The Differentiability of the Upper Envelop

Yao-Liang Yu
Department of Computing Science
University of Alberta
Edmonton, AB T6G 2E8, Canada
yaoliang@cs.ualberta.ca

November 7, 2012

Abstract

We present the proof of the Danskin-Valadier theorem, *i.e.* when the directional derivative of the supremum of a collection of functions admits a natural representation.

1 Preliminary

Consider a collection of extended real-valued functions $f_i : \mathcal{X} \mapsto \bar{\mathbb{R}}$, where $i \in \mathcal{I}$ is some index set, $\mathcal{X}$ is some real vector space, and $\bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. Define the supremum (*i.e.* upper envelop) of the collection as

$$f(x) := \sup_{i \in \mathcal{I}} f_i(x). \tag{1}$$

We are interested in studying the directional derivative of f, hopefully relating it to the directional derivatives of f_i.

Recall that the directional derivative of g, along direction d, is defined as

$$g'(x; d) := \lim_{t \downarrow 0} \frac{g(x + td) - g(x)}{t}.$$

It should be clear that $g'(x; \alpha d) = \alpha g'(x; d), \forall \alpha \geq 0$. We can similarly define the left directional derivative

$$g'_l(x; d) := \lim_{t \uparrow 0} \frac{g(x + td) - g(x)}{t} = -g'(x; -d).$$

The last equality enables us to focus exclusively on the usual directional derivative (but make immediate claims for the left directional derivative as well). Note that $g'_l(x; d) = g'(x; d)$, *i.e.* $g'(x; d) = -g'(x; -d)$ iff $g'(x; \alpha d) = \alpha g'(x; d)$ for all $\alpha \in \mathbb{R}$. We say g is Gâteaux differentiable at x if $g'(x; d)$ is a linear functional of d. (In fact, we usually require $g'(x; d) \in \mathcal{X}^*$ if $\mathcal{X}$ is a topological vector space (t.v.s.).)

2 General case

Throughout this section we will tacitly assume the directional differentiability of f_i at the point x for all $i \in \mathcal{I}$, along direction d.

Let us start with an easy proposition. Note that no topology on the index set $\mathcal{I}$ is needed.

Proposition 1 *Define $f(x) := \sup_{i \in \mathcal{I}} f_i(x)$ and $\mathcal{I}_x := \{i \in \mathcal{I} : f_i(x) = f(x)\}$. Fix a direction $d \in \mathcal{X}$ and a point $x \in \mathcal{X}$ with $f(x) \in \mathbb{R}$, then*

$$\liminf_{t \downarrow 0} \frac{f(x + td) - f(x)}{t} \geq \sup_{i \in \mathcal{I}_x} f'_i(x; d) \tag{2}$$

In particular, if for some $i_x \in \mathcal{I}_x$, f_{i_x} and f are both Gâteaux differentiable at x, then $f'(x; d) = f'_{i_x}(x; d)$.

Proof: If $\mathcal{I}_x = \emptyset$, then nothing needs to prove (since $\inf \emptyset = \infty, \sup \emptyset = -\infty$ by definition). Fix $i_x \in \mathcal{I}_x$. By definition $f_{i_x}(x) = f(x) \geq f_i(x)$, hence

$$\liminf_{t \downarrow 0} \frac{f(x+td) - f(x)}{t} \geq \liminf_{t \downarrow 0} \frac{f_{i_x}(x+td) - f_{i_x}(x)}{t} = f'_{i_x}(x; d).$$

Since i_x is chosen arbitrarily from the set $\mathcal{I}_x$, the proof is complete. ■

The last claim in Proposition 1 is nice, however, it depends on the differentiability of the envelop function, a property that is usually hard to verify in the first place. Nevertheless, for absolutely continuous functions, we have an almost free lunch:

Proposition 2 (Milgrom-Segal) *Suppose $\forall i,\ t \mapsto f_i(x+td)$ is absolutely continuous on some interval $[a,b]$ and $\sup_{i \in \mathcal{I}} |f'_i(x+td; d)| \in L_1([a,b])$, then $t \mapsto f(x+td) := \sup_{i \in \mathcal{I}} f_i(x+td)$ is absolutely continuous on $[a,b]$. If in addition, $\forall i, f_i$ is differentiable[1] and $\mathcal{I}_x := \{i \in \mathcal{I} : f_i(x) = f(x)\} \neq \emptyset$ a.e., then choose (arbitrarily!) $i_t \in \mathcal{I}_{x+td}$,*

$$f(x+td) = f(x+ad) + \int_a^t f'_{i_s}(x+sd; d) \mathrm{d}s. \tag{3}$$

Proof: Note that

$$f(x+td) - f(x+sd) \leq \sup_{i \in \mathcal{I}} |f_i(x+td) - f_i(x+sd)| = \sup_{i \in \mathcal{I}} |\int_s^t f'_i(x+rd; d) \mathrm{d}r| \leq \int_s^t \sup_{i \in \mathcal{I}} |f'_i(x+rd; d)| \mathrm{d}r,$$

hence the absolute continuity of the envelop function. (3) follows from the last claim in Proposition 1. ■

The reverse part of Proposition 1 is more interesting hence deserves a name.

Theorem 1 (Danskin) *Define $f(x) := \sup_{i \in \mathcal{I}} f_i(x)$. Fix a direction $d \in \mathcal{X}$ and a point $x \in \mathcal{X}$ with $f(x) \in \mathbb{R}$, $f_i(x+td) \in \mathbb{R}$, $\forall i$ and $\forall |t|$ sufficiently small. Suppose*

1. *$\mathcal{I}$ is countably compact[2];*
2. *$\exists\ t_0 > 0$ such that $\forall i \in \mathcal{I}, f_i(x+td)$ is absolutely continuous on $[0, t_0]$ (for instance, when $\int_0^{t_0} |f'_i(x+td; d)| \mathrm{d}t < \infty$);*
3. *The map $i \mapsto f_i(x)$ is upper semicontinuous (u.s.c.) and $(t,i) \mapsto f'_i(x+td; d)$ is u.s.c. at $(0,i)$;*

then the directional derivative of f exists and is given by the formula

$$f'(x; d) = \max_{i \in \mathcal{I}_x} f'_i(x; d), \quad \textit{where} \quad \mathcal{I}_x := \{i \in \mathcal{I} : f_i(x) = f(x)\}. \tag{4}$$

Proof: By assumption 3, $f_i(x)$, as a function of i, is u.s.c., hence the supremum in the definition of f is attained (for $\mathcal{I}$ is assumed countably compact). This proves that $\mathcal{I}_x \neq \emptyset$. Applying the u.s.c. of $f_i(x)$ again we know $\mathcal{I}_x$ is closed hence also countably compact. A similar argument then justifies our notation, *i.e.* max instead of sup in (4), and establishes the finiteness of the right-hand side in (4).

Thanks to Proposition 1, it suffices to prove

$$S := \limsup_{t \downarrow 0} \frac{f(x+td) - f(x)}{t} \leq \max_{i \in \mathcal{I}_x} f'_i(x; d),$$

from which the theorem will follow. Let $0 < t_n \downarrow 0$ such that $\Delta(t_n) := \frac{f(x+t_n d) - f(x)}{t_n} \to S$. Choose $i_n \in \mathcal{I}$ such that[3] $f_{i_n}(x+t_n d) \geq f(x+t_n d) - \epsilon_n$ where $\epsilon_n > 0, \epsilon_n / t_n \to 0$. Since $\mathcal{I}$ is countably compact, the

[1] if $|\mathcal{I}| \leq \aleph_0$, then a.e. differentiability is enough.

[2] Those who are not familiar with this notion can safely treat it as compact set in a metric space (and consequently replace all nets in the proof with sequences), while those who are really curious about it might want to read [Yu, 2012].

[3] Using (**??**) and the finiteness of $f(x)$ to argue that $f(x+t_n d) < \infty$ for all $|t_n|$ sufficiently small.

product space $\{t_n, 0\} \times \{\Delta(t_n), S\} \times \mathcal{I}$ remains countably compact, hence we can choose a convergent subnet $(t_\alpha, \Delta(t_\alpha), i_\alpha)$ such that $t_\alpha \to 0, \Delta(t_\alpha) \to S, i_\alpha \to i_* \ni \mathcal{I}$. Clearly

$$\frac{f(x+t_\alpha d) - f(x)}{t_\alpha} = \frac{f_{i_\alpha}(x+t_\alpha d) - f_{i_\alpha}(x)}{t_\alpha} + \frac{f_{i_\alpha}(x) - f(x)}{t_\alpha} + \epsilon_\alpha/t_\alpha \leq \frac{f_{i_\alpha}(x+t_\alpha d) - f_{i_\alpha}(x)}{t_\alpha} + \epsilon_\alpha/t_\alpha.$$

Assumption 2 implies that the map $t \mapsto f_{i_\alpha}(x+td)$ is absolutely continuous on $[0, t_\alpha]$, hence $\exists 0 \leq \hat{t}_\alpha \leq t_\alpha$ such that[4]

$$\frac{f_{i_\alpha}(x+t_\alpha d) - f_{i_\alpha}(x)}{t_\alpha} = \frac{1}{t_\alpha}\int_0^{t_\alpha} f'_{i_\alpha}(x+td; d)\mathrm{d}t \leq f'_{i_\alpha}(x+\hat{t}_\alpha d; d).$$

Taking limit we obtain

$$S = \limsup_{t_\alpha \to 0} \frac{f(x+t_\alpha d) - f(x)}{t_\alpha} \leq \limsup_{t_\alpha \to 0} \frac{f_{i_\alpha}(x+t_\alpha d) - f_{i_\alpha}(x)}{t_\alpha} \leq \limsup_{\hat{t}_\alpha \to 0} f'_{i_\alpha}(x+\hat{t}_\alpha d; d) \leq f'_{i_*}(x; d), \quad (5)$$

where the last inequality is due to assumption 3.

The proof will be complete once we show $i_* \in \mathcal{I}_x$. Note first that from (5), we have

$$\limsup_{t_\alpha \to 0} f_{i_\alpha}(x+t_\alpha d) - \limsup_{t_\alpha \to 0} f_{i_\alpha}(x) \leq \limsup_{t_\alpha \to 0} \big(f_{i_\alpha}(x+t_\alpha d) - f_{i_\alpha}(x)\big) = 0,$$

hence

$$f(x) \geq f_{i_*}(x) \geq \limsup_{t_\alpha \to 0} f_{i_\alpha}(x) \geq \limsup_{t_\alpha \to 0} f_{i_\alpha}(x+t_\alpha d) \geq \liminf_{t_\alpha \to 0} f_{i_\alpha}(x+t_\alpha d) = \liminf_{t_\alpha \to 0} f(x+t_\alpha d) \geq f(x),$$

where the last inequality is due to (2). ■

Remark 1 *It is clear that if $f := \inf_{i \in \mathcal{I}} f_i$, then $-f = \sup_{i \in \mathcal{I}} -f_i$, hence only assumption 3 (and the formula (4)) need some obvious change.*

Next we give a simplified version of Theorem 1 that is hopefully easier to apply.

Corollary 1 *Suppose $f(x, y) : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$ is u.s.c. on y for each x and its partial derivative $\nabla_x f(x, y)$ is jointly continuous, where $\mathcal{X}$ is an open subset of some t.v.s., $\mathcal{Y}$ is countably compact, then*

$$\phi(x) := \max_{y \in \mathcal{Y}} f(x, y)$$

is u.s.c. and admits directional derivative in all directions, in particular,

$$\phi'(x; d) = \max_{y \in \mathcal{Y}_x} \langle d; \nabla_x f(x, y)\rangle, \quad \text{where} \quad \mathcal{Y}_x := \{y \in \mathcal{Y} : f(x, y) = \phi(x)\}. \quad (6)$$

Remark 2 *It is clear that we may replace* max *by* min *in the corollary (but change u.s.c. to l.s.c.). If $f(x, y)$ is jointly continuous and $\mathcal{Y}$ is actually compact[5], then the envelop function ϕ is also continuous (Berge's maximum theorem).*

Theorem 1 needs the (somewhat annoying) countable compactness assumption. Fortunately, we can remove it by (slightly) strengthening assumptions 2 and 3. Note that when compactness is lost, the maximum is not necessarily attained, hence we need to introduce maximizing sequences. Let us denote $\hat{\mathcal{I}}_x$ as the *collection* of all maximizing sequences $\{i_n\} \subseteq \mathcal{I}$ such that $f_{i_n}(x) \to f(x)$.

Theorem 2 *Fix a direction $d \in \mathcal{X}$ and a point $x \in \mathcal{X}$ with $f(x) \in \mathbb{R}$, $f_i(x+td) \in \mathbb{R}$, $\forall i$ and $\forall |t|$ sufficiently small. Suppose*

[4] Suppose not, then $\frac{1}{t_\alpha}\int_0^{t_\alpha} f'_{i_\alpha}(x+td; d)\mathrm{d}t = \sup_{0 \leq t \leq t_\alpha} f'_{i_\alpha}(x+td; d)$, hence $f'_{i_\alpha}(x+td; d) = \sup_{0 \leq t \leq t_\alpha} f'_{i_\alpha}(x+td; d)$ a.e..

[5] The stronger compactness assumption is necessary: There exists some countably compact space $\mathcal{X}$ whose square $\mathcal{X} \times \mathcal{X}$ is not even pseudocompact, see [Gillman and Jerison, 1960][page 135, Example 9.15]. Augmenting with [Engelking, 1989][page 238, Problem 3.12.21] we know there exists a continuous function $f : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ whose pointwise supremum over the countably compact space $\mathcal{X}$ is not continuous. This argument is due to AliReza Olfati. We conjecture that sequentially compactness is not sufficient for the continuity of ϕ either (but note that AliReza's counterexample will not work).

1. *The maps* f_i *are equi-directionally differentiable at* x *along direction* d, i.e.

$$\forall \epsilon > 0, \exists \tau > 0, \text{s.t. } \forall 0 \le t \le \tau, \forall i \in \mathcal{I}, \left| \frac{f_i(x+td) - f_i(x)}{t} - f_i'(x;d) \right| \le \epsilon;$$

2. $\sup_{i\in\mathcal{I}} f_i'(x;d) < \infty$;

then the directional derivative of f *exists and is given by the formula*

$$f'(x;d) = \sup_{\{i_n\}\in\hat{\mathcal{I}}_x} \limsup_{n\to\infty} f_{i_n}'(x;d). \tag{7}$$

If in addition

3. $\forall \bar{x}$ *near* x, $\mathcal{I}_{\bar{x}} := \{i \in \mathcal{I} : f_i(\bar{x}) = f(\bar{x})\} \neq \emptyset$,

then

$$f'(x;d) = \lim_{t\downarrow 0} f_{i_t}'(x;d), \quad \textit{where} \quad i_t \in \mathcal{I}_{x+td} \textit{ is arbitrary!} \tag{8}$$

Proof: The proof is almost the same as that of Theorem 1 (in fact easier).

Let us first prove Proposition 1 again. Let $0 < t_n \to 0, 0 < \epsilon_n/t_n \to 0$. Let $\delta > 0$ and choose $N > 0$ such that $\forall n \ge N, \epsilon_n/t_n \le \delta$, and (by assumption 1)

$$\forall i \in \mathcal{I}, \quad \frac{f_i(x+t_nd) - f_i(x)}{t_n} \ge f_i'(x;d) - \delta.$$

Fix (arbitrarily) $\{i_m\} \in \hat{\mathcal{I}}_x$, then $f_{i_m}(x) \to f(x)$ hence for m large

$$\frac{f(x+t_nd) - f(x)}{t_n} \ge \frac{f(x+t_nd) - f_{i_m}(x)}{t_n} - \epsilon_n/t_n \ge \frac{f_{i_m}(x+t_nd) - f_{i_m}(x)}{t_n} - \delta \ge f_{i_m}'(x;d) - 2\delta. \tag{9}$$

Since $\{i_m\}$ and $\delta > 0$ is chosen arbitrarily, we have proved

$$\liminf_{t_n\downarrow 0} \frac{f(x+t_nd) - f(x)}{t_n} \ge \sup_{\{i_m\}\in\hat{\mathcal{I}}_x} \limsup_{m\to\infty} f_{i_m}'(x;d).$$

Next we prove the other half inequality, *i.e.*

$$\limsup_{t_n\downarrow 0} \frac{f(x+t_nd) - f(x)}{t_n} \le \sup_{\{i_m\}\in\hat{\mathcal{I}}_x} \limsup_{m\to\infty} f_{i_m}'(x;d),$$

from which (7) will follow.

Let $i_n \in \mathcal{I}$ be such that[6] $f_{i_n}(x+t_nd) \ge f(x+t_nd) - \epsilon_n$ where $\epsilon_n/t_n \to 0$. (Don't confuse i_n with the sequence $\{i_m\}$ in the previous paragraph.) By the definition of i_n,

$$\frac{f(x+t_nd) - f(x)}{t_n} \le \frac{f_{i_n}(x+t_nd) - f(x)}{t_n} + \frac{\epsilon_n}{t_n} \le \frac{f_{i_n}(x+t_nd) - f_{i_n}(x)}{t_n} + \frac{\epsilon_n}{t_n} \le f_{i_n}'(x;d) + \frac{\epsilon_n}{t_n} + \delta,$$

where $\delta > 0$ is arbitrary, and the last inequality is due to assumption 1.

The only thing left to prove is to show that $\{i_n\} \in \hat{\mathcal{I}}_x$. Indeed, for n large,

$$f(x) \ge f_{i_n}(x) \ge f_{i_n}(x+t_nd) - t_nf_{i_n}'(x) - t_n\delta.$$

But the latter part converges to $f(x)$ due to assumption 2 and the fact that $f(x+t_nd) \ge f_{i_n}(x+t_nd) \ge f(x+t_nd) - \epsilon_n$ while $f(x+t_nd) \to f(x)$ (see (9)).

Similar arguments as in the previous paragraph shows that the right-hand side of (8) is upper bounded by the right-hand side in (7). On the other hand,

$$\frac{f(x+td) - f(x)}{t} \le \frac{f_{i_t}(x+td) - f_{i_t}(x)}{t} = f_{i_t}'(x;d) + o(t),$$

due to assumption 1 and 3. Therefore (8) follows by sandwiching. ■

[6] Argue similarly as in the proof of Theorem 1 that $f(x+t_nd) < \infty$ for all $|t|$ sufficiently small.

Remark 3 *Had* $f := \inf_{i\in\mathcal{I}} f_i$*, we only need to change assumption 2 to* $\inf_{i\in\mathcal{I}} f_i'(x;d) > -\infty$*. Theorem 2 was proved first by [Bernhard and Rapaport, 1995] under some unnecessary assumptions. Our treatment here combines some idea presented in [Milgrom and Segal, 2002]. Note that Corollary 1 also follows from Theorem 2: the joint continuity of* $\nabla_x f(x,y)$ *over the compact set* $\mathcal{Y}$ *implies the equi-differentiability of* $f(x,y)$*, hence Theorem 2 holds, in particular, (8), the u.s.c. of the envelop function, and Proposition 1 yield the corollary. This argument appeared in [Milgrom and Segal, 2002] but seems to rely on the stronger compactness assumption of* $\mathcal{Y}$*.*

Corollary 2 *Suppose* $f_i : \mathcal{X} \mapsto \mathbb{R}, i \in \mathcal{I}, |\mathcal{I}| < \infty$ *all have directional derivative at* x *along direction* d*, then their pointwise supremum* $f := \max_{i\in\mathcal{I}} f_i$ *has directional derivative given by*

$$f'(x;d) = \max_{i\in\mathcal{I}_x} f_i'(x;d), \quad \textit{where} \quad \mathcal{I}_x = \{i \in \mathcal{I} : f_i(x) = f(x)\}.$$

3 Convex case

In this section we put an additional assumption on f_i, that is, they are all convex functions. Again, we start with an easy proposition.

Proposition 3 *Define* $f(x) := \sup_{i\in\mathcal{I}} f_i(x)$*. Then*[7]

$$\partial f(x) \supseteq \overline{\text{conv}}\left(\bigcup_{i\in\mathcal{I}_x} \partial f_i(x)\right), \quad \textit{where} \quad \mathcal{I}_x := \{i \in \mathcal{I} : f_i(x) = f(x)\}. \tag{10}$$

Proof: Take $i \in \mathcal{I}_x$, $g \in \partial f_i(x)$, then $\forall y \in \mathcal{X}$

$$f(y) \ge f_i(y) \ge f_i(x) + \langle y - x; g\rangle = f(x) + \langle y - x; g\rangle ,$$

hence $g \in \partial f(x)$. The proof is complete by noticing that the subdifferential is always convex and weak-* closed. ■

Remark 4 *Note that in Proposition 3, we do not require* f_i *to be convex. From a practical point of view, this easy proposition is enough for many purposes, for instance, when one needs a subgradient for* $f(x)$*. The reverse inclusion is more difficult but of theoretical value, as we shall see in an example.*

Theorem 3 *Define* $f(x) := \sup_{i\in\mathcal{I}} f_i(x)$ *where* f_i *are convex. Fix a point* $x \in \mathcal{X}$ *with* $f(x) \in \mathbb{R}$*. Suppose*

1. $\mathcal{I}$ *is countably compact;*
2. $\exists$ *a neighborhood* U *of* x *such that* $\forall y \in U$, $i \mapsto f_i(y)$ *is u.s.c.;*
3. $\forall i \in \mathcal{I}$*, the convex function* f_i *is u.s.c. at* x*;*

then the directional derivative of f *is given by the formula*

$$f'(x;d) = \sup_{i\in\mathcal{I}_x} f_i'(x;d), \quad \textit{where} \quad \mathcal{I}_x := \{i \in \mathcal{I} : f_i(x) = f(x)\}. \tag{11}$$

Moreover, if assumption 3 is strengthened to "continuous at x*", then*

$$\partial f(x) = \overline{\text{conv}}\left(\bigcup_{i\in\mathcal{I}_x} \partial f_i(x)\right). \tag{12}$$

Proof: The envelop f is apparently convex hence the existence of the directional derivative. By assumption 1 and 2, we know $\mathcal{I}_x$ is not empty. Since Proposition 1 remains true, we only need to prove

$$f'(x;d) \le \sup_{i\in\mathcal{I}_x} f_i'(x;d). \tag{13}$$

[7]The closure is always taken w.r.t. the weak-* topology on $\mathcal{X}^*$ induced by $\mathcal{X}$.

Fix $\epsilon > 0, 0 < t_n \downarrow 0$ and consider the set

$$\mathcal{I}_n := \left\{ i \in \mathcal{I} : \frac{f_i(x + t_n d) - f(x)}{t_n} \geq f'(x; d) - \epsilon \right\}.$$

Since $x + t_n d \in U$ eventually, we know $\mathcal{I}_n$ is a non-empty closed set (assumption 2). Also, since

$$t \mapsto \frac{f_i(x + td) - f(x)}{t_n} = \frac{f_i(x + td) - f_i(x)}{t} + \frac{f_i(x) - f(x)}{t}$$

is apparently nondecreasing (due to convexity of f_i), $\exists i_* \in \cap_n \mathcal{I}_n \neq \emptyset$ (assumption 1). Hence

$$\frac{f_{i_*}(x + t_n d) - f(x)}{t_n} \geq f'(x; d) - \epsilon.$$

Multiplying both sides by t_n and then letting $t_n \downarrow 0$ we obtain $i_* \in \mathcal{I}_x$ (assumption 3), hence the required inequality (13).

The lefthand side in (12) always contains the righthand side (cf. Proposition 3), while the other direction follows from the l.s.c. of $f'(x; \cdot)$ (being a pointwise supremum of continuous functions $f_i'(x; \cdot)$, whose continuity is guaranteed by the continuity of f_i at x). ■

Remark 5 *The above beautiful proof is taken from [Aubin, 1998], see also [Hiriart-Urruty and Lemaréchal, 1993]. Comparing Theorem 1 and 3 we see that the extra convexity assumption dispenses the assumptions on the directional derivatives, which justifies our separate treatment for the convex case.*

Not surprisingly, our next step is to trade the countable compactness assumption for "uniform continuity".

Theorem 4 *Define $f(x) := \sup_{i \in \mathcal{I}} f_i(x)$. Fix a point $x \in \mathcal{X}$ with $f(x) \in \mathbb{R}$. Suppose*

1. *The maps f_i are equi-directionally differentiable at x along any direction d;*
2. *$\exists$ a neighborhood U of x such that $\forall y \in U$, $i \mapsto f_i(y)$ is u.s.c.;*
3. *$\forall i \in \mathcal{I}$, the convex function f_i is u.s.c. at x;*

then the directional derivative of f is given by the formula

$$f'(x; d) = \sup_{i \in \mathcal{I}_x} f_i'(x; d), \quad \textit{where} \quad \mathcal{I}_x := \{i \in \mathcal{I} : f_i(x) = f(x)\}. \tag{14}$$

Moreover, if assumption 3 is strengthened to "continuous at x", then

$$\partial f(x) = \overline{\text{conv}} \left(\bigcup_{i \in \mathcal{I}_x} \partial f_i(x) \right). \tag{15}$$

Theorem 5 *Define $f(x) := \sup_{i \in \mathcal{I}} f_i(x)$. Fix a point $x \in \mathcal{X}$ with $f(x) \in \mathbb{R}$. Suppose*

1. *$\exists$ a neighborhood U of x such that $\forall y \in U$, $i \mapsto f_i(y)$ is u.s.c.;*
2. *$\forall i \in \mathcal{I}$, the convex function f_i is u.s.c. at x;*

then the directional derivative of f is given by the formula

$$f'(x; d) = \sup_{i \in \mathcal{I}_x} f_i'(x; d), \quad \textit{where} \quad \mathcal{I}_x := \{i \in \mathcal{I} : f_i(x) = f(x)\}. \tag{16}$$

Moreover, if assumption 2 is strengthened to "continuous at x", then

$$\partial f(x) = \overline{\text{conv}} \left(\bigcup_{i \in \mathcal{I}_x} \partial f_i(x) \right). \tag{17}$$

4 Minimax case

The next theorem about the "stability" of Nash equilibria is well-known in game theory. Nevertheless, we include a proof (for the sake of the writer, who has just started to learn game theory ☺).

Theorem 6 *Consider a game with k players, action spaces $\mathcal{X}^i$, and payoff functions $f_i(\cdot, p)$ where $p \in \mathcal{P}$ is a perturbation parameter. Suppose*

1. *$\forall i$, $\mathcal{X}^i$ is compact;*
2. *$\forall p \in \mathcal{P}$, the Nash equilibrium set $\Pi_{i=1}^k \mathcal{X}_p^i$ is nonempty;*
3. *$\forall i, f_i : \Pi_{i=1}^k \mathcal{X}^i \times \mathcal{P} \mapsto \mathbb{R}$ is jointly continuous;*

then the equilibrium correspondence $\varphi : p \twoheadrightarrow \Pi_{i=1}^k \mathcal{X}_p^i$ is upper hemicontinuous.

Proof: Take $(p_\alpha, x_\alpha) \in \mathrm{Gr}(\varphi)$ such that $p_\alpha \to \bar{p}$. Due to compactness, we can assume $x_\alpha \to \bar{x}$ (by passing to a subnet if necessary). The proof is complete once we show $\bar{x} \in \varphi(\bar{p})$.

Suppose not, then $\exists j, \exists \tilde{x}^j \in \mathcal{X}^j$ such that $f_j(\bar{x}^j, \bar{x}^{-j}, \bar{p}) < f_j(\tilde{x}^j, \bar{x}^{-j}, \bar{p})$. By assumption 2 we have $f_j(x_\alpha^j, x_\alpha^{-j}, p_\alpha) < f_j(\tilde{x}^j, x_\alpha^{-j}, p_\alpha)$, $\forall \alpha$ "large". This contradicts $(p_\alpha, x_\alpha) \in \mathrm{Gr}(\varphi)$. ■

Now we are ready for a wonderful theorem.

Theorem 7 (Milgrom-Segal) *Define $V(p) := \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} f(x, y, p)$ and fix a direction d. Suppose*

1. *$\mathcal{X}$ and $\mathcal{Y}$ are compact, $\mathcal{P}$ is an open subset of some real vector space;*
2. *$f : \mathcal{X} \times \mathcal{Y} \times \mathcal{P} \to \mathbb{R}$ and its (partial) directional derivative $f'_p(\cdot, \cdot, \cdot; d) : \mathcal{X} \times \mathcal{Y} \times \mathcal{P} \mapsto \mathbb{R}$ are continuous;*
3. *$\forall p \in \mathcal{P}$, the Nash equilibrium (saddle-point) $\mathcal{X}_p \times \mathcal{Y}_p \neq \emptyset$;*

then the value function V is continuous and its directional derivative (along direction d) satisfies

$$V'(p; d) = \min_{x \in \mathcal{X}_p} \max_{y \in \mathcal{Y}_p} f'_p(x, y, p; d) = \max_{y \in \mathcal{Y}_p} \min_{x \in \mathcal{X}_p} f'_p(x, y, p; d). \tag{18}$$

Proof: Under the stated assumptions, the value function satisfies

$$V(p) = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y, p) = \max_{x \in \mathcal{X}} \min_{y \in \mathcal{Y}} f(x, y, p)$$

and is continuous (due to Berge's maximum theorem).

Denote $(x_p, y_p) \in \mathcal{X}_p \times \mathcal{Y}_p$, by definition, for $t > 0$

$$\frac{f(x_{p+td}, y_p, p + td) - f(x_{p+td}, y_p, p)}{t} \leq \frac{V(p + td) - V(p)}{t} \leq \frac{f(x_p, y_{p+td}, p + td) - f(x_p, y_{p+td}, p)}{t}.$$

Apply the mean value theorem on both sides:

$$f'_p(x_{p+td}, y_p, p + r(t)d; d) \leq \frac{V(p + td) - V(p)}{t} \leq f'_p(x_p, y_{p+td}, p + s(t)d; d),$$

where $0 \leq r(t), s(t) \leq t$. Note that both $r(t)$ and $s(t)$ are (right) continuous at $t = 0$. Since $(x_p, y_p) \in \mathcal{X}_p \times \mathcal{Y}_p$ is arbitrary,

$$\max_{y \in \mathcal{Y}_p} f'_p(x_{p+td}, y, p + r(t)d; d) \leq \frac{V(p + td) - V(p)}{t} \leq \min_{x \in \mathcal{X}_p} f'_p(x, y_{p+td}, p + s(t)d; d).$$

The function $g(y, \tilde{p}) := \min_{x \in \mathcal{X}_p} f'_p(x, y, \tilde{p}; d)$ is continuous due to Berge's maximum theorem. Combining with Theorem 6, we know the function $h(\tilde{p}) := \max_{y \in \mathcal{Y}_{\tilde{p}}} g(y, \tilde{p})$ is upper semicontinuous. Hence

$$\limsup_{t \downarrow 0} \min_{x \in \mathcal{X}_p} f'_p(x, y_{p+td}, p + s(t)d; d) = \max_{y \in \mathcal{Y}_p} \min_{x \in \mathcal{X}_p} f'_p(x, y, p; d),$$

and similarly (by considering $-f'_p(x,y,p;d)$)

$$\liminf_{t\downarrow 0} \max_{y\in\mathcal{Y}_p} f'_p(x_{p+td}, y, p+r(t)d; d) = \min_{x\in\mathcal{X}_p}\max_{y\in\mathcal{Y}_p} f'_p(x,y,p;d).$$

The proof is complete by invoking the weak duality. ■

Remark 6 *When $\mathcal{X},\mathcal{Y}$ and $\mathcal{P}$ are all convex, f is (jointly) convex in (x,p), then the value function V is also convex. However,* (18) *does not provide us a formula for the subdifferential of V (unless $|\mathcal{Y}_p| = 1$). The strong duality appeared in* (18) *is surprising and probably hard to come up with (before seeing the proof). Assumption 3 is satisfied when the payoff function f is quasiconvex in x and quasiconcave in y for each $p \in \mathcal{P}$ (cf. Sion's minimax theorem).*

Next we present an application of Theorem 7, where we are interested in studying how the value function of the optimization problem

$$V(p) := \sup_{x\in\mathcal{X}:g(x,p)\geq 0} f(x,p) \tag{19}$$

behaves when the perturbation parameter p changes. The tool we use is the Lagrangian multipliers.

Corollary 3 *Fix a direction d. Suppose that*

1. *$\mathcal{X}$ is compact convex, $\mathcal{P}$ is comapct;*
2. *$f : \mathcal{X}\times\mathcal{P} \mapsto \mathbb{R}$ and $g : \mathcal{X}\times\mathcal{P} \mapsto \mathbb{R}^k$ are (jointly) continuous and concave in x for each $p \in \mathcal{P}$, $f'_p(\cdot,\cdot;d)$ and $g'(\cdot,\cdot;d)$ are (jointly) continuous;*
3. *$\exists \tilde{x} \in \mathcal{X}$ such that $\min_{p\in\mathcal{P}} g(\tilde{x},p) > 0$ (Slater's condition);*

then the value function (19) *admits directional derivative (along direction d), given by*

$$\forall p \in \text{int}\mathcal{P},\ V'(p;d) = \max_{x\in\mathcal{X}_p}\min_{y\in\mathcal{Y}_p} L'_p(x,y,p;d) = \min_{y\in\mathcal{Y}_p}\max_{x\in\mathcal{X}_p} L'_p(x,y,p;d), \tag{20}$$

where $L(x,y,p) := f(x,p) + \sum_{i=1}^{k} y_i g_i(x,p)$ is the Lagrangian and $\mathcal{Y}_p := \operatorname{Argmin}_{y\in\mathbb{R}^k_+}\left(\sup_{x\in\mathcal{X}} L(x,y,t)\right)$.

Proof: The theory of Lagrangian multipliers implies

$$V(p) = \sup_{x\in\mathcal{X}}\inf_{y\in\mathbb{R}^k_+} L(x,y,p).$$

All assumptions in Theorem 7 are met except the compactness of $\mathcal{Y}$. Fix $y_p \in \mathcal{Y}_p$, then

$$V(p) \geq L(\tilde{x}, y_p, p) \geq f(\tilde{x},p) + y^i_p g^i(\tilde{x},p),$$

hence $y^i_p \leq \sup_{p\in\mathcal{P}} \frac{V(p)-f(\tilde{x},p)}{g^i(\tilde{x},p)} < \infty$ due to continuity and compactness. Therefore $\mathcal{Y} = \mathbb{R}^k_+$ can be replaced by some compact set in $\mathbb{R}^k_+$. ■

It is also possible to discuss the absolute continuity of the value function. We record this result from [Milgrom and Segal, 2002] for completeness.

Proposition 4 *Suppose that $\forall x \in \mathcal{X}, \forall y \in \mathcal{Y}$, the map $t \mapsto f(x,y,p+td)$ is absolutely continuous on some interval $[a,b]$, that the saddle-point set $\mathcal{X}_t \times \mathcal{Y}_t \neq \emptyset$ a.e., and that $t \mapsto \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} f(x,y,p+td) \in L_1([a,b])$, then the value function $V(t) := \inf_{x\in\mathcal{X}}\sup_{y\in\mathcal{Y}} f(x,y,p+td)$ is absolutely continuous. If in addition, $\cup_{t\in[a,b]}\{t\}\times\mathcal{X}_t\times\mathcal{Y}_t$ has at most $\aleph_0$ many isolated points, $f'_p(x,y,p+td;d)$ is (separately) continuous in x and y, and the family $\{f(x,y,p+td\}_{(x,y)\in\mathcal{X}\times\mathcal{Y}}$ is equidifferentiable, then pick (arbitrarily!) $(x_t,y_t) \in \mathcal{X}_t \times \mathcal{Y}_t$,*

$$V(t) = V(a) + \int_a^t f'_p(x_s, y_s, p+sd; d)\mathrm{d}s. \tag{21}$$

Proof: The absolute continuity of V follows from repeated application of Proposition 2.

To prove (21), fix s and pick any $(x_s, y_s) \in \mathcal{X}_s \times \mathcal{Y}_s$, since there are only at most $\aleph_0$ many isolated points, we know (s, x_s, y_s) is a limit point a.s., hence $\exists$ distinct $(t_\alpha, x_{t_\alpha}, y_{t_\alpha}) \to (s, x_s, y_s)$. W.l.o.g., take $t_\alpha \geq s$, then

$$\frac{f(x_{t_\alpha}, y_s, p + t_\alpha d) - f(x_{t_\alpha}, y_s, p + sd)}{t_\alpha - s} \leq \frac{V(t_\alpha) - V(s)}{t_\alpha - s} \leq \frac{f(x_s, y_{t_\alpha}, p + t_\alpha d) - f(x_s, y_{t_\alpha}, p + sd)}{t_\alpha - s}.$$

Using equidifferentiability,

$$f'_p(x_{t_\alpha}, y_s, p + sd; d) + o(t_\alpha - s) \leq \frac{V(t_\alpha) - V(s)}{t_\alpha - s} \leq f'_p(x_s, y_{t_\alpha}, p + sd; d) + o(t_\alpha - s).$$

Applying the separate continuity in x and y then completes the proof. ∎

References

Jean-Pierre Aubin. *Optima and Equilibria: An Introduction to Nonlinear Analysis*. Springer, 2nd edition, 1998.

Pierre Bernhard and Alain Rapaport. On a theorem of danskin with an application to a theorem of von Neumann-Sion. *Nonlinear Analysis, Theory, Methods & Applications*, 24(8):1163–1181, 1995.

Ryszard Engelking. *General Topology*. Heldermann Verlag Berlin, revised and completed edition, 1989.

Leonard Gillman and Meyer Jerison. *Rings of Continuous Functions*. Springer, 1960.

Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex Analysis and Minimization Algorithms, Vol.1*. Springer-Verlag, 1993.

Paul Milgrom and Ilya Segal. Envelop theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601, 2002.

Yao-Liang Yu. Various notions of compactness. `http://webdocs.cs.ualberta.ca/~yaoliang/mynotes/compact.pdf`, 2012.