ON THE INTERCHANGE OF SUBDIFFERENTIATION AND CONDITIONAL EXPECTATION FOR CONVEX FUNCTIONALS

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

July 1981
WP-81-89

Working Papers are interim reports on work of the International Institute for Applied Systems Analysis and have received only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute or of its National Member Organizations.

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.
We show that the operators $E^G$ (conditional expectation given a $\tau$-field $G$) and $\partial$ (subdifferentiation), when applied to a normal convex integrand $f$, commute if the effective domain multifunction $\omega \mapsto \{ x \in \mathbb{R}^n | f(\omega, x) < +\infty \}$ is $G$-measurable.
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 $(\Omega, A, P)$ be a probability space, $G$ a $\tau$-field contained in $A$, and $f$ an $A$-normal convex integrand defined on $\Omega \times \mathbb{R}^n$ with values in $\mathbb{R} \cup \{\infty\}$. The latter means that the map

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

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 $x : \Omega \rightarrow \mathbb{R}^n$, the function

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

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 $x: \Omega \rightarrow \mathbb{R}^n$ such that $I_f(x)$ is finite,

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

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

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

Finally, we assume that $A$ -- and hence also $G$ -- is countably generated, and that there exists a regular conditional probability (given $G$), $P^G: A \times \Omega \rightarrow [0,1]$. 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

$$(E^G f)(\omega, x) = \int f(\xi, x) P^G(d\xi | \omega) .$$

Also given $\Gamma: \Omega \rightarrow \mathbb{R}^n$, 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

$L^1_\Gamma = \{ u \in L^1(\Omega, A, P; \mathbb{R}^n) | u(\omega) \in \Gamma(\omega) \text{ a.s.} \} \subset L^1_n(A)$

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

$$E^G \Gamma(\omega) = \text{cl}\{ \int u(\xi) P^G(d\xi | \omega) | u \in L^1_n(A), u(\omega) \in \Gamma(\omega) \text{ a.s.} \} .$$

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 $l^\infty$ with $l^1$ and $(L^\infty)^*$ yield for each functional two different subgradient multifunctions. We shall use $\mathcal{A} I_f$ and $\mathcal{A} I_{E^G_f}$ for designating $L^1$-subgradients and $\mathcal{A}^* I_f$ and $\mathcal{A}^* I_{E^G_f}$ 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:

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

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

$$\forall x \in L_n^\infty(G), \{v \in L_n^1(A) \mid v(\omega) \in \mathcal{A} f(\omega, x(\omega)) \ a.s. \}.$$
We are interested in the relationship between $\partial I_f$ and $\partial I_{\text{E}^G_f}$. 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 $x^0 \in L_n^\infty(G)$ at which $I_f$ is finite and norm continuous,

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

(8) $\partial I_{\text{E}^G_f}(x) = E^G(\partial I_f(x)) + \text{rc}[\partial I_{\text{E}^G_f}(x)]$,

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

(9) $\partial I_{\text{E}^G_f}(x) = E^G \partial I_f(x)$.

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

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

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

(10) $\omega \mapsto D(\omega) = \text{cl dom } f(\omega, \cdot) = \text{cl } \{x \in \mathbb{R}^n | f(\omega,x) < +\infty\}$
is $G$-measurable. Assume that $I_f(x) < +\infty$ for every $x \in L^\infty_n(G)$ such that $x(\omega) \in \text{dom } f(\omega, \cdot)$ a.s., and that there exists $x^0 \in L^\infty_n(G)$ at which $I_f$ is finite and norm continuous. Then for every $x \in L^\infty_n(G)$ one has

\begin{equation}
\exists E^Gf(\cdot, x(\cdot)) = E^G\exists f(\cdot, x(\cdot)) \text{ a.s.},
\end{equation}

or in other words, the closed-valued $G$-measurable multi-functions

$$
\omega \mapsto \exists E^Gf(\omega, x(\omega))
$$

and

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

are almost surely equal.

**Proof.** From (8) it follows that

$$
\exists I_{E^Gf}(x) \subseteq E^G\exists f(x).
$$

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

$$
\exists E^Gf(\cdot, x(\cdot)) \subseteq E^G\exists f(\cdot, x(\cdot)) \text{ a.s.}.
$$

It thus suffices to prove the reverse inclusion. Let us suppose that $u \in \exists E^Gf(\cdot, x(\cdot))$. For every $y \in \mathbb{R}^n$, define

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

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

Since almost surely $0 \in \exists E^Gg(\omega, x(\omega))$, we know that

$$
0 \in \exists I_{E^Gg}(x) \subseteq \exists^* I_{E^Gg}(x). \text{ Hence } x \text{ minimizes } I_{E^Gg} \text{ on } L^\infty_n(G). \text{ Let }
$$

Let
inj denote the natural injection of $L_n^\infty(G)$ into $L_n^\infty(A)$ with

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

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

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

where $\psi_W$ is the indicator function of $W$, or equivalently:

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

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

$$v(\omega) \in \partial g(\omega, x(\omega)) \ \text{a.s.},$$

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

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

$$v + v_s \perp W.$$  

This last relation can also be expressed as

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

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

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

where $\text{inj}^*: (L_n^\infty(A))^* \rightarrow (L_n^\infty(G))^*$ 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) \( \text{inj}^* v_s = -\text{inj}^* v = -E^G v \).

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

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

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

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

\[
0 = E^G (v - E^G v) \in E^G \partial g(\cdot, x(\cdot))
\]

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

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

for all \( y \in \text{dom } I_g \subset L^\infty_n(A) \). 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 \( y \in \text{dom } I_g \cap \omega \).

Suppose to the contrary that (17) holds for every \( y \in \text{dom } I_g \cap \omega \)-- or equivalently because of the \( \leq \) inequality that (17) holds for every \( y \in \text{cl } \text{dom } I_g \cap \omega \)-- but there exists \( \hat{y} \in L^1_n(A) \) such that \( I_g(\hat{y}) < +\infty \) and for which (17) fails, i.e. we have

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

Because \(-E^G v\) and \( x \) are \( G \)-measurable, this inequality implies that

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

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

\[
\hat{y}(\omega) \in \text{dom } g(\omega, \cdot) \subset D(\omega)
\]
Taking conditional expectation on both sides, we see that

\[(E^G \hat{y})(\omega) \in E^G \hat{D}(\omega) = D(\omega) \]

because \( D \) is a closed-valued \( G \)-measurable multifunction. Naturally \( E^G \hat{y} \in \mathcal{W} \). Because \( I_g \) is by assumption finite on \( \{ z \in L_\infty^0(\mathbb{G}) \mid z(\omega) \in \text{dom } g(\omega, \cdot) \text{ a.s.} \} \), and \( D(\omega) = \text{cl dom } g(\omega, \cdot) \), it follows from (19) that \( E^G \hat{y} \in \text{cl dom } I_g \). Hence (17) cannot hold for every \( y \in \text{dom } I_g \cap \mathcal{W} \) since \( E^G \hat{y} \) belongs to \( (\text{cl dom } I_g) \cap \mathcal{W} \) and satisfies (18).

There remains only to show that (17) holds for every \( y \in L_\infty^0(\mathbb{G}) \) such that \( \text{inj } y = y \in \text{dom } I_g \). But now from (13) we have that for each such \( y \)

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

or again equivalently: for each \( y \in \text{dom } I_g \cap L_\infty^0(\mathbb{G}) \),

\[ (\text{inj } v_s)[x-y] \geq 0 \]

But this is precisely (17), since we know from (15) that on \( L_\infty^0(\mathbb{G}) \), \( \text{inj } v_s = -E^G v \).

\[ \square \]

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

\[ F(x) = E\{f(\omega, x)\} \]

Suppose moreover that there exists \( x^0 \in \mathbb{R}^n \) 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 \( x \in \mathbb{R}^n \),

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

where the expectation of the closed-valued measurable multi-
function $\Gamma$ is defined by

$$E\Gamma = \text{cl}\{(v(\omega)P(d\omega) | v \in L^1_n(A), v(\omega) \in \Gamma(\omega) \text{ a.s.}\}.$$  

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

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 $\Gamma \rightarrow \mathbb{E}\Gamma$ 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 $x_1 \in L^\infty_n(G), x_2 \in L^\infty_n(A)$,

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

(22) there exists $(x_1^0, x_2^0) \in L^\infty_n(G) \times L^\infty_n(A)$

at which $I_f$ is finite and norm continuous.

Suppose also that the effective domain multifunction

$$\omega \rightarrow \text{dom} f(\omega, \cdot, \cdot) = \{(x_1, x_2) \in \mathbb{R}^n_1 \times \mathbb{R}^n_2 | f(\omega, x_1, x_2) < +\infty\}$$

is uniformly bounded and that there exists a summable function $h \in L^1(A)$ such that $(x_1, x_2) \in \text{dom} f(\omega, \cdot, \cdot)$ implies that
Finally suppose that the multifunction
\[ \omega \mapsto D_1(\omega) = \text{cl} \{ x_1 \in \mathbb{R}^n_1 | \exists x_2 \in \mathbb{R}^n_2 \text{ such that } f(\omega, x_1, x_2) < +\infty \} \]
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
\[
\text{(23) } \text{find } \inf E[g(\omega, x_1(\omega))] \text{ over all } x_1 \in \mathbb{R}^n_1(G) ,
\]
where
\[ q(\omega, x_1) = E^G(\inf_{x_2 \in \mathbb{R}^n_2} f(\cdot, x_1, x_2)) \]
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
\[ \omega \mapsto \text{epi}(\inf_{x_2} f(\omega, x_1, x_2)) \]
is closed-convex-valued and $A$-measurable. Its effective domain multifunction, or more precisely
\[ \omega \mapsto D_1(\omega): = \text{cl} \text{ dom } q(\omega, \cdot) , \]
is $G$-measurable. Combining (11) with the representation for the subgradients of infimal functions [13, VIII.4], we have that for every $x_1 \in \mathbb{R}^n_1(G)$
\[ \partial q(\cdot, x_1(\cdot)) = E^G(\partial f(\omega, x_1(\omega), \cdot) , \partial f(\omega, x_1(\omega), x_2) \text{ for some } x_2 \in \mathbb{R}^n_2(\cdot) , \]
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 \( \mathbb{R}^n \) such that the \( \mathcal{A} \)-measurable set

\[
\{ \omega \mid \text{dom } f(\omega, \cdot) \cap C \neq \emptyset \}
\]

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


