equiv \beta \in R^n$, we see that

$$l(a_0^\nu + \alpha, c_0^\nu + \alpha + \beta) - l(a_0^\nu, c_0) \geq \langle b_0^\nu, \alpha\rangle + \langle d^\nu, \beta\rangle - r^\nu(|\alpha|^2 + |\beta|^2),$$

which shows that $(b_0^\nu - d^\nu, d^\nu) \in \partial_p l(a_0^\nu, c_0)$.

We further observe that $v_0^\nu(t)$must be a constant, for otherwise, taking $\alpha = 0$, we get

$$0 \geq \sup\Big\{ \int_0^1 \langle v_0^\nu(t) - d^\nu | \eta(t)\rangle dt - r^\nu \|\eta(\cdot)\|^2 : \int_0^1 \eta(t) dt = 0 \Big\} > 0.$$

Thus, for $i = 0$, we have: there is a $d^\nu \in R^n$ such that

$$\begin{aligned} &(b_0^\nu - d^\nu, d^\nu) \in \partial_p l(a_0^\nu, c_0^\nu); \\ &v_0^\nu(t) \equiv d^\nu, \\ &w_0^\nu(t) \equiv 0. \end{aligned}$$

For $i = 1$ we have the inequality

$$\int_0^1 [L(t, z_1^\nu(t) + \xi(t), y_1^\nu(t) + \eta(t)) - L(t, z_1^\nu(t), y_1^\nu(t))]dt$$
$$\geq \int_0^1 \Big[\langle w_1^\nu(t), \xi(t)\rangle + \langle v_1^\nu(t), \eta(t)\rangle - r^\nu(|\xi(t)|^2 + |\eta(t)|^2)\Big]dt + \langle b_1^\nu, \alpha\rangle - r^\nu|\alpha|^2$$

for all sufficiently small $(\alpha, \xi(\cdot), \eta(\cdot)) \in R^n \times L_2 \times L_2$. It is obvious that b_1^ν must be zero. Standard measurable selection arguments imply that for almost every t

$$L(t, z_1^\nu(t) + \xi, y_1^\nu(t) + \eta) - L(t, z_1^\nu(t), y_1^\nu(t)) \geq \langle w_1^\nu(t), \xi\rangle + \langle v_1^\nu(t), \eta\rangle - r^\nu(|\xi|^2 + |\eta|^2)$$

for all (ξ, η). In other words

$$(w_1^\nu(t), v_1^\nu(t)) \in \partial_p L(t, z_1^\nu(t), y_1^\nu(t)) \text{ a.e.} \tag{24}$$

Finally, for $i = 2$, we have the inequality

$$\int_0^1 k(t)\Big[|z_2^\nu(t) + \xi(t) - a_2^\nu - \alpha - \int_0^t (y_2^\nu(\tau) + \eta(\tau))d\tau| - |z_2^\nu(t) - a_2^\nu \ - \int_0^t y_2^\nu(\tau)d\tau|\Big]dt$$
$$\geq \int_0^1 \Big[\langle w_2^\nu(t), \xi(t)\rangle + \langle v_2^\nu(t), \eta(t)\rangle - r^\nu(|\xi(t)|^2 + |\eta(t)|^2)\Big]dt + \langle b_2^\nu, \alpha\rangle - r^\nu|\alpha|^2)$$

for all sufficiently small $(\alpha, \xi(\cdot), \eta(\cdot)) \in R^n \times L_2 \times L_2$.

This is all the more valid for $(\xi(\cdot), \eta(\cdot)) \in L_\infty$. So considering the inequality in $R^n \times L_\infty \times L_\infty$ and taking into account that $k(t)z_2^\nu(t)$ is a summable function (which is implied by the fact that the proximal subdifferential is nonempty) we conclude that there is a measurable $q^\nu(t)$ (with values in R^n) which satisfies $|q^\nu(t)| \leq 1$ almost everywhere and actually is a unit vector proportional to

$$z_2^\nu(t) - a_2^\nu \ - \int_0^t y_2^\nu(\tau)d\tau$$

(if the latter is not equal to zero) and such that

$$w_2^\nu(t) = k(t)q^\nu(t); \quad v_2^\nu(t) = -\int_t^1 k(\tau)q^\nu(\tau)d\tau; \quad b_2^\nu = -\int_0^1 k(t)q^\nu(t)dt.$$

Summarizing, we conclude: there are (for sufficiently large ν)

$$\begin{array}{lll} a^\nu \in R^n; & z^\nu(\cdot) \in L_2; & y^\nu(\cdot) \in L_2; \\ b^\nu \in R^n; & w^\nu(\cdot) \in L_2; & v^\nu(\cdot) \in L_2; \\ d^\nu \in R^n; & q^\nu(\cdot) \in L_\infty & \end{array}$$

(namely, $a^\nu = a_0^\nu$, $b^\nu = b_0^\nu$, $z^\nu(t) = z_1^\nu(t)$, $y^\nu(t) = y_1^\nu(t)$, $w^\nu(t) = w_1^\nu(t)$, $v^\nu(t) = v_1^\nu(t)$) such that

$$|a^\nu| \leq 1/\nu; \quad \|z^\nu(\cdot)\|_2 \leq 1/\nu; \quad \|y^\nu(\cdot)\|_2 \leq 1/\nu; \tag{25}$$

$$(b^\nu - d^\nu, d^\nu) \in \partial_p l(a^\nu, c^\nu); \tag{26}$$

$$(w^\nu(t), v^\nu(t)) \in \partial_p f(t, z^\nu(t), y^\nu(t))\text{a.e.}; \tag{27}$$

$$|b^\nu - \int_0^1 k(t)q^\nu(t)dt| \le 1/\nu, \quad \|q^\nu(\cdot)\|_\infty \le 1; \tag{28}$$

$$\|w^\nu(\cdot) + k(\cdot)q^\nu(\cdot)\|_2 \le 1/\nu; \tag{29}$$

$$\|v^\nu(\cdot) + d^\nu - \int_t^1 k(\tau)q^\nu(\tau)d\tau\|_2 \le 1/\nu \tag{30}$$

We also observe that the inequalities

$$|I_i(a_i^\nu, z_i^\nu(\cdot), y_i^\nu(\cdot)) - I_i(0,0,0)| < 1/\nu, \quad i = 0,1,2$$

along with lower semicontinuity of the terminal function and the integrand on the one hand, and (A_3) on the other (which makes possible the application of the Fatou lemma) allows to state that $l(a^\nu, c^\nu) \to l(0,0)$ and(up to selection of a subsequence) $f(t, z^\nu(t), y^\nu(t)) \to f(t,0,0)$ almost everywhere.

Step 5 – Conclusion. Observe that $\{d^\nu\}$ must be a bounded sequence. Indeed, as $k(\cdot) \in L_1$, the functions $t \to \int_t^1 k(\tau)q^\nu(\tau)d\tau$ form a precompact system in the topology of uniform convergence. Thus, if d^ν are unbounded (in which case we can just assume that $|d^\nu| \mapsto \infty$), then also $|v^\nu(t)|$ goes to infinity a.e. as $\nu \mapsto \infty$. It follows that the Lipschitz constant of $f(t,\cdot,\cdot)$ at $(z^\nu(t)$,$y^\nu(t))$ is not smaller than $|v^\nu(t)|$ which, however, contradicts (A_4) and the condition of the theorem, as both $z^\nu(t)$ and $y^\nu(t)$ go to zero almost everywhere.

Set

$$p^\nu(t) = -d^\nu + \int_t^1 k(\tau)q^\nu(\tau)d\tau.$$

As d^ν are uniformly bounded and the functions $t \to \int_t^1 k(\tau)q^\nu(\tau)d\tau$ form a precompact system in the topology of uniform convergence, we may assume that $p^\nu(t)$ converge uniformly to a certain function $p(t)$. By (30)

$$\|v^\nu(\cdot) - p(\cdot)\|_2 \to 0. \tag{31}$$

We notice further that $p^\nu(1) = -d^\nu$, $|p^\nu(0) + (b^\nu - d^\nu)| \le 1/\nu$ by (28), so there is a vector s^ν with $|s^\nu| \le 1/\nu$ such that by (27)

$$(p^\nu(0) + s^\nu, -p^\nu(1)) \in \partial_p l(a^\nu, c^\nu). \tag{32}$$

Therefore $p(\cdot)$ satisfies the transversality condition. On the other hand, we notice (taking (28) into account) that the sequence of $k(\cdot)q^\nu(\cdot)$ is weak precompact in L_1. By the Mazur theorem, there is a sequence of convex combinations of $k(\cdot)q^\nu(\cdot)$ norm converging in L_1, and even almost everywhere. As we assume that $p^\nu(\cdot)$ converge to $p(\cdot)$ uniformly, the limit must be $\dot{p}(\cdot)$. By (29) the sequence of corresponding convex combinations of $w^\nu(\cdot)$ contains a subsequence converging almost everywhere to $\dot{p}(\cdot)$.

The proof of the theorem is now concluded with the following simple lemma.

Lemma 3 *Let $\phi(x,y)$ for $x,y\in R^n$ be l.s.c., and let*

$$(w^\nu,v^\nu)\in\partial_p\phi(x^\nu,y^\nu),\quad \text{and}\quad \phi(w^\nu,v^\nu)\to\phi(\bar{x},\bar{y})$$

where $x^\nu\mapsto\bar{x}$, $y^\nu\mapsto\bar{y}$, $v^\nu\mapsto\bar{v}$ and w^ν *are uniformly bounded. Let u^ν be a sequence of convex combinations of w^k, for $k\geq\nu$, converging to a certain u. Then*

$$u\in\operatorname{conv}\{w:(w,\bar{v})\in\partial\phi(\bar{x},\bar{y})\},$$

Proof. We have

$$u^\nu=\sum_{i=1}^{n+1}\alpha_i^\nu w^{k_i(\nu)}.$$

As every sequence $\{w^{k_i(\nu)}\}$, $i=1,...,n+1$, is bounded we can choose a sequence if integers, say $\{\nu_s\}$ such that for each i the sequences of $w^{k_i(\nu_s)}$ and $\alpha_i^{\nu_s}$ converge respectively to certain w_i and α_i as $s\to\infty$. Obviously, $\alpha_i\geq 0$ and $\sum\alpha_i=1$, and $(w_i,\bar{v})\in\partial\phi(\bar{x},\bar{y})$ by definition.

6 Proof of Theorem 1

With no loss of generality we can suppose, as in the proof of Theorem 5, that $x_*(t)\equiv 0$. Fix some $N>0$, $\epsilon>0$, and let B_N denote the ball of radius N around the origin in R^n. We set

$$\begin{aligned} L_\epsilon(t,x,y) &= L(t,x,y)+\epsilon|y|^2;\\ L_{N\epsilon}(t,x,y) &= L_\epsilon(t,x,y)+\delta(y,B_N);\\ \bar{L}_{N\epsilon}(t,x,y) &= \operatorname{conv}_y L_{N\epsilon}(t,x,y). \end{aligned}$$

The latter, as above, is the convex hull of $L_{N\epsilon}$ with respect to y. Because $L_{N\epsilon}$ satisfies conditions (B) (which is obvious) and (C) (because of (A_3)), we shall be able to apply Theorem 3 to it when the time comes.

We further define functionals J_ϵ, $J_{N\epsilon}$ and $\bar{J}_{N\epsilon}$ by replacing the integrand in the definition of J by $L_\epsilon, L_{N\epsilon}$ or $\bar{L}_{N\epsilon}$ respectively. It is clear that zero is the local minimum in W_1^1 of both J_ϵ and $J_{N\epsilon}$. As the condition of Theorem 4 is obviously satisfied for J_N (with $R(t)\equiv B_N$), we conclude that

$$L(t,0,0)=L_{N\epsilon}(t,0,0)=\bar{L}_{N\epsilon}(t,0,0)\quad\text{a.e.}\tag{33}$$

and zero is also a local minimum of $\bar{J}_{N\epsilon}$ in W_1^1.

We further notice that $\bar{L}_{N\epsilon}$ satisfies (A_4) as, being finite and convex on B_N, the function $\bar{L}_{N\epsilon}(t,x,\cdot)$ is Lipschitz continuous at every y with $|y|<N$. This means that Theorem 5 can be applied and therefore there is a function $p(\cdot)\in W_1^1$ such that

$$\dot{p}(t)\in\operatorname{conv}\{w:(w,p(t))\in\partial\bar{L}_{N\epsilon}(t,0,0)\},\ \text{a.e.}\tag{34}$$

and

$$(p(0), -p(1)) \in \partial l(0,0). \tag{35}$$

As follows from (33) and Lemma 1, zero is an exposed point of $L_{N\epsilon}(t, 0, \cdot)$. Therefore by Corollary 1 (after Theorem 3), for every w with $(w, p(t)) \in \partial \bar{L}_{N\epsilon}(t, 0, 0)$ we have that

$$w \in \mathrm{conv}\{u : (u, p(t)) \in \partial L_{N\epsilon}(t, 0, 0)\}, \text{ a.e.} \tag{36}$$

and

$$L_{N\epsilon}(t, 0, y) - L_{N\epsilon}(t, 0, 0) - \langle p(t), y\rangle \geq 0$$

or equivalently,

$$L(t, 0, y) + \epsilon|y|^2 - L(t, 0, 0) - \langle p(t), y\rangle \geq 0, \text{ when } |y| \leq N. \tag{37}$$

It is clear that $\partial L_{N\epsilon}(t, 0, 0) = \partial L_\epsilon(t, 0, 0)$ as $L_{N\epsilon}(t, x, y) = L_\epsilon(t, x, y)$ near the origin. Furthermore, L_ϵ is the sum of L and a smooth function whose derivative is zero at zero. Therefore, combining (34) and (36), we get

$$\dot{p}(t) \in \mathrm{conv}\{w : (w, p(t)) \in \partial L(t, 0, 0)\}. \tag{38}$$

Denote by $\mathbf{P}_{N\epsilon}$ the collection of $p(\cdot) \in W_1^1$ satisfying (35), (37), (38). It is clear that

$$\mathbf{P}_{N'\epsilon'} \subset \mathbf{P}_{N\epsilon}, \text{ if } N' > N, \ \epsilon' < \epsilon. \tag{39}$$

As follows from (A_3) and (38), there is a summable function $k(t)$ such that $|w| \leq k(t)$ if $(w, p(t)) \in \partial L(t, 0, 0)$. Therefore $\{\dot{p}(\cdot) : p(\cdot) \in \mathbf{P}_{N\epsilon}\}$ is a weakly compact set in L_1 for any N, ϵ. Now the same argument as in the proof of Theorem 5 shows that $\{p(0) : p(\cdot) \in \mathbf{P}_{N\epsilon}\}$ must be bounded: if $|p^\nu(0)| \to \infty$ for a sequence of $p^\nu(\cdot) \in \mathbf{P}_{N\epsilon}$, then $|p^\nu(t)| \to \infty$ for any t because of the weak compactness of derivatives which, however, contradicts (37).

Thus, every $\mathbf{P}_{N\epsilon}$ is relatively weak compact in W_1^1. But it is also weak closed which, again, can be proved as in the proof of Theorem 5 with the help of Lemma 3. Therefore $\mathbf{P}_{N\epsilon}$ are actually weak compact and from (39) we can now deduce that

$$\bigcap_{N,\epsilon} \mathbf{P}_{N\epsilon} \neq \emptyset.$$

For any $p(\cdot)$ of the intersection, both (35) and (36) are valid, while (37) gives

$$L(t, 0, y) - L(t, 0, 0) - \langle p(t), y\rangle \geq 0, \ \forall y$$

almost everywhere on $[0, 1]$.

The final statement of the theorem is justified by the observation that (36) and the implication (36) $\Rightarrow$ (38) hold for arbitrarily small N.

7 Examples

Two examples will illustrate the main features or our result. The first demonstrates that, for a non-differentiable integrand, there may be a solution of the Euler inclusion not satisfying the Weierstrass condition even in the case of a local minimum relative to the W_1^1-norm. The second brings out contrasts with previous works of Clarke.

Example 1. Let $n = 1$ and $L(x, y) = \min\{2|y|, 1 + |y|\} - |x|$. Then $x_*(t) \equiv 0$ is an absolute minimum in the problem

$$\text{minimize} \quad J(x(\cdot)) = \int_0^1 L(x(t), \dot{x}(t))dt, \quad x(0) = x(1) = 0.$$

Indeed, for $x(\cdot) \neq 0$ we have

$$J(x(\cdot)) \geq \int_0^1 (|\dot{x}(t)| - |x(t)|)dt \geq \int_0^1 (|\dot{x}(t)| - \int_0^t |\dot{x}(\tau)|d\tau)dt = \int_0^1 |\dot{x}(t)|(1-t)dt > 0.$$

The Euler condition gives

$$\dot{p}(t) \in \operatorname{conv}\{w : (w, p(t)) \in \{-1, 1\} \times [-2, 2]\} = [-1, 1],$$

that is, $|p(t)| \leq 2$, $|\dot{p}(t)| \leq 1$. In particular, $p(t) \equiv \alpha$ satisfies the relations if $|\alpha| \leq 2$ whereas the Weierstrass condition will be satisfied only for $|\alpha| \leq 1$.

Next we provide an example demonstrating the relative strength of necessary conditions of three types, so as to be able to compare our result with those of Clarke; here $\bar{\partial}$ stands for Clarke's generalized gradient. The three types are:

1. Our necessary conditions in Theorem 1,
2. Clarke's Euler condition (4) (given in [6] only for integrands that are locally Lipschitz continuous in (x, y)) plus the Weierstrass and the transversality conditions, and
3. Clarke's "separated" Euler condition (5) (given in [9, 10] only for locally Lipschitz integrands that are independent of t) plus the Weierstrass and the transversality conditions.

Example 2. Again we take $n = 1$ and set

$$J(x(\cdot)) = x(0) - \gamma x(1) + \int_0^1 \max\{|\dot{x}(t)| - |x(t)|, 0\}dt.$$

We shall test the above conditions by the curves

$$x_\alpha(t) = \alpha e^t, \quad \alpha \ge 0$$

none of which is optimal as we shall see. We have $\dot{x}(t) = x(t)$ along every such curve, so we have to work with the subdifferentials of $L(x,y) = \max\{|y| - |x|, 0\}$ only at points (x,y) with $y = x \ge 0$. At any such point we have (leaving easy calculations to the reader):

$$\begin{array}{lcl}
\bar{\partial}_x L(x,y) & = & \begin{cases} [-1,0], & \text{if} \quad x > 0, \\ \{0\}, & \text{if} \quad x = 0; \end{cases} \\
\bar{\partial}_y L(x,y) & = & \begin{cases} [0,1], & \text{if} \quad x > 0, \\ [-1,1], & \text{if} \quad x = 0; \end{cases} \\
\bar{\partial} L(x,y) & = & \begin{cases} \{(-\lambda,\lambda):\ 0 \le \lambda \le 1\}, & \text{if} \quad x > 0; \\ [-1,1] \times [-1,1] & \text{if} \quad x = 0; \end{cases} \\
\partial L(x,y) & = & \begin{cases} \{(-\lambda,\lambda):\ 0 \le \lambda \le 1\}, & \text{if} \quad x > 0; \\ \{(\pm\lambda,\lambda):\ |\lambda| \le 1\} & \text{if} \quad x = 0. \end{cases} \\
\{w:\ (w,p) \in \partial L(x,y)\} & = & \begin{cases} \{-p\}, & \text{if} \quad x > 0, \\ \{\pm|p|, |p|\}, & \text{if} \quad x = 0. \end{cases}
\end{array}$$

The Weierstrass condition for x_α

$$L(x_\alpha(t), y) - L(x_\alpha(t), \dot{x}_\alpha(t)) - \langle p, y - \dot{x}_\alpha(t)\rangle \ge 0, \ \forall\ y \tag{40}$$

is valid whenever $p(t) \in [0,1]$ for $\alpha > 0$ and $p(t) \in [-1,1]$ for $\alpha = 0$. The transversality condition is

$$p(0) = 1, \quad p(1) = \gamma. \tag{41}$$

Now we see that (5), (40), (41) is satisfied for $x_*(\cdot) = x_\alpha(\cdot)$ for $\alpha > 0,\ 1 \ge \gamma \ge 0$: take, say $p(t) = 1 - (1-\gamma)t$. The Euler condition in Clarke's form (4) and the one in Theorem 1 do not have solutions satisfying also (40), (41), for positive α and $\gamma \ne e^{-1}$. Indeed, in this case (4) takes the form $\dot{p} = -p$.

For $\alpha = 0$, Clarke's separated Euler condition (5) recognizes $x(t) \equiv 0$ as being nonoptimal for all $0 \le \gamma \le 1$ whereas (4) fails to disqualify $x_\alpha(\cdot)$: the same $p(t) = 1 - (1-\gamma)t$ satisfies all the conditions in this case.

Finally, Theorem 1 confirms non-optimality of $x(t) \equiv 0$ if $\gamma < e^{-1}$ and fails to do so if $e^{-1} \le \gamma \le 1$ Indeed, the Euler condition of the theorem takes the form

$$-|p(t)| \le \dot{p}(t) \le |p(t)|,$$

any solution of the inequality satisfies $e^{-t}p(0) \le |p(t)| \le e^t p(0)$

We see that the Euler condition of Theorem 1 is strictly sharper than Clarke's Euler Lagrange condition (4) and, on the other hand, each of the other two conditions can in certain situations work better than the other. We observe however that in the example the set of values of (α, γ) for which Theorem 1 works successfully whereas the separated condition of Clarke does not is much more massive. We believe this is a typical case and the situations in which Clarke's separated condition would work better than that of Theorem 1 are rather exceptional and highly unstable for the set

$$\limsup_{(u,v)\to(x,y)} \bar{\partial}_x L(u,v) \times \bar{\partial}_y L(u,v)$$

always contains the generalized gradient of Clarke at (x, y) (leaving aside $\partial L(x, y)$), typically even as a proper subset.

We also finally observe that also the Hamiltonian condition of [10] for generalized Bolza problems cannot be applied in the situation of the last example for its Hamiltonian in this case is extended–real–valued (not everywhere finite) as required by the condition.

References

[1] J. Borwein and A. Ioffe, Proximal analysis on smooth spaces, to appear.

[2] J. Borwein and D. Preiss, A smooth variational principle with applications, Trans. Amer. Math. Soc. **303**, 517–527 (1987).

[3] F. H. Clarke, Necessary conditions for nonsmooth problems of optimal control and the calculus of variations, Ph.D. thesis, University of Washington, Seattle, 1973.

[4] F. H. Clarke, The Euler-Lagrange differential inclusion, J. Differential Eq. **19**, 80–90 (1975).

[5] F. H. Clarke, Admissible relaxation in variational and control problems, J. Math. Anal. Appl. **51**, 557–576 (1975).

[6] F. H. Clarke, The generalized problem of Bolza, SIAM J. Control Opt. **14**, 469–478 (1976).

[7] F. H. Clarke, The maximum principle under minimal hypotheses, SIAM J. Control Optim. **14**, 1078–1091.

[8] F. H. Clarke, Extremal arcs and extended Hamiltonian systems, Trans. Amer. Math. Soc. **231**, 349–367 (1977).

[9] F. H. Clarke, The Erdmann condition and Hamiltonian inclusions in optimal control and the calculus of variations, Canadian J. Math. **23** (1980), 494–509.

[10] F. H. Clarke, Optimization and Nonsmooth Analysis, Wiley-Interscience, New York, 1983; reprinted as Vol. 5 of Classics in Applied Mathematics, SIAM (Philadelphia), 1990.

[11] F. H. Clarke, Methods of Dynamic and Nonsmooth Optimization, (CBMS–NSF Regional conference series in applied mathematics, vol 57) SIAM Publications, Philadelphia, 1989.

[12] F. H. Clarke, The decoupling principle in the calculus of variations, J. Math. Analysis Appl. **172** (1993), 92-105.

[13] I. Ekeland and R. Temam, Convex Analysis and Variational Problems, Dunod, Paris, 1972.

[14] A. Ioffe, On subdifferentiability spaces, Ann. New York Acad. Sci. **410**, 107–121 (1983).

[15] A. Ioffe, Calculus of Dini subdifferentials and contingent derivatives of set-valued maps, Nonlinear Anal. Theory Meth. Appl. **8**, 517–539 (1984).

[16] A. Ioffe, Proximal analysis and approximate subdifferentials, J. London Math. Soc. **41**, 175–192 (1990).

[17] A. Ioffe, Euler–Lagrange and Hamiltonian formalisms in dynamic optimization (submitted).

[18] A. Ioffe and V. Tikhomirov, Extension of variational problems, Trans. Moscow Math. Soc. **18**, 186–246 (1968).

[19] A. Ioffe and V. Tikhomirov, Theory of Extremal Problems, Nauka, Moscow, 1974 (Russian); translation North-Holland, Amsterdam, New York, 1979.

[20] P. Loewen and R. T. Rockafellar, The adjoint arc in nonsmooth optimization, Trans. Amer. Math. Soc. **325**, 39–72 (1991).

[21] P. Loewen and R. T. Rockafellar, Optimal control of unbounded differential inclusions, SIAM J. Control Opt. **32**, 442–470 (1994).

[22] B. Mordukhovich, Approximation Methods in Problems of Optimization and Control, Nauka, Moscow, 1988 (Russian); translation to appear in Wiley Interscience.

[23] B. Mordukhovich, On variational analysis of differential inclusions, in Optimization and Nonlinear Analysis, A. Ioffe, M. Marcus and S. Reich (eds.), (Pitman Research Notes in Math. vol. 244), 1992.

[24] B. Mordukhovich, Discrete approximations and refined Euler-Lagrange conditions for nonconvex differential inclusions, SIAM J. Control Opt., to appear.

[25] R. T. Rockafellar, Conjugate convex functions in optimal control and the calculus of variations, J. Math. Analysis Appl. **32**, 174–222 (1970).

[26] R. T. Rockafellar, Generalized Hamiltonian equations for convex problems of Lagrange, Pacific J. Math. **33**, 411–428 (1970).

[27] R. T. Rockafellar, Existence and duality theorems for convex problems of Bolza, Trans. Amer. Math. Soc. **159**, 1–40 (1971).

[28] R. T. Rockafellar, Dualization of subgradient conditions for optimality, Nonlinear Anal. Theory Meth. Appl. **20**, 627–646 (1993).

[29] R. T. Rockafellar, Equivalent subgradient versions of Hamiltonian and Euler-Lagrange equations in variational analysis, SIAM J. Control Opt., pending.

[30] J. Warga, Relaxed variational problems, J. Math. Anal. Appl. **4**, 111-128 (1962).