Affine Gateaux Differentials and the von Mises Statistical Calculus

Simone Cerreia-Vioglio, Fabio Maccheroni, Massimo Marinacci
Luigi Montrucchio, Lorenzo Stanca

March 1, 2024

Abstract

This paper presents a general study of one-dimensional differentiability for functionals on convex domains that are not necessarily open. The local approximation is carried out by affine functionals, rather than linear ones as in standard Gateaux differentiability. This affine notion of differentiability naturally arises in many applications and, here and there, it appeared in the literature. Our systematic analysis aims to give a general perspective on it.

Key words Affine Gateaux differentiability; Gateaux differentiability; influence functions; robust statistics; Envelope theorem; multi-utility representation.

1 Introduction

To study the asymptotic behavior of statistical functionals, Richard von Mises elaborated a notion of directional derivability for functionals defined over spaces of probability distributions.[1] Indeed, these domains have no interior points and so standard Gateaux differentiability is no longer adequate.

In this paper we build upon von Mises's idea by studying functionals defined over abstract convex sets. For a function $f : C \to \mathbb{R}$ defined on a convex set C, the starting point is the directional derivative

$$Df\,(x;y) = \lim_{t\downarrow 0} \frac{f\,((1-t)\,x + ty) - f\,(x)}{t}$$

at a point x of C along the direction y, where y is another point of C. We use the term weak affine differential when the function $Df(x;\cdot)$ is affine on C (Definition 6). For this basic notion it is already possible to prove a number of results. Yet, we reserve the term affine differential to the important special case when the affine functional $Df\,(x;\cdot)$ can be extended to the whole space, yielding a notion of affine gradient. To exemplify, the von Mises differential is just our affine differential for functionals whose domain is a space of distributions.

The early notion of differentiability introduced by von Mises has been used in statistics (see Hampel [14], Huber [15, pp. 34-40], and Fernholz [23]). A similar approach, again on spaces of distributions, was also used in risk theory (see Chew et al. [8]). More recently, a general formulation of Roy's identity in consumer theory has been established through affine derivatives by [7]. Summing up, affine differentiability naturally pops up in different applications. Our purpose is to provide a systematic analysis of this notion.

[1] See [25, Part II] as well as Reeds [19] and Fernholz [23].

Outline In Section 3 we introduce affine differentiability. The main result, Lemma 17, shows that weakly affine differentiable functions are hemidifferentiable, that is, differentiable over line segments. A mean value theorem, Theorem 18, is a consequence of this lemma. In Section 4 we discuss applications to optimization. In particular, we establish a Danskin-type envelope theorem, Theorem 31, for weakly affine differentiable functions.

Applications to statistics and economics are considered in Section 5. In risk theory, Theorem 34 provides a global perspective to the local expected utility analysis of Machina [17] (see also [6]). As an illustration, we compute the local utilities for the quadratic model of Chew et al. [9] and for the prospect theory model of Tversky and Kahneman [24]. In a final example we apply the envelope theorem of Section 4 to a Bayesian statistical problem.

Basic affine differentiation often proves to be too weak in applications. For this reason, in Section 6 we offer some stronger variants, non-trivial versions of the classical Hadamard and Frechet differentiation in normed vector spaces. We illustrate these notions with some applications in risk theory and statistics.

2 Preliminaries

Throughout $\langle X, X^* \rangle$ is a dual pair between two vector spaces X and X^*, with generic elements x and x^*. When X is normed, X^* is its topological dual, unless otherwise stated. The pairing map is denoted by $\langle x, x^* \rangle$, with $x \in X$ and $x^* \in X^*$. We refer to the $\sigma(X, X^*)$-topology as the weak topology of X, while we refer to the $\sigma(X^*, X)$-topology as the weak* topology of X^*. Let A be a subset of X. Its *annihilator*, denoted by $A^\perp$, is the set of all $x^* \in X^*$ that vanish on A, i.e., $\langle x, x^* \rangle = 0$ for all $x \in A$. Clearly, $A^\perp$ is a weak*-closed vector subspace of X^*.

Throughout C denotes a convex subset of X. A point $x \in C$ is an *algebraic interior point* of C if for every $y \in X$ there is $\varepsilon > 0$ such that $x + \varepsilon y \in C$. The *algebraic interior* of C is denoted by $\operatorname{cor} C$. A point $x \in C$ is an (*algebraic*) *relative interior point* of C if, for every $y \in \operatorname{aff} C$ there is $\varepsilon > 0$ such that $x + \varepsilon y \in C$. The *relative interior* of C is denoted by $\operatorname{ri} C$. For short, we call *internal* the elements of $\operatorname{ri} C$. Clearly, $\operatorname{cor} C \subseteq \operatorname{ri} C$.

A function $f : C \to \mathbb{R}$ is *affine* if

$$f(tx + (1-t)y) = tf(x) + (1-t)f(y)$$

for all $x, y \in C$ and all $t \in [0,1]$.

Definition 1 *An affine function $f : C \to \mathbb{R}$ is* extendable *when it admits a weakly continuous linear extension to the whole space X, that is, when there exist $x^* \in X^*$ and $\gamma \in \mathbb{R}$ such that $f(\cdot) = \langle \cdot, x^* \rangle + \gamma$.*

As well-known, affine functionals on C are not always extendable, even when weakly continuous (cf. Example 13).

Let Y be a topological space, typically assumed to be metrizable. The set of all finite signed Borel measures is denoted by $ca(Y)$. The subset $\Delta(Y)$ of $ca(Y)$ consists of all Borel probability measures. We denote by $C_b(Y)$ the space of all bounded and continuous functions on Y.

The extension of an affine function may be not unique (when exists).

Proposition 2 *Let $f(\cdot) = \langle \cdot, x^* \rangle + \gamma$ on X, with $\gamma \in \mathbb{R}$. An affine functional $g : X \to \mathbb{R}$ agrees with f on C if and only if $g(\cdot) = \langle \cdot, y^* \rangle + \gamma_{y^*}$, where $y^* \in x^* + (C - C)^\perp$ and $\gamma_{y^*} = f(\bar{x}) - \langle \bar{x}, y^* \rangle$, with $\bar{x} \in C$ arbitrarily fixed.*

Proof Let $g(\cdot) = \langle \cdot, y^* \rangle + \gamma_1$ be an affine functional on X such that, for each $x \in C$, $f(x) = g(x)$. This implies that $\langle x, x^* - y^* \rangle$ is constant when x runs over C. That is, $\langle x - y, y^* - x^* \rangle = 0$ when x and y are distinct points of C. Hence, $y^* - x^* \in (C - C)^\perp$. Equivalently, $y^* \in x^* + (C - C)^\perp$. From $f(\bar{x}) = \langle \bar{x}, y^* \rangle + \gamma_1$ it follows that $\gamma_1 = f(\bar{x}) - \langle \bar{x}, y^* \rangle$.

To prove the converse, let $g(\cdot) = \langle \cdot, y^* \rangle + \gamma_{y^*}$ be with y^* and γ_{y^*} given above. Tedious algebra shows that

$$g(x) = \langle x - \bar{x} + \bar{x}, y^* - x^* + x^* \rangle + f(\bar{x}) - \langle \bar{x}, y^* \rangle = \langle x, x^* \rangle + \gamma = f(x)$$

as desired. ■

Thus, the representation of an affine functional on C is unique if and only if $(C - C)^\perp = \{0\}$. Since

$$(C - C)^\perp = (C - x)^\perp \tag{1}$$

holds for any arbitrary point $x \in C$, we have $(C - C)^\perp = \{0\}$ when $\operatorname{cor} C \neq \emptyset$.

Uniquely extendable affine functionals can be characterized in finite-dimensional spaces, as the next known result shows.

Proposition 3 *An affine functional $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ is extendable. The extension is unique if and only if* $\operatorname{int} C \neq \emptyset$.

In normed vector spaces a less general result holds.

Proposition 4 *An affine functional $f : C \to \mathbb{R}$ defined on a convex set C of a normed vector space X is uniquely extendable if* $\operatorname{int} C \neq \emptyset$. *It is weakly continuous when f is locally bounded at some* $\operatorname{int} C$.

3 Affine differentiability

3.1 Differential

We begin with the protagonist of our analysis.

Definition 5 *The* affine directional derivative *of a functional $f : C \to \mathbb{R}$ at a point $x \in C$ along the direction $y \in C$ is given by*

$$Df(x; y) = \lim_{t \downarrow 0} \frac{f((1 - t)x + ty) - f(x)}{t} \tag{2}$$

when this limit exists finite.

When this limit exists finite for all $y \in C$, the map $Df(x; \cdot) : C \to \mathbb{R}$ is well defined. It clearly satisfies the following properties:

(i) $Df(x; x) = 0$;

(ii) $Df(x; \cdot)$ is homogeneous, i.e.,

$$Df(x; (1 - \alpha)x + \alpha y) = \alpha Df(x; y) \tag{3}$$

for all $y \in C$ and all $\alpha \in [0, 1]$.

The map $Df(x;\cdot)$ is, in general, not affine (see Example 7 below), a failure that motivates the following taxonomy.[2]

Definition 6 *Let $x \in X$. A functional $f : C \to \mathbb{R}$ is:*

(i) weakly affinely differentiable *at x if $Df(x;\cdot) : C \to \mathbb{R}$ is an affine functional;*

(ii) affinely differentiable *at x if $Df(x;\cdot) : C \to \mathbb{R}$ is an extendable affine functional.*

Weak affine differentiability (for short, wa-differentiability) does not require any vector topology, which is instead needed for affine differentiability (for short, a-differentiability). We call the map $Df(x;\cdot) : C \to \mathbb{R}$ the *wa-differential* of f at x when it is affine. When $Df(x;\cdot)$ is extendable, we call it the *a-differential* of f at x; in this case, there is a pair $(x^*, \gamma) \in X^* \times \mathbb{R}$ such that

$$Df(x;\cdot) = \langle \cdot, x^* \rangle + \gamma \tag{4}$$

This pair is not unique unless $(C - C)^{\perp} = \{0\}$. Since $Df(x;x) = 0$, it follows that $\gamma = -\langle x, x^* \rangle$ and so the affine differential admits the intrinsic representation

$$Df(x;y) = \langle y - x, x^* \rangle \tag{5}$$

where the inessential scalar γ has been dropped. In light of (1), equation (5) is independent of the element x^* chosen. With this, we call *gradient* of f at x any such equivalent x^*. As it is unique up to elements in $(C - C)^{\perp}$, we have an equivalence class

$$[\nabla_a f(x)] = x^* + (C - C)^{\perp} = x^* + (C - x)^{\perp} \in X^* / (C - C)^{\perp} \tag{6}$$

which is a weak*-closed affine space of X^*. Here $\nabla_a f(x)$ is a representative element of this equivalent class, so it stands for any element x^* of this equivalence class.

Example 7 Define $f : \mathbb{R}^2_+ \to \mathbb{R}$ by $f(x_1, x_2) = x_1^{\alpha} x_2^{\beta}$, with $\alpha, \beta > 0$. When $\alpha + \beta < 1$, the directional derivative at the origin $\mathbf{0} = (0,0)$ does not exist. Instead, it exists when $\alpha + \beta = 1$, with $Df(\mathbf{0};y) = f(y)$. But, f is not wa-differentiable at $\mathbf{0}$ since $Df(\mathbf{0};\cdot)$ is not affine. It is easy to see that f is a-differentiable at $\mathbf{0}$ when $\alpha + \beta > 1$ and $\nabla_a f(\mathbf{0}) = \{\mathbf{0}\}$ is its unique gradient. ▲

It is desirable to have criteria ensuring a-differentiability. The following conditions are a direct consequence of Propositions 3 and 4.

Proposition 8 *Let $f : C \to \mathbb{R}$ be wa-differentiable at a point $x \in C$. Then, f is a-differentiable at x_0 if either $X = \mathbb{R}^n$ or X is normed, $x \in \operatorname{int} C$ and f is locally Lipschitz at x.*

Proof The case $X = \mathbb{R}^n$ follows from Proposition 3. When X is normed, the local Lipschitz condition at $x \in \operatorname{int} C$ implies that $|Df(x;y)| \leq L \|y - x\|$ and so that $Df(x;\cdot)$ is locally bounded at x. Therefore, Proposition 4 yields the desired result. ■

When $f : C \to \mathbb{R}$ is Gateaux differentiable at $x \in C$, the *Gateaux gradient* $\nabla_G f(x) \in X^*$ is given by

$$\lim_{t \to 0^+} \frac{f(x + ty) - f(x)}{t} = \langle y, \nabla_G f(x) \rangle \qquad \forall y \in X$$

The next result shows that affine and Gateaux differentiability are equivalent only on $\operatorname{cor} C$, a set that in most relevant cases is empty.

[2] Proposition 57 in Appendix provides conditions under which $Df(x;\cdot)$ is affine.

Proposition 9 *The function $f : C \to \mathbb{R}$ is a-differentiable at $x \in \operatorname{cor} C$ if and only if it is Gateaux differentiable at x. In this case, $\nabla_a f(x) = \nabla_G f(x)$.*

Here the affine gradient inherits the uniqueness of the Gateaux one.

Proof Let f be Gateaux differentiable at $x \in \operatorname{cor} C$ and let $x^* = \nabla_G f(x)$. For each $y \in C$,

$$\langle y - x, x^* \rangle = \lim_{t \to 0^+} \frac{f(x + t(y - x)) - f(x)}{t} = Df(x; y)$$

Hence, $x^* \in [\nabla_a f(x)]$. As $x \in \operatorname{cor} C$, the set $[\nabla_a f(x)]$ is a singleton and so we can write $x^* = \nabla_a f(x)$. Conversely, let f be a-differentiable at x, with $x^* \in [\nabla_a f(x)]$. Take $y \in X$. It follows that $x + t_0 y \in C$ for some $t_0 > 0$ small enough. It holds

$$t_0 \langle v, x^* \rangle = \langle t_0 v, x^* \rangle = Df(x; x + t_0 v) = \lim_{t \to 0^+} \frac{f((1-t)x + t(x + t_0 v))}{t} = t_0 \langle v, \nabla_G f(x) \rangle$$

Hence, $x^* = \nabla_G f(x)$. This proves that $[\nabla_a f(x)] = \{\nabla_G f(x)\}$. ■

Observe that when f has an extension $\tilde{f} : X \to \mathbb{R}$ which is Gateaux differentiable on C, it is a-differentiable on C and

$$\nabla_a f(x) = \nabla_G \tilde{f}(x) + (C - C)^{\perp}$$

for all $x \in C$.

3.2 Examples

We present a few examples to illustrate the concepts introduced so far.

Example 10 Given an interval $I = [a, b]$, let $NBV(I)$ be the vector space of the functions $F : I \to \mathbb{R}$ of bounded variation that are right continuous on (a, b) and normalized, i.e., $F(a) = 0$. We consider the standard duality between $NBV(I)$ and the space $C(I)$ of continuous function $u : I \to \mathbb{R}$, with pairing defined by $\langle F, u \rangle = \int_a^b u \mathrm{d}F$ where the integral is Riemann-Stieltjes.

If a functional $T : C \to \mathbb{R}$ defined on a convex subset of $NBV(I)$ is a-differentiable at F, there is a function $u_F \in C(I)$ such that

$$DT(F; G) = \int_a^b u_F \mathrm{d}(G - F) \qquad \forall G \in C$$

For instance, consider the convex set $\mathcal{D} = \mathcal{D}[a, b]$ of all probability distributions on the interval (i.e., of all decreasing $F \in NBV(I)$ with $\int_a^b \mathrm{d}F = 1$). In this case,

$$(\mathcal{D} - \mathcal{D})^{\perp} = \operatorname{span}\left\{1_{[a,b]}\right\}$$

and so the gradient $u_F = \nabla_a T(F)$ is unique up to an additive constant. ▲

Example 11 Replace I with a metric space Y and consider the dual pair $\langle ca(Y), C_b(Y) \rangle$. A functional $T : \Delta(Y) \to \mathbb{R}$ is affinely differentiable at $\mu \in \Delta(Y)$ if there is $u_\mu \in C_b(Y)$ such that

$$DT(\mu; \lambda) = \int_X u_\mu \mathrm{d}(\lambda - \mu) \qquad \forall \mu \in \Delta(Y)$$

The gradient $u_\mu \in C_b(Y)$ is unique up to an additive constant. Indeed, if $u \in (\Delta(Y) - \Delta(Y))^{\perp}$, then $\langle u, \lambda - \mu \rangle = 0$ for all $\lambda, \mu \in \Delta(Y)$. By setting $\lambda = \delta_x$ and $\mu = \delta_y$, we then get $u(x) = u(y)$ for all $x, y \in Y$. Hence, u is constant. ▲

Example 12 Let S be a metric space and A a convex subset of $\mathbb{R}^n$. Denote by C the convex set of all norm-bounded (i.e., $\sup_S \|u\|_n < \infty$) and continuous maps $u : S \to A$. Given $\mu \in \Delta(S)$ and $F : S \times A \to \mathbb{R}$, define $I : C \to \mathbb{R}$ by

$$I(u) = \int F(s, u(s)) \mathrm{d}\mu(s)$$

Assume the following conditions:

(i) $F(\cdot, x)$ is Borel measurable for every $x \in A$;

(ii) $F(s, \cdot)$ is wa-differentiable for every $s \in S$, with differential $DF(s, x; \cdot)$;

(iii) there exists $\kappa : S \to \mathbb{R}$, with $\int \kappa \mathrm{d}\mu < \infty$, such that

$$|F(\cdot, x) - F(\cdot, y)| \leq \kappa(\cdot) \|x - y\|_n \qquad \forall x, y \in A$$

Under these conditions, I is well defined and wa-differentiable at each $u \in C$, with

$$DI(u; g) = \int DF(s, u(s); g(s)) d\mu(s) \qquad \forall g \in C$$

Indeed, by (iii) we have

$$\frac{1}{t}|F(s, (1-t)u(s) + tg(s)) - F(s, u(s))| \leq L(s) \|g(s) - u(s)\|_n \qquad \forall s \in S$$

The desired result thus follows from the Dominated Convergence Theorem. ▲

Inspired by an example in Phelps [18], next we present a wa-differentiable function which is not a-differentiable.

Example 13 Let $C = \left\{\{x_n\} \in \mathbb{R}^{\mathbb{N}} : \forall n, |x_n| \leq a_n\right\}$, where $\{a_n\}$ is a given scalar sequence with $0 < a_n < 1$ and $\sum_n a_n < \infty$. Clearly, C is a closed and bounded convex subset of ℓ^2. The convex function $f : C \to \mathbb{R}$ defined by

$$f(x) = \left(\sum_n x_n\right)^2$$

is wa-differentiable at each $x \in C$, with

$$Df(x; y) = \left(2 \sum_n x_n\right) \cdot \sum_n y_n - 2f(x) \qquad \forall y \in C$$

Clearly, $Df(x; \cdot)$ is affine on C. But, f is not a-differentiable at any $x \in C$ since there is no element u in ℓ^2 such that $(u, y) = \sum_n y_n$ for all $y \in C$. Indeed, for each m consider the point $y_m = (0, 0, .., a_m, ...) \in C$. It follows that $u_m a_m = (u, y_m) = a_m$, namely, $u_m = 1$ for all n. But then $u \notin \ell^2$.

The map $y \mapsto \sum_n y_n$ is continuous. Indeed, if the sequence $\{y_n^m\}$ converges to $\{y_n\}$ as $m \to \infty$, then it converges pointwise, i.e., $y_n^m \to y_n$ for every n. By the Dominated Convergence Theorem, $\sum_n y_n^m \to \sum_n y_n$. Hence, $Df(x; \cdot)$ is weakly continuous. Finally, observe that $\operatorname{cor} C$ is empty, something not surprising in light of Proposition 4. ▲

A function $B : C \times C \to \mathbb{R}$ is a *biaffine form* when it is affine in each argument. It is *extendable* when it can be extended to a bilinear form on $X \times X$. An example of a biaffine form is the map $B : C \times C \to \mathbb{R}$ defined by

$$B(x,y)=\left(\sum_n x_n\right)\cdot\left(\sum_n y_n\right)$$

where C is the convex subset of ℓ^2 in Example 13. As already seen, this form is not extendable.

A biaffine form B is *symmetric* when $B(x,y)=B(y,x)$ for all $x,y\in C$. The symmetrization of a form B is given by

$$B_S(x,y)=\frac{1}{2}\left[B(x,y)+B(y,x)\right]$$

Associated with a biaffine form B there is the (*affine*) *quadratic form* $Q:C\to\mathbb{R}$ given by

$$Q(x)=B(x,x)=B_S(x,x)$$

The wa-differentiability of biaffine forms and of quadratic forms is easily checked, with wa-derivatives

$$DB(x_1,x_2;y_1,y_2)=B(x_1,y_2)+B(y_1,x_2)-2B(x_1,x_2) \tag{7}$$

and

$$DQ(x;y)=2B_S(x,y)-Q(x)$$

When $DQ(x;\cdot)$ is extendable we get the gradient $\nabla_a Q(x)=2B_S(x,\cdot)$.

Example 14 The *Mann-Whitney* biaffine form $B:\Delta(\mathbb{R})\times\Delta(\mathbb{R})\to\mathbb{R}$ is given by

$$B(\mu,\lambda)=\int F_\mu(t)\,\mathrm{d}F_\lambda(t)=\int_{\mathbb{R}}F_\mu\mathrm{d}\lambda$$

Here F_μ is the cumulative distribution function associated with $\mu\in\Delta(\mathbb{R})$ and $\int F_\mu dF_\lambda$ is a Lebesgue-Stieltjes integral. This biaffine form is used in statistics (see [13]). By (7),

$$DB(\mu,\lambda;\mu_1,\lambda_1)=\int(F_{\mu_1}-F_\mu)\,\mathrm{d}\lambda+\int F_\mu\mathrm{d}(\lambda_1-\lambda)=\int F_{\mu_1}\mathrm{d}\lambda+\int F_\mu\mathrm{d}\lambda_1+\gamma$$

where γ is a scalar independent of μ_1 and λ_1. This wa-differential is not always extendable. It is, however, extendable when F_μ and F_λ are continuos with bounded support, say contained in an interval $[a,b]$. Indeed,

$$\int F_{\mu_1}\mathrm{d}\lambda+\int F_\mu\mathrm{d}\lambda_1=\int F_{\mu_1}\mathrm{d}F_\lambda+\int F_\mu\mathrm{d}F_{\lambda_1}=\int_a^b F_{\mu_1}\mathrm{d}F_\lambda+\int F_\mu\mathrm{d}F_{\lambda_1}=1-\int F_\lambda\mathrm{d}F_{\mu_1}+\int F_\mu\mathrm{d}F_{\lambda_1}$$

where, by the continuity of the distribution functions, the integral $\int_a^b F_{\mu_1}dF_\lambda$ is Riemann-Stieltjes.[3] Therefore, B is a-differentiable at (μ,λ), with gradient

$$\nabla_a B(\mu,\lambda)=(-F_\lambda,F_\mu)\in C_b(\mathbb{R})\times C_b(\mathbb{R})$$

▲

Example 15 A quadratic functional on $\Delta(Y)$ is

$$Q(\mu)=\int_{Y\times Y}\psi(x,y)\,\mathrm{d}\mu(x)\otimes\mathrm{d}\mu(y) \tag{8}$$

where $\psi\in C_b(Y\times Y)$. This functional is a-differentiable, with

$$\nabla_a Q(\mu)=\int_Y\left[\psi(\cdot,y)+\psi(y,\cdot)\right]\mathrm{d}\mu(y)$$

When Y is a compact interval and the kernel ψ is symmetric, this functional has been studied by [9]. It will be further discussed in Example 35. ▲

[3] In the last equality we used integration by parts (see, e.g., Theorem 14.10 of [4]).

3.3 Mean value theorem

Let $x, y \in C$ and, for each $t \in [0,1]$, set

$$x_t = (1-t)\,x + ty$$

Each function $f : C \to \mathbb{R}$ has a scalar auxiliary function $\varphi_{x,y} : [0,1] \to \mathbb{R}$ defined by $\varphi_{x,y}(t) = f(x_t)$.

Definition 16 *A function $f : C \to \mathbb{R}$ is* hemidifferentiable *if, for all $x, y \in C$, its auxiliary function $\varphi_{x,y} : [0,1] \to \mathbb{R}$ is differentiable on $[0,1]$.*[4]

We begin with a non-trivial lemma. To ease notation, when no confusion may arise we often omit subscripts and just write φ.

Lemma 17 *If $f : C \to \mathbb{R}$ is wa-differentiable,*[5] *then it is hemidifferentiable, with*

$$\varphi'(t) = \frac{1}{1-t} Df(x_t; y) = -\frac{1}{t} Df(x_t; x) \qquad \forall t \in (0,1) \tag{9}$$

and

$$\varphi'_+(0) = Df(x; y) \quad ; \quad \varphi'_-(1) = -Df(y; x)$$

As $Df(x_t; \cdot)$ is affine, (9) is equivalent to

$$\varphi'(t) = \frac{1}{\tau - t} Df(x_t; x_\tau) \tag{10}$$

for all $(t, \tau) \in (0,1) \times [0,1]$ with $\tau \neq t$. With $\tau = 0, 1$ we get relations (9).

When f is a-differentiable on C, from (5) it follows immediately that, for each $t \in (0,1)$,[6]

$$\varphi'(t) = \langle y - x, \nabla_a f(x_t) \rangle \tag{11}$$

Proof For each $t \in [0,1]$ we have the following obvious algebraic relations

$$y - x_t = (1-t)(y-x) \quad \text{and} \quad x - x_t = t(x-y)$$

Now fix $x, y \in C$ and $t \in [0,1)$. The limit

$$\lim_{h \downarrow 0} \frac{\varphi(t + (1-t)h) - \varphi(t)}{h}$$

exists. Indeed,

$$\frac{\varphi(t + (1-t)h) - \varphi(t)}{h} = \frac{f(x_t + h(1-t)(y-x)) - f(x_t)}{h} = \frac{f(x_t + h(y - x_t)) - f(x_t)}{h}$$

Hence,

$$\lim_{h \downarrow 0} \frac{\varphi(t + (1-t)h) - \varphi(t)}{h} = Df(x_t; y)$$

We also have

$$\lim_{h \downarrow 0} \frac{\varphi(t + (1-t)h) - \varphi(t)}{h} = (1-t) \lim_{h \downarrow 0} \frac{\varphi(t + (1-t)h) - \varphi(t)}{(1-t)h} = (1-t)\,\varphi'_+(t)$$

[4] Differentiable on $[0,1]$ means right-differentiable at $t = 0$ and left-differentiable at $t = 1$.

[5] A function is "wa-differentiable" ("a-differentiable") when wa-differentiable (a-differentiable) at all points of its domain.

[6] Recall that $\nabla_a f(x_t)$ that is a representative of the equivalence class $[\nabla_a f(x_t)]$, so (11) means $\varphi'(t) = \langle y - x, x^* \rangle$ for all $x^* \in [\nabla_a f(x_t)]$.

Hence,

$$\varphi'_+(t) = \frac{1}{1-t} Df(x_t; y)$$

Use the same method for the left derivative φ'_-. Specifically, begin with the limit

$$\begin{aligned}\lim_{h\uparrow 0} \frac{\varphi(t+th)-\varphi(t)}{h} &= -\lim_{k\downarrow 0} \frac{\varphi(t-tk)-\varphi(t)}{k} = -\lim_{k\downarrow 0} \frac{f(x_t + kt(x-y)) - f(x_t)}{k} \\ &= -\lim_{k\downarrow 0} \frac{f(x_t + k(x-x_t)) - f(x_t)}{k} = -Df(x_t; x)\end{aligned}$$

At the same time, we have

$$\varphi'_-(t) = \lim_{h\uparrow 0} \frac{\varphi(t+th)-\varphi(t)}{th} = \frac{1}{t} \lim_{h\uparrow 0} \frac{\varphi(t+th)-\varphi(t)}{h} = -\frac{1}{t} Df(x_t; x)$$

Since $Df(x_t; \cdot)$ is affine, we obtain

$$0 = Df(x_t; x_t) = Df(x_t; (1-t)x + ty) = (1-t) Df(x_t; x) + tDf(x_t; y)$$

which implies

$$\frac{1}{1-t} Df(x_t; y) = -\frac{1}{t} Df(x_t; x) = \varphi'_-(t) = \varphi'_+(t)$$

Hence, $\varphi'_+ = \varphi'_-$ on $(0,1)$ and the proof is complete. ■

It is apparent from the proof of this lemma that when f is not wa-differentiable, but has one-sided directional derivatives, we can still infer that

$$\varphi'_+(t) = \frac{1}{1-t} Df(x_t; y) \quad \text{and} \quad \varphi'_-(t) = -\frac{1}{t} Df(x_t; x)$$

A remarkable consequence of the last lemma is a mean value theorem.

Theorem 18 (Mean Value Theorem) *Let $f : C \to \mathbb{R}$ be wa-differentiable. For each $x, y \in C$ there exists $t \in (0,1)$ such that*

$$f(y) - f(x) = \frac{1}{1-t} Df(x_t; y) \tag{12}$$

When f is a-differentiable, we get

$$f(y) - f(x) = \langle y - x, \nabla_a f(x_t) \rangle$$

Proof Let $x, y \in C$. The auxiliary function $\varphi(t) = f(x_t)$ is continuous on $[0,1]$ and differentiable on $(0,1)$. By the basic Mean Value Theorem, there exists $t \in (0,1)$ such that

$$\varphi'(t) = \frac{\varphi(1) - \varphi(0)}{1-0} = \varphi(1) - \varphi(0)$$

By Lemma 17,

$$\frac{1}{1-t} Df(x_t; y) = f(y) - f(x)$$

as desired. ■

3.4 Affine calculus

To develop an effective affine calculus we need a slightly stronger notion of affine differentiability.

Definition 19 *A wa-differentiable function $f : C \to \mathbb{R}$ is (*radially*) *continuously wa-differentiable* if, for each $x, y \in C$, the functions*

$$t \mapsto Df(x_t; y) \quad \text{and} \quad t \mapsto Df(x_t; x) \tag{13}$$

are both continuous on $[0,1]$.

For instance, it is easy to check that all biaffine forms as well as all quadratic functionals are radially continuously wa-differentiable.

By (9), we have

$$Df(x_t; x) = -\frac{t}{1-t} Df(x_t; y)$$

As a consequence, in (13) the continuity of $t \mapsto Df(x_t; y)$ implies that of $t \mapsto Df(x_t; x)$ at all $t \in [0,1)$. Only at $t = 1$, the continuity of $t \mapsto Df(x_t; x)$ is a genuine assumption.

Proposition 20 *If $f : C \to \mathbb{R}$ is continuously wa-differentiable, then*

$$f(y) - f(x) = \int_0^1 \frac{1}{1-t} Df(x_t; y)\,\mathrm{d}t$$

for all $x, y \in C$.

When f is a-differentiable, we get

$$f(y) - f(x) = \int_0^1 \frac{1}{1-t} \langle y - x, \nabla_a f(x_t) \rangle \,\mathrm{d}t$$

Proof Consider the auxiliary function $\varphi(t) = f(x_t)$. By Lemma 17, $\varphi'(t)$ exists for all $t \in (0,1)$. Moreover, φ' is continuous on $(0,1)$ by the relation $\varphi'(t) = (1-t)^{-1} Df(x_t; y)$ and by the continuity of the first map in (13). On the other hand, by the continuity of first map in (13)

$$\lim_{t \to 0^+} \varphi'(t) = \lim_{t \to 0^+} (1-t)^{-1} Df(x_t; y) = Df(x; y) = \varphi'_+(0)$$

Analogously, $\lim_{t \to 1-} \varphi'(t) = \varphi'_-(1)$. As a result, φ' is continuous on $(0,1)$ and bounded on $[0,1]$. Hence,

$$f(y) - f(x) = \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)\,\mathrm{d}t = \int_0^1 \frac{1}{1-t} Df(x_t; y)\,\mathrm{d}t$$

as desired. ■

Example 21 According to Example 11, when $T : \Delta(Y) \to \mathbb{R}$ is a-differentiable by the Mean Value Theorem (Theorem 18) there is $t \in (0,1)$ such that

$$T(\lambda) - T(\mu) = \frac{1}{1-t} \int_Y u_{\mu_t} \mathrm{d}(\lambda - \mu)$$

By Proposition 20, we then have

$$T(\lambda) - T(\mu) = \int_0^1 \left(\frac{1}{1-t} \int_Y u_{\mu_t} \mathrm{d}(\lambda - \mu) \right) \mathrm{d}t$$

when T is radially continuously a-differentiable. ▲

To characterize convexity for wa-differentiable functions we need the following monotonicity notion.

Definition 22 *A wa-differentiable function $f : C \to \mathbb{R}$ has a* monotone *wa-differential if*

$$Df(x;y) + Df(y;x) \leq 0 \tag{14}$$

for all $x, y \in C$.

This definition has a sharper form for a-differentiable functions because (14) becomes

$$\langle y - x, \nabla_a f(y) - \nabla_a f(x) \rangle \geq 0$$

With this, we can characterize convexity for wa-differentiable functions.

Proposition 23 *Let $f : C \to \mathbb{R}$ be wa-differentiable. The following properties are equivalent:*

(i) f is convex;

(ii) for all $x, y \in C$,

$$f(y) \geq f(x) + Df(x;y) \tag{15}$$

(iii) the wa-differential of f is monotone.

The convexity of a quadratic functional $Q : C \to \mathbb{R}$ is a simple illustration of this result. In view of (14), we obtain the condition

$$Q(x) + Q(y) - 2B_S(x,y) \geq 0 \qquad \forall x, y \in C$$

When the biaffine form is extendable, it becomes the familiar condition $Q(x-y) \geq 0$ of semidefinite positivity for quadratic functionals.

Proof (i) implies (ii). By the Jensen inequality,

$$Df(x;y) = \lim_{t \downarrow 0} \frac{f((1-t)x + ty) - f(x)}{t} \leq \lim_{t \downarrow 0} \frac{t(f(y) - f(x))}{t} = f(y) - f(x)$$

as desired.

(ii) implies (iii). By adding up $f(y) \geq f(x) + Df(x;y)$ and $f(y) \geq f(x) + Df(x;y)$, we get $0 \geq Df(x;y) + Df(x;y)$.

(iii) implies (i). Consider the auxiliary function $\varphi(t) = f(x_t)$ on $[0,1]$ for arbitrary points $x, y \in C$. Let

$$x_{t_1} = (1-t_1)x + t_1 y \quad \text{and} \quad x_{t_2} = (1-t_2)x + t_2 y$$

with $0 < t_1 < t_2 < 1$. It is easy to check that

$$x_{t_2} = \frac{1-t_2}{1-t_1} x_{t_1} + \frac{t_2 - t_1}{1-t_1} y \quad \text{and} \quad x_{t_1} = \frac{t_1}{t_2} x_{t_2} + \frac{t_2 - t_1}{t_2} x$$

By (3),

$$\begin{aligned} Df(x_{t_2}; x_{t_1}) &= Df\left(x_{t_2}; \frac{t_1}{t_2} x_{t_2} + \frac{t_2 - t_1}{t_2} x\right) = \frac{t_2 - t_1}{t_2} Df(x_{t_2}; x) \\ Df(x_{t_1}; x_{t_2}) &= Df\left(x_{t_1}; \frac{1-t_2}{1-t_1} x_{t_1} + \frac{t_2 - t_1}{1-t_1} y\right) = \frac{t_2 - t_1}{1-t_1} Df(x_{t_1}; y) \end{aligned}$$

By adding up,

$$Df\left(x_{t_2};x_{t_1}\right)+Df\left(x_{t_1};x_{t_2}\right)=\left(t_2-t_1\right)\left[\frac{1}{t_2}Df\left(x_{t_2};x\right)+\frac{1}{1-t_1}Df\left(x_{t_1};y\right)\right]$$

By (9),

$$Df\left(x_{t_2};x_{t_1}\right)+Df\left(x_{t_1};x_{t_2}\right)=\left(t_2-t_1\right)\left[\varphi'\left(t_1\right)-\varphi'\left(t_2\right)\right]\leq 0$$

The first derivative φ' is thus nondecreasing, so φ is convex. Since this is true for any pair of points $x, y \in C$, we conclude that f is convex. ■

Thanks to the previous results we can also relate gradients and subdifferentials for convex functionals. The *subdifferential* of a convex function $f : C \to \mathbb{R}$ at $x \in C$ is the set

$$\partial f(x)=\left\{x^*\in X^*:\forall y\in C, f(y)\geq f\left(x\right)+\left\langle y-x,x^*\right\rangle\right\}$$

while the (*negative conical*) *polar* $A^- \subseteq X^*$ of a set A in X is given by

$$A^-=\left\{x^*\in X^*:\forall x\in A,\left\langle x,x^*\right\rangle\leq 0\right\}$$

For a convex set C, the *normal cone* at a point $x \in C$ is defined as $N_C\left(x\right) = \left(C - x\right)^-$.

Proposition 24 *Let $f : C \to \mathbb{R}$ be a-differentiable and convex. For each $x \in C$,*

$$\partial f\left(x\right)=x^*+N_C\left(x\right) \tag{16}$$

for all $x^* \in \left[\nabla_a f\left(x\right)\right]$. Moreover, for each $x \in \operatorname{ri} C$,

$$\partial f\left(x\right)=\left\{\nabla_a f\left(x\right)\right\}$$

At non-internal points of C the subdifferential of a convex function can thus be strictly larger than its a-differential because the vector space $\left(C - x\right)^\perp$ can be strictly included in the normal cone $N_C\left(x\right)$.

Proof Let $x^* \in \left[\nabla_a f\left(x\right)\right]$. By (15),

$$f\left(y\right)\geq f\left(x\right)+Df\left(x;y\right)=f\left(x\right)+\left\langle y-x,x^*\right\rangle \qquad \forall y\in C$$

For an arbitrary element p^* of $\left(C - x\right)^-$, we have $\left\langle y - x, p^*\right\rangle \leq 0$. Consequently,

$$f\left(y\right)\geq f\left(x\right)+\left\langle y-x,x^*\right\rangle\geq f\left(x\right)+\left\langle y-x,x^*+p^*\right\rangle$$

Hence, $x^* + p^*$ is a subdifferential. We thus proved the inclusion $x^* + \left(C - x\right)^- \subseteq \partial f\left(x\right)$.

Conversely, let $p^* \in \partial f\left(x\right)$. Then, $f\left(y\right) - f\left(x\right) \geq \left\langle y - x, p^*\right\rangle$ and so

$$f\left(x_t\right)-f\left(x\right)\geq t\left\langle y-x,p^*\right\rangle$$

for all $y \in C$ and $t \in [0, 1]$. Dividing by t and letting $t \to 0$, we get $Df\left(x;y\right) \geq \left\langle y - x, p^*\right\rangle$. If x^* is a gradient,

$$\left\langle y-x,x^*\right\rangle=Df\left(x;y\right)\geq\left\langle y-x,p^*\right\rangle\Longrightarrow 0\geq\left\langle y-x,p^*-x^*\right\rangle$$

That is, $p^* \in x^* + \left(C - x\right)^-$. This proves the converse inclusion $\partial f\left(x\right) \subseteq x^* + \left(C - x\right)^-$.

In sum, $\partial f\left(x\right) = x^* + N_C\left(x\right)$. The equality $\partial f\left(x\right) = \nabla_a f\left(x\right)$ is a consequence of the fact that $\left(C - x\right)^- = \left(C - x\right)^\perp$ when $x \in \operatorname{ri} C$. ■

This proposition implies that subdifferentials are not empty for a-differentiable convex functions. Thus, a-differentiable convex functions are weakly lower semicontinuous.

Next we turn to the algebra of affine differentiability, which is similar to the standard one (we omit the routine proof).

Proposition 25 *Let $f, g : C \to \mathbb{R}$ be wa-differentiable at $x \in C$. Then, the two functions $f + g$ and fg are wa-differentiable at x_0, with*

$$D(f+g)(x;y) = Df(x;y) + Dg(x;y)$$

and

$$D(fg)(x;y) = f(x)\,Dg(x;y) + g(x)\,Df(x;y)$$

Moreover, if $h : \mathbb{R} \to \mathbb{R}$ is differentiable at $f(x)$, then $h \circ f$ is wa-differentiable at x, with

$$D(h \circ f)(x;y) = h'(f(x))\,Df(x;y)$$

A consequence is the following sum rule for subdifferentials.

Proposition 26 *If two convex functions $f, g : C \to \mathbb{R}$ are a-differentiable at $x \in C$, then*

$$\partial(f+g)(x) = \partial f(x) + \partial g(x)$$

Proof It follows from the sum rule and from (16) by observing that $(C-x)^- + (C-x)^- = (C-x)^-$. Hence, if $x^* \in [\nabla_a f(x)]$ and $y^* \in [\nabla_a g(x)]$, we have

$$\partial(f+g)(x) = x^* + y^* + (C-x)^- = x^* + (C-x)^- + y^* + (C-x)^- = \partial f(x) + \partial g(x)$$

as desired. ■

We close by considering the wa-differentiability of quasiconvex functionals.

Proposition 27 *A wa-differentiable $f : C \to \mathbb{R}$ is quasiconvex if and only if*

$$Df(x;y) > 0 \Longrightarrow f(y) > f(x) \tag{17}$$

for all $x, y \in C$.

Proof Let $f : C \to \mathbb{R}$ be wa-differentiable. The function f is quasiconvex if and only if, for all $x, y \in C$, its restrictions $\varphi_{x,y}(t) = f(x_t)$ are quasiconvex. By Lemma 17, f is hemidifferentiable. Thus, f is quasiconvex if and only if, for all $x, y \in C$,

$$\varphi_{x,y}(t_2) \le \varphi_{x,y}(t_1) \Longrightarrow \varphi'_{x,y}(t_1)(t_2 - t_1) \le 0 \tag{18}$$

for all $t_1, t_2 \in [0,1]$. By (10), it becomes

$$f(x_{t_2}) \le f(x_{t_1}) \Longrightarrow \frac{1}{t_2 - t_1} Df(x_{t_1}; x_{t_2})(t_2 - t_1) \le 0$$

for all $t_1, t_2 \in [0,1]$. By taking $t_1 = 1$ and $t_2 = 1$ we get the desired result. ■

4 Optimization

We begin with a first-order condition for a (global) minimizer.

Proposition 28 *Let $f : C \to \mathbb{R}$ be wa-differentiable. If $\hat{x} \in C$ is a minimizer of f, then*

$$Df(\hat{x};y) \ge 0 \qquad \forall y \in C \tag{19}$$

with equality when $\hat{x} \in \operatorname{ri} C$.

When f is a-differentiable, the first-order condition (19) takes the variational inequality form

$$\langle y - \hat{x}, \nabla_a f(\hat{x}) \rangle \geq 0 \qquad \forall y \in C$$

Proof Let $\hat{x}$ be a minimizer. Then, $f(\hat{x}_t) \geq f(\hat{x})$ for all $t \in [0,1]$ and $y \in C$. Hence,

$$Df(\hat{x}, y) = \lim_{t \downarrow 0} \frac{f(\hat{x}_t) - f(\hat{x})}{t} \geq 0$$

It remains to show that if $\hat{x} \in \operatorname{ri} C$ the wa-differential vanishes. Fix a point $y \in C$ that differs from $\hat{x}$. As $\hat{x} \in \operatorname{ri} C$, there exist $x \in C$ and $\bar{t} \in (0,1)$ such that $(1-\bar{t})x + \bar{t}y = \hat{x}$. By Lemma 17, the function $\varphi(t) = f(x_t)$ is differentiable at $\bar{t}$ and

$$\varphi'(\bar{t}) = \frac{1}{1-\bar{t}} Df(x_{\bar{t}}; y) = \frac{1}{1-\bar{t}} Df(\hat{x}; y)$$

On the hand φ has a minimizer at the interior point $\bar{t}$. Hence, $\varphi'(\bar{t}) = 0$ and so $Df(\hat{x}; y) = 0$ for all $y \in C$. ■

Example 29 Let C be a convex set in a pre-Hilbert space H. Given $h \in H$, define $f : C \to \mathbb{R}$ by $f(x) = \|x - h\|^2$. The function f is convex and a-differentiable, with $Df(\hat{x}; y) = 2\langle \hat{x} - h, y - \hat{x} \rangle$. By Proposition 28, at the minimizer $\hat{x} \in C$ we have

$$\langle h - \hat{x}, y - \hat{x} \rangle \leq 0 \qquad \forall y \in C$$

This the well-known characterization of the projection onto a convex set. ▲

Let A be a topological space and $f : A \times C \to \mathbb{R}$ a parametric objective function (here C is interpreted a set of parameters). Define the *value function* $v : C \to \mathbb{R}$ by

$$v(x) = \sup_{a \in A} f(a, x)$$

and the *solution correspondence* $\sigma : C \rightrightarrows A$ by

$$\sigma(x) = \arg\max_{a \in A} f(a, x)$$

We say that σ is *viable* when $\sigma(x) \neq \emptyset$ for all $x \in C$.

Our first result provides an estimate for the wa-differential of the value function. A piece of notation: for a fixed element $a \in A$, we denote by $Df_a(x; y)$ the wa-differential at $x \in C$ of the section $f_a : C \to \mathbb{R}$ of the objective function f.

Proposition 30 *Let $f_a : C \to \mathbb{R}$ be wa-differentiable for each $a \in A$. If v is wa-differentiable and σ is viable, then for each $x \in C$ and each $a \in \sigma(x)$,*

$$Dv(x; y) \geq Df_a(x; y) \qquad \forall y \in C \tag{20}$$

with equality when $x \in \operatorname{ri} C$.

Proof Fix $x \in C$ and $a \in \sigma(x)$. The function $\Phi : C \to \mathbb{R}$ given by $\Phi(y) = v(y) - f(a, y)$ is wa-differentiable. We have $\Phi \geq 0$ and $\Phi(x) = 0$. Hence, x is a minimizer and so the first-order condition (19) gives the inequality (20). If $x \in \operatorname{ri} C$ we then get the desired equality. ■

The next result, a variant of the classic Danskin Theorem (see [10]), provides conditions under which the directional derivative of the value function exists.

Theorem 31 *Let σ be viable. Given $x, y \in C$, assume that f satisfies the following properties:*

(i) the affine directional derivative $Df_a(x;y)$ exists for all $a \in \sigma(x)$;

(ii) for every sequence $\{t_n\} \subseteq [0,1]$ with $t_n \downarrow 0$, there is a sequence in A, with terms $a_n \in \sigma(x_{t_n})$, such that

$$\limsup_{n\to\infty} \frac{f(a_n, x_{t_n}) - f(a_n, x)}{t_n} \leq \sup_{a\in\sigma(x)} Df_a(x;y)$$

Then, the affine directional derivative $Dv(x;y)$ exists, with

$$Dv(x;y) = \sup_{a\in\sigma(x)} Df_a(x;y)$$

In Section 5.4 we will apply this result.

Proof We first show that, for each $y \in C$,

$$\liminf_{t\downarrow 0} \frac{v(x_t) - v(x)}{t} \geq \sup_{a\in\sigma(x)} Df_a(x;y) \tag{21}$$

Fix $y \in C$ and pick a point $a_0 \in \sigma(x)$. Take any sequence $t_n \in (0,1)$ with $t_n \downarrow 0$. By definition,

$$\frac{v(x_{t_n}) - v(x)}{t_n} \geq \frac{f(a_0, x_{t_n}) - f(a_0, x)}{t_n}$$

Hence,

$$\lim \inf_{n\to\infty} \frac{v(x_{t_n}) - v(x)}{t_n} \geq Df_{a_0}(x;y)$$

This is true for every sequence $t_n \downarrow 0$. Consequently,

$$\liminf_{t\downarrow 0} \frac{v(x_t) - v(x)}{t} \geq Df_{a_0}(x;y)$$

As this is true for every $a_0 \in \sigma(x)$, we get (21). To end the proof, let us prove that

$$\limsup_{t\downarrow 0} \frac{v(x_t) - v(x)}{t} \leq \sup_{a\in\sigma(x)} Df_a(x;y) \tag{22}$$

Given a sequence $t_n \downarrow 0$, define $\{a_n\}$ via assumption (ii). We have:

$$\frac{v(x_t) - v(x)}{t_n} = \frac{f(a_n, x_t) - v(x)}{t_n} = \frac{f(a_n, x_t) - f(a_n, x)}{t_n} + \frac{f(a_n, x) - v(x)}{t_n} \leq \frac{f(a_n, x_t) - f(a_n, x)}{t_n}$$

In light of (ii) we have

$$\lim \sup_{n\to\infty} \frac{v(x_t) - v(x)}{t_n} \leq \lim \sup_{n\to\infty} \frac{f(a_n, x_t) - f(a_n, x)}{t_n} \leq \sup_{a\in\sigma(x)} Df_a(x;y)$$

We get (22) that, in turn, yields the desired result. ∎

5 Applications

5.1 Influence curve

Let $T : \Delta(Y) \to \mathbb{R}$ be wa-differentiable. Using the identification $x \mapsto \delta_x$, based on Hampel [14] we define the *influence function* $\mathcal{I}(\cdot; T, \mu) : Y \to \mathbb{R}$ by

$$\mathcal{I}(x; T, \mu) = \lim_{t\downarrow 0} \frac{T((1-t)\mu + t\delta_x) - T(\mu)}{t} = DT(\mu; \delta_x)$$

When T is a-differentiable,

$$\mathcal{I}(x;T,\mu)=\int u_\mu \mathrm{d}(\delta_x-\mu)=u_\mu(x)-\int u_\mu d\mu$$

where $u_\mu\in[\nabla_a T(\mu)]$. Since the gradient u_μ is unique up to an additive constant, under the normalization $\int u_\mu \mathrm{d}\mu=0$ we get

$$\mathcal{I}(x;T,\mu)=u_\mu(x) \tag{23}$$

The influence function thus completely pins down the a-differential of the functional T as μ varies. Conversely, the influence function agrees with the gradient under the normalization condition $\int u_\mu \mathrm{d}\mu=0$.

Example 32 The functional $T:\mathcal{D}[0,1]\to\mathbb{R}$ defined by $T(F)=F(x_0)$, with $x_0\in(0,1)$, is wa-differentiable. Indeed,

$$\lim_{t\downarrow 0}\frac{T((1-t)F+tG)-T(F)}{t}=G(x_0)-F(x_0)$$

and the map $G\mapsto DT(F;G)=G(x_0)-F(x_0)$ is affine. The influence function is given by

$$\mathcal{I}(x;T,F)=G_x(x_0)-F(x_0)$$

where G_x is the distribution function of the Dirac measure δ_x. Clearly, it is discontinuous at x_0. By (23), it should coincide with the continuous function u_F. Hence, T is not a-differentiable at any point $F\in\mathcal{D}[0,1]$. ▲

Thus, the existence of the influence function does not ensure a-differentiability. Next we show that this is the case even the existence of a continuous influence function (see also Example 2.2.3 of [23]).

Example 33 Define the convex functional $T:\mathcal{D}[0,1]\to\mathbb{R}$ by

$$T(F)=\sum_{x\in[0,1]}[F(x)-F(x-)]^\alpha \tag{24}$$

with $\alpha>1$. This functional measures the jumps of F. The series is well defined since there are at most countably many jumps. Clearly,

$$\sum_{x\in[0,1]}[F(x)-F(x-)]^\alpha\le\sum_{x\in[0,1]}[F(x)-F(x-)]\le 1$$

and the sum is finite. It is easy to see that T is wa-differentiable, with

$$DT(F;G_x)=-\alpha T(F)+\alpha(F(x)-F(x-))^{\alpha-1}$$

If the distribution F is continuous, then $DT(F;G_x)=0$. We infer that the gradient should vanish everywhere and, by Theorem 18, F would be constant, a contradiction. Hence, T is not a-differentiable. ▲

5.2 Multi-utility representations

Let us now interpret an element $\mu\in\Delta(Y)$ as a lottery with prizes in a metric space Y. A decision maker (DM) has a preference (binary) relation $\succsim$ over $\Delta(Y)$ represented by a utility function $U:\Delta(Y)\to\mathbb{R}$, namely,

$$\mu\succsim\lambda\Longleftrightarrow U(\mu)\ge U(\lambda)$$

We introduce the auxiliary subrelation $\succsim^*$ defined by

$$\mu \succsim^* \lambda \Longleftrightarrow \alpha\mu + (1-\alpha)\,\nu \succsim \alpha\lambda + (1-\alpha)\,\nu$$

for all $\alpha \in [0,1]$ and all $\nu \in \Delta(Y)$. It captures the comparison over which the DM feels sure (see [12] and [5]).

In the next theorem $\nabla_a U(\mu)$ is understood to be a normalized gradient. For instance, $\nabla_a U(\mu) = u_\mu$ with $\int u_\mu \mathrm{d}\mu = 0$, that is, the influence function associated with the probability measure μ. Moreover, we denote by $\operatorname{Im} \nabla_a U \subseteq C_b(Y)$ the image $\{\nabla_a U(\lambda) : \lambda \in \Delta(Y)\}$.

Theorem 34 *If $\succsim$ is a preference relation with a-differentiable utility function U, then*

$$\mu \succsim^* \lambda \Longleftrightarrow \int u \mathrm{d}\mu \geq \int u \mathrm{d}\lambda \qquad \forall u \in \operatorname{Im} \nabla_a U$$

In the theory of risk, since Machina [17] the functions $u \in \operatorname{Im} \nabla_a U$ are known as *local utilities.* Theorem 34 formalizes the idea that, individually, each local utility models a local expected utility behavior of $\succsim$, but jointly all local utilities characterize a global expected utility feature of $\succsim$.

Proof Define $\hat{\succsim}$ by

$$\mu \hat{\succsim} \lambda \Longleftrightarrow \int u \mathrm{d}\mu \geq \int u \mathrm{d}\lambda \qquad \forall u \in \operatorname{Im} \nabla_a U$$

Assume that $\mu \hat{\succsim} \lambda$. Let $\nu \in \Delta(Y)$ and $\alpha \in (0,1)$. By the Mean Value Theorem (Theorem 18),

$$\Delta = U(\alpha\mu + (1-\alpha)\,\nu) - U(\alpha\lambda + (1-\alpha)\,\nu) = \alpha \langle \nabla_a U(\zeta), \mu - \lambda \rangle$$

where $\zeta = \alpha[(1-t)\,\mu + t\lambda] + (1-\alpha)\,\nu$ for some $t \in (0,1)$. Since $\mu \hat{\succsim} \lambda$, by setting $u_\zeta = \nabla_a U(\zeta)$ we have

$$\Delta = \alpha \left[\int u_\zeta \mathrm{d}\mu - \int u_\zeta \mathrm{d}\lambda \right] \geq 0$$

that is, $\mu \succsim^* \lambda$. Conversely, assume $\mu \succsim^* \lambda$ and pick any element $\nu \in \Delta(Y)$. It follows that, for each $t \in [0,1]$,

$$(1-t)\,\nu + t\mu \succsim (1-t)\,\nu + t\lambda$$

Hence, for $t > 0$,

$$\frac{U((1-t)\,\nu + t\mu) - U(\nu)}{t} \geq \frac{U((1-t)\,\nu + t\lambda) - U(\nu)}{t}$$

Letting $t \downarrow 0$, we get $\langle \nabla_a U(\nu), \mu - \nu \rangle \geq \langle \nabla_a U(\nu), \lambda - \nu \rangle$. Namely, $\langle \nabla_a U(\nu), \mu \rangle \geq \langle \nabla_a U(\nu), \lambda \rangle$. That is, $\mu \hat{\succsim} \lambda$ as desired. ■

The next example is based on [9].

Example 35 Define the utility function $U : \Delta(Y) \to \mathbb{R}$ by

$$U(\mu) = \int \psi(x,y)\, \mathrm{d}\mu(x) \otimes \mathrm{d}\mu(y)$$

where $\psi(x,y)$ is a symmetric, continuous and bounded function on $Y \times Y$. In Example 15 we saw that, up to a constant,

$$\nabla_a U(\mu) = u_\mu = \int \psi(\cdot, y)\, \mathrm{d}\mu(y)$$

Therefore,

$$\operatorname{Im} \nabla_a U = \left\{ \int \psi(\cdot, y)\, \mathrm{d}\nu(y) : \nu \in \Delta(X) \right\}$$

By Theorem 34,

$$\mu \succsim^* \lambda \Longleftrightarrow \int \psi(x,y)\,\mathrm{d}\mu(x) \otimes \mathrm{d}\nu(y) \geq \int \psi(x,y)\,\mathrm{d}\lambda(x) \otimes \mathrm{d}\nu(y)$$

for all $\nu \in \Delta(Y)$. Clearly, this is equivalent to

$$\mu \succsim^* \lambda \Longleftrightarrow \int \psi(x,y)\,\mathrm{d}\mu(x) \geq \int \psi(x,y)\,\mathrm{d}\lambda(x) \qquad \forall y \in X$$

▲

5.3 Prospect theory

To each monetary lottery $\mu \in \Delta(\mathbb{R})$ we associate its distribution function $F_\mu(x) = \mu((-\infty, x])$. Fix a nonatomic positive Borel measure ρ on the real line and consider the utility function $U : \Delta(\mathbb{R}) \to \mathbb{R}$ given by

$$U(\mu) = \int_{[0,\infty]} w_+(1 - F_\mu(x))\,\mathrm{d}\rho(x) - \int_{[-\infty,0]} w_-(F_\mu(x))\,\mathrm{d}\rho(x) \tag{25}$$

where $w_+, w_- : [0,1] \to [0,1]$ are two strictly increasing and continuously differentiable maps. Instances of this utility function appear in the Prospect Theory of Tverski and Kahneman [24] (see Wakker [26]). The utility function U is wa-differentiable:[7]

$$\begin{aligned}
DU(\mu;\lambda) &= \lim_{t\downarrow 0} \frac{U(\mu + t(\lambda - \mu)) - U(\mu)}{t} \\
&= -\int_{[0,+\infty]} w_+'(1 - F_\mu(x))(F_\lambda - F_\mu)\,\mathrm{d}\rho(x) - \int_{[-\infty,0]} w_-'(F_\mu(x))(F_\lambda - F_\mu)\,\mathrm{d}\rho(x) \\
&= \int_{\mathbb{R}} \varphi_\mu(x)(F_\mu(x) - F_\lambda(x))\,\mathrm{d}\rho(x)
\end{aligned}$$

where φ is the bounded scalar function

$$\varphi_\mu(x) = \begin{cases} w_+'(1 - F_\mu(x)) & \text{if } x \geq 0 \\ w_-'(F_\mu(x)) & \text{otherwise} \end{cases}$$

It is not obvious whether $DU(\mu;\cdot)$ is extendable or not. To prove this we must use integration by parts for Riemann-Stieltjes integrals. Set $\mathrm{d}\eta = \varphi_\mu \mathrm{d}\rho$ and integrate first on a finite interval. By the Dominated Convergence Theorem,

$$\int_{\mathbb{R}} \varphi_\mu(x)(F_\mu(x) - F_\lambda(x))\,\mathrm{d}\rho(x) = \lim_{R\to+\infty} \int_{[-R,R]} \varphi_\mu(x)(F_\mu(x) - F_\lambda(x))\,\mathrm{d}\rho(x)$$

Hence,

$$\int_{[-R,R]} \varphi_\mu(x)(F_\mu(x) - F_\lambda(x))\,\mathrm{d}\rho(x) = \int_{[-R,R]} (F_\mu(x) - F_\lambda(x))\,\mathrm{d}\eta(x) = \int_{-R}^{R} (F_\mu(x) - F_\lambda(x))\,\mathrm{d}\Phi_\mu(x)$$

where the last integral is Riemann-Stieltjes, and the function $\Phi_\mu : \mathbb{R} \to \mathbb{R}$ defined by

$$\Phi_\mu(x) = \int_{[-\infty,x]} \varphi_\mu(t)\,\mathrm{d}\rho(t) \tag{26}$$

[7] Continuous differentiability of w_+ and w_- implies their Lipschitzianity, so we can apply the Dominated Convergence Theorem.

is bounded and absolutely continuous. By the integration-by-parts formula,

$$\int_{-R}^{R} (F_\mu(x) - F_\lambda(x))\, \mathrm{d}\Phi_\mu(x) = [(F_\mu(x) - F_\lambda(x))\, \Phi_\mu(x)]_{-R}^{R} - \int_{-R}^{R} \Phi_\mu(x)\, \mathrm{d}(F_\mu - F_\lambda)$$
$$= [(F_\lambda(x) - F_\mu(x))\, \Phi_\mu(x)]_{-R}^{R} + \int_{[-R,R]} \Phi_\mu(x)\, \mathrm{d}(\lambda - \mu)$$

Letting $R \to \infty$, we finally get

$$DV(\mu;\lambda) = \int_{\mathbb{R}} \Phi_\mu(x)\, \mathrm{d}(\lambda - \mu) = \langle \Phi_\mu, \lambda - \mu \rangle$$

since $[(F_\lambda(x) - F_\mu(x))\, \Phi_\mu(x)]_{-R}^{R} \to 0$ as Φ_μ is bounded.

In sum, U is a-differentiable and its gradient is

$$\nabla_a U(\mu) = \Phi_\mu \in C_b(\mathbb{R})$$

Let $\succsim$ be a preference relation on $\Delta(\mathbb{R})$ represented by U. For its subrelation $\succsim^*$ we thus have the following consequence of Theorem 34,

Proposition 36 *Let $\mu, \lambda \in \Delta(\mathbb{R})$. It holds, for each $\nu \in \Delta(\mathbb{R})$,*

$$\mu \succsim^* \lambda \Longleftrightarrow \int_{\mathbb{R}} \varphi_\nu(x)\,(1 - F_\mu(x))\, \mathrm{d}\rho(x) \geq \int_{\mathbb{R}} \varphi_\nu(x)\,(1 - F_\lambda(x))\, \mathrm{d}\rho(x) \tag{27}$$

Proof By Theorem 34, $\mu \succsim^* \lambda$ if and only if, for each $\nu \in \Delta(\mathbb{R})$, $\int \Phi_v(t)\, \mathrm{d}\mu(t) \geq \int \Phi_v(t)\, \mathrm{d}\lambda(t)$. By Fubini's Theorem,

$$\int \Phi_v(t)\, \mathrm{d}\mu(t) = \int \mathrm{d}\mu(t) \int I_{[-\infty,t]}(x)\, \varphi_\nu(x)\, \mathrm{d}\rho(x) = \int \mathrm{d}\rho(x)\, \varphi_\nu(x) \int I_{[x,+\infty]}(t)\, \mathrm{d}\mu(t)$$

where we used the relation $I_{[-\infty,t]}(x) = I_{[x,+\infty]}(t)$. Hence,

$$\int \Phi_v(t)\, \mathrm{d}\mu(t) = \int \varphi_\nu(x)\,(1 - F_\mu(x-))\, \mathrm{d}\rho(x) = \int \varphi_\nu(x)\,(1 - F_\mu(x))\, \mathrm{d}\rho(x)$$

where the last equality is true because ρ is nonatomic. This proves (27). ■

5.4 Bayesian robustness

Let Θ be a parameter space and $\mathcal{X}$ a sample space. For any given sample $x \in \mathcal{X}$, a *posterior functional* $\rho_x : \Delta(\Theta) \to \mathbb{R}$ maps a prior distribution $\mu \in \Delta(\Theta)$ to a scalar representing a posterior statistic of interest. For example, the posterior mean is given by $\rho_x(\mu) = \int \theta \mathrm{d}\mu_x(\theta)$, with μ_x being the Bayesian update of μ.

Bayesian robustness (see, e.g., [2] and [3]) investigate how posterior outcomes vary under different prior specifications, often by examining the range of ρ_x over a set of priors $\mathcal{M} \subseteq \Delta(\Theta)$. When $\mathcal{M}$ is convex, our methods can be used to compute such a range. Indeed, if $\mu_1, \mu_2 \in \Delta(\Theta)$ are such that $\rho_x(\mathcal{M}) = [\rho_x(\mu_1), \rho_x(\mu_2)]$, then by Proposition 28 it follows that

$$\inf_{\nu \in \mathcal{M}} D\rho_x(\mu_1;\nu) = \sup_{\nu \in \mathcal{M}} D\rho_x(\mu_2;\nu) = 0$$

These two necessary conditions can be used to develop numerical algorithms (see [1]).

Another statistical functional of interest is the expected posterior loss of a Bayesian estimator. Formally, denote by $a \in A$ the choice of the estimator. Given a loss function $\ell : A \times \Theta \to \mathbb{R}$ and a prior $\mu \in \Delta(\Theta)$, the expected loss $L(\mu)$ of the Bayesian estimator is

$$L(\mu) = \inf_{a \in A} \int \ell(a, \theta) \mathrm{d}\mu(\theta)$$

Given an alternative prior $\nu \in \Delta(\Theta)$, the directional derivative $DL(\mu;\nu)$ captures the sensitivity of the estimator to the prior μ (see [16] and also [22] for applications in economics). In the rest of this subsection we show that, thanks to Theorem 31, we are able to compute $DL(\mu;\nu)$ in important cases of interest, allowing for an unbounded set of parameters (as common in applications).

So, let $A = \Theta = \mathbb{R}$. Consider $\mu, \nu \in \Delta(\mathbb{R})$ with finite first moment. To apply Theorem 31 we study the related optimization problem

$$L^*(\mu) = \sup_{a\in A} \int -\ell(a,\theta)\mathrm{d}\mu(\theta) = \sup_{a\in\mathbb{R}} U(a,\mu)$$

where $U(a,\mu) = \int -|a-\theta|\, d\mu(\theta)$. The next result is an application of the earlier Danskin-type theorem (Theorem 31).

Proposition 37 *The affine directional derivative $DL^*(\mu,\nu)$ exists, with*

$$DL^*(\mu,\nu) = \sup_{a\in\sigma(\mu)} DU_a(\mu,\nu) = \sup_{a\in\sigma(\mu)} \int |a-\theta|\mathrm{d}\mu(\theta) - \int |a-\theta|\mathrm{d}\nu(\theta) \tag{28}$$

Proof As well-known, the set $\sigma(\lambda)$ of maximizers of the section U_a for $\lambda \in \Delta(\mathbb{R})$ is a median value of the distribution F_λ. Specifically, a point $a \in \sigma(\lambda)$ is characterized by the equations

$$F_\lambda(a) \geq \frac{1}{2} \quad \text{and} \quad F_\lambda(a^-) \leq \frac{1}{2} \tag{29}$$

The set $\bigcup_{t\in[0,1]}\sigma((1-t)\mu + t\nu)$ is contained into a compact interval K of $\mathbb{R}$ (just pick any compact interval that contains the sets of points $1/2-\eta \leq F_\mu(x) \leq 1/2+\eta$ and $1/2-\eta \leq F_\nu(x) \leq 1/2+\eta$ for some $0<\eta<1/2$). By convexity, $\sigma((1-t)\mu+t\nu) \subseteq K$ for all $t\in[0,1]$. Let us show that the two assumptions of Theorem 31 are satisfied. Clearly, U_a is wa-differentiable and its gradient is

$$DU_a(\mu,\nu) = \int |a-\theta|\mathrm{d}\mu(\theta) - \int |a-\theta|\mathrm{d}\nu(\theta)$$

Hence, (i) holds. Take $t_n \downarrow 0$ and let $\{a_n\} \subseteq K$ be any sequence with $a_n \in \sigma((1-t_n)\mu + t_n\nu)$. In view of Theorem 31-(ii), consider any subsequence $\{a_{n_k}\}$ such that the sequence

$$\frac{U(a_{n_k},(1-t_{n_k})\mu + t_{n_k}\nu) - U(a_{n_k},\mu)}{t_{n_k}} \tag{30}$$

converges. As K is compact, without loss of generality (passing if needed to a further subsequence) we can assume that the sequence $\{a_{n_k}\}$ converges to a point $\bar{a}$.

Claim It holds $\bar{a} \in \sigma(\mu)$.

Proof For any $\varepsilon > 0$ we have $a_n \leq \bar{a}+\varepsilon$ for n sufficiently large. By monotonicity,

$$\frac{1}{2} \leq (1-t_n)F_\mu(a_n) + t_nF_\nu(a_n) \leq (1-t_n)F_\mu(\bar{a}+\varepsilon) + t_nF_\nu(\bar{a}+\varepsilon)$$

As $n\to\infty$ we get $1/2 \leq F_\mu(\bar{a}+\varepsilon)$. Since F_μ is right-continuos we have $1/2 \leq F_\mu(\bar{a})$ which is the first condition of (29). As to the second one, begin now with $a_n \geq \bar{a}-\varepsilon$. Hence, $a_n - \eta \geq \bar{a}-\varepsilon-\eta$ for any $\eta>0$. Thus

$$(1-t_n)F_\mu(\bar{a}-\varepsilon-\eta) + t_nF_\nu(\bar{a}-\varepsilon-\eta) \leq (1-t_n)F_\mu(a_n-\eta) + t_nF_\nu(a_n-\eta) \leq \frac{1}{2}$$

As $n\to\infty$, we get $F_\mu(\bar{a}-\varepsilon-\eta) \leq 1/2$ for all $\varepsilon+\eta>0$. Hence, condition (29) is true and so $\bar{a}\in\sigma(\mu)$. □

By this Claim, we can write

$$\lim_{k\to\infty}\frac{U\left(a_{n_k},\left(1-t_{n_k}\right)\mu+t_{n_k}\nu\right)-U\left(a_{n_k},\mu\right)}{t_{n_k}}=\lim_{k\to\infty}\int|a_{n_k}-\theta|\mathrm{d}\mu(\theta)-\int|a_{n_k}-\theta|\mathrm{d}\nu(\theta)$$
$$=\int|\bar{a}-\theta|\mathrm{d}\mu(\theta)-\int|\bar{a}-\theta|\mathrm{d}\nu(\theta)=DU_{\bar{a}}\left(\mu,\nu\right)\leq\sup_{a\in\sigma(\mu)}DU_a\left(\mu,\nu\right)$$

This is the case for every subsequence $\{a_{n_k}\}$ for which (30) has a limit. Consequently,

$$\limsup_{n\to\infty}\frac{U\left(a_n,\left(1-t_n\right)\mu+t_n\nu\right)-U\left(a_n,\mu\right)}{t_n}\leq\sup_{a\in\sigma(\mu)}DU_a\left(\mu,\nu\right)$$

and so assumption (ii) holds. By Theorem 31 we thus have (28). ■

6 Variations and extensions

6.1 Hadamard and strict differentiability

Endow the convex set C with a topology finer than (or equal to) than the relative weak topology. This finer topology is assumed to be metrizable by a metric ρ.[8] This allows us to restrict ourself to the following sequential formulation of the Hadamard directional derivative, adapted to our setting.[9]

Definition 38 *The (*affine*)* Hadamard directional derivative *of $f:C\to\mathbb{R}$ at $x\in C$ along the direction $y\in C$ is given by*

$$D_Hf\left(x;y\right)=\lim_{n\to\infty}\frac{f\left(\left(1-t_n\right)x+t_ny_n\right)-f\left(x\right)}{t_n}$$

when the limit exists finite for every sequence $t_n\downarrow 0$ and every sequence $\{y_n\}$ in C with $\rho\left(y_n,y\right)\to 0$.

The function f is called *Hadamard wa-differentiable* at x if $D_Hf\left(x;\cdot\right)$ is affine. It is called *Hadamard a-differentiable* at x if $D_Hf\left(x;\cdot\right)$ is extendable. Analogously, $\nabla_Hf\left(x\right)\in X^*$ denotes a representative of the equivalence class of Hadamard gradients.

Definition 39 *A function $f:C\to\mathbb{R}$ is* strictly wa-differentiable *at x if there exists an affine functional $D_Sf\left(x;\cdot\right):C\to\mathbb{R}$ such that, for each $y\in C$,*

$$D_Sf\left(x;y\right)=\lim_{n\to\infty}\frac{f\left(\left(1-t_n\right)x_n+t_ny\right)-f\left(x\right)}{t}$$

for every sequence $t_n\downarrow 0$ and every sequence $\{x_n\}\subseteq C$ with $\rho\left(x_n,x\right)\to 0$.

It is worth mentioning a stronger notion of differentiability that combines the two previous ones.

Definition 40 *A function $f:C\to\mathbb{R}$ is* strictly Hadamard wa-differentiable *at x if there exists an affine functional $D_{SH}f\left(x;\cdot\right):C\to\mathbb{R}$ such that, for each $y\in C$,*

$$D_{SH}f\left(x;y\right)=\lim_{n\to\infty}\frac{f\left(\left(1-t_n\right)x_n+t_ny_n\right)-f\left(x_n\right)}{t_n}$$

for every sequence $t_n\downarrow 0$ and all sequences $\{x_n\}$ and $\{y_n\}$ in C with $\rho\left(x_n,x\right)\to 0$ and $\rho\left(y_n,x\right)\to 0$.

[8] Usually the metric ρ can be extended to the vector space X. In this case the topology τ is compatible with the vector structure of the space and ρ inherits several nice properties. However, this does not significantly affect our analysis.

[9] We omit alternative Hadamard formulations related to compact differentiability (see [20]).

Example 41 The quadratic functional $Q : \Delta(Y) \to \mathbb{R}$ given by

$$Q(\mu) = \int \psi(x, y) \, \mathrm{d}\mu(x) \otimes \mathrm{d}\mu(y)$$

with ψ symmetric (see Example 15), is strictly Hadamard wa-differentiable under the Prokhorov metric when Y is a Polish space. Indeed, let $\mu_n \Longrightarrow \mu$, $\lambda_n \Longrightarrow \lambda$ and $t_n \downarrow 0$. Then,

$$\frac{Q((1-t_n)\mu_n + t_n\lambda_n) - Q(\mu_n)}{t_n} = t_n Q(\lambda_n) + (t_n - 2) Q(\mu_n) + 2(1 - t_n) \int \psi(x, y) \, \mathrm{d}\mu_n \otimes d\lambda_n$$

As Y is Polish, $\mu_n \Longrightarrow \mu$ and $\lambda_n \Longrightarrow \lambda$ imply $\mu_n \otimes \lambda_n \Longrightarrow \mu \otimes \lambda$. Hence, the limit of this ratio exists, with $D_{SH} Q(\mu; \lambda) = -2Q(\mu) + 2 \int \psi(x, y) \, \mathrm{d}\mu \otimes \mathrm{d}\lambda$. ▲

This example can be easily generalized by showing that a quadratic functional $Q : C \to \mathbb{R}$ is strictly Hadamard wa-differentiable under a metric ρ when the associated biaffine form B_S is ρ-continuous on $C \times C$. To further elaborate, we need to introduce some classes of metrics.

Definition 42 *A metric ρ on C is:*

(i) convex *if*

$$\rho(x, \alpha y_1 + (1-\alpha) y_2) \le \alpha \rho(x, y_1) + (1-\alpha) \rho(x, y_2) \tag{31}$$

for all $x, y_1, y_2 \in C$ and all $\alpha \in [0, 1]$;

(ii) homogeneous *if*

$$\rho(x, x + \alpha(y - x)) = \alpha \rho(x, y)$$

for all $x, y \in C$ and all $\alpha \in [0, 1]$.

For instance, the Prokhorov metric on $\Delta(Y)$, with Y metric separable, is equivalent to the convex Dudley metric (see Theorem 11.3.3 of [11]). With this, next we establish couple of differentiability criteria.

Proposition 43 *Let ρ be convex. Let $f : C \to \mathbb{R}$ be wa-differentiable in a ρ-neighborhood of a point $\bar{x} \in C$.*

(i) If the map $x \mapsto D_a f(x; y)$ is ρ-continuous at $\bar{x}$ for every $y \in C$, then f is strictly wa-differentiable at $\bar{x}$.

(ii) If the map $(x, y) \mapsto D_a f(x; y)$ is ρ-continuous at $(\bar{x}, y)$ for every $y \in C$, then f is strictly Hadamard wa-differentiable at $\bar{x}$.

Proof Let $x, y \in C$ and f be wa-differentiable on $U_\varepsilon(\bar{x}) \cap C$, where $U_\varepsilon(\bar{x})$ is a ρ-neighborhood. Since ρ is convex,

$$\rho(x_t, \bar{x}) \le (1-t)\rho(x, \bar{x}) + t\rho(y, \bar{x}) \le \rho(x, \bar{x}) + t\rho(y, \bar{x})$$

Consequently, taking the points x_n such that $\rho(x_n, \bar{x}) < \varepsilon/2$ and $0 < t_n < \varepsilon/[2\rho(y, \bar{x})]$, then the points $x_n + t_n(y - x_n)$ and x_n belong to the neighborhood $U_\varepsilon(\bar{x})$. By Theorem 18,

$$f(x_n + t_n(y - x_n)) - f(x_n) = \frac{1}{1 - \tau_n} Df(x_n + \tau_n t_n(y - x_n); x_n + t_n(y - x_n))$$

for some $\tau_n \in (0, 1)$. On the other hand,

$$x_n + t_n(y - x_n) = \left(\frac{1 - t_n}{1 - \tau_n t_n}\right)[x_n + \tau_n t_n(y - x_n)] + \left(\frac{(1-\tau_n) t_n}{1 - \tau_n t_n}\right) y$$

Therefore, by (3) we get

$$\frac{f(x_n+t_n(y-x_n))-f(x_n)}{t_n}=\frac{1}{1-\tau_n t_n}Df(x_n+\tau_n t_n(y-x_n);y)$$

Letting $t_n\downarrow 0$ and $\rho(x_n,\bar{x})\to 0$ the limit exists and is equal to $D_S f(\bar{x};y)$. The second statement is similarly proved. ■

Here is a kind of converse of the previous statement.

Proposition 44 *Let $f:C\to\mathbb{R}$ be strictly Hadamard wa-differentiable at $\bar{x}\in C$. If the directional derivative $Df(x;y)$ exists in a ρ-neighborhood of the point $\bar{x}$, then $Df(\cdot;\cdot)$ is ρ-continuous at $(\bar{x},y)$ for all $y\in C$.*

Likewise, if f is strictly wa-differentiable at $\bar{x}$, then $Df(\cdot\,;y)$ is ρ-continuous at $\bar{x}$.

Proof Fix $y\in C$ and consider two sequences $\{x_n\},\{y_n\}\subseteq C$ so that $\rho(y_n,y)\to 0$ and $\rho(x_n,\bar{x})\to 0$ and such that $\{x_n\}$ is contained into the claimed neighborhood of $\bar{x}$. Fix $\varepsilon>0$. For every n there is $t_n\in(0,1)$ so that

$$\left|\frac{f((1-t_n)x_n+t_ny_n)-f(x_n)}{t_n}-Df(x_n;y_n)\right|\leq\frac{\varepsilon}{2}.$$

The sequence $\{t_n\}$ can be chosen so that $t_n\downarrow 0$. The hypothesis of strict Hadamard wa-differentiability implies that, for all n sufficiently large,

$$\left|\frac{f((1-t_n)x_n+t_ny_n)-f(x_n)}{t_n}-D_Hf(\bar{x};y)\right|\leq\frac{\varepsilon}{2}$$

This yields $|Df(x_n;y_n)-D_Hf(\bar{x};y)|\leq\varepsilon$. Hence, $Df(x_n;y_n)\to D_Hf(\bar{x};y)$ as $\rho(y_n,y)\to 0$ and $\rho(x_n,\bar{x})\to 0$, which is the desired continuity property. ■

Next we establish a noteworthy continuity consequence of Proposition 43.

Proposition 45 *If $f:C\to\mathbb{R}$ is strictly Hadamard wa-differentiable, then it is ρ-continuous.*

Proof Let $x\in C$. Take a sequence $\{x_n\}\subseteq C$ such that $\rho(x_n,x)\to 0$. By applying the Mean Value Theorem (Theorem 18) to the two points x and x_n, there exists a sequence $\{t_n\}\subseteq(0,1)$ such that

$$f(x_n)-f(x)=\frac{1}{1-t_n}Df((1-t_n)x+t_nx_n;x_n)=-\frac{1}{t_n}Df((1-t_n)x+t_nx_n;x)\tag{32}$$

Partition the elements of the sequence $\{t_n\}$ so to obtain two finite or infinite subsequences for which either $t_n\in(0,1/2)$ or $t_n\in[1/2,1)$. For the elements that fall into $(0,1/2)$ we have $(1-t_n)^{-1}<2$. Analogously, it holds $\left|-t_n^{-1}\right|\leq 2$ for the points of the sequence that fall into $[1/2,1)$. By Proposition 43,

$$Df((1-t_n)x+t_nx_n;x_n)\to Df(x;x)=0$$

as well as

$$Df((1-t_n)x+t_nx_n;x)\to Df(x;x)=0$$

Passing, if needed, to subsequences, in light of (32) we infer that $f(x_n)-f(x)\to 0$ as $\rho(x_n,x)\to 0$. ■

Let us illustrate our Hadamard results with some examples.

Example 46 (i) The functional $T(F) = F(x_0)$ of Example 32 is wa-differentiable, with $DT(F;G) = G(x_0) - F(x_0)$. By Proposition 43, T is strictly wa-differentiable at every point F which is continuous at x_0 under the Prokhorov metric. Instead, under the *Kolmogorov metric* ρ_∞,[10] the functional T is strictly Hadamard wa-differentiable at any point F. Indeed,

$$|DT(F;G) - DT(F_1;G_1)| \leq |G(x_0) - G_1(x_0)| + |F(x_0) - F_1(x_0)| \leq \|G - G_1\|_\infty + \|F - F_1\|_\infty$$

By Proposition 43, the continuity of $DT(\cdot\,;\cdot)$ delivers the desired result.

(ii) Define $T : \Delta(Y) \to \mathbb{R}$ by $T(\mu) = \int f \mathrm{d}\mu$, with $f : Y \to \mathbb{R}$ bounded and Borel measurable. Endow $\Delta(Y)$ with the metric $\rho_V(\lambda, \mu) = \|\lambda - \mu\|_V$ generated by the variation norm.[11] Clearly, T is wa-differentiable (but not a-differentiable if f is not continuos), with $DT(\lambda;\mu) = \int f \mathrm{d}(\mu - \lambda)$. By Proposition 43, T is strictly Hadamard wa-differentiable at every $\mu \in \Delta(Y)$. Indeed,

$$\begin{aligned}
&|DT(\lambda;\mu) - DT(\lambda_1;\mu_1)| = \left| \int f \mathrm{d}(\mu - \lambda) - \int f \mathrm{d}(\mu_1 - \lambda_1) \right| \\
&= \left| \int f \mathrm{d}(\mu - \mu_1) + \int f \mathrm{d}(\lambda_1 - \lambda) \right| \leq \left| \int f \mathrm{d}(\mu - \mu_1) \right| + \left| \int f \mathrm{d}(\lambda_1 - \lambda) \right| \\
&\leq \|f\|_\infty \|\mu - \mu_1\|_V + \|f\|_\infty \|\lambda_1 - \lambda\|_V
\end{aligned}$$

(iii) In light of Example 12, it is easy to check that, if $F(s, \cdot)$ is strictly Hadamard wa-differentiable for the Euclidean metric of $\mathbb{R}^n$, then the functional $I : C \to \mathbb{R}$ is strictly Hadamard wa-differentiable for the uniform metric $\rho_\infty(u, v) = \sup_S \|u - v\|_n$. ▲

6.2 Frechet differentiability

We begin with a notion of wa-differentiability closely related to the uniform convergence over bounded sets of the directional derivative.

Definition 47 *A function* $f : C \to \mathbb{R}$ *is* boundedly wa-differentiable *at* $x \in C$ *if there exists a* ρ*-continuous and affine functional* $A : C \to \mathbb{R}$ *such that*

$$\lim_{n\to\infty} \left(\frac{f((1 - t_n)x + t_n y_n) - f(x)}{t_n} - A(y_n) \right) = 0 \tag{33}$$

for every sequence $t_n \downarrow 0$ *and every* ρ*-bounded sequence* $\{y_n\}$ *in* C*.*

By setting $y_n = x$, one gets $A(x) = 0$. Moreover, since ρ-convergent sequences are ρ-bounded, a boundedly wa-differentiable function at $x \in C$ is also Hadamard wa-differentiable, with $D_H f(x, \cdot) = A(\cdot)$.

Proposition 48 *Let* ρ *be convex. A function* $f : C \to \mathbb{R}$ *is boundedly wa-differentiable at* $x \in C$ *if and only if the limit (33) exists finite for all sequences* $\{y_n\}$ *in some circular neighborhood* $B_\varepsilon(x)$ *of the point* x*.*

Proof An implication is obvious as $B_\varepsilon(x)$ is a bounded set. Suppose now that the limit exists finite for each sequence in some $B_\varepsilon(x)$. Fix a sequence $t_n \downarrow 0$ and any ρ-bounded sequence $\{y_n\} \subseteq C$. From

$$\rho(x, (1 - \alpha)x + \alpha y_n) \leq \alpha \rho(x, y_n)$$

[10] That is, $\rho_\infty(F, G) = \sup |F(t) - G(t)|$.

[11] Recall that $\|\nu\|_V = \sup \left\{ \left| \int f d\nu \right| : f \in C_b(Y) \right\}$ for $\nu \in ca(Y)$.

it follows that, for $\alpha > 0$ is sufficiently small, the sequence $z_n = (1-\alpha)\,x + \alpha y_n$ belongs to $B_\varepsilon(x) \cap C$ for all n. By setting $\tau_n = t_n/\alpha$, we can write

$$\lim_{n\to\infty}\left[\frac{f\left(x+\tau_n\left(z_n-x\right)\right)-f\left(x\right)}{\tau_n}-A\left(z_n\right)\right]=0$$

Coming back to the old variables, we get the limit (33). Observe that $A(z_n) = A((1-\alpha)\,x + \alpha y_n) = \alpha A(y_n)$. ■

We now introduce Frechet wa-differentiability. The convexity of the metric ρ is a convenient assumption as it will be readily seen (see Proposition 50).

Definition 49 *A function $f : C \to \mathbb{R}$ is* Frechet wa-differentiable *at $x \in C$ if there exists a ρ-continuous and affine functional $A : C \to \mathbb{R}$, with $A(x) = 0$, such that*

$$\lim_{n\to\infty}\frac{f\left(x_n\right)-f\left(x\right)-A\left(x_n\right)}{\rho\left(x_n,x\right)}=0 \tag{34}$$

for every sequence $\{x_n\}$ in C with $\rho(x_n, x) \to 0$.

We can also formulate a "strict" version in which we have

$$\lim_{n\to\infty}\frac{f\left(y_n\right)-f\left(x_n\right)-A\left(y_n\right)}{\rho\left(x_n,y_n\right)}=0 \tag{35}$$

for all $\rho(x_n, x) \to 0$ and $\rho(y_n, x) \to 0$.

Proposition 50 *Let ρ be convex. If $f : C \to \mathbb{R}$ is Frechet wa-differentiable at $x \in C$, then it is boundedly wa-differentiable at x.*

Proof Assume (34). Given a sequence $t_n \downarrow 0$ and any bounded sequence $\{x_n\}$, set $z_n = (1-t_n)\,x + t_n x_n$. Thanks to the convexity of ρ, we have $\rho(x, z_n) \le t_n \rho(x, x_n)$. Therefore, $\rho(x, z_n) \to 0$. In light of (34),

$$f\left(z_n\right)-f\left(x\right)-A\left(z_n\right)=\rho\left(z_n,x\right)o\left(1\right)$$

and

$$f\left(\left(1-t_n\right)x+t_nx_n\right)-f\left(x\right)-t_nA\left(x_n\right)=t_n\left[\frac{\rho\left(z_n,x\right)}{t_n}\right]o\left(1\right)=o\left(t_n\right)$$

But $t_n^{-1}\rho(z_n, x) \le \rho(x, x_n)$, which is bounded. Therefore,

$$f\left(\left(1-t_n\right)x+t_nx_n\right)-f\left(x\right)-t_nA\left(x_n\right)=o\left(t_n\right)$$

Hence, f is boundedly wa-differentiable at x. ■

A consequence of this proposition is that the affine function A in (34) is unique (if exists), with $D_F f(x;\cdot) = A(\cdot)$. Moreover, a Frechet wa-differentiable function $f : C \to \mathbb{R}$ is Hadamard wa-differentiable, with $D_F f(x;\cdot) = D_H f(x;\cdot)$.

Condition (34) is a classical Frechet condition in which a distance replaces a norm. In fact, (34) can be written, as $\rho(y, x) \to 0$,

$$f\left(y\right)=f\left(x\right)+D_Ff\left(x;y\right)+o\left(\rho\left(y,x\right)\right)$$

When $D_F f(x;\cdot)$ is extendable, it becomes

$$f\left(y\right)=f\left(x\right)+\nabla_Ff\left(x\right)\left(y-x\right)+o\left(\rho\left(y,x\right)\right)$$

Next we give a converse of the last result.

Proposition 51 *Let ρ be homogeneous. If $f : C \to \mathbb{R}$ is boundedly wa-differentiable at $x \in \operatorname{int}_\rho C$, then it is Frechet wa-differentiable at x.*

Proof Let f be boundedly wa-differentiable at x. As $x \in \operatorname{int}_\rho C$, there is $\varepsilon > 0$ such that $\rho(x, y) \leq \varepsilon$ implies $y \in C$. Let $\{x_n\}$ be any sequence in C such that $\rho(x_n, x) \to 0$. Without loss of generality, we can assume that $x_n \neq x$ for all n. Define the new sequence

$$y_n = \frac{1}{t_n} x_n + \left(1 - \frac{1}{t_n}\right) x \quad ; \quad t_n = \frac{\rho(x_n, x)}{\varepsilon}$$

By construction, $x_n = (1 - t_n) x + t_n y_n$. Moreover, since ρ is homogeneous,

$$\rho(x, x_n) = t_n \rho(x, y_n)$$

Hence, $\rho(x, y_n) = \varepsilon$ and thus it is a ρ-bounded sequence contained in C. Using this sequence in (33) we get

$$\lim_{n\to\infty} \frac{f((1 - t_n) x + t_n y_n) - f(x) - t_n A(y_n)}{t_n} = 0$$

As $A(x_n) = t_n A(y_n)$, we obtain

$$\lim_{n\to\infty} \frac{f(x_n) - f(x) - A(x_n)}{\rho(x_n, x)} = 0$$

Therefore, f is Frechet wa-differentiable, with $D_F f(x; \cdot) = A(\cdot)$. ■

Example 52 The quadratic functional $Q : \Delta(Y) \to \mathbb{R}$ given by

$$Q(\mu) = \int \psi(x, y) \, \mathrm{d}\mu(x) \otimes \mathrm{d}\mu(y)$$

is boundedly wa-differentiable for any metric when Y is Polish. Indeed, for any sequence $\{\lambda_n\} \subseteq \Delta(X)$ and $t_n \downarrow 0$,

$$\frac{Q((1 - t_n) \mu + t_n \lambda_n) - Q(\mu)}{t_n} - DQ(\mu; \lambda_n) = t_n Q(\mu) + t_n Q(\lambda_n) - 2t_n \int \psi(x, y) \, \mathrm{d}\mu \otimes \mathrm{d}\lambda_n$$

This quantity vanishes as $n \to \infty$. This fact does not imply that Q is Frechet wa-differentiable, as it will be momentarily seen.

The quadratic functional Q is Frechet wa-differentiable for the variation metric. Indeed,

$$Q(\lambda_n) - Q(\mu) - DQ(\mu; \lambda_n) = Q(\lambda_n - \mu)$$

and so

$$\begin{aligned} |Q(\lambda_n - \mu)| &= \left| \int \psi(x, y) \, \mathrm{d}(\lambda_n - \mu) \otimes \mathrm{d}(\lambda_n - \mu) \right| = \left| \mathrm{d}(\lambda_n - \mu)(y) \int \psi(x, y) \, \mathrm{d}(\lambda_n - \mu)(x) \right| \\ &\leq \|\psi\|_\infty \|\lambda_n - \mu\|_V \left| \int \mathrm{d}(\lambda_n - \mu) \right| \leq \|\psi\|_\infty \|\lambda_n - \mu\|_V^2 \end{aligned}$$

This implies

$$\frac{Q(\lambda_n) - Q(\mu) - DQ(\mu; \lambda_n)}{\rho_V(\lambda_n, \mu)} \to 0$$

as $\rho_V(\lambda_n, \mu) \to 0$ and so (34) holds.

The quadratic functional may fail to be Frechet wa-differentiable. Take $Y = \mathbb{R}$ and set $\psi(x, y) = f(x) \vee f(y)$, where $f : \mathbb{R} \to \mathbb{R}$ is non-constant and differentiable. Consider the Prokhorov metric ρ_P.

Pick a point $x \in \mathbb{R}$ for which $f'(x) \neq 0$, say $f'(x) > 0$ (the case $f'(x) < 0$ is similar). Let $y > x$ be sufficiently close to x. Then,

$$Q(\delta_y) - Q(\delta_x) - DQ(\delta_x; \delta_y) = Q(\delta_y - \delta_x) = f(x) + f(y) - 2f(y) = f(x) - f(y)$$

Since $\rho_P(\delta_y, \delta_x) = \min\{|y - x|; 1\}$, it follows that

$$\lim_{x \downarrow x} \frac{Q(\delta_y - \delta_x)}{\rho_P(\delta_y, \delta_x)} = \lim_{x \downarrow x} \frac{f(x) - f(y)}{y - x} = -f'(x) \neq 0$$

Hence, Q is not Frechet wa-differentiable at δ_x. ▲

Example 53 Define $T : \mathcal{D} \to \mathbb{R}$ by

$$T(F) = \int [F(x) - F_0(x)]^2 \, \mathrm{d}F_0(x)$$

where $\mathcal{D}$ is the collection of all probability distribution on $\mathbb{R}$. For an empirical distributions F_n, $T(F_n)$ is the Cramer-von Mises test statistic for the test problem H_0: $F = F_0$ versus H_1: $F \neq F_0$.

The wa-differential of T is

$$DT(F; G) = 2 \int (F - F_0)(G - F) \, \mathrm{d}F_0 \qquad \forall G \in \mathcal{D}$$

Notice that $DT(F_0; \cdot) = 0$. Indeed, T is a convex functional with F_0 as a minimizer. We show that T is strictly Frechet wa-differentiable at every point F under the Kolmogorov metric ρ_∞ (see also [21]). Let $\rho_\infty(F_n, F) \to 0$ and $\rho_\infty(G_n, F) \to 0$ be two sequences. Some tedious algebra shows that

$$R = T(G_n) - T(F_n) - DT(F_n; G_n) = \int (G_n - F_n)^2 \, \mathrm{d}F_0$$

Hence,

$$|R| \leq [\rho_\infty(G_n, F_n)]^2 \leq \rho_\infty(G_n, F_n) [\rho_\infty(G_n, F) + \rho_\infty(F, F_n)] = o(\rho_\infty(G_n, F_n))$$

which is (35). ▲

Interpret the functional $\theta = T(F)$ as the parameter of an unknown population − more generally $\theta = T(\mu)$, with $\mu \in \Delta(Y)$. Inferences about θ are usually based on the statistic $\hat{\theta} = T(F_n)$, where F_n is the empirical distribution

$$F_n = \frac{1}{n} \sum_{i=1}^{n} G_{x_i}$$

If T is a-differentiable, we have

$$T(F_n) = T(F) + \int \nabla_a T(F) \, \mathrm{d}(F_n - F) + R(F; F_n)$$

where $R(F; F_n)$ is the remainder. By setting $\nabla_a T(F) = u_F(x)$ (the influence function, i.e., the normalized gradient), we obtain

$$T(F_n) = T(F) + \int u_F \mathrm{d}F_n - \int u_F \mathrm{d}F + R(F; F_n) = T(F) + \frac{1}{n} \sum_{i=1}^{n} u_F(x_i) + R(F; F_n)$$

Suppose that an appropriately defined notion of differentiability guarantees that

$$R(F; F_n) = o_p\left(\frac{1}{\sqrt{n}}\right)$$

The normalization ensures that the mean value $\int u_F(x)\,\mathrm{d}F(x)$ is zero. By assuming that $\sigma^2 = \int u_F^2(x)\,\mathrm{d}F(x)$ is finite, then

$$\sqrt{n}\left[T(F_n) - T(F)\right] = \sqrt{n}\sum_{i=1}^{n} u_F(x_i) + o_p(1)$$

Slutsky's Lemma and the central limit theorem imply the asymptotic normality, i.e.,

$$\sqrt{n}\left[T(F_n) - T(F)\right] \to_d \mathcal{N}\left(0, \sigma^2\right)$$

Next we consider a simple case (see, e.g., [23]).

Proposition 54 *Assume that for some metric ρ we have*

$$\rho(F_n, F) = O_p\left(\frac{1}{\sqrt{n}}\right)$$

If T is Frechet differentiable at F, with normalized gradient $\nabla T(F) = u_F$, then

$$\sqrt{n}\left[T(F_n) - T(F)\right] \to_d \mathcal{N}\left(0, \sigma^2\right)$$

Proof By (34),

$$T(F_n) - T(F) = \int u_F dF_n + \rho(F_n, F)\,o(1)$$

$$\sqrt{n}\left[T(F_n) - T(F)\right] = \sqrt{n}\sum_{i=1}^{n} u_F(x_i) + \frac{\sqrt{n}O_p(1)o(1)}{\sqrt{n}} = \sqrt{n}\sum_{i=1}^{n} u_F(x_i) + o(1)$$

as desired. ■

7 Appendix

We begin with a well-known result, proved for the sake of completeness. Here $D^+ f(t)$ indicates the right upper Dini derivative.

Proposition 55 *Let $f : [a, b] \to \mathbb{R}$ be continuous. Assume that $t \to D^+ f(t)$ exists and is continuous on (a, b). Then,*

(i) there is $\tau \in (a, b)$ such that $f(b) - f(a) = D^+ f(\tau)(b - a)$;

(ii) f is continuously differentiable on (a, b), i.e.,

$$D^+ f(t) = D^- f(t) = D_+ f(t) = D_- f(t) \qquad \forall t \in (a, b)$$

Proof We prove only (ii) since (i) then follows from the standard Mean Value Theorem. Define the scalar function

$$F(t) = f(t_0) + \int_{t_0}^{t} D^+ f(\tau)\, d\tau \qquad \forall t \in (a, b)$$

where t_0 is a fixed element in (a, b). By the continuity of $D^+ f(t)$, the function F is well defined, continuous and differentiable on (a, b), with $DF(t) = D^+ f(t)$. By construction, on (a, b) we have

$$D^+\left[f - F\right] = D^+ f - DF = 0$$

A standard result then implies that $f - F$ is constant on $[a, b]$. As $F(t_0) = f(t_0)$, we get $f = F$. Point (ii) is proved. ■

An implication of the previous result is the following version of Mean Value Theorem without the assumption of a-differentiability.

Proposition 56 *Let $f : C \to \mathbb{R}$. Assume that:*

(i) the limit

$$Df(x;y) = \lim_{t \downarrow 0} \frac{f((1-t)x + ty) - f(x)}{t}$$

exists and is finite for all $x, y \in C$;

(ii) f is hemicontinuous on C;[12]

(iii) the function $t \mapsto Df(x_t; y)$ is continuous on $(0,1)$.
Then, for every $x, y \in C$ there is $t \in (0,1)$ such that

$$f(y) - f(x) = \frac{1}{1-t} Df(x_t; y) = -\frac{1}{t} Df(x_t; x) \tag{36}$$

Proof Set $\varphi(t) = f(x_t)$. By (ii), φ is continuous on $[0,1]$. As previously remarked, inspection of the the proof of Lemma 17 shows that the right derivative of φ exists and

$$\varphi'_+(t) = \frac{1}{1-t} Df(x_t; y) \qquad \forall t \in [0,1)$$

By (iii), φ'_+ is continuous. We can then apply Proposition 55. Hence,

$$f(y) - f(x) = \varphi(1) - \varphi(0) = \varphi'_+(\tau) = \frac{1}{1-\tau} Df(x_\tau; y).$$

In light of (ii) of Proposition 55, it holds $\varphi'_-(t) = \varphi'_+(t)$ for every $t \in (0,1)$. Hence,

$$\frac{1}{1-t} Df(x_t; y) = -\frac{1}{t} Df(x_t; x)$$

as desired. ■

Proposition 57 *Let $f : C \to \mathbb{R}$. Under conditions (i)-(iii) of Proposition 56 and the additional condition*

$$\lim_{t \downarrow 0} Df(x_t; z) = Df(x; z) \qquad \forall x, y, z \in C \tag{37}$$

then

$$Df(x; \alpha y + (1-\alpha)z) = \alpha Df(x; y) + (1-\alpha) Df(x; z) \tag{38}$$

for all $x, y, z \in C$ and all $\alpha \in [0,1]$.

Proof Set $\bar{\alpha} = 1 - \alpha$. Write

$$f(x + t[\alpha y + \bar{\alpha} z - x]) - f(x) = f(x + t\alpha[y - x] + t\bar{\alpha}[z - x]) - f(x) = A + B$$

where

$$A = f(x + t\alpha[y-x] + t\bar{\alpha}[z-x]) - f(x + t\bar{\alpha}[z-x]) \quad ; \quad B = f(x + t\bar{\alpha}[z-x]) - f(x)$$

Clearly,

$$Df(x; \alpha y + \bar{\alpha} z) = \lim_{t \downarrow 0} \frac{A}{t} + \lim_{t \downarrow 0} \frac{B}{t} = \lim_{t \downarrow 0} \frac{A}{t} + \lim_{t \downarrow 0} \frac{f(x + t\bar{\alpha}[z-x]) - f(x)}{t} = \lim_{t \downarrow 0} \frac{A}{t} + \bar{\alpha} Df(x; z)$$

[12] Hemicontinuity is the continuity analog of hemidiffentiability.

provided the limit exists. On the other hand, if we apply (36) to the increment A, we obtain that

$$A = \frac{1}{1-\tau} Df\left(x + \tau t\alpha\left[y-x\right] + t\bar{\alpha}\left[z-x\right]; x + t\alpha\left[y-x\right] + t\bar{\alpha}\left[z-x\right]\right)$$

holds for some $\tau \in (0,1)$. Tedious algebra shows that

$$x + t\alpha\left[y-x\right] + t\bar{\alpha}\left[z-x\right] = (1-\beta)\left[x + t\alpha\left[y-x\right] + t\bar{\alpha}\left[z-x\right]\right] + \beta y$$

where

$$\beta = \frac{t\alpha(1-\tau)}{1-\tau t\alpha} \in (0,1)$$

By (3),

$$A = \frac{1}{1-\tau} \cdot \frac{t\alpha(1-\tau)}{1-\tau t\alpha} Df\left(x + \tau t\alpha\left[y-x\right] + t\bar{\alpha}\left[z-x\right]; y\right)$$

Therefore,

$$\frac{A}{t} = \frac{\alpha}{1-\tau t\alpha} Df\left(x + t\left[\tau\alpha\left[y-x\right] + \bar{\alpha}\left[z-x\right]\right]; y\right).$$

Thanks to (37), $\lim_{t\downarrow 0} A/t = \alpha Df(x;y)$ and this completes the proof. ■

References

[1] S. Basu, Posterior sensitivity to the sampling distribution and the prior: more than one observation, *Annals of the Institute of Statistical Mathematics*, 51, 499-513, 1999.

[2] J.O. Berger, Robust Bayesian analysis: sensitivity to the prior, *Journal of Statistical Planning and Inference*, 25, 303-328, 1990.

[3] J.O. Berger, D. Rios Insua and F. Ruggeri, Bayesian robustness, in *Robust Bayesian Analysis*, Springer, New York, 2000.

[4] N.L. Charothers, *Real analysis*, Cambridge University Press, Cambridge.

[5] S. Cerreia-Vioglio, D. Dillenberger and P. Ortoleva, Cautious expected utility and the certainty effect, *Econometrica*, 83, 693-728, 2015.

[6] S. Cerreia-Vioglio, F. Maccheroni and M. Marinacci, Stochastic dominance without the independence axiom, *Management Science*, 63, 1097-1109, 2017.

[7] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci and L. Montrucchio, On the equality of Clarke and Greenberg-Pierskalla differentials, *Journal of Convex Analysis*, 29, 459-480, 2022.

[8] S.H. Chew, E. Karni, and Z. Safra, Risk aversion in the theory of expected utility with rank dependent probabilities, *Journal of Economic Theory*, 42, 370-381, 1987.

[9] S.H. Chew, L.G. Epstein and U. Segal, Mixture symmetry and quadratic utility *Econometrica*, 59, 139-163, 1991.

[10] J.M. Danskin, The theory of max-min, with applications, *SIAM Journal on Applied Mathematics*, 14, 641-664, 1966.

[11] R.M. Dudley, *Real analysis and probability*, Cambridge University Press, Cambridge, 2004.

[12] P. Ghirardato, F. Maccheroni and M. Marinacci, Differentiating ambiguity and ambiguity attitude, *Journal of Economic Theory*, 118, 133-173, 2004.

[13] R.D. Gill, Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part 1), *Scandinavian Journal of Statistics*, 16, 97-128, 1989.

[14] F.R. Hampel, The influence curve and its role in robust estimation, *Journal of the American Statistical Association*, 69, 383-393, 1974.

[15] P.J. Huber, *Robust statistics*, Springer, Berlin, 2011.

[16] J.B. Kadane and D.T. Chuang, Stable decision problems, *Annals of Statistics*, 1095-1110, 1978.

[17] M.J. Machina, Expected Utility analysis without the independence axiom, *Econometrica*, 50, 277-323, 1982.

[18] R.R. Phelps, *Lectures on Choquet's theorem,* 2$^{\text{nd}}$ ed., Springer, Berlin, 2001.

[19] J.A. Reeds, *On the definition of von Mises functionals*. Ph.D. dissertation, Harvard University, Cambridge, 1976.

[20] A. Shapiro, On concepts of directional differentiability, *Journal of Optimization Theory and Applications*, 66, 477-487, 1990.

[21] J. Shao, Differentiability of statistical functionals and consistency of the jackknife, *Annals of Statistics*, 21, 61-75, 1993.

[22] L. Stanca, Robust Bayesian choice, *Mathematical Social Sciences*, 126, 94-106, 2023.

[23] L.T. Fernholz, *von Mises calculus for statistical functionals*, Lecture Notes in Statistics, vol. 19, Springer, Berlin, 1983.

[24] A. Tversky and D. Kahneman, Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and Uncertainty*, 5, 297-323, 1992.

[25] R. von Mises, On the asymptotic distribution of differentiable statistical functions, *Annals of Mathematical Statistics*, 18, 309-348, 1947.

[26] P. P. Wakker, *Prospect theory: For risk and ambiguity*, Cambridge University Press, Cambridge, 2010.