Journal of Convex Analysis
Volume 13 (2006), No. 3+4, 759–772

Three Theorems on Subdifferentiation of Convex Integral Functionals*

Alexander Ioffe
Department of Mathematics,
Technion – Israel Institute of Technology, Haifa 32000, Israel
ioffe@math.technion.ac.il

Dedicated to the memory of Simon Fitzpatrick.

Received: June 16, 2005
Revised manuscript received: December 16, 2005

The paper is devoted to calculation of subdifferential of the integral functional $\int_T f(t,x)d\mu(t)$ with f being a normal convex integrand. The first theorem gives a precise formula for the case when x is finite dimensional and some qualification condition, which reduces to the general qualification condition for the sum rule of convex analysis, is satisfied. The two other theorems deal with infinite dimensional case and the absence of any qualification conditions and offer some approximations, one for the case when the space of x is reflexive and separable and the other, when it is just separable.

Keywords: Convex function, normal convex integrand, subdifferential, normal cone, fuzzy calculus

1991 Mathematics Subject Classification: 90C25, 52A41, 26E15

1. Introduction. Statements of the results.

Let (T,Σ,μ) be a space with positive σ-finite measure and X a Banach space, and let $f(t,x)$ be a function on $T\times X$ which is measurable in t for every x and convex closed function of x for almost every t.

We shall be interested in subdifferentials of integral functionals

$$I_f(x)=\int_T f(t,x)d\mu(t)$$

(in what follows we write simply $d\mu$ rather than $d\mu(t)$).

To make sure that I_f is well-defined we impose the following (actually standard) assumptions:

($\mathbf{A}_1$) there are μ-summable[1] $a(t): T\to X^*$ and $\alpha(t): T\to I\!R$ such that

$$f(t,x)\geq\langle a(t),x\rangle+\alpha(t),\ \forall x\in X \text{ a.e.}.$$

($\mathbf{A}_2$) there is an $\overline{x}\in X$ such that $f(t,\overline{x})$ is summable.

*Supported in part by the US-Israel BSF grant no. 2000157.
[1]that is measurable in the strong sense (uniform limit of measurable functions with countably many values) with $\|a(t)\|$ μ-integrable

ISSN 0944-6532 / $ 2.50

Then (as easily follows from the Fatou lemma) I_f is a proper closed convex functions on X. In what follows we set

$$D = \text{dom } I_f = \{x \in X : \ I_f(x) < \infty\}.$$

If X is a separable Banach space, the function f is called a *normal convex integrand* if it is measurable with respect to the σ-algebra $\Sigma \otimes \mathcal{B}(X)$ and is a convex closed function of x. For a separable X it can be shown that whenever $f(t,x)$ is as stated in the beginning: convex closed in x, measurable in t for any x and satisfies $(\mathbf{A}_1)$, then there is a normal convex integrand $g(t,x)$ such that for any x the equality $f(t,x) = g(t,x)$ holds almost everywhere.

For non-separable spaces no comparable concept has yet been developed, as well as a comparable integration theory for set-valued mappings (although certain results on integral functionals in general Banach spaces can be found e.g. in [13]). Therefore we here consider only separable spaces.

Functionals of the I_f-type can naturally be considered "continuous" extensions of sums of convex functions, although they have a definite independent interest. In particular they appear in dual problems for optimal control of linear systems and certain problems of mathematical economics (see e.g. [8]), and information about the structure of their subdifferentials is instrumental in proving the existence and characterizing solutions.

In this paper we prove the following theorems

Theorem 1.1. *Let* $\dim X < \infty$. *If* $D \bigcap \text{ri } (\text{dom } f_t) \neq \emptyset$ *for almost every* t, *then for every* $x \in D$

$$\partial I_f(x) = \int_T \partial f_t(x) d\mu + N(D,x).$$

Here and below f_t stands for the function $x \to f(t,x)$ and $N(D,x)$ is the normal cone to D at x.

In the theorem below $L^p(T,X)$ $(p \in [1,\infty))$ is the Banach space of X-valued measurable mappings such that

$$\int_T \|x(t)\|^p d\mu < \infty$$

and $L^1_{weak}(T,X^*)$ is the collection of measurable and weakly integrable X^*-valued mappings, that is to say, such that $\langle x^*(t), x\rangle$ is μ- summable for any $x \in X$ and there is an $x^* \in X^*$ such that $\int_T \langle x^*(t), x\rangle d\mu = \langle x^*, x\rangle$ for all $x \in X$. In this case we also write $\int_T x^*(t) d\mu = x^*$

Theorem 1.2. *Suppose that* X *is a separable reflexive Banach space. Let* f *be a normal convex integrand on* $T \times X$ *satisfying* $(\mathbf{A}_1)$, $(\mathbf{A}_2)$. *Then* $x^* \in \partial I_f(x)$ *if and only if there are sequences of* μ*-measurable mappings* $x_n(\cdot) \in L^1(T,X)$ *and* $x_n^*(\cdot) \in L^1_{weak}(T,X^*)$ *taking values in* X *and* X^* *respectively with the following properties*

(a) $\quad x_n^*(t) \in \partial f_t(x_n(t))$ *almost everywhere;*

(b) $\quad x_n^* = \displaystyle\int_T x_n^*(t) d\mu$ *norm converge to* x^*;

(c) $\quad \displaystyle\int_T f(t, x_n(t)) d\mu \to I_f(x)$;

(*d*) $\lim_{n\to\infty}\int_T \langle x_n^*(t), x - x_n(t)\rangle d\mu = 0.$

Moreover, for any $p \in (1,\infty)$ *it is possible to choose* $x_n(\cdot)$ *belonging to* $L^p(T,X)$ *and to guarantee that*

(*e*) *there is a summable nonnegative* $\psi(t)$ *such that* $\|a(t)\|\cdot\|x_n(t)\| \le \psi(t)$ *a.e. for each* n*;*

(*f*) *there is a sequence* $(u_n) \subset X$ *norm converging to* x *and such that*

(f_1) $n\int_T \|a(t)\|\cdot\|u_n - x_n(t)\|^p d\mu \to 0$*, in particular*

(f_2) $\int_{T'} \|x_n(t) - x\|^p d\mu \to 0$ *for any set* T' *of finite measure, and*

(f_3) $\int_T \|x_n^*(t)\|\cdot\|u_n - x_n(t)\| d\mu \to 0.$

Theorem 1.3. *Suppose that* X *is a separable Banach space. Let* f *be a normal convex integrand on* $T \times X$ *satisfying* ($\mathbf{A}_1$), ($\mathbf{A}_2$)*. Then* $x^* \in \partial I_f(x)$ *if and only if there are nets of* μ*-measurable mappings* $x_\nu(\cdot) \in L^1(T,X)$ *and* $x_\nu^*(\cdot) \in L^1_{weak}(T,X^*)$ *with values in* X *and* X^* *respectively with the following properties*

(*a*) $x_\nu^*(t) \in \partial f_t(x_\nu(t))$ *almost everywhere;*

(*b*) $x_\nu^* = \int_T x_\nu^*(t) d\mu$ *weak* converge to* x^**;*

(*c*) $\int_T f(t, x_\nu(t)) d\mu \to I_f(x)$*;*

(*d*) $\lim_\nu \int_T \langle x_\nu^*(t), x - x_\nu(t)\rangle d\mu = 0.$

Moreover, for any $p \in (1,\infty)$ *it is possible to choose* $x_\nu(\cdot)$ *belonging to* $L^p(T,X)$ *and to guarantee that*

(*e*) *there is a summable nonnegative* $\psi(t)$ *such that* $\|a(t)\|\cdot\|x_\nu(t)\| \le \psi(t)$ *a.e. for each* ν*;*

(*f*) *there is a net of* $u_\nu \in X$ *norm converging to* x *and such that*

(f_1) *for any increasing sequence* $(\nu(n))$ *of indices,*

$$n\int_T \|a(t)\|\cdot\|u_{\nu(n)} - x_{\nu(n)}(t)\|^p d\mu \to 0$$

with the integrals bounded by the same constant for all ν*;*

(f_2) $x_\nu(\cdot)$ *have uniformly bounded norms and* $\int_{T'} \|x_\nu(t) - x\|^p d\mu \to 0$ *for any set* T' *of finite measure,*

(f_3) $\int_T \|x_\nu^*(t)\|\cdot\|u_\nu - x_\nu(t)\| d\mu \to 0.$

Theorem 1.1 is a strengthened version the 1968 result by Ioffe and Tikhomirov [10] (see also [11, 19]). It was assumed there that int $D \neq \emptyset$ (and under this assumption the result of [10] was extended to separable Banach spaces in [9]). Theorem 1.1 is an easy consequence of this earlier result. It is rather surprising that the possibility of such an extension has not been noticed up to now.

Meanwhile, the qualification condition of Theorem 1.1 (D meets relative interiors of dom f_t) is in a sense more natural, in particular because for the case of measures consisting of a finite number of atoms (in which case the integral becomes a sum of finitely many functions) it reduces to the standard qualification condition $\bigcap \mathrm{ri}\,(\mathrm{dom}\, f_i) \neq \emptyset$, so that the subdifferential additivity rule of finite dimensional convex analysis in its strongest form:

$$\mathrm{ri}(\mathrm{dom}\, f_1) \bigcap \mathrm{ri}(\mathrm{dom}\, f_2) \neq \emptyset \;\Rightarrow\; \partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x) \tag{1}$$

(see [15]) follows from Theorem 1.1.

But there is one substantial difference between the conclusion of Theorem 1.1 and (1) showing an additional subtleness of the first. Adding to the right-hand side of (1) the normal cone to the intersection of the domains of f_1 and f_2 does not make any change. On the contrary in the equality of Theorem 1.1 the normal cone cannot be omitted. As a simple example, consider

$$I(x) = \int_0^1 \frac{x^2}{t} dt \quad (x \in I\!R).$$

Here $\partial I(0) = I\!R$ but $\int_0^1 \partial f_t(0) dt = \{0\}$.

Theorems 1.2 and 1.3 contain descriptions of subdifferentials of integral functionals without any a priori qualification conditions. Results of that sort for finite sums and compositions of convex functions were actively discussed since mid-nineties [1, 5, 6, 12, 14, 16, 17, 18] and a variety of techniques was exploited. We also mention a paper by Fitzpatrick and Simon [4] in which subdifferentials and conjugates of marginal functions were considered along with compositions. In [6] the subdifferential $\partial(f_1 + f_2)(x)$ was represented as the intersection of sums of the ε-subdifferentials of the functions at the same point. Alternatively, in [1, 14, 17, 18] and other studies the subdifferential was represented by limits of sums of exact subdifferentials at nearby points. The latter formulations opened a possibility to get more precise information by specifying conditions that would narrow down the class of sequences needed to get the elements of e.g. $\partial(f_1 + f_2)(x)$.

Thibault [18] was first to observe that the results obtained in the course of these studies look very similar to the fuzzy calculus rules of modern variational analysis and applied the techniques developed in [2, 3] to convex subdifferentials. (The possibility of using general nonsmooth calculus in the purely convex context was also mentioned in [14].) Interestingly enough, it turned out that this approach led to the best results in the sense that the class of sequences needed to reproduce the subdifferential at a given point was further narrowed down. In this paper we adopt the same idea of applying the techniques of nonsmooth subdifferential calculus to study subdifferentials of convex functions. The method we use is (rather a substantial) modification of the approach developed in [7] for nonconvex Lipschitz (in x) integrands on $T \times I\!R^n$.

The following are a few comments I would like to add to the statements of the theorems before we pass to proofs. The first is that it is not clear whether it is possible to include the case $p = \infty$ into the statements. The proofs below do not cover this case for at a certain step of calculation normal cones to intersections of dom f_t with balls around x inevitably appear and cannot be later eliminated. On the other hand, in (finite dimensional) examples I looked upon it was possible to consider only uniformly converging $x_n(\cdot)$.

The second remark is that the assumptions on f are actually minimal for making I_f well defined. Although we deal with $x(\cdot) \in L^p$, we do not assume that $a(\cdot) \in L^q$. Convex analysis and fuzzy techniques do all the job of making meaningful all the products $\langle a(t), x(t)\rangle$ which appear in the course of calculations.

As in the case of finite sums the results of different power appear in the cases of a reflexive and non-reflexive X: it is possible to prove norm convergence of subdifferentials in the first case and only weak*-convergence of nets (rather than sequences) in the second. This happens because if X is reflexive we can use certain powerful tools of convex analysis such as the Mazur theorem while in the non-reflexive case we need to consider certain restrictions to finite dimensional subspaces to reach the same goal. It seems however important to emphasize that this is the only difference between reflexive and non-reflexive cases for all the estimates for elements of the sequences (nets) and their integrals remain the same. And actually the proofs of Theorems 1.2 and 1.3 are almost identical.

Finally, it is interesting to look what kind of results we can get for finite sums of convex functions from Theorems 1.2 and 1.3. In other words which corollaries we can get from the theorems under the additional assumption that the measure consists of finitely many atoms. It turns out that the result so obtained coincides with the strongest statement of [18]. It has to be mentioned however that [18] applies to all Banach space, not only separable, but, on the other hand, our result covers also norm convergence in the reflexive case, while the proof of [18] is specially designed for weak convergence[2]. We state here only the corresponding corollary from Theorem 1.2. There is no difficulty in stating and proving the corresponding corollary of Theorem 1.3 for non-reflexive case.

Corollary 1.4. *Let X be a separable reflexive Banach space, and let f_1 and f_2 be closed convex proper functions such that* dom $f_1 \bigcap$ dom $f_2 \neq \emptyset$. *Then $x^* \in \partial(f_1 + f_2)(x)$ if and only if there are sequences x_{ni} and x^*_{ni} ($i = 1, 2,\ n = 1, 2, ...$) such that*

(a) $x^*_{ni} \in \partial f_i(x_{ni})$;
(b) $x^*_{n1} + x^*_{n2} \to x^*$ *and* $x_{ni} \to x$;
(c) $f_i(x_{ni}) \to f_i(x)$;
(d) $\langle x^*_{ni}, x - x_{ni}\rangle \to 0$;
(e) $(\|x^*_{n1}\| + \|x^*_{n2}\|) \cdot \|x_{n1} - x_{n2}\| \to 0$.

Proof. (a) and (b) follow from parts (a), (b) and (f) of Theorem 1.2; (c) follows from the (c) of Theorem 1.2 if we take into account that f_i are lower semicontinuous. The same consideration is essential in obtaining (d): the (d) of Theorem 1.2 gives $\langle x^*_{n1}, x - x_{n1}\rangle + \langle x^*_{n2}, x - x_{n2}\rangle \to 0$ and (a) implies that $\langle x^*_{ni}, x - x_{ni}\rangle \leq f_i(x) - f_i(x_{ni})$. Finally, the last statement of the theorem gives $\|x^*_{ni}\| \cdot \|u_n - x_{ni}\| \to 0$. On the other hand by ($b$) the norms of x^*_{ni} differ by uniformly bounded quantities which, together with the fact that x_{ni} and u_n converge to zero, allow us to be sure that $\|x^*_{nj}\| \cdot \|u_n - x_{ni}\| \to 0$. This implies

[2] It was pointed out by the referee that certain results of [14], associated with Theorem 1.3, can be improved up to inclusion convergence of sequences rather than nets.

that

$$(\|x_{n1}^*\| + \|x_{n2}^*\|)\|x_{n1} - x_{n2}\| \leq (\|x_{n1}^*\| + \|x_{n2}^*\|)(\|u_n - x_{n1}\| + \|u_n - x_{n2}\|) \to 0.$$

□

2. Proof of Theorem 1.1.

The inclusion

$$\int_T \partial f_t(x) d\mu + N(D, x) \subset \partial I_f(x)$$

is elementary, so only the opposite inclusion must be proved. Without loss of generality we may assume that $0 \in D$.

As X is finite dimensional, we naturally, consider it a Euclidean space as well as all subspaces of X. Let L denote the subspace spanned by D, and let $\hat{f}$ be the restriction of f to $T \times L$. Let further $\hat{D}$ denote the same set D but considered as a set in L, so that say $N(\hat{D}, x)$ is the intersection with L of the normal cone to D. Obviously, dom $I_{\hat{f}} = \hat{D}$.

Clearly, int $\hat{D} \neq \emptyset$, so by [10] for $p \in \partial I_{\hat{f}}(x)$ we have

$$p \in \int_T \partial \hat{f}_t(x) d\mu + N(\hat{D}, x).$$

Up to now we have considered L as a separate space. But it is also a subspace of X. Let $L^\perp$ be the orthogonal complement of L. Then for $x \in D$ we have $N(D, x) = N(\hat{D}, x) + L^\perp$.

Likewise, for $q \in \partial I_f(x)$ there is a unique $u \in L^\perp$ such that $p = q + u \in \partial I_{\hat{f}}(x)$ and conversely, if $p \in \partial I_{\hat{f}}(x)$, then $p + L^\perp \subset \partial I_f(x)$.

Finally, let δ_L stands for the *indicator* of L, that is the function equal to zero on L and ∞ outside of L. We can view $\hat{f}_t$ as the function on X equal to f_t on L and ∞ outside of L, that is with $f_t + \delta_L$. As by the assumption $L \bigcap$ ri (dom f_t) $\neq \emptyset$,

$$\partial \hat{f}_t(x) = \partial f_t(x) + L^\perp.$$

Combining all these facts together, we conclude that

$$\begin{aligned} \partial I_f(x) &= \partial I_{\hat{f}}(x) + L^\perp \\ &= \int_T \partial \hat{f}_t(x) d\mu + N(\hat{D}, x) + L^\perp \\ &= \int_T (\partial f_t(x) + L^\perp) d\mu + N(D, x) \\ &= \int_T \partial f_t(x) d\mu + N(D, x) \end{aligned}$$

and the proof has been completed.

3. Proof of Theorem 1.2.

We shall first prove the theorem under the additional assumption that the measure is finite. The passage to the general case is elementary - we discuss it at the end of the prove. We can further choose a strictly convex and smooth norm in X for future calculations.

1. Set

$$\xi(t) = \|a(t)\| + 1; \qquad \beta(t) = \alpha(t) - \xi(t)$$

with $a(t)$ and $\alpha(t)$ from $(\mathbf{A}_1)$, and for $r = 1, 2, ...$ and $p > 1$ set

$$f_{r,p}(t,u) = \inf_x (f(t,x) + r\xi(t)\|x-u\|^p)$$

Then $f_{r,p}$ is also a normal convex integrand. As $f_{r,p}(t,u) \le f(t,u)$, every $f_{r,p}$ satisfies $(\mathbf{A}_2)$. The following simple calculation shows that $f_{r,p}$ also satisfies $(\mathbf{A}_1)$ with the same $a(\cdot)$ and $\beta(t) = \alpha(t) - \xi(t)$ instead of $\alpha(\cdot)$:

$$\begin{array}{rcl} f_{r,p}(t,u) & \ge & \inf_x (\langle a(t), x\rangle + r\xi(t)\|x-u\|^p) + \alpha(t) \\ & \ge & \inf_x (-\xi(t)\|x-u\| + r\xi(t)\|x-u\|^p) + \langle a(t), u\rangle + \alpha(t) \\ & \ge & \langle a(t), u\rangle + \alpha(t) - \xi(t) = \langle a(t), u\rangle + \beta(t), \end{array} \tag{2}$$

the last inequality due to the obvious fact that for all $r \ge 1$

$$\inf_{p\ge 1}\inf_{\lambda>0}(r\lambda^p - \lambda) = -1.$$

On the other hand, for any u and any z

$$f_{r,p}(t,u) \le f(t,z) + r\xi(t)\|z-u\|^p, \quad \text{a.e.}. \tag{3}$$

2. From now on up to the Step **5** of the proof we fix a $z \in \operatorname{dom} I_f$ assuming that I_f <u>attains at z its minimal value</u>. Set

$$J^r_{f,p}(x(\cdot), u) = \int_T (f(t,x(t)) + r\xi(t)\|x(t)-u\|^p)\, d\mu + \|z-u\|^p.$$

We first note that *for any $\gamma \in I\!R$ there is an $N = N(\gamma) > 0$ such that $\|u\| \le N$ whenever there is a measurable $x(\cdot)$ such that $J^r_{f,p}(x(\cdot), u) \le \gamma$.*

Indeed, set $\beta = \int_T \beta(t)d\mu$, $\xi = \int_T \xi(t)d\mu$. Then by (2)

$$\gamma \ge J^r_{f,p}(x(\cdot), u) \ge \int_T (\beta(t) - \xi(t)\|u\|)d\mu + \|z-u\|^p = \beta - \xi\|u\| + \|z-u\|^p$$

from which the claim immediately follows as $p > 1$.

In what follows we set $\gamma = I_f(z)$ and $N = N(\gamma)$.

3. Since $f_{r,p}$ is a normal convex integrand satisfying $(\mathbf{A}_1)$, $(\mathbf{A}_2)$, $I_{f_{r,p}}(u)$ is a closed convex and proper function on X, hence weakly lower semicontinuous. Therefore, it follows from (2) that $I_{f_{r,p}}(u) + \|z-u\|^p$ attains its minimum at a unique (due to strict convexity of the norm) point u_r. We have

$$I_{f_{r,p}}(u_r) + \|z-u_r\|^p \le I_{f_{r,p}}(z) \le I_f(z) = \gamma.$$

Hence the inequality

$$\|u_r\| \le N, \quad \forall\, r, \tag{4}$$

follows from the fact that for any u the infimum in the definition of $f_{r,p}$ is attained at a unique point. So if we define $x_r(t)$ by

$$f(t, x_r(t)) + r\xi(t)\|x_r(t) - u_r\|^p = f_{r,p}(t, u_r), \tag{5}$$

then $J^r_{f,p}(x_r(\cdot), u_r) \le \gamma$.

As f is a normal convex integrand, $x_r(\cdot)$ is measurable and by (2), (3)

$$f(t, z) + r\xi(t)\|z - u_r\|^p \ge \xi(t)\left(r\|x_r(t) - u_r\|^p - \|x_r(t) - u_r\|\right) + \beta(t) - \xi(t)\|u_r\|. \tag{6}$$

almost everywhere. It follows that for $\omega(t) = |f(t, z) - \beta(t) + N\xi(t)|$, we have

$$\omega(t) + r\xi(t)\|z - u_r\|^p \ge \xi(t)\left(r\|x_r(t) - u_r\|^p - \|x_r(t) - u_r\|\right), \tag{7}$$

Observe that $r\lambda^p - \lambda \le \eta$ implies that $\lambda \le \max\{\lambda_{p,r}, \eta/r\}$, where

$$\lambda_{p,r} = \left(\frac{r+1}{r}\right)^{\frac{1}{p-1}}$$

(for $r\lambda^p - \lambda \ge r\lambda$ if $\lambda \ge \lambda_{p,r}$). Therefore (7) implies that

$$\|x_r(t) - u_r\| \le \max\{2^{\frac{1}{p-1}}, \frac{\omega(t)}{r\xi(t)} + \|z - u_r\|^p\}$$

and, consequently,

$$\xi(t)\|x_r(t) - u_r\| \le \max\{2^{\frac{1}{p-1}}\xi(t), \frac{\omega(t)}{r} + \|z - u_r\|^p\xi(t)\}.$$

This implies that there is a summable $\Psi(t)$ such that

$$\xi(t)\|x_r(t)\| \le \Psi(t) \text{ a.e.}, \quad \forall \text{ r} \tag{8}$$

and also by (4) that

$$\int_T \xi(t)\|x_r(t) - u_r\|^p \to 0, \quad \text{as } r \to \infty. \tag{9}$$

We can therefore assume that $\|x_r(t) - u_r\| \to 0$ almost everywhere.

4. We claim that actually

$$\|u_r - z\| \to 0; \quad \int_T r\xi(t)\|x_r(t) - u_r\|^p d\mu \to 0, \tag{10}$$

so that, effectively, we can assume that $x_r(t) \to z$ for almost every t.

Set

$$K_r = r\int_T \xi(t)\|x_r(t) - u_r\|^p d\mu + \|u_r - z\|^p.$$

To prove (10) we shall show that $K_r \to 0$. Assume the contrary: $\lim K_r = \rho > 0$ along certain subsequence of indices r. We can assume that this is true for the entire sequence as the subsequent discussion is insensitive to whether we deal with the whole sequence or some its subsequence.

As the sequence (u_r) is bounded, by Mazur's theorem there is a sequence of convex combinations of its elements norm converging to a certain v. More specifically, for any $r = 1, 2, ...$ we can find finite collections of indices $r \le r_1 < r_2 < ... < r_{k(r)}$ and nonnegative numbers $\alpha_1, ..., \alpha_{k(r)}$ with sum one such that

$$v_r = \sum_{i=1}^{k(r)} \alpha_i u_{r_i} \ \to \ v.$$

Clearly, by (9) the corresponding sequence of convex combinations of $x_r(\cdot)$ will converge in $L^p(T, X)$ and almost everywhere to the function identically equal to v:

$$z_r(\cdot) = \sum_{i=1}^{k(r)} \alpha_i x_{r_i}(\cdot) \to x(\cdot) \equiv v \quad \text{in } L^p(T, X).$$

We have

$$\begin{aligned}
\gamma \ &\ge \ \sum_{i=1}^{k(r)} \alpha_i \left(I_{f_{r_i,p}}(u_{r_i}) + \|z - u_{r_i}\|^p \right) \\
&= \ \sum_{i=1}^{k(r)} \alpha_i \int_T f(t, x_{r_i}(t)) d\mu + \sum_{i=1}^{k(r)} \alpha_i \left(\int_T r_i \xi(t) \|x_{r_i}(t) - u_{r_i}\|^p) d\mu + \|z - u_{r_i}\|^p \right) \\
&= \ \sum_{i=1}^{k(r)} \alpha_i \int_T f(t, x_{r_i}(t)) d\mu + \sum_{i=1}^{k(r)} \alpha_i K_{r_i} \\
&\ge \ \int_T f(t, z_r(t)) d\mu + \sum_{i=1}^{k(r)} \alpha_i K_{r_i}.
\end{aligned}$$

Applying the Fatou theorem (available thanks to $(\mathbf{A}_1)$ and (8)) we get from the above inequality

$$\gamma = I_f(z) \ge \int_T \liminf_{r \to \infty} f(t, z_r(t)) d\mu + \rho^p \ge I_f(v) + \rho^p.$$

But $I_f(v) \ge I_f(z)$, hence $\rho = 0$ in contradiction with the assumption. The inequality also shows that

$$\int_T f(t, x_r(t)) d\mu \to I_f(z). \tag{11}$$

This proves the claim

We can now summarize the results of our discussion as follows: *if I_f attains its minimum at z, then the $(x_r(\cdot), u_r)$ at which $J^r_{f,p}$ attains its minimal value satisfies* (8), (10), (11).

5. Assume now that $x^* \in \partial I_f(z)$. This means that $I_f(x) - \langle x^*, x\rangle$ attains minimum at z. Or equivalently, if

$$g(t,x) = f(t,x) - \frac{1}{\mu(T)}\langle x^*, x\rangle,$$

then I_g attains minimum at z on X.

Let $(x_r(\cdot), u_r)$ be the minimizer of $J_g^r(x(\cdot), u)$. Then zero belongs to the subdifferential of J_g^r at $(x_r(\cdot), u_r)$, and as we have seen, (10) and (11) hold, the latter with f replaced by g. We have

$$J_g^r(x(\cdot), u) = \int \left(f(t, x(t)) - \frac{1}{\mu(T)}\langle x^*, x(t)\rangle + r\xi(t)\|x(t) - u\|^p \right) d\mu + \|z - u\|^p.$$

Subdifferentiation of J_g^r makes no difficulty. If $x^*(\cdot) \in L^q(T, X^*)$ belongs to the subdifferential of $\int g(t, x(t))d\mu$ at a certain $y(\cdot)$, then $\int (g(t, x(t)) - \langle x^*(t), x(t)\rangle)d\mu$ attains minimum at $y(\cdot)$, and as g is a normal integrand (which means that the infimum of the integral is equal to the integral of the infimum of the integrand in x at every t,), $g(t,x) - \langle x^*(t), x\rangle \geq g(t, y(t)) - \langle x^*(t), y(t)\rangle$ for all x almost everywhere, that is to say, $x^*(t) \in \partial g(t, y(t))$ for almost every t.

The other two terms in J_g^r are differentiable with respect to u with easily calculable derivatives. To differentiate the second term, set

$$y_r^*(t) = pr\xi(t)\|x_r(t) - u_r\|^{p-1} w_r^*(t),$$

where $w_r^*(t)$ is the derivative of $\|\cdot\|$ at $x_r(t) - u_r$ if $x_r(t) \neq u_r$ or any element of the unit ball if $x_r(t) = u_r$. Then the integral of $-y_r^*(t)$ is the derivative with respect to u of $\int_T (r/2)\xi(t)\|x_r(t) - u\|^p d\mu$. The derivative of the third term $\|\cdot\|^p$ with respect to u at u_r is

$$z_r^* = p\|z - u_r\|^{p-1} v_r^*,$$

where v_r^* is the derivative of the norm at $z - u_r$ if $z \neq u_r$ or any element of the unit ball if $z = u_r$.

As $J_g^r(x_r(\cdot), u)$, as a function of u, attains minimum at u_r, the derivative of is zero, that is

$$\int_T y_r^*(t)d\mu = -z_r^*. \tag{12}$$

Furthermore as $J_g^r(x(\cdot), u_r)$ attains at $x_r(\cdot)$ minimum in $x(\cdot)$, the integrand $g(t,x) + r\xi(t)\|x - u_r\|^p$ attains minimum at $x_r(t)$ for almost every t. This means that $-y_r^*(t) \in \partial g_t(x_r(t))$ almost everywhere, or equivalently, there is a measurable $x^*(t)$ such that

$$x_r^*(t) \in \partial f(t, x_r(t)) \quad \text{a.e..} \tag{13}$$

and

$$x_r^*(t) - \frac{1}{\mu(T)}x^* + y_r^*(t) = 0 \quad \text{a.e..} \tag{14}$$

Observe that the integral in (12) has to be understood in the weak sense: for any $u \in X$

$$\int_T \langle y_r^*(t), u\rangle d\mu = \langle z_r^*, u\rangle.$$

Clearly $\|z_r^*\| \to 0$ (as $u_r \to z$ and $\|v_r^*\| \leq 1$). Therefore

$$\| \int_T x_r^*(t) d\mu - x^* \| \to 0. \tag{15}$$

Furthermore, it follows from (8), (14) and the second relation of (10) that

$$\int_T \langle x_r^*(t), u_r - x_r(t) \rangle d\mu \to 0$$

and, moreover that

$$\int_T \|x_r^*(t)\| \cdot \|x_r(t) - u_r\| d\mu = r \int_T \xi(t) \|x_r(t) - u_r\|^p d\mu \to 0 \tag{16}$$

by (10). Together with (12) and the first relation of (10), this also implies that

$$\int_T \langle x_r^*(t), z - u_r \rangle d\mu \to 0,$$

whence

$$\int_T \langle x_r^*(t), z - x_r(t) \rangle d\mu \to 0. \tag{17}$$

Thus, *if* $x^* \in \partial I_f(z)$ *then there are sequences* $(x_r(\cdot)) \subset L^p(T, X)$*,* $(u_r) \subset X$ *and a sequence* $(x_r^*(\cdot))$ *of measurable* X^**-valued functions satisfying* (8), (10), (11), (13), (15)-(17). This proves the only if part of the theorem.

Indeed, if in the statement of the theorem, we write r instead of n and z instead of x, then (a) is (13), (b) is (15), (c) is (11), (e) follows from (8), (f_1) is (10), (f_3) is (16) and (d) and (f_2) are immediate from (f_3).

6. To prove the theorem in the opposite direction, we shall verify that *if there are a summable* $x_r(t)$ *and a weakly integrable* $x_r^*(t)$ *such that*

- $x_r^*(t) \in \partial f_t(x_r(t))$ *almost everywhere,*
- $x_r^* = \int_T x_r^*(t) d\mu$ *weakly converge to some* x^**;*
- $\liminf_{r\to\infty} \int_T \langle x_r^*(t), z - x_r(t) \rangle d\mu \geq 0$, a.e.,*;*
- *the limiting relations* (11)*,* (17) *hold,*

then $x^* \in \partial I_f(z)$*.*

Indeed, for any $u \in X$ we have

$$\begin{aligned}
I_f(u) - I_f(z) &= \lim_{r\to\infty} \int_T (f(t,u) - f(t, x_r(t))) d\mu \\
&\geq \limsup_{r\to\infty} \int_T \langle x_r^*(t), u - x_r(t) \rangle d\mu \\
&= \limsup_{r\to\infty} \left(\int_T \langle x_r^*(t), u - z \rangle d\mu + \int_T \langle x_r^*(t), z - x_r(t) \rangle d\mu \right) \\
&= \limsup_{r\to\infty} \int_T \langle x_r^*(t), u - z \rangle d\mu \\
&= \limsup_{r\to\infty} \langle x_r^*, u - z \rangle = \langle x^*, u - z \rangle
\end{aligned}$$

This is true for any u, hence $x^* \in \partial I_f(z)$. This completes the proof of the theorem in case when the measure is finite.

7. Assume now that μ is σ-finite with $\mu(T) = \infty$. To prove the theorem in this case we have to

(a) take any integrable positive function $\lambda(t)$;
(b) replace $f(t,x)$ and $a(t)$ by $\lambda(t)^{-1} f(t,x)$ and $\lambda(t)^{-1} a(t)$ respectively;
(c) replace $d\mu$ by $\lambda(t) d\mu$;
(d) define $\xi(t)$ by $\|a(t)\| + \lambda(t)^{-1}$;

(We observe that in the proof we only needed that $\xi(t)$ be a summable function and never used its specific structure. So the proof will go through, provided ξ is a summable positive function.)

(e) define $g(t,x)$ by $g(t,x) = f(t,x) - \eta(t)\langle x^*, x\rangle$, where $\eta(t)$ is any positive function on T with the integral equal one. (14) in this case assumes the form

$$x_r^*(t) - \eta(t)x^* + y_r^*(t) = 0 \quad \text{a.e..}$$

Remark. The requirement that (11) hold can be replaced by the following: $x_r(t) \to z$ almost everywhere and there is a summable $\Psi(t)$ such that (8) holds for all r. Indeed, in this case by $(\mathbf{A}_1)$

$$f(t, x_r(t)) \geq -\|\xi(t)\|\|x_r(t)\| + \alpha(t) \geq \alpha(t) - \Psi(t).$$

By Fatou lemma it follows that

$$\liminf_{r\to\infty} \int_T f(t, x_r(t))d\mu \geq \int_T f(t,z)d\mu = I_f(z)$$

and the above calculation again works (with the inequality instead equality in the first line).

4. Proof of Theorem 1.3.

The proof of the third theorem follows, with some variations, the lines of the proof of Theorem 1.2. The key difference is that we cannot any more apply the Mazur theorem and have to use another device to guarantee convergence of u_r to z. Namely, we fix a strictly convex and Gâteaux differentiable norm in X and instead of $J^r_{f,p}$, consider the functionals

$$J^r_{f,L}((x(\cdot), u) = J^r_{f,p}((x(\cdot), u) + \delta_L(z - u),$$

where L is a finite dimensional subspace of X and δ_L is the indicator of L, that is the function equal to zero on L and $+\infty$ outside of L.

We note that all results of the discussions at the first two steps of the proof of Theorem 1.2 apply to our case without any change. Step **3** needs a small change: we no longer can be sure that $J^r_{f,p}$ attains minimum and have to apply a variational principle to define u_r and $x_r(\cdot)$. Specifically, we apply Ekeland's principle and define u_r and $x_r(\cdot)$ as the point at which the function

$$M^r_{f,p}(x(\cdot), u) = J^r_{f,p}(x(\cdot), u) + \frac{1}{r}\left(\int_T \|x(t) - x_r(t)\|d\mu + \|u - u_r\|\right) + \delta_L(u - z)$$

attains absolute minimum on $L^1(T,X)$ and hence on $L^p(T,X)$.

In the subsequent analysis, we replace $f(t,x)$ by

$$f^r(t,x) = f(t,x) + \frac{1}{r}\|x - x_r(t)\|$$

and $J^r_{f,p}$ by $M^r_{f,p}$ and proceed up to the end of Step **3** with small technical changes caused by the difference in functions and not affecting the results including (8), (9). In particular, it follows from (9), that $x_r(\cdot) \in L^p(T,X)$.

The main change comes at Step **4**. As u_r belong to a bounded subset of a finite dimensional linear manifold, it is possible to choose a subsequence converging to a certain v. By (9) $x_r(\cdot)$ will converge in L^p along the same subsequence of indices to the function identically equal to v, and from this point the argument similar to that in the proof of Theorem 1.2 leads to (10), (11).

We then proceed as in the beginning of the fifth step of the proof of Theorem 1.2 (keeping in mind the differences in the functionals). We define, however, $y_r^*(t)$ and z_r^* exactly as in the proof of Theorem 1.2. But instead of (12), we shall have

$$\int_t y_r^*(t)d\mu - v_r^* \in \frac{1}{r}B + L^\perp,$$

where B stands for the unit ball in X. The analogues of (13) and (14) now look as follows:

$$x_r^*(t) \in \partial f^r(t, x_r(t)) \subset \partial f(t, x_r(t)) + \frac{1}{r}B$$

and

$$\|x_r^*(t) - \frac{1}{\mu(T)}x^* + y_r^*(t)\| \le \frac{1}{r}.$$

Thus for any finite dimensional subspace L and any $\varepsilon > 0$, we can find $x(t)$, u and $x^*(t)$ such that

$$\xi(t)\|x(t)\| \le \Psi(t); \quad \|u - z\| < \varepsilon; \quad \int_T r\xi(t)\|x(t) - z\|^p d\mu < \varepsilon; \tag{18}$$

$$|\int_T f(t,x(t))d\mu - I_f(z)| < \varepsilon; \quad x^*(t) \in \partial f(t, x(t) + \varepsilon B; \tag{19}$$

$$\int_t x^*(t)d\mu - x^* \in \varepsilon B + L^\perp. \tag{20}$$

The collection of all pairs (L,ε) has a natural structure of a directed set: $\nu = (L,\varepsilon) \succ \nu' = (L',\varepsilon')$ if L' is a proper subspace of L and $\varepsilon < \varepsilon'$. If we use this collection as the index set for pairs (x_ν, x_ν^*) satisfying (18)-(20) for $\nu = (L,\varepsilon)$, then (20) gives weak* convergence of $\int x_\nu(t)d\mu$ to x^* and the other claims of Theorem 1.3 follow from (18), (19) exactly as in the proof of Theorem 1.2.

Acknowledgements. I wish to thank the referees for useful comments. I am especially grateful to one of the referees for the extremely careful reading, for many suggestions and remarks that helped to improve the paper and for the subsequent discussion.

References

[1] H. Attouch, J.-B. Baillon, M. Théra: Variational sum of monotone operators, J. Convex Analysis 1 (1994) 1–29.

[2] J. M. Borwein, A. D. Ioffe: Proximal analysis in smooth spaces, Set-Valued Anal. 4 (1996) 1–24.

[3] J. M. Borwein, Q. J. Zhu: Viscosity solutions and viscosity subderivatives in smooth Banach spaces with applications to metric regularity, SIAM J. Control Optimization 34 (1996) 1576–1991.

[4] S. Fitzpatrick, S. Simon: The conjugates, compositions and marginals of convex functions, J. Convex Analysis 8 (2001) 423–446.

[5] J.-B. Hiriart-Urruty, M. Moussaoui, A. Seeger, M. Volle: Subdifferential calculus without qualification conditions, using approximate subdifferentials: A survey, Nonlinear Anal., Theory Methods Appl. 24 (1995) 1727–1754.

[6] J.-B. Hiriart-Urruty, R. R. Phelps: Suddifferential calculus using ε-subdifferentials, J. Funct. Analysis 118 (1993) 154–166.

[7] A. D. Ioffe: Calculus of Dini subdifferentials of functions and contingent coderivatives of set-valued maps, Nonlinear Anal., Theory Methods Appl. 5 (1984) 517–539.

[8] A. D. Ioffe: A variational problem associated with a model of welfare economics with a measure space of agents, Adv. Math. Economics 8 (2006) 297–314.

[9] A. D. Ioffe, V. L. Levin: Subdifferentials of convex functions, Tr. Mosk. Mat. Obshch. 26 (1972) 3–73.

[10] A. D Ioffe, V. M. Tikhomirov: On minimization of integral functionals, Funct. Anal. Appl. 3 (1969) 218–227.

[11] A. D. Ioffe, V. M. Tikhomirov: Theory of Extremal Problems, Nauka, Moscow (1974), (in Russian); English translation: North Holland, Amsterdam (1979).

[12] F. Jules, M. Lassonde: Formulas for subdifferentials of sums of convex functions, J. Convex Analysis 9 (2002) 519–533.

[13] V. L. Levin: Convex Analysis in Spaces of Measurable Functions and its Application to Mathematics and Economics, Nauka, Moscow (1985), (in Russian).

[14] J.-P. Penot: Subdifferential calculus without qualification assumptions, J. Convex Analysis 3 (1996) 207–219.

[15] R. T. Rockafellar: Convex Analysis, Princeton University Press, Princeton (1970).

[16] L. Thibault: A generalized sequential formula for subdifferentials of sums of convex functions defined on Banach spaces, in: Recent Developments in Optimization, R. Durier (ed.) et al., Lecture Notes in Economics and Math. Systems 429, Springer, Berlin (1995) 340–345.

[17] L. Thibault: Sequential convex subdifferential calculus and sequential Lagrange multipliers, SIAM J. Control Optimization 35 (1997) 1434–1444.

[18] L. Thibault: Limiting convex subdifferential calculus with applications to integration and maximal monotonicity of subdifferential, in: Constructive, Experimental, and Nonlinear Analysis, M. Théra (ed.), CMS Conf. Proc. 27, Amer. Math. Soc., Providence (2000) 279–289.

[19] M. Valadier: Sous-differentiels d'une borne supérieure et d'une somme continue de fonctions convexes, C. R. Acad. Sci., Paris, Sér. A 268 (1969) 39–42.