ON THE INTERCHANGE OF SUBDIFFERENTIATION AND CONDITIONAL EXPECTATION FOR CONVEX FUNCTIONALS

R.T. Rockafellar
R. J-B. Wets

July 1981
WP-81-89

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

R.T. Rockafellar's research was supported in part by the Air Force Office of Scientific Research, Air Force Systems Command, USAF under grant no. 77-3204.

R. J-B. Wets' research was supported in part by a grant of the National Science Foundation.

ABSTRACT

We show that the operators E^{G} (conditional expectation given a τ-field G) and ∂ (subdifferentiation), when applied to a normal convex integrand f, commute if the effective domain multifunction $\omega \to \{x \in R^n \mid f(\omega,x) < +\infty\}$ is G-measurable.

ON THE INTERCHANGE OF SUBDIFFERENTIATION AND CONDITIONAL EXPECTATION FOR CONVEX FUNCTIONALS

R.T. Rockafellar and R. J-B. Wets

We deal with interchange of conditional expectation and subdifferentiation in the context of stochastic convex analysis. The purpose is to give a condition that allows the commuting of these two operators when applied to convex integral functionals.

Let (Ω, A, P) be a probability space, G a τ-field contained in A, and f an A-*normal convex integrand* defined on $\Omega \times R^n$ with values in $R \cup \{\infty\}$. The latter means that the map

$$\omega \to \text{epi } f(\omega,\cdot) = \{(x,\alpha) \in R^{n+1} \mid \alpha \geq f(\omega,x)\}$$

is a closed-convex-valued A-measurable multifunction. See [2] and [9] for more on normal integrands and their properties. In particular recall that for any A-measurable function $x:\Omega \to R^n$, the function

$$\omega \to f(\omega, x(\omega))$$

is a A-measurable and the *integral functional* associated with f is defined by

$$I_f(x) = \int f(\omega, x(\omega)) P(d\omega) \quad .$$

To bypass some trivialities we impose the following summability conditions:

(1) there exists a G-measurable $x:\Omega \to R^n$ such that $I_f(x)$ is finite,

(2) there exists $v \in L_n^1(G) = L^1(\Omega, G, P; R^n)$ such that $I_{f*}(v)$ is finite,

where f^* is the (A-normal) conjugate convex integrand, i.e.

$$f^*(\omega,x) = \sup_{x \in R^n} [v \cdot x - f(\omega,x)] \quad .$$

Finally, we assume that A -- and hence also G -- is countably generated, and that there exists a *regular* conditional probability (given G), $P^G: A \times \Omega \to [0,1]$. Whenever we refer to the conditional expectation given G, we always mean the version obtained by integrating with respect to P^G. Consequently all conditional expectations will be regular.

In particular the conditional expectation $E^G f$ of f is the G-normal integrand defined by

$$(E^G f)(\omega,x) = \int f(\zeta,x) P^G(d\zeta \mid \omega) \quad .$$

Also given $\Gamma: \Omega \rightrightarrows R^n$, a closed-valued A-measurable multifunction, its conditional expectation given G is a closed-valued G-measurable multifunction obtained via a projection-type operation from a set

$$L_\Gamma^1 = \{u \in L^1(\Omega, A, P; R^n) \mid u(\omega) \in \Gamma(\omega) \text{ a.s.}\} \subset L_n^1(A)$$

onto $L_n^1(G) = L^1(\Omega, G, P; R^n)$. Valadier has shown that a regular version $E^G\Gamma: \Omega \rightrightarrows R^n$ is given by the expression

$$E^G\Gamma(\omega) = \text{cl}\{\int u(\zeta) P^G(d\zeta \mid \omega) \mid u \in L_n^1(A), u(\omega) \in \Gamma(\omega) \text{ a.s.}\} \ .$$

We refer to [12] and the references given therein for the properties of $E^G f$; in particular to the article of Dynkin and Estigneev [3], which specifically deals with regular conditional expectations of measurable multifunctions.

We consider I_f and $I_{E^G f}$ as (integral) functionals on $L_n^\infty(A)$ and $L_n^\infty(G)$ respectively. The natural pairings of L^∞ with L^1 and $(L^\infty)^*$ yield for each functional two different subgradient multifunctions. We shall use ∂I_f and $\partial I_{E^G f}$ for designating L^1-subgradients and $\partial^* I_f$ and $\partial^* I_{E^G f}$ for $(L^\infty)^*$-subgradients. Rockafellar [8, Corollary 1B] shows that when the summability conditions (1) and (2) are satisfied, one has the following representation for $(L^\infty)^*$-subgradients:

(3) $$\partial^* I_f(x) = \{v+v_s \mid v \in \partial I_f(x), v_s \in S_n(A) \text{ with } v_s[x-x'] \geq 0 \ \forall x' \in \text{dom } I_f\}$$

where $S_n(A)$ is the space of *singular* continuous linear functionals on $L_n^\infty(A)$, and

$$\text{dom } I_f = \{x \in L_n^\infty(A) \mid I_f(x) < +\infty\}$$

is the effective domain of I_f. (For the decomposition of $(L_n^\infty)^*$ consult [2, Chapter VIII]). Furthermore the L^1-subgradient set is given by

(4) $$\partial I_f(x) = \{v \in L_n^1(A) \mid v(\omega) \in \partial f(\omega, x(\omega)) \text{ a.s.}\} \quad .$$

The summability conditions (1) and (2) on f imply similar properties for $E^G f$, so the formulas above also apply to $I_{E^G f}$. Thus for $x \in L_n^\infty(G)$ we get

(5) $$\partial^* I_{E^G f}(x) = \{u+u_s \mid u \in \partial I_{E^G f}(x), u_s \in S_n(G) \text{ with } u_s[x-x'] \geq 0 \ , \forall x' \in \text{dom } I_{E^G f}\}$$

and

(6) $$\partial I_{E^G f}(x) = \{u \in L_n^1(G) \mid u(\omega) \in \partial E^G f(\omega, x(\omega)) \text{ a.s.}\} \quad .$$

We are interested in the relationship between ∂I_f and $\partial I_{E^{\mathcal{G}} f}$. Relying on the formulas just given, Castaing and Valadier [2, Theorem VIII.37] show that if in place of the summability conditions (1) and (2), one makes the stronger assumption:

(7) there exists $x^0 \in L_n^\infty(\mathcal{G})$ at which I_f is finite and norm continuous,

then for every $x \in L_n^\infty(\mathcal{G})$ one gets:

$$(8) \qquad \partial I_{E^{\mathcal{G}} f}(x) = E^{\mathcal{G}}(\partial I_f(x)) + \mathrm{rc}[\partial I_{E^{\mathcal{G}} f}(x)] \quad ,$$

where rc denotes the recession (or asymptotic) cone [2,7]. If $x \in \mathrm{int\,dom}\, I_{E^{\mathcal{G}} f}$, $\partial I_{E^{\mathcal{G}} f}(x)$ is weakly compact and then $\mathrm{rc}[\partial I_{E^{\mathcal{G}} f}(x)] = \{0\}$, in which case

$$(9) \qquad \partial I_{E^{\mathcal{G}} f}(x) = E^{\mathcal{G}} \partial I_f(x) \quad .$$

This was already observed by Bismut [1, Theorem 4]. For the subspace of L_n^∞ of constant functions, Hiriart-Urruty [4] obtains a similar result for the ε-subdifferentials of convex functions.

Here we shall go one step further and provide a condition under which the rc term can be dropped from the identity (8) without requiring that $x \in \mathrm{int\,dom}\, I_f$. Very simple examples show that the rc term is sometimes inescapable in (8). For instance, suppose $\mathcal{G} = \{\phi, \Omega\}$ (so $E^{\mathcal{G}} = E$) and consider $f(\omega,\cdot) = \psi_{(-\infty,\xi(\omega)]}$, the indicator of the unbounded interval $(-\infty,\xi(\omega)]$, where ξ is a random variable uniformly distributed on $[0,1]$. In this case $\psi_{(-\infty,0]} = Ef = E^{\mathcal{G}} f = I_{E^{\mathcal{G}} f}$, so that $\partial I_{E^{\mathcal{G}} f}(0) = R_+$ but $E^{\mathcal{G}}(\partial I_f(0)) = E\{0\} = \{0\}$. Thus (8) would fail without the rc term.

THEOREM. Suppose f *is an* A*-normal convex integrand such that the closure of its effective domain multifunction*

$$(10) \qquad \omega \mapsto D(\omega) := \mathrm{cl\,dom}\, f(\omega,\cdot) = \mathrm{cl}\, \{x \in R^n \mid f(\omega,x) < +\infty\}$$

is G-measurable. Assume that $I_f(x) < +\infty$ *for every* $x \in L_n^\infty(G)$ *such that* $x(\omega) \in \mathrm{dom}\, f(\omega,\cdot)$ *a.s., and that there exists* $x^0 \in L_n^\infty(G)$ *at which* I_f *is finite and norm continuous. Then for every* $x \in L_n^\infty(G)$ *one has*

$$\partial E^G f(\cdot,x(\cdot)) = E^G \partial f(\cdot,x(\cdot)) \text{ a.s.} \quad , \tag{11}$$

or in other words, the closed-valued G-measurable multifunctions

$$\omega \mapsto \partial E^G f(\omega,x(\omega))$$

and

$$\omega \mapsto E^G[\partial f(\cdot,x(\cdot))](\omega)$$

are almost surely equal.

Proof. From (8) it follows that

$$\partial I_{E^G f}(x) \subset E^G(\partial I_f(x)) \quad .$$

In view of (6) and (4) this holds if and only if

$$\partial E^G f(\cdot,x(\cdot)) \subset E^G \partial f(\cdot,x(\cdot)) \text{ a.s.} \quad .$$

It thus suffices to prove the reverse inclusion. Let us suppose that $u \in \partial E^G f(\cdot,x(\cdot))$. For every $y \in R^n$, define

$$g(\omega,y) = f(\omega,y) - u(\omega)\cdot y \quad .$$

This is an A-normal convex integrand which inherits all the properties assumed for f in the Theorem (recall that $u \in L_n^1(G)$). Moreover $0 \in \partial E^G g(\cdot,x(\cdot))$. We shall show that $0 \in E^G \partial g(\cdot,x(\cdot))$, which in turn will imply that $u \in E^G \partial f(\cdot,x(\cdot))$ and thereby complete the proof of the Theorem.

Since almost surely $0 \in \partial E^G g(\omega,x(\omega))$, we know that $0 \in \partial I_{E^G g}(x) \subset \partial^* I_{E^G g}(x)$. Hence x minimizes $I_{E^G g}$ on $L_n^\infty(G)$. Let inj denote the natural injection of $L_n^\infty(G)$ into $L_n^\infty(A)$ with

$$W = \text{inj}\ [L_n^\infty(G)] \quad .$$

Now note that inj $\bar{x} = \bar{x}$ also minimizes $I_{E^G g}$ on $W \subset L_n^\infty(A)$, or equivalently I_g on W, since the two integral functionals coincide on W (by the definition of conditional expectation.) Thus

$$0 \in \partial^*(I_g + \psi_W)(x) \quad ,$$

where ψ_W is the indicator function of W, or equivalently:

$$0 \in \partial^* I_g(x) + \partial^* \psi_W(x) \quad ,$$

since g is (norm) continuous at some $x^0 = \text{inj } x^0 \in W$. By (3), this means that there exist $v \in L_n^1(A)$, $v_s \in S_n(A)$, such that

(12) $\quad v(\omega) \in \partial g(\omega, x(\omega))$ a.s. ,

(13) $\quad v_s[x - x'] \geq 0 \quad$ for all $x' \in \text{dom } I_g$,

and $-(v + v_s)$ is orthogonal to W, i.e.

(14) $\quad (v + v_s)[x'] = 0 \quad$ for all $x' \in W$.

This last relation can also be expressed as

$$(v + v_s)[\text{inj } y] = 0 \quad \text{for all } y \in L_n^\infty(G) \quad ,$$

or still for all $y \in L_n^\infty(G)$

$$\text{inj}^*(v + v_s)[y] = 0 \quad ,$$

where $\text{inj}^* : (L_n^\infty(A))^* \to (L_n^\infty(G))^*$ is the adjoint of inj. Thus the continuous linear functional $\text{inj}^*(v + v_s)$ must be identically 0 on $L_n^\infty(G)$, i.e. on $L_n^\infty(G)$ one has

(15) $\quad \mathrm{inj}^* v_s = -\mathrm{inj}^* v = -E^G v \quad .$

The last equality follows from the observation that $E^G = \mathrm{inj}^*$ when inj^* is restricted to $L_n^1(A)$, cf. [2, p.265] for example.

We shall complete the proof by showing that the assumptions (12), (13) and (15) imply that

(16) $\quad (v - E^G v)(\omega) \in \partial g(\omega, x(\omega))$ a.s.

This will certainly do, since it trivially yields the sought-for relation

$$0 = E^G(v - E^G v) \in E^G \partial g(\cdot, x(\cdot)) \quad .$$

To obtain (16), it will be sufficient to show that

(17) $\quad E\{(-E^G v)(\omega) \cdot [x(\omega) - y(\omega)]\} \geq 0$

for all $y \in \mathrm{dom}\ I_g \subset L_n^\infty(A)$. To see this, recall that the relations (17) and $v \in \partial I_g(x)$ (cf. (12)) imply that $v - E^G v \in \partial I_g(x)$, from which (16) follows via the representation of L^1-subgradients given by (4). In fact, because the effective domain multifunction, or more precisely its closure $\omega \mapsto D(\omega)$, is G-measurable, it is sufficient to show that (17) holds for every $y \in \mathrm{dom}\ I_g \cap W$. Suppose to the contrary that (17) holds for every $y \in \mathrm{dom}\ I_g \cap W$ -- or equivalently because of the $\leq$ inequality that (17) holds for every $y \in \mathrm{cl\ dom}\ I_g \cap W$ -- but there exists $\hat{y} \in L_n^1(A)$ such that $I_g(\hat{y}) < +\infty$ and for which (17) fails, i.e. we have

$$E\{(-E^G v)(\omega) \cdot [x(\omega) - \hat{y}(\omega)]\} < 0 \quad .$$

Because $-E^G v$ and x are G-measurable, this inequality implies that

(18) $\quad E\{(-E^G v)(\omega) \cdot [x(\omega) - E^G \hat{y}(\omega)]\} < 0 \quad .$

Moreover, since $I_g(\hat{y}) < +\infty$, it follows that almost surely

$$\hat{y}(\omega) \in \mathrm{dom}\ g(\omega, \cdot) \subset D(\omega) \quad .$$

Taking conditional expectation on both sides, we see that

(19) $$(E^{\mathcal{G}}\hat{y})(\omega) \in E^{\mathcal{G}}D(\omega) = D(\omega) \quad ,$$

because D is a closed-valued $\mathcal{G}$-measurable multifunction. Naturally $E^{\mathcal{G}}\hat{y} \in \mathcal{W}$. Because I_g is by assumption finite on $\{z \in L_n^\infty(\mathcal{G}) \mid z(\omega) \in \text{dom } g(\omega,\cdot) \text{ a.s.}\}$, and $D(\omega) = \text{cl dom } g(\omega,\cdot)$, it follows from (19) that $E^{\mathcal{G}}\hat{y} \in \text{cl dom } I_g$. Hence (17) cannot hold for every $y \in \text{dom } I_g \cap \mathcal{W}$ since $E^{\mathcal{G}}\hat{y}$ belongs to $(\text{cl dom } I_g) \cap \mathcal{W}$ and satisfies (18).

There remains only to show that (17) holds for every $y \in L_n^\infty(\mathcal{G})$ such that inj $y = y \in \text{dom } I_g$. But now from (13) we have that for each such y

$$v_s[x-y] = v_s[\text{inj } x - \text{inj } y] \geq 0 \quad ,$$

or again equivalently: for each $y \in \text{dom } I_g \cap L_n^\infty(\mathcal{G})$,

$$(\text{inj}^* v_s)[x-y] \geq 0 \quad .$$

But this is precisely (17), since we know from (15) that on $L_n^\infty(\mathcal{G})$, $\text{inj}^* v_s = -E^{\mathcal{G}}v$. □

<u>*COROLLARY*</u>. *Suppose* f *is a* A*-normal convex integrand such that* $F(x) < +\infty$ *whenever* $x \in \text{dom } f(\omega,\cdot)$ a.s., *where*

$$F(x) = E\{f(\omega,x)\} \quad .$$

Suppose moreover that there exists $x^0 \in R^n$ *at which* F *is finite and continuous, and that the multifunction*

$$\omega \mapsto D(\omega) = \text{cl dom } f(\omega,\cdot)$$

is almost surely constant. Then for all $x \in R^n$,

(20) $$E[\partial f(\cdot,x)] = \partial F(x) \quad ,$$

where the expectation of the closed-valued measurable multi-

function Γ *is defined by*

$$E\Gamma = \mathrm{cl}\{\int v(\omega)P(d\omega) \mid v \in L_n^1(A), v(\omega) \in \Gamma(\omega) \text{ a.s.}\} \quad .$$

PROOF. Just apply the Theorem with $G = \{\phi, \Omega\}$, and identify the class of constant functions -- the G-measurable functions -- with R^n. □

This Corollary was first derived by Ioffe and Tikhomirov [5] and later generalized by Levin [6]. Note that our definition of the expectation of a closed-valued measurable multifunction is at variance with the definition now in vogue for the integral of a measurable multifunction, which does not involve the closure operation. (Otherwise the definition of the integral of a multifunction would be inconsistent with that of its conditional expectation, in particular with respect to $G = \{\phi, \Omega\}$, and also when $\Gamma \to E\Gamma$ is viewed as an integral on a space of closed sets it could generate an element that it is not an element of that space.)

APPLICATION

Consider the *stochastic optimization problem:*

(21) find $\inf E[f(\omega, x_1(\omega), x_2(\omega))]$ over all $x_1 \in L^\infty_{n_1}(G)$, $x_2 \in L^\infty_{n_2}(A)$,

where A and G are as before, and f is an A-normal convex integrand which satisfies the norm-continuity condition:

(22) there exists $(x_1^0, x_2^0) \in L^\infty_{n_1}(G) \times L^\infty_{n_2}(A)$

at which I_f is finite and norm continuous.

Suppose also that the effective domain multifunction

$$\omega \to \mathrm{dom}\, f(\omega,\cdot,\cdot) = \{(x_1, x_2) \in R^{n_1} \times R^{n_2} \mid f(\omega, x_1, x_2) < +\infty\}$$

is uniformly bounded and that there exists a summable function $h \in L^1(A)$ such that $(x_1, x_2) \in \mathrm{dom}\, f(\omega,\cdot\cdot)$ implies that

$|f(\omega,x_1,x_2)| \leq h(\omega)$. Finally suppose that the multifunction

$$\omega \mapsto D_1(\omega) = \mathrm{cl}\ \{x_1 \in R^{n_1} | \exists x_2 \in R^{n_2} \text{ such that } f(\omega,x_1,x_2) < +\infty\}$$

is G-measurable. For a justification and discussion of these assumptions cf. [11, Section 2]. From Theorem 1 of [11], it follows that the problem

(23) find inf $E[g(\omega,x_1(\omega))]$ over all $x_1 \in L^{\infty}_{n_1}(G)$,

where

$$q(\omega,x_1) = E^{G}\{\inf_{x_2 \in R^{n_2}} f(\cdot,x_1,x_2)\}(\omega) ,$$

is equivalent to (21) in the sense that if $(\bar{x}_1,\bar{x}_2)$ solves (21), then $\bar{x}_1$ solves (23), and similarly any solution x_1 of (23) can be "extended" to a solution (x_1,x_2) of (21). Both problems also have the same optimal value.

The hypotheses imply that

$$(\omega,x_1) \mapsto \inf_{x_2} f(\omega,x_1,x_2)$$

is an A-normal convex integrand, since the multifunction $\omega \mapsto \mathrm{epi}(\inf_{x_2} f(\omega,x_1,x_2))$ is closed-convex-valued and A-measurable. Its effective domain multifunction, or more precisely

$$\omega \mapsto D_1(\omega) := \mathrm{cl\ dom}\ q(\omega,\cdot) ,$$

is G-measurable. Combining (11) with the representation for the subgradients of infimal functions [13, VIII.4], we have that for every $x_1 \in L^{\infty}_{n_1}(G)$

$$\partial q(\cdot,x_1(\cdot)) = E^{G}\{v(\omega) \,|\, (v(\omega),0) \in_{a.s.} \partial f(\omega,x_1(\omega),x_2) \text{ for some } x_2 \in R^{n_2}\}(\cdot) ,$$

from which Theorem 2, the main result of [11], follows directly.

REMARK. If the underlying probability measure P has finite support, then $(L_n^\infty)^* = L_n^1$, and (11) and (20) are satisfied without any other restriction.

On the other hand, if P is nonatomic, and the effective domain multifunction (or its closure) is not G-measurable, then the identities (11) and (20) do not apply. More precisely, suppose that there exists a subset C of R^n such that the A-measurable set

$$\{\omega \mid \text{dom } f(\omega,\cdot) \cap C \neq \phi\}$$

has (strictly) positive mass and is not G-measurable. Then the term $rc[\partial I_{E^G f}(x)]$ can never be dropped from the representation of $\partial I_{E^G f}$ given by (8), as can be seen from an adaptation of the arguments in Section 4 of [10]. In those cases the inclusion $E^G \partial f \subset \partial E^G f$ will be strict for at least some $x \in L_n^\infty(G)$.

REFERENCES

[1] Bismut, J.-M. Intégrales convexes et probabilités. *J. Mathematical Analysis and Applications*, 42(1973): 639-673.

[2] Castaing, C. and M. Valadier. *Convex Analysis and Measureable Multifunctions.* Springer-Verlag Lecture Notes in Mathematics 580. Berlin, 1977.

[3] Dynkin, E.B. and I.V. Estigneev. Regular conditional expectations of correspondences. *Theory of Probability and its Applications.* 21(1976):325-338.

[4] Hiriart-Urruty, J.-B. About properties of the mean value functional and the continuous inf-convolution in stochastic convex analysis. In *Optimization Techniques Modeling and Optimization in the Service of Man.* Ed. J. Cea, Springer-Verlag Lecture Notes in Computer Science. Berlin, 1976, 763-789.

[5] Ioffe, A.D. and V.M. Tikhomirov. On the minimization of integral functionals. *Funz. Analiz.* 3(1969):61-70.

[6] Levin, V.L. On the subdifferentiability of convex functionals. *Uspekhi Mat. Nauk.* 25(1970):183-184.

[7] Rockafellar, R.T. *Convex Analysis.* Princeton University Press, Princeton, 1970.

[8] Rockafellar, R.T. Integrals which are convex functionals, II, *Pacific J. Mathematics*, 39(1971):439-469.

[9] Rockafellar, R.T. Integral functionals, normal integrands and measurable selections. In *Nonlinear Operators and the Calculus of Variations*. Springer-Verlag Lecture Notes in Mathematics 543, Berlin, 1976, 157-207.

[10] Rockafellar, R.T. and R. Wets. Stochastic convex programming: relatively complete recourse and induced feasibility. *SIAM J. Control Optimization*, 14(1976), 574-589.

[11] Rockafellar, R.T. and R. Wets. Nonanticipativity and L^1-martingales in stochastic optimization problems. *Mathematical Programming Study*, 6(1976), 170-187.

[12] Valadier, M. Sur l'espérance conditionnelle multivoque non convexe. *Ann. Inst. Henri Poincaré*, 16(1980), 109-116.

[13] Wets, R. J-B. *Grundlage konvexer Optimierung*. Springer-Verlag Lecture Notes in Economics and Mathematical Systems 137, Berlin, 1976.