SUBDIFFERENTIATION AND SMOOTHING OF NONSMOOTH INTEGRAL FUNCTIONALS[*]

JAMES V. BURKE[†], XIAOJUN CHEN[‡], AND HAILIN SUN[§]

Abstract. The subdifferential calculus for the expectation of nonsmooth random integrands involves many fundamental and challenging problems in stochastic optimization. It is known that for Clarke regular integrands, the Clarke subdifferential equals the expectation of their Clarke subdifferential. In particular, this holds for convex integrands. However, little is known about calculation of Clarke subgradients for the expectation for non-regular integrands. The focus of this contribution is to approximate Clarke subgradients for the expectation of random integrands by smoothing methods applied to the integrand. A framework for how to proceed along this path is developed and then applied to a class of *measurable composite max integrands*, or CM integrands. This class contains non-regular integrands from stochastic complementarity problems as well as stochastic optimization problems arising in statistical learning.

Key words. Stochastic optimization, Clarke subgradient, smoothing, non-regular integrands.

AMS subject classifications. 90C15, 90C46

1. Introduction. Let $X \subseteq \mathbb{R}^n$ be a convex compact set with non-empty interior and $\Xi \subseteq \mathbb{R}^\ell$ be a Lebesgue measurable closed set with non-empty interior. In this paper, we consider the stochastic optimization problem

$$\min_{x\in X} \quad F(x) := \mathbb{E}[f(\xi, x)], \tag{1.1}$$

where $\xi : \Omega \to \Xi$ is a random variable on the induced probability space $(\Omega, \mathcal{A}, \rho)$, $f : \Xi \times X \to \mathbb{R}$ is continuous on X and measurable in Ξ for every $x \in X$, and $\mathbb{E}[\cdot]$ denotes the expected value over Ξ. A point $x \in \mathbb{R}^n$ is called a Clarke stationary point for (1.1) if it satisfies

$$0 \in \partial F(x) + \mathcal{N}_X(x), \tag{1.2}$$

where ∂ denotes the Clarke subdifferential and $\mathcal{N}_X(x)$ is the normal cone to X at $x \in X$. Condition (1.2) is a first-order necessary condition for optimality of problem (1.1).

The subdifferential $\partial F(x) = \partial\mathbb{E}[f(\xi, x)]$ does not in general have a closed form and is difficult to calculate. Consequently, much the existing literature [21, 22, 23] employs the first-order necessary condition

$$0 \in \mathbb{E}[\partial_x f(\xi, x)] + \mathcal{N}_X(x), \tag{1.3}$$

where

$$\mathbb{E}[\partial_x f(\xi, x)] := \{\mathbb{E}[\varphi(\xi, x)] \,|\, \varphi(\xi, x) \text{ is a measurable selection from } \partial_x f(\xi, x)\}$$

[*]Submitted to the editor 12 May, 2017.

Funding: The second author's work is supported partly by Hong Kong Research Grant Council grant PolyU153000/15p, the third author's work is supported partly by the National Natural Science Foundation of China (Grant No. 11401308)

[†]Department of Applied Mathematics, University of Washington, Seattle, (jvburke@uw.edu).

[‡]Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, (maxjchen@polyu.edu.hk).

[§]School of Economics and Management, Nanjing University of Science and Technology, Nanjing, (hlsun@njust.edu.cn).

is the Aumann (set-valued) expectation of $\partial_x f(\xi, x)$ with respect to ξ defined in [2]. Points x satisfying (1.3) are called *weak stationary points* for problem (1.1) [16, 24]. In some cases, elements of $\mathbb{E}[\partial_x f(\xi, x)]$ can be computed. However, as the name implies, condition (1.3) is much weaker than condition (1.2). In particular, it is always the case that

$$\partial F(x) \subseteq \mathbb{E}[\partial_x f(\xi, x)], \tag{1.4}$$

and so if x satisfies (1.2), then x satisfies (1.3), but the converse is not in general true. On the other hand, [12, Theorem 2.7.2] tells us that if $f(\xi, \cdot)$ is Clarke regular [12, Definition 2.3.4] on X for almost all $\xi \in \Xi$, then equality holds in (1.4). Unfortunately, in many applications, Clarke regularity fails to hold, and the set $\mathbb{E}[\partial_x f(\xi, x)]$ is much larger than the set $\partial\mathbb{E}[f(\xi, x)]$. In such cases, condition (1.3) is too weak for assessing optimality. For example, consider $f(\xi, x) = \xi|x|$ with $\xi \sim N(0, 1)$ and $x \in R$. Then $\mathbb{E}[f(\xi, x)] = \mathbb{E}[\xi|x|] \equiv 0$ for $x \in R$, but $\mathbb{E}[\partial f(\xi, 0)] = \sqrt{\pi/2}[-1, 1]$.

The main contributions of the paper are the development of a framework for the study of smoothing methods for the expectation of random integrands $F(x) = \mathbb{E}[f(\xi, x)]$ based on the smoothing of the integrand f, a smoothing approach to the approximation of the Clarke subgradients of expectation $F(x) = \mathbb{E}[f(\xi, x)]$, and the application of these techniques to the class of measurable composite max (CM) integrands. CM intergrands arise in several important applications including stochastic nonlinear complementarity problems [10, 11] and optimization problems in statistical learning [1, 3].

The paper is organized as follows. In Sections 2 and 3, we recall some basic definitions and properties from variational analysis, the theory of measurable multifunctions, and the variational properties of the expectation function, respectively. In Section 4, we define measurable smoothing functions and present some useful properties. In Section 5, we present an approximation theory of smoothing functions for measurable CM functions, and prove the gradient sub-consistency of CM integrands. Finally we show that the subgradient of the expectation function can be computed via smoothing without regularity.

2. Background.

2.1. Finite-dimensional variational analysis. Since we allow mappings to have infinite values, it is convenient to define the extended reals $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$. The *effective domain* of $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, denoted $\operatorname{dom} f \subseteq \mathbb{R}^n$, is the set on which f is finite. To avoid certain pathological mappings the discussion is restricted to *proper* not everywhere infinite *lower semi-continuous* (lsc) functions. Of particular importance is the *epigraph* of such functions: $\operatorname{epi} f := \{(x, \mu) \mid f(x) < \mu\}$. We have that f is lsc if and only if $\operatorname{epi} f$ is closed, and f is convex if and only if $\operatorname{epi} f$ is convex.

DEFINITION 2.1 (Subderivatives). *[19, Exercise 9.15] For a locally Lipschitz function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ near a point $u_* \in \mathbb{R}^n$ with $f(u_*)$ finite,*

(i) the subderivative $df(u_*) : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is defined by*

$$df(u_*)(w) := \liminf_{\tau\downarrow 0} \frac{f(u_* + \tau w) - f(u_*)}{\tau};$$

(ii) the regular subderivative *(or the Clarke generalized directional derivative when f is locally Lipschitz)* $\widehat{d}f(u_*) : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is defined by*

$$\widehat{d}f(u_*)(w) := \limsup_{u\to u_*,\, \tau\downarrow 0} \frac{f(u + \tau w) - f(u)}{\tau}.$$

DEFINITION 2.2 (Clarke Subgradients and Subdifferential Regularity). *Consider a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, a point $v \in \mathbb{R}^n$, and a point $u_* \in \mathbb{R}^n$ with $f(u_*)$ finite.*

(i) *[19, Theorem 8.49] The vector v is a* Clarke subgradient *of f at u_* if f is l.s.c. on a neighborhood of u_* and v satisfies*

$$\widehat{d}f(u_*)(w) \geq \langle v\,,\, w\rangle \;\; \forall\; w \in \mathbb{R}^n.$$

We call the set of Clarke subgradients v the Clarke subdifferential *of f at u_* and denote this set by $\partial f(u_*)$.*

(ii) *[19, Corollary 8.19] A locally Lipschitz function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be* subdifferentially regular *(or Clarke regular) at $u_* \in dom\, f$ with $\partial f(u_*) \neq \emptyset$ if*

$$df(u_*)(w) = \widehat{d}f(u_*)(w) \quad \forall\, w \in \mathbb{R}^n.$$

Remark 2.3. In [19], the notion of subdifferential regularity is defined in [19, Definition 7.25]. In the definition given above we employ characterizations of this notion given by the cited results. Note that subdifferential mappings are multi-functions.

DEFINITION 2.4 (Strict Continuity and Strict Differentiability). *Let $H : D \to \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, and $h : \mathbb{R}^n \to \overline{\mathbb{R}}$.*

(a) *[Strict Continuity [19, Definition 9.1]] We say that H is strictly continuous at $\bar{x} \in int(D)$ if*

$$\operatorname{lip} H(\bar{x}) := \limsup_{\substack{x,x' \to \bar{x} \\ x \neq x'}} \frac{\|H(x') - H(x)\|}{\|x' - x\|} < \infty\,.$$

(b) *[Strict Differentiability [19, Definition 9.17]] We say that h is strictly differentiable at a point $\bar{x} \in dom\, h$ if h is differentiable at $\bar{x}$ and*

$$\lim_{\substack{x,x' \to \bar{x} \\ x \neq x'}} \frac{h(x') - h(x) - \langle \nabla h(\bar{x})\,,\, x' - x\rangle}{\|x' - x\|} = 0\,.$$

It is easily seen that if h is continuously differentiable on an open set U, then h is strictly differentiable and subdifferentially regular on U with $\partial h(x) = \{\nabla f(x)\}$ for all $x \in U$ ([19, Theorem 9.18 and Exercise 9.64]).

The notion of strict continuity of f at a point $\bar{x}$ implies the existence of a neighborhood of $\bar{x}$ on which f is Lipschitz continuous, that is, f is locally Lipschitz continuous at $\bar{x}$ where the local Lipschitz modulus is lower bounded by $\operatorname{lip} H(\bar{x})$. In this light, Definition 2.1 and Definition 2.2(ii) combine to tell us that

$$df(u_*)(w) = \widehat{d}f(u_*)(w) = \lim_{\tau \downarrow 0} \frac{f(u_* + \tau w) - f(u_*)}{\tau} \quad \forall\, w \in \mathbb{R}^n, \tag{2.1}$$

wherever f is strictly continuous and subdifferentially regular at u_*. Moreover, in this case, [19, Theorem 8.30] tells us that

$$d_x f(\xi, x)(v) = \sup\{\langle g\,,\, v\rangle \mid g \in \partial_x f(\xi, x)\}\,. \tag{2.2}$$

Remark 2.5 (Subdifferentials of Compositions). If $g : X \to \overline{\mathbb{R}}$ is given as the composition of two functions $f : Y \to \overline{\mathbb{R}}$ and $h : X \to Y$, i.e. $g(x) = (f \circ h)(x) = f(h(x))$, then we write $\partial g(x) = \partial(f \circ h)(x)$. On the other hand, we write $\partial f(h(x))$ to denote the subdifferential of f evaluated at $h(x)$.

THEOREM 2.6 (Strict Differentiability and the Subdifferential). *[19, Theorem 9.18] [12, Proposition 2.2.4] Let $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ with $\bar{x} \in dom\, h$. Then h is strictly differentiable at $\bar{x}$ if and only if h is strictly continuous at $\bar{x}$ and $\partial h(\bar{x}) = \{\nabla h(\bar{x})\}$.*

2.2. Measurable multi-functions. We review some of the properties of measurable multi-functions used in this paper [2, 13, 15, 19]. For more information on this topic, we refer the interested reader to [19, Chapter 14] and [18].

A multi-function, or multi-valued mapping, S from $\mathbb{R}^k$ to $\mathbb{R}^s$ is a mapping that takes points in $\mathbb{R}^k$ to sets in $\mathbb{R}^s$, and is denoted by $\Psi : \mathbb{R}^k \rightrightarrows \mathbb{R}^s$. The *outer limit* of S at $\bar{x} \in \mathbb{R}^k$ *relative to* $X \subseteq \mathbb{R}^k$ is

$$\operatorname*{Limsup}_{x \to_X \bar{x}} S(x) := \left\{ v \mid \exists \{x^k\} \to_X \bar{x}, \{v^k\} \to v : v^k \in S(x^k) \quad \forall k \in \mathbb{N} \right\} \tag{2.3}$$

and the *inner limit* of S at $\bar{x}$ *relative to* X is

$$\operatorname*{Liminf}_{x \to_X \bar{x}} S(x) := \left\{ v \mid \forall \{x^k\} \to_X \bar{x}, \exists \{v^k\} \to v : v^k \in S(x^k) \quad \forall k \in \mathbb{N} \right\}.$$

Here the notation $\{x^k\} \to_X \bar{x}$ means that $\{x^k\} \subseteq X$ with $x^k \to \bar{x}$. If $X = \mathbb{R}^k$, we write $x \to \bar{x}$ instead of $x \to_{\mathbb{R}^k} \bar{x}$. We say that S is *outer semicontinuous (osc)* at $\bar{x}$ *relative to* X if

$$\operatorname*{Limsup}_{x \to_X \bar{x}} S(x) \subseteq S(\bar{x}).$$

When the outer and inner limits coincide, we write

$$\operatorname*{Lim}_{x \to_X \bar{x}} S(x) := \operatorname*{Limsup}_{x \to_X \bar{x}} S(x),$$

and say that S is *contiuous* at $\bar{x}$ *relative to* X.

Let Ω be a nonempty subset of $\mathbb{R}^\ell$ and let $\mathcal{A}$ be a σ-field of subsets of Ω, called the *measurable* subsets of Ω or the $\mathcal{A}$*-measurable* subsets. Let $\rho : \mathcal{A} \to [0, 1]$ be a σ-finite Borel regular, complete, non-atomic, probability measure on $\mathcal{A}$. The corresponding measure space is denoted $(\Omega, \mathcal{A}, \rho)$. A multi-function $\Psi : \Omega \rightrightarrows \{ R^n \}$ is said to be $\mathcal{A}$-measurable, or simply measurable, if for all open sets $\{ V \} \subseteq \mathbb{R}^n$ the set $\{\xi \mid \{ V \} \cap \Psi(\xi) \neq \emptyset\}$ is in $\mathcal{A}$. The multi-function Ψ is said to be $\mathcal{A} \otimes \mathcal{B}^n$*-measurable* if $\operatorname{gph}(\Psi) = \{(\xi, v) \mid v \in \Psi(\xi)\} \in \mathcal{A} \otimes \mathcal{B}^n$, where $\mathcal{B}^n$ denotes the Borel σ-field on $\mathbb{R}^n$ and $\mathcal{A} \otimes \mathcal{B}^n$ is the σ-field on $\Omega \times \mathbb{R}^n$ generated by all sets $A \times D$ with $A \in \mathcal{A}$ and $D \in \mathcal{B}^n$. If $\Psi(\xi)$ is closed for each ξ then Ψ is *closed-valued*. Similarly, Ψ is said to be *convex-valued* if $\Psi(\xi)$ is convex for each ξ. Finally, we note that the completeness of the measure space guarantees the measurability of subsets of Ω obtained as the projections of measurable subsets $\{ G \}$ of $\Omega \times \mathbb{R}^n$:

$$\{ G \} \in \mathcal{A} \otimes \mathcal{B}^n \quad \Longrightarrow \quad \{\xi \in \Omega \mid \exists\, \omega \in \mathbb{R}^n \text{ with } (\xi, \omega) \in \{ G \}\} \in \mathcal{A}.$$

In particular, this implies that the multi-function Ψ is $\mathcal{A}$-measurable if and only if $\operatorname{gph}(\Psi)$ is $\mathcal{A} \otimes \mathcal{B}^n$-measurable [19, Theorem 14.8].

Let $\Psi : \Omega \rightrightarrows \{ R^n \}$, and denote by $\mathcal{S}(\Psi)$ the set of ρ-measurable functions $f : \Omega \to \mathbb{R}^n$ that satisfy $f(\xi) \in \Psi(\xi)$ a.e. in Ω $(\xi \in \Omega)$. We call $\mathcal{S}(\Psi)$ the *set of measurable selections* of Ψ.

Theorem 2.7 (Measurable Selections). *[19, Corollary 14.6] A closed-valued measurable map* $\Psi : \Omega \rightrightarrows \mathbb{R}^n$ *always admits a measurable selection.*

We say that the measurable multi-function $\Psi : \Omega \rightrightarrows \mathbb{R}^n$ is *integrably bounded*, or for emphasis ρ-integrably bounded, if there is a ρ-integrable function $a : \Omega \to \mathbb{R}^n_+$ such that

$$(|v_1|, \ldots, |v_n|) \leq a(\xi) \tag{2.4}$$

for all pairs $(\xi, v) \in \Omega \times \mathbb{R}^n$ satisfying $v \in \Psi(\xi)$. Here and elsewhere we interpret vector inequalities as element-wise inequalities. Let $1 \leq p \leq \infty$. When $\Omega = \mathbb{R}^\ell$, we let $L^p_m(\mathbb{R}^\ell, \mathcal{A}, \rho)$ denote the Banach space of functions mapping $\mathbb{R}^\ell$ to $\mathbb{R}^m$. When $p = 2$, $L^2_m(\mathbb{R}^\ell, \mathcal{A}, \rho)$ is a Hilbert space with the inner product on the measure space $(\mathbb{R}^\ell, \mathcal{A}, \rho)$ given by

$$\langle \psi, \varphi \rangle_\rho = \int_{\mathbb{R}^\ell} \langle \psi(\xi)\, , \, \varphi(\xi) \rangle \, d\rho,$$

where $\langle \cdot \, , \, \cdot \rangle$ denotes the usual finite-dimensional vector inner product. If $\rho(\mathbb{R}^\ell) < \infty$, then

$$L^q_m(\mathbb{R}^\ell, \mathcal{A}, \rho) \subseteq L^p_m(\mathbb{R}^\ell, \mathcal{A}, \rho) \quad \text{whenever} \quad 1 \leq p \leq q \leq \infty.$$

If the function a in (2.4) is such that $\|a(\xi)\|_p$ is integrable with respect to the measure ρ on the measure space $(\Omega, \mathcal{A}, \rho)$, then the multi-function Ψ is said to be L^p-bounded.

Proposition 2.8. *[7, Proposition 2.2][Weak compactness of measurable selections] Let the multi-function $\Psi : \mathbb{R}^\ell \rightrightarrows \mathbb{R}^m$ be closed- and convex-valued, and L^2-bounded on $L^2_m(\mathbb{R}^\ell, \mathcal{M}^n, \lambda_n)$, where $\mathcal{M}^n$ is the Lebesgue field on $\mathbb{R}^n$ and λ_n is n-dimensional Lebesgue measure. Then the set of measurable selections $\mathcal{S}(\Psi)$ is a weakly compact, convex set in $L^2_m(\mathbb{R}^\ell, \mathcal{M}^n, \lambda_n)$.*

Given a measurable multi-function $\Psi : \Omega \rightrightarrows \mathbb{R}^n$, we define the integral of Ψ over Ω with respect to the measure ρ by

$$\int \Psi d\rho := \left\{ \int_\Omega f d\rho \mid f \in \mathcal{S}(\Psi) \right\}.$$

The next theorem, due to Hildenbrand [15], is a restatement of Theorems 3 and 4 of Aumann [2] for multi-functions on the non-atomic measure space $(\Omega, \mathcal{A}, \rho)$. These results are central to the theory of integrals of multi-valued functions.

Theorem 2.9 (Integrals of Multi-Functions). *[15, Theorem 4 and Proposition 7] The following properties hold for integrably bounded multi-functions $\Psi : \Omega \rightrightarrows \mathbb{R}^n$ on non-atomic measure spaces $(\Omega, \mathcal{A}, \rho)$.*

(i) If Ψ is $\mathcal{A} \otimes \mathcal{B}^n$-measurable, then $\int \Psi d\rho = \int \operatorname{conv} \Psi d\rho$.

(ii) If Ψ is closed (not necessarily $\mathcal{A} \otimes \mathcal{B}^n$-measurable), then $\int \Psi d\rho$ is compact.

We conclude this section with a very elementary, but useful lemma on measurable *tubes*, i.e. multi-valued mappings $\Psi : \Omega \rightrightarrows \mathbb{R}^n$ of the form

$$\Psi(\xi) := \kappa(\xi)\mathbb{B}, \tag{2.5}$$

where $\mathbb{B} := \{x \mid \|x\| \leq 1\}$ is the closed unit ball in $\mathbb{R}^n$ and $\kappa : \Omega \to \mathbb{R}_+$ is measurable.

Lemma 2.10 (Tubes). *Let $\Psi : \Omega \rightrightarrows \mathbb{R}^n$ be a measurable tube as in (2.5) with $\kappa \in L^2_1(\Omega, \mathcal{A}, \rho)$ non-negative a.e. on Ω. Then, for every $E \in \mathcal{A}$, $\int_E \Psi(\xi) d\rho \subseteq \left[\int_E \kappa(\xi) d\rho\right] \mathbb{B} \subseteq \|\kappa\|_2 \, \rho(E) \mathbb{B}$.*

Proof. The mapping Ψ in (2.5) is obviously closed valued and measurable. Therefore, Theorem 2.7 tells us that $\mathcal{S}(\Psi)$ is non-empty. Let $E \in \mathcal{A}$ and $s \in \mathcal{S}(\Psi)$. Then

$$|\int_E s(\xi) d\rho| \leq \int_E |s(\xi)| d\rho \leq \int_E \kappa(\xi) d\rho,$$

so that $\int_E s(\xi) d\rho \in \left[\int_E \kappa(\xi) d\rho\right] \mathbb{B}$. This proves the lemma since $\int_E \kappa(\xi) d\rho = \langle \kappa \, , \, \mathcal{X}_E \rangle \leq \|\kappa\|_2 \, \rho(E)$. ∎

3. The Variational Properties of $F(x) := \mathbb{E}[f(\xi, x)]$**.** We examine the relationship between the variational properties of $f(\xi, x)$ and those of $F(x) = \mathbb{E}[f(\xi, x)]$. Our approach is motivated by the case where f is specified during the modeling process in stochastic optimization, and we are asked to optimize its expectation. For this reason it is important to understand the properties that f should satisfy in order that the optimization of F is in some sense numerically tractable. For example, it has already been mentioned in (1.4) that, in general, we only have $\partial F(x) \subseteq \mathbb{E}[\partial_x f(\xi, x)]$. But there are situations in which equality holds. We begin by reviewing these results. The first step is to recall the standard conditions on f that imply the local Lipschitz continuity of F (e.g. see [12, Hypothesis 2.7.1]).

3.1. LL integrands. Let λ denote Lebesgue measure on $\mathbb{R}^n$ and let ρ be a probability measure on $\mathbb{R}^\ell$ that is absolutely continuous with respect to Lebesgue measure with support Ξ. Let $\hat{\rho}$ denote the induced product measure on $\mathbb{R}^\ell \times \mathbb{R}^n$. We consider the following class of functions.

Definition 3.1 (Carathéodory Mappings). *[19, Example 14.15] We say that the function $f : \Xi \times X \to \mathbb{R}$ is a Carathéodory mapping on $\Xi \times X$ if $f(\xi, \cdot)$ is continuous on an open set containing X for all $\xi \in \Xi$, and $f(\cdot, x)$ is measurable on Ξ for all $x \in X$.*

Definition 3.2 (Locally Lipschitz (LL) Integrands). *Let U be an open subset of $\mathbb{R}^n$. We say that $f : \Xi \times U \to \mathbb{R}$ is an LL integrand on $\Xi \times U$ if f is a Carathéodory mapping on $\Xi \times U$ and for each $\bar{x} \in U$ there is an $\epsilon(\bar{x}) > 0$ and an integrable mapping $\kappa_f(\cdot, \bar{x}) \in L_1^2(\mathbb{R}^\ell, \mathcal{M}, \rho)$ such that*

$$|f(\xi, x_1) - f(\xi, x_2)| \le \kappa_f(\xi, \bar{x})\|x_1 - x_2\| \quad \forall\; x_1, x_2 \in \mathbb{B}_{\epsilon(\bar{x})}(\bar{x}) \;\; \textit{and almost all } \xi \in \Xi,$$

where $\mathbb{B}_\epsilon(\bar{x}) := \{x \mid \|x - \bar{x}\| \le \epsilon\} \subseteq U$.

Lemma 3.3 (Properties of LL Integrands). *Let U be an open subset of $\mathbb{R}^n$, and let $f : \Xi \times U \to \mathbb{R}$ be an LL integrand on $\Xi \times U$ with $f(\cdot, x) \in L_1^1(\mathbb{R}^\ell, \mathcal{M}, \rho)$ for all $x \in U$. Then the following statements hold.*

(a) The function $f(\xi, \cdot)$ is strictly continuous on U (see Definition 2.4) for almost all $\xi \in \Xi$ with

$$\mathrm{lip}_x f(\xi, \bar{x}) \le \kappa_f(\xi, \bar{x}) \quad \textit{a.e. } \xi \in \Xi.$$

(b) The mapping $F(x) := \mathbb{E}[f(\xi, x)]$ is locally Lipschitz continuous on U with local Lipschitz modulus $\kappa_F(\bar{x}) := \mathbb{E}[\kappa_f(\xi, \bar{x})]$. In particular, F is strictly continuous on U.

(c) The function $\widehat{d_x} f(\xi, x)(v)$ is measurable in ξ for every $(x, v) \in U \times \mathbb{R}^n$.

(d) The set of measurable selections $\mathcal{S}(\partial_x f(\cdot, x))$ is a weakly compact, convex set in $L_n^2(\mathbb{R}^\ell, \mathcal{M}, \rho)$.

(e) The Clarke subdifferential $\partial F(x)$ is a nonempty, convex, compact subset of $\mathbb{R}^n$ contained in $\kappa_F(\bar{x})\mathbb{B}$ for every $x \in U$.

(f) For every $E \in \mathcal{M}$ such that $E \subseteq \Xi$ and every $\bar{x} \in U$

$$\int_E f(\xi, x) d\rho \in \int_E f(\xi, \bar{x}) d\rho + \|\kappa_f(\cdot, \bar{x})\|_2\, \rho(E)\mathbb{B} \quad \textit{and}$$
$$\int_E \partial_x f(\xi, x) d\rho \subseteq \|\kappa_f(\cdot, \bar{x})\|_2\, \rho(E)\mathbb{B}$$

for all $x \in \mathbb{B}_{\epsilon(\bar{x})}(\bar{x})$.

Proof. (a) This follows immediately from the definition of strict continuity.
(b) This follows immediately from the inequality

$$|F(x') - F(x)| \le \mathbb{E}[|f(\xi, x') - f(\xi, x)|] \le \mathbb{E}[\kappa_f(\xi, \bar{x})] \, \|x' - x\| \,.$$

(c) This follows from the well known fact that the limsup of measurable functions is measurable, e.g. [14, Theorem 2.7].
(d) This follows immediately from Proposition 2.8 since $\kappa_f(\cdot, \bar{x}) \in L^2_1(\mathbb{R}^\ell, \mathcal{M}, \rho)$.
(e) This is an immediate consequence of [12, Proposition 2.1.2].
(f) By definition, $f(\xi, x) - f(\xi, \bar{x}) \in \kappa_f(\xi, \bar{x})\mathbb{B}$ for all $x \in \epsilon(\bar{x})\mathbb{B}$ and $\xi \in \Xi$. Hence, for all $x \in \epsilon(\bar{x})\mathbb{B}$, $f(\xi, x) - f(\xi, \bar{x})$ is a measurable selection from the tube $\kappa_f(\xi, \bar{x})\mathbb{B}$ on Ξ. Similarly, by [19, Theorem 9.13], any measurable selection $s(\xi)$ from $\partial_x f(\xi, x)$ satisfies $s(\xi) \in \kappa_f(\xi, \bar{x})\mathbb{B}$ for all $x \in \epsilon(\bar{x})\mathbb{B}$. Therefore, both inclusions follows from Lemma 2.5. ∎

3.2. Subdifferential regularity. If f is an LL integrand on $\Xi \times U$, then, by Lemma 3.3(e), $\partial F(x)$ is a nonempty, convex, compact subset of $\mathbb{R}^n$ for every $x \in U$. But this does not say that $\partial F(x)$ is representable in terms of $\partial f(\xi, x)$.

THEOREM 3.4 (The Subdifferential of F). *[12, Theorem 2.7.2] Let U be an open subset of $\mathbb{R}^n$, and let $f : \Xi \times U \to \mathbb{R}$ be an LL integrand on $\Xi \times U$ with $f(\cdot, x) \in L^1_1(\mathbb{R}^\ell, \mathcal{M}, \rho)$ for all $x \in U$. Then*

$$\partial F(x) \subseteq \mathbb{E}[\partial_x f(\xi, x)] \quad \forall\, x \in U. \tag{3.1}$$

If, in addition, $\bar{x} \in U$ is such that $f(\xi, \cdot)$ is subdifferentially regular in x at $\bar{x}$ for almost all $\xi \in \Xi$, then F is subdifferentially regular at $\bar{x}$ and equality holds in (3.1).

Remark 3.5. In [12, Theorem 2.7.2], Clarke uses the hypothesis that $f(\xi, \cdot)$ *is subdifferentially regular in x at $\bar{x}$ for all $\xi \in \Xi$.* However, the above result holds with essentially the same proof.

COROLLARY 3.6. *Let U be an open subset of $\mathbb{R}^n$, and let $f : \Xi \times U \to \mathbb{R}$ be an LL integrand on $\Xi \times U$ with $f(\cdot, x) \in L^1_1(\mathbb{R}^\ell, \mathcal{M}, \rho)$ for all $x \in U$. If $\bar{x} \in U$ is such that either $f(\xi, \cdot)$ is subdifferentially regular at $\bar{x} \in U$ for almost all $\xi \in \Xi$ or $-f(\xi, \cdot)$ is subdifferentially regular at $\bar{x} \in U$ for almost all $\xi \in \Xi$, then equality holds in (3.1).*

Proof. If $f(\xi, \cdot)$ is subdifferentially regular in x at $\bar{x} \in U$ for almost all $\xi \in \Xi$, then the result follows from Theorem 3.4. If $-f(\xi, \cdot)$ is subdifferentially regular in x at $\bar{x}$ for almost all $\xi \in \Xi$, then, by [12, Proposition 2.3.1] and Theorem 3.4,

$$\partial F(\bar{x}) = -\partial(-F)(\bar{x}) = -\mathbb{E}[\partial(-f)(\xi, \bar{x})] = \mathbb{E}[\partial f(\xi, \bar{x})].$$

Note that, in opposition to Theorem 3.4, the corollary does not say that the hypotheses imply that F is subdifferentially regular at $\bar{x}$. Indeed, this may not be the case. The following example illustrates this possibility.

Example 3.7. Consider the Carathéodory function $f(\xi, x) := -|\xi||x|$, where $\xi \sim N(0, 1)$, $x \in \mathbb{R}$. It is easy to see that this function is not Clarke regular in x at $(\xi, 0)$ for all $\xi \neq 0$. In addition, the function F is not Clarke regular at $x = 0$. To see this, consider the subderivative and regular subderivative of F at $x = 0$:

$$dF(0)(w) = \liminf_{\tau \downarrow 0} \frac{\mathbb{E}[f(\xi, \tau w)] - \mathbb{E}[f(\xi, 0)]}{\tau} = -\mathbb{E}[|\xi|]\, |w| = -\sqrt{\frac{\pi}{2}}\, |w|$$

and

$$\widehat{d}F(0)(w) = \limsup_{x' \to 0,\, \tau \downarrow 0} \frac{\mathbb{E}[f(\xi, x' + \tau w)] - \mathbb{E}[f(\xi, x')]}{\tau} = \mathbb{E}[|\xi|]\, |w| = \sqrt{\frac{\pi}{2}}\, |w| \neq dF(0)(w).$$

Nonetheless, by Corollary 3.6, $\partial F(0) = \mathbb{E}[\partial f(\xi, 0)]$. This can also be verified by direct computation.

Before leaving this section we provide an elementary lemma useful in the analysis to follow.

LEMMA 3.8. *Let $h : \Xi \times X \to \mathbb{R}$ be a Carathéodory function, and let $\xi \in \Xi$ be such that $h(\xi, \cdot)$ is strictly continuous and subdifferentially regular at $x \in X$. Given $v \in \mathbb{R}^n$, set*

$$\ell_1(\xi, x; v) := \lim_{t \downarrow 0} \frac{h(\xi, x + 2tv) - h(\xi, x + tv)}{t} \quad \text{and}$$
$$\ell_2(\xi, x; v) := \lim_{t \downarrow 0} \frac{h(\xi, x - tv) - h(\xi, x - 2tv)}{t}.$$

Then, for any $v \in \mathbb{R}^n$, the limits $\ell_1(\xi, x; v)$ and $\ell_2(\xi, x; v)$ exist and we have

$$\begin{aligned} \ell_1(\xi, x; v) &= d_x h(\xi, x)(v) = \sup \{ \langle g , v \rangle \mid g \in \partial_x h(\xi, x) \} \qquad \text{and} \\ \ell_2(\xi, x; v) &= -d_x h(\xi, x)(-v) = \inf \{ \langle g , v \rangle \mid g \in \partial_x h(\xi, x) \}. \end{aligned} \tag{3.2}$$

Proof. Strict continuity (Definition 2.4) tells us that

$$|d_x h(\xi, x)(v)| \le \|v\| \operatorname{lip}_x h(\xi, x) < \infty \quad \forall\, v \in \mathbb{R}^n,$$

so that $d_x h(\xi, x)(v)$ is finite for all $v \in \mathbb{R}^n$. Therefore, by (2.1), the limit $\ell_1(\xi, x; v)$ exists and

$$\begin{aligned} d_x h(\xi, x)(v) &= 2 d_x h(\xi, x)(v) - d_x h(\xi, x)(v) \\ &= \lim_{t \downarrow 0} \left(2 \frac{h(\xi, x + 2tv) - h(\xi, x)}{2t} - \frac{h(\xi, x + tv) - h(\xi, x)}{t} \right) \\ &= \lim_{t \downarrow 0} \frac{h(\xi, x + 2tv) - h(\xi, x + tv)}{t}. \end{aligned}$$

The first equvalence in (3.2) now follows from (2.2). The second equivalence follows from the first by replacing v by $-v$. ∎

4. Smoothing Functions for $F(x) := \mathbb{E}[f(\xi, x)]$.

4.1. Measurable smoothing functions.

DEFINITION 4.1 (Smoothing Functions). *[9, Definition 1] Let $F : U \to \mathbb{R}$, where $U \subseteq \mathbb{R}^n$ is open. We say that $\widetilde{F} : U \times \mathbb{R}_{++} \to \mathbb{R}$ is a smoothing function for F on U if*

(i) $\widetilde{F}(\cdot, \mu)$ converges continuously to F on U in the sense of [19, Definition 5.41], i.e.,

$$\lim_{\mu \downarrow 0, x \to \bar{x}} \widetilde{F}(x, \mu) = F(\bar{x}) \quad \forall\, \bar{x} \in U, \text{ and}$$

(ii) $\widetilde{F}(\cdot, \mu)$ is continuously differentiable on U for all $\mu > 0$.

We now construct a class of smoothing functions for the Carathéordory function f that generate smoothing functions for F.

DEFINITION 4.2 (Measurable Smoothing Functions). *Let $U \subseteq \mathbb{R}^n$ be open and let $f : \Xi \times U \to \mathbb{R}$ be a Carathéodory mapping on $\Xi \times U$ with $f(\cdot, x) \in L^1_1(\mathbb{R}^\ell, \mathcal{M}, \rho)$ for all $x \in U$. A mapping $\tilde{f} : \Xi \times U \times \mathbb{R}_{++} \to \mathbb{R}$ is said to be a* measurable smoothing function *for f on $\Xi \times U \times \mathbb{R}_{++}$ with smoothing parameter $\mu > 0$ if, for all*

$\mu > 0$, $\tilde{f}(\cdot,\cdot,\mu)$ is a Carathéodory map on $\Xi \times U$ with $f(\cdot, x, \mu) \in L_1^1(\mathbb{R}^\ell, \mathcal{M}, \rho)$ for all $(x,\mu) \in U \times \mathbb{R}_{++}$ and the following conditions hold:

(i) The function $\tilde{f}(\xi,\cdot,\mu)$ converges continuously *to $f(\xi,\cdot)$ on U as $\mu \downarrow 0$ for almost all $\xi \in \Xi$ in the sense of [19, Definition 5.41], i.e.,*

$$\lim_{\mu\downarrow 0, x\to\bar{x}} \tilde{f}(\xi, x, \mu) = f(\xi, \bar{x}) \quad \forall\, \bar{x} \in U \text{ and } \xi \in \Xi, \tag{4.1}$$

and, for every $(\bar{x},\bar{\mu}) \in U \times \mathbb{R}_{++}$, there is an open neighborhood $V \subseteq U$ of $\bar{x}$ and a function $\kappa_f(\cdot,\bar{x},\bar{\mu}) \in L_1^2(\Xi, M, \rho)$ such that

$$|\tilde{f}(\xi, x, \mu)| \le \kappa_f(\xi, \bar{x}, \bar{\mu}) \quad \forall\, (\xi, x, \mu) \in \Xi \times V \times (0, \bar{\mu}]\ . \tag{4.2}$$

(ii) For all $\mu > 0$, the gradient $\nabla_x \tilde{f}(\xi, x, \mu)$ exists, is continuous on U for all $\xi \in \Xi$, and, for every $(\bar{x},\bar{\mu}) \in U \times \mathbb{R}_{++}$, there is an open neighborhood $V \subseteq U$ of $\bar{x}$ and a function $\hat{\kappa}_f(\cdot,\bar{x},\bar{\mu}) \in L_1^2(\Xi, \mathcal{M}, \rho)$ such that

$$\left\|\nabla_x \tilde{f}(\xi, x, \mu)\right\| \le \hat{\kappa}_f(\xi, \bar{x}, \bar{\mu}), \quad \forall\, (\xi, x, \mu) \in \Xi \times V \times (0, \bar{\mu}]\ . \tag{4.3}$$

Remark 4.3. Just as in Lemma 3.3, Lemma 2.10 can be applied to show that (4.3) implies that

$$\int_E \nabla_x \tilde{f}(\xi, x, \mu) d\rho \in \left\|\hat{\kappa}_f(\cdot, \bar{x}, \bar{\mu})\right\|_2 \rho(E) \mathbb{B} \quad \forall\, (x, \mu) \in V \times (0, \bar{\mu}] \tag{4.4}$$

for all $E \in \mathcal{M}$.

The conditions in (4.2) and (4.3) are added to the usual notion of smoothing function in Definition 4.1 to facilitate the application of the Dominated Convergence Theorem when needed.

THEOREM 4.4 (Measurable Smoothing Functions Yield Smoothing Functions). *Let $U \subseteq \mathbb{R}^n$ be open with $X \subseteq U$, and let $f : \Xi \times U \to \mathbb{R}$ be a Carathéodory mapping on $\Xi \times U$ such that $f(\cdot, x) \in L_1^1(\Xi, \mathcal{M}, \rho)$ for all $x \in U$. Let $\tilde{f} : \Xi \times U \times \mathbb{R}_{++} \to \mathbb{R}$ be a measurable smoothing function for f on $\Xi \times U \times \mathbb{R}_{++}$. Then the functions $F(x) := \mathbb{E}[f(\xi, x)]$ and $\widetilde{F}(x, \mu) := \mathbb{E}[\tilde{f}(\xi, x, \mu)]$ are well defined on U and $U \times \mathbb{R}_{++}$, respectively, and $\widetilde{F}$ is a smoothing function for F on U satisfying*

$$\nabla_x \widetilde{F}(x, \mu) = \mathbb{E}[\nabla_x \tilde{f}(\xi, x, \mu)] \quad \forall\ (x, \mu) \in U \times \mathbb{R}_{++}. \tag{4.5}$$

Proof. The fact that F and $\widetilde{F}$ are well defined follows from the definitions. It remains only to show that $\widetilde{F}$ is a smoothing function for F. By (4.1), (4.2) and the Dominated Convergence Theorem, for all $x \in U$,

$$\lim_{\mu\downarrow 0, x\to\bar{x}} \widetilde{F}(x, \mu) = \lim_{\mu\downarrow 0, x\to\bar{x}} \mathbb{E}[\tilde{f}(\xi, x, \mu)] = \mathbb{E}[\lim_{\mu\downarrow 0, x\to\bar{x}} \tilde{f}(\xi, x, \mu)] = \mathbb{E}[f(\xi, x)] = F(x)$$

which establishes (i) in Definition 4.1.

Next let $(\bar{x},\bar{\mu}) \in U \times \mathbb{R}_{++}$ and $d \in \mathbb{R}^n$ with $d \neq 0$. By (4.3) and the Mean Value Theorem (MVT), for all small $t > 0$ and $\xi \in \Xi$ there is a $z_t(\xi)$ on the line segment joining $\bar{x} + td$ and $\bar{x}$ such that

$$\left|\frac{\tilde{f}(\xi, \bar{x} + td, \bar{\mu}) - \tilde{f}(\xi, \bar{x}, \bar{\mu})}{t}\right| = |\nabla_x \tilde{f}(\xi, z_t(\xi), \bar{\mu})^T d| \le \hat{\kappa}_f(\xi, \bar{x}, \bar{\mu})\, \|d\|\, .$$

Hence, by the Dominated Convergence Theorem,

$$\begin{aligned}\lim_{t\downarrow 0}\frac{\widetilde{F}(\bar{x}+td,\bar{\mu})-\widetilde{F}(\bar{x},\bar{\mu})}{t} &= \lim_{t\downarrow 0}\mathbb{E}\left[\frac{\tilde{f}(\xi,\bar{x}+td,\bar{\mu})-\tilde{f}(\xi,\bar{x},\bar{\mu})}{t}\right]\\ &= \mathbb{E}\left[\lim_{t\downarrow 0}\frac{\tilde{f}(\xi,\bar{x}+td,\bar{\mu})-\tilde{f}(\xi,\bar{x},\bar{\mu})}{t}\right]\\ &= \left\langle \mathbb{E}[\nabla_x\tilde{f}(\xi,\bar{x},\bar{\mu})]\,,\,d\right\rangle.\end{aligned}$$

Since this is true for all choices of $d\in\mathbb{R}^n$, we have $\nabla_x\widetilde{F}(\bar{x},\bar{\mu})=\mathbb{E}[\nabla_x\tilde{f}(\xi,\bar{x},\bar{\mu})]$ which establishes (4.5).

Finally we show that $\nabla_x\widetilde{F}(\cdot,\mu)$ is continuous on U for all $\mu>0$. Let $(\bar{x},\bar{\mu})\in U\times\mathbb{R}_{++}$. By (4.5), (4.3) and the Dominated Convergence Theorem,

$$\begin{aligned}\lim_{x\to\bar{x}}\nabla_x\widetilde{F}(x,\bar{\mu}) &= \lim_{x\to\bar{x}}\mathbb{E}[\nabla_x\tilde{f}(\xi,x,\bar{\mu})]\\ &= \mathbb{E}[\lim_{x\to\bar{x}}\nabla_x\tilde{f}(\xi,x,\bar{\mu})]=\mathbb{E}[\nabla_x\tilde{f}(\xi,\bar{x},\bar{\mu})]=\nabla_x\widetilde{F}(\bar{x},\bar{\mu}).\end{aligned}$$ ◻

4.2. Gradient consistency of $F(x)=\mathbb{E}[f(\xi,x)]$**.** A key concept relating smoothing to the variational properties of F is the notion of *gradient consistency* introduced in [9].

Definition 4.5 (Gradient Consistency of Smoothing Functions). *Let $U\subseteq\mathbb{R}^n$ be open and let $F:U\to\mathbb{R}$ be such that $\widetilde{F}:U\times\mathbb{R}_{++}\to\mathbb{R}$ is a smoothing function for F on U. We say that $\widetilde{F}$ is* gradient consistent *at $\bar{x}\in U$ if*

$$\mathrm{co}\left\{\operatorname*{Limsup}_{\mu\downarrow 0,x\to\bar{x}}\nabla_x\widetilde{F}(x,\mu)\right\}=\partial F(\bar{x}),$$

where the limit supremum is taken in the multi-valued sense (2.3).

As a first step toward understanding how the gradient consistency of a measurable smoothing function for f can be used to construct a smoothing function for F, we give the following result.

Theorem 4.6. *Let $U\subseteq\mathbb{R}^n$, $\bar{x}\in U$, and $f:\Xi\times U\to\mathbb{R}$ be as in Corollary 3.6, and suppose that $\tilde{f}:\Xi\times U\times\mathbb{R}_{++}\to\mathbb{R}$ is a measurable smoothing function for f on $\Xi\times U\times\mathbb{R}_{++}$. If, for almost all $\xi\in\Xi$, $\tilde{f}(\xi,\cdot,\cdot)$ is gradient consistent at $\bar{x}$, i.e.*

$$\mathrm{co}\left\{\operatorname*{Limsup}_{x\to\bar{x},\mu\downarrow 0}\nabla_x\tilde{f}(\xi,x,\mu)\right\}=\partial_x f(\xi,\bar{x})\quad a.e.\ \xi\in\Xi,\tag{4.6}$$

then $\widetilde{F}(x,\mu):=\mathbb{E}[\tilde{f}(\xi,x,\mu)]$ is a smoothing function for $F(x):=\mathbb{E}[f(\xi,x)]$ satisfying

$$\partial F(\bar{x})=\mathbb{E}\left[\mathrm{co}\left\{\operatorname*{Limsup}_{x\to\bar{x},\mu\downarrow 0}\nabla_x\tilde{f}(\xi,x,\mu)\right\}\right].\tag{4.7}$$

Proof. The fact that $\widetilde{F}$ is a smoothing function for F comes from Theorem 4.4. Therefore, the result is an immediate consequence of Corollary 3.6. ◻

The pointwise condition (4.6) does not imply the gradient consistency of $\widetilde{F}$ at $\bar{x}$. To obtain such a result from (4.7) we also need to know that

$$\mathbb{E}\left[\operatorname{co}\left\{\operatorname*{Limsup}_{x\to\bar{x},\mu\downarrow 0}\nabla_x\tilde{f}(\xi,x,\mu)\right\}\right]=\operatorname{co}\left\{\operatorname*{Limsup}_{x\to\bar{x},\mu\downarrow 0}\mathbb{E}[\nabla_x\tilde{f}(\xi,x,\mu)]\right\}. \tag{4.8}$$

The equivalence (4.8) is nontrivial requiring much stronger hypotheses to achieve.

Since $\partial_x f(\xi,x)$ compact-, convex-valued in x, the left-hand side of (4.6) is contained in the right-hand side if and only if for almost all $\xi\in\Xi$ and $\epsilon>0$ there is a $\delta(\xi,\bar{x},\epsilon)>0$ such that

$$\nabla_x\tilde{f}(\xi,x,\mu)\in\partial_x f(\xi,\bar{x})+\epsilon\mathbb{B}\quad\forall\,(x,\mu)\in(\bar{x},0)+\delta(\xi,\bar{x},\epsilon)\mathbb{B}\text{ with }\mu>0.$$

This motivates the hypotheses employed in the following theorem.

THEOREM 4.7 (Gradient Sub-Consistency). *Let $U\subseteq\mathbb{R}^n$ and $f:\Xi\times U\to\mathbb{R}$ be as in Corollary 3.6, and suppose that $\tilde{f}:\Xi\times U\times\mathbb{R}_{++}\to\mathbb{R}$ is a measurable smoothing function for f on $\Xi\times U\times\mathbb{R}_{++}$. If $\bar{x}\in U$ is such that there exists $\bar{\nu}\in(0,1)$ such that for all $\nu\in(0,\bar{\nu})$ there exist $\delta(\nu,\bar{x})>0$ and $\Xi(\nu,\bar{x})\in\mathcal{M}$ with $\rho(\Xi(\nu,\bar{x}))\geq 1-\nu$ for which*

$$\nabla_x\tilde{f}(\xi,x,\mu)\in\partial_x f(\xi,\bar{x})+\nu\mathbb{B}\quad\forall\,(x,\mu)\in[(\bar{x},0)+\delta(\nu,\bar{x})(\mathbb{B}\times(0,1))]\quad a.e.\ \xi\in\Xi(\nu,\bar{x}), \tag{4.9}$$

then

$$\operatorname{co}\left\{\operatorname*{Limsup}_{x\to\bar{x},\mu\downarrow 0}\nabla\widetilde{F}(x,\mu)\right\}\subseteq\partial F(\bar{x})=\mathbb{E}\left[\operatorname{co}\left\{\operatorname*{Limsup}_{x\to\bar{x},\mu\downarrow 0}\nabla_x\tilde{f}(\xi,x,\mu)\right\}\right]. \tag{4.10}$$

Proof. Since $\partial F(x)$ is convex, we need only show that the inclusion, without the convex hull, is satisfied. Let $g\in\operatorname{Limsup}_{x\to\bar{x},\mu\downarrow 0}\nabla\widetilde{F}(x,\mu)$. Then there is a sequence $(x^k,\mu_k)\to(\bar{x},0)$ with $\mu_k>0$ such that $\nabla\widetilde{F}(x^k,\mu_k)\to g$. By Lemma 3.3(d), there is a measurable selection s from $\partial_x f(\cdot,\bar{x})$. Let $\nu\in(0,\bar{\nu})$, and let $\bar{k}$ be such that $(x^k,\mu_k)\in(\bar{x},0)+\delta(\nu,\bar{x})(\mathbb{B}\times(0,1))$ for all $k\geq\bar{k}$. For all $k\geq\bar{k}$, define

$$q_k(\xi):=\begin{cases}\nabla_x\tilde{f}(\xi,x^k,\mu_k), & \xi\in\Xi(\nu,\bar{x})\\ s(\xi), & \xi\in\Xi\setminus\Xi(\nu,\bar{x}).\end{cases}$$

Then

$$\begin{aligned}
g&=\lim_{k\to\infty}\nabla_x\widetilde{F}(x^k,\mu_k)\\
&=\lim_{k\to\infty}\int_{\xi\in\Xi(\nu,\bar{x})}\nabla_x\tilde{f}(\xi,x^k,\mu_k)d\xi+\int_{\xi\in\Xi/\Xi(\nu,\bar{x})}\nabla_x\tilde{f}(\xi,x^k,\mu_k)d\xi\quad\text{(Theorem 4.4)}\\
&=\lim_{k\to\infty}\int_{\xi\in\Xi}q_k(\xi)d\xi-\int_{\xi\in\Xi\setminus\Xi(\nu,\bar{x})}s(\xi)d\xi+\int_{\xi\in\Xi\setminus\Xi(\nu,\bar{x})}\nabla_x\tilde{f}(\xi,x^k,\mu_k)d\xi\\
&\in\mathbb{E}[\partial_x f(\xi,\bar{x})]+\nu\mathbb{B}+\nu\|\kappa_f(\cdot,\bar{x})\|_2\mathbb{B}+\nu\|\hat{\kappa}_f(\cdot,\bar{x},\bar{\mu})\|_2\mathbb{B}\quad\text{((4.9), Theorem 3.3(f), (4.4))}\\
&=\partial F(\bar{x})+\nu(1+\|\kappa_f(\cdot,\bar{x})\|_2+\|\hat{\kappa}_f(\cdot,\bar{x},\bar{\mu})\|_2)\mathbb{B}.\quad\text{(Corollary 3.6)}
\end{aligned}$$

Since $\nu\in(0,\bar{\nu})$ was chosen arbitrarily, this proves the inclusion in (4.10) since $\partial F(\bar{x})$ is closed. The equivalence in (4.10) follows from Theorem 4.6 since (4.9) implies (4.6). ∎

In what follows, we refer to (4.10) as the *gradient sub-consistency property* for the smoothing function $\widetilde{F}$ at $\bar{x}$, and we refer to (4.9) as the *uniform subgradient approximation property* for the measurable smoothing function $\tilde{f}$ at $\bar{x}$.

5. Composite Max (CM) Integrands. We introduce the class of CM integrands and associated smoothing functions and show that they satisfy the properties required for the application of the results of the previous sections. The nonsmoothness of CM integrands arises through composition with finite piecewise linear convex functions on $\mathbb{R}$. The simplest such piecewise linear functions is given by $(t)_+ := \max\{0, t\}$. Indeed, all piecewise linear convex functions can be built up from this basic function. Integral smoothing techniques based on $(t)_+$ first appear in the work of Chen and Mangasarian [8] and were later expanded by Chen [9] to a broader class of nonsmooth functions under composition. In [6] it is shown that certain economies are possible by using the piecewise linear convex functions directly in the construction of smoothers. We use these here. As in [6, 8, 9], we convolve these piecewise linear functions with a density to obtain a rich class of measurable smoothing mappings useful in applications. We begin with the following definition.

Definition 5.1 (Measurable Mappings with Amenable Derivatives). *Let $\Xi \times X \subseteq \mathbb{R}^\ell \times \mathbb{R}^n$ and let U be an open set containing X. We say that the mapping $g : \Xi \times U \to \mathbb{R}^m$ is a measurable mapping with amenable derivative if the following two conditions are satisfied:*

- *(i) Each component of g is a Carathéodory mapping and, for all $\xi \in \Xi$, $g(\xi, \cdot)$ is continuously differentiable in x on U;*
- *(ii) For all $(\xi, x) \in \Xi \times U$, the gradient $\nabla_x g(\xi, x)$ is locally L^2 bounded in x uniformly in ξ in the sense that there is a function $\hat{\kappa}_g : \Xi \times U \to \mathbb{R}$ satisfying $\hat{\kappa}_g(\cdot, x) \in L^2_1(\mathbb{R}^\ell, \mathcal{M}, \rho)$ for all $x \in U$ and*

$$\forall\ \bar{x} \in X\ \exists\ \epsilon(\bar{x}) > 0 \quad \text{such that} \quad \|\nabla_x g(\xi, x)\| \le \hat{\kappa}_g(\xi, \bar{x}) \quad \forall\ x \in \mathbb{B}_{\epsilon(\bar{x})}(\bar{x}).$$

Define $\hat{\kappa}^E_g(\bar{x}) := \mathbb{E}[\hat{\kappa}_g(\xi, \bar{x})]$.

We now define CM integrands.

Definition 5.2 (Composite Max (CM) Integrands). *A CM integrand on $\Xi \times X$ is any mapping of the form*

$$f(\xi, x) := \mathsf{q}(\mathsf{c}(\xi, x) + C(g(\xi, x))) \tag{5.1}$$

for which there exists an open set U containing X such that

1. *$C : \mathbb{R}^m \to \mathbb{R}^m$ is of the form $C(y) := [p_1(y_1),\ p_2(y_2), \ldots,\ p_m(y_m)]^T$, where $p_i : \mathbb{R} \to \mathbb{R}\, (i = 1, \ldots, m)$ are finite piecewise linear convex functions having finitely many points of nondifferentiability,*
2. *the mappings c and g are measurable mappings with amenable derivatives mapping $\Xi \times \mathbb{R}^n$ to $\mathbb{R}^m$ and having common underlying open set U containing X on which $\mathsf{c}(\xi, \cdot)$ and $g(\xi, \cdot)$ are continuously differentiable in x on U for all $\xi \in \Xi$, and*
3. *the mapping $\mathsf{q} : \mathbb{R}^m \to \mathbb{R}$ is continuously differentiable with Lipschitz continuous derivative on the set*

$$\mathcal{Q} := cl(\text{co}\ \{\mathsf{c}(\xi, x) + C(g(\xi, x)) \mid (\xi, x) \in \Xi \times U\}).$$

Let $\bar{\kappa}_q$ be a Lipschitz constant for ∇q on $\mathcal{Q}$.

Remark 5.3. The family of CM integrands is designed to include many important classes of functions useful in applications, e.g. the gap functions of the Nonlinear Complementarity Problem (NCP); and the difference of two Clarke regular functions where nonsmoothness occurs due the presence of compositions with piecewise convex functions.

Following [6, Section 4], we assume with no loss in generality that for each $i = 1,\dots,m$, there is a positive integer r_i and scalar pairs $(a_{ij}, b_{ij}),\ i = 1,\dots,m,\ j = 1,\dots,r_i$ such that

$$p_i(t) := \max\{a_{ij}t + b_{ij} \mid j = 1,\dots,r_i\},$$

where $a_{i1} < a_{i2} < \cdots < a_{i(r_i-1)} < a_{ir_i}$. Again with no loss in generality, we assume that the scalar pairs $(a_{ij}, b_{ij}),\ i = 1,\dots,m,\ j = 1,\dots,r_i$ are coupled with a scalar partition of the real line

$$-\infty = t_{i1} < t_{i2} < \cdots < t_{ir_i} < t_{i(r_i+1)} = \infty$$

such that for all $j = 1,\dots,r_i-1$, $a_{ij}t_{i(j+1)} + b_{ij} = a_{i(j+1)}t_{i(j+1)} + b_{i(j+1)}$ and

$$p_i(t) = \begin{cases} a_{i1}t + b_{i1}, & t \le t_{i2},\\ a_{ij}t + b_{ij}, & t \in [t_{ij}, t_{i(j+1)}] \qquad (j \in \{2,\dots,r_i-1\}),\\ a_{ir_i}t + b_{ir_i}, & t \ge t_{ir_i}.\end{cases}$$

This representation for the functions p_i gives

$$\partial p_i(t) = \begin{cases} a_{ij}, & t_{ij} < t < t_{i(j+1)},\ j = 1,\dots,r_i\\ [a_{i(j-1)}, a_{ij}], & t = t_{ij},\ j = 2,\dots,r_i,\end{cases} \qquad i = 1,\dots,m. \tag{5.2}$$

It is easily shown that the functions p_i and C (5.1) are globally Lipschitz continuous with common Lipschitz constant

$$\bar\kappa_C := \max\{|a_{ij}| \mid i = 1,\dots,m, j = 1\dots,r_i\}. \tag{5.3}$$

Clearly CM integrands on $\Xi\times X$ are Carátheodory functions on $\Xi\times X$, Moreover, CM integrands are explicitly constructed so that they are also LL integrands on $\Xi\times X$. We record this easily verified result in the next lemma.

Lemma 5.4 (CM Integrands are LL Integrands). *Let $f : \Xi\times X \to \mathbb{R}$ be an CM integrand as in (5.1). Then f is an LL integrand on $\Xi\times X$, where, for all $\bar x \in X$, one may take*

$$\kappa_f(\cdot,\bar x) := \bar\kappa_{\mathsf{q}}[\hat\kappa_{\mathsf{c}}(\cdot,\bar x) + \bar\kappa_C\hat\kappa_g(\cdot,\bar x)], \tag{5.4}$$

where $\bar\kappa_q$ and $\bar\kappa_C$ are defined in Definition 5.2 and (5.3), respectively, and $\hat\kappa_{\mathsf{c}}$ and $\hat\kappa_g$ are given in Definition 5.1.

Since the functions p_i are continuously differentiable on the open set $\mathbb{R}\backslash\{t_{i2},\dots,t_{ir_i}\}$ and the functions q, $\mathsf{c}(\xi,\cdot)$, and $g(\xi,\cdot)$ are continuously differentiable, the set on which the CM integrand $f(\xi,\cdot)$ is continuously differentiable is easily identified.

Proposition 5.5. *Let $f : \Xi\times X \to \mathbb{R}$ be a CM integrand as in (5.1), and, for each $i = 1,\dots,m$, set $q_i(\xi,x) := p_i(g_i(\xi,x))$. Given $(\xi,x) \in \Xi\times U$, set*

$$\begin{aligned}\widetilde U_i(\xi) &:= \left\{x \in U \,\middle|\, x \notin (g_i(\xi,\cdot))^{-1}(\{t_{i2},\dots,t_{ir_i}\})\right\},\ i = 1,\dots,m,\\ \widetilde\Xi_i(x) &:= \left\{\xi \in \Xi \,\middle|\, \xi \notin (g_i(\cdot,x))^{-1}(\{t_{i2},\dots,t_{ir_i}\})\right\},\ i = 1,\dots,m,\\ \widetilde U(\xi) &:= \bigcap_{i=1}^m \widetilde U_i(\xi) \quad \text{and} \quad \widetilde\Xi(x) := \bigcap_{i=1}^m \widetilde\Xi_i(x).\end{aligned}$$

Then $q_i(\xi,\cdot)$ is continuously differentiable on the open set $\widetilde{U}_i(\xi)$ with

$$\nabla_x q_i(\xi,x) = \nabla_t p_i(g_i(\xi,x))\nabla g_i(\xi,x),\ \ i=1,\ldots,m,$$

and $f(\xi,\cdot)$ is continuously differentiable and subdifferentially regular on the open set $\widetilde{U}(\xi)$ with

$$\nabla_x f(\xi,x) = (\nabla_x \mathsf{c}(\xi,x) + \mathrm{diag}(\nabla_t p_i(g_i(\xi,x)))\nabla_x g(\xi,x))^T \nabla \mathsf{q}(\mathsf{c}(\xi,x) + \tilde{C}(g(\xi,\bar{x})))$$

and $\partial_x f(\xi,x) = \{\nabla_x f(\xi,x)\}$. Therefore, given $x\in U$, $f(\xi,\cdot)$ is continuously differentiable and subdifferentially regular at x for all $\xi\in\widetilde{\Xi}(x)$. In particular, if $x\in U$ is such that $\rho(\widetilde{\Xi}(x))=1$, then $f(\xi,\cdot)$ is continuously differentiable and subdifferentially regular at x for almost all $\xi\in\Xi$.

Proof. Observe that each of the sets $\widetilde{U}_i(\xi)$ is open due to the continuity of $g_i(\xi,\cdot)$. In addition, given $x\in\widetilde{U}_i(\xi)$, $t:=g_i(\xi,x)$ is a point of continuous differentiability for p_i. Hence, by a standard chain rule (e.g. see [20, Theorem 9.15]), $q_i(\xi,\cdot)$ is continuously differentiable at every $x\in\widetilde{U}_i(\xi)$. Therefore, every $q_i(\xi,\cdot), i=1,\ldots,m$ is continuously differentiable for $x\in\widetilde{U}(\xi)$, and so, by the same standard chain rule, $f(\xi,\cdot)$ is continuously differentiable on $\widetilde{U}(\xi)$ with it's gradient as given. The subdifferential regularity follows from [19, Theorem 9.18 and Exercise 9.64].

Given $x\in U$ and $\xi\in\widetilde{\Xi}(x)$ we have $g_i(\xi,x)\notin\{t_{i2},\ldots,t_{ir_i}\}$ for $i=1,\ldots,m$. Hence $x\in\widetilde{U}(\xi)$ so that $f(\xi,\cdot)$ is continuously differential and subdifferentially regular at x as required. The final statement of the proposition is now self-evident. ∎

5.1. Smoothing CM integrands. We use the techniques described in [6] to smooth CM integrands. Let $\beta:\mathbb{R}\to\mathbb{R}_+$ be a piecewise continuous *density function* satisfying

$$\beta(t)=\beta(-t) \qquad \text{and} \qquad \omega := \int_{\mathbb{R}} |t|\beta(t)\,dt < \infty. \tag{5.5}$$

We denote the *distribution function* for the density β by φ, i.e., $\varphi:\mathbb{R}\to[0,1]$ is given by $\varphi(x)=\int_{-\infty}^x \beta(t)\,dt$. Since β is symmetric and $\beta(\cdot)\geq 0$, φ is a non-decreasing continuous function satisfying

$$\varphi(0)=\frac{1}{2},\ \ 1-\varphi(x)=\varphi(-x),\ \ \lim_{x\to\infty}\varphi(x)=1 \ \text{ and } \lim_{x\to-\infty}\varphi(x)=0. \tag{5.6}$$

Moreover, for every $\alpha\in(0,1)$, $\varphi^{-1}(\alpha)$ is a bounded interval in $\mathbb{R}$, and so

$$-\infty < \varphi^{-1}_{\min}(\alpha) := \inf\left\{\zeta \;\middle|\; \zeta\in\varphi^{-1}(\alpha)\right\} \leq \varphi^{-1}_{\max}(\alpha) := \sup\left\{\zeta \;\middle|\; \zeta\in\varphi^{-1}(\alpha)\right\} < +\infty. \tag{5.7}$$

Finally, we note that since β is non-negative, piecewise continuous, and $\int_{\mathbb{R}}\beta(t)\,dt=1$, it must be bounded on $\mathbb{R}$, so that $\beta_{\max}:=\sup\{\beta(t)\mid t\in\mathbb{R}\}<+\infty$ which implies that φ is Lipschitz continuous on $\mathbb{R}$ with modulus $\beta_{\max}$, i.e., $|\varphi(t_1)-\varphi(t_2)|\leq\beta_{\max}|t_1-t_2|$.

LEMMA 5.6. *[6, Lemma 4.1] For each $i=1,\ldots,m$, let $p_i:\mathbb{R}\to\mathbb{R}$ be the finite max-function defined above. Furthermore, let $\beta:\mathbb{R}\to\mathbb{R}_+$ be a piecewise continuous function satisfying (5.5). Then, for each $i=1,\ldots,m$, the convolution*

$$\tilde{p}_i(t,\mu) := \int_{\mathbb{R}} p_i(t-\mu s)\beta(s)\,ds$$

is a (well-defined) smoothing function with

$$
\begin{aligned}
\nabla_t \tilde{p}_i(t,\mu) = a_{i1}\Big(1-\varphi\Big(\frac{t-t_{i2}}{\mu}\Big)\Big) \\
+ \sum_{j=2}^{r_i-1} a_{ij}\Big(\varphi\Big(\frac{t-t_{ij}}{\mu}\Big)-\varphi\Big(\frac{t-t_{i(j+1)}}{\mu}\Big)\Big) + a_{ir_i}\varphi\Big(\frac{t-t_{ir_i}}{\mu}\Big),
\end{aligned}
\tag{5.8}
$$

$$
\eta_i(t) := \lim_{\mu\downarrow 0} \nabla_t \tilde{p}_i(t,\mu) = \begin{cases} a_{ij} & t_{ij} < t < t_{i(j+1)},\ j=1,\dots,r_i \\ \frac{1}{2}(a_{i(j-1)}+a_{ij}) & t = t_{ij},\ j=2,\dots,r_i \end{cases}
\tag{5.9}
$$

is an element of $\partial p_i(\bar{t})$*, and*

$$
\operatorname*{Limsup}_{t\to\bar{t},\mu\downarrow 0} \nabla_t \tilde{p}_i(t,\mu) = \partial p_i(\bar{t}) \quad \forall \bar{t}\in\mathbb{R}.
$$

In addition, for $\hat{t},\bar{t}\in\mathbb{R}$ *and* $0<\hat{\mu}\le\bar{\mu}$*, we have*

$$
|\tilde{p}_i(\hat{t},\hat{\mu})-\tilde{p}_i(\bar{t},\bar{\mu})| \le \bar{\kappa}_C[\,|\hat{t}-\bar{t}|+|\hat{\mu}-\bar{\mu}|\,] \quad \textit{and}
\tag{5.10}
$$

$$
|\nabla_t\tilde{p}_i(\hat{t},\hat{\mu})-\nabla_t\tilde{p}_i(\bar{t},\bar{\mu})| \le \frac{2r_i}{\hat{\mu}}[\,|\hat{t}-\bar{t}|+(1-\hat{\mu}/\bar{\mu})\max_j|\bar{t}-t_{ij}|\,].
\tag{5.11}
$$

Remark 5.7. The bounds (5.10) and (5.11) do not appear in [6], but are straightforward to verify directly from the definitions and (5.8).

Theorem 5.8 (Smoothing of CM Integrands). *[6, Theorem 4.6] Let* f *be a CM integrand. Then* $\tilde{f}:\Xi\times U\times\mathbb{R}_{++}\to\mathbb{R}$ *given by*

$$
\tilde{f}(\xi,x,\mu) := \mathsf{q}(\mathsf{c}(\xi,x)+\tilde{C}(g(\xi,x),\mu)),
\tag{5.12}
$$

where $\tilde{C}(y,\mu) := [\tilde{p}_1(y_1,\mu),\tilde{p}_2(y_2,\mu),\dots,\tilde{p}_m(y_m,\mu)]^T$ *with each* $\tilde{p}_i$ *is as given in Lemma 5.6, is a measurable smoothing function for* f*. If, furthermore,* $(\xi,\bar{x})\in\Xi\times U$ *is such that* $\operatorname{rank}\nabla_x g(\xi,\bar{x})=m$*, then, for all* $\mu>0$*,*

$$
\nabla_x\tilde{f}(\xi,\bar{x},\mu) = (\nabla_x\mathsf{c}(\xi,\bar{x})+\operatorname{diag}(\nabla_t\tilde{p}_i(g_i(\xi,\bar{x}),\mu))\nabla_x g(\xi,\bar{x}))^T\nabla\mathsf{q}(\mathsf{c}(\xi,\bar{x})+\tilde{C}(g(\xi,\bar{x}))),
\tag{5.13}
$$

and

$$
\operatorname*{Limsup}_{x\to\bar{x},\mu\downarrow 0}\nabla_x\tilde{f}(\xi,x,\mu)\subseteq\partial_x f(\xi,\bar{x}) \quad \textit{and} \quad \operatorname{co}\left\{\operatorname*{Limsup}_{x\to\bar{x},\mu\downarrow 0}\nabla_x\tilde{f}(\xi,x,\mu)\right\} = \partial_x f(\xi,\bar{x}),
$$

where

$$
\partial_x f(\xi,\bar{x}) = (\nabla_x\mathsf{c}(\xi,\bar{x})+\operatorname{diag}(\partial_t p_i(g_i(\xi,\bar{x}),\mu))\nabla_x g(\xi,\bar{x}))^T\nabla\mathsf{q}(\mathsf{c}(\xi,\bar{x})+C(g(\xi,\bar{x}))).
$$

We now show that the function $\tilde{f}$ defined in (5.12) is a measurable smoothing function for f in the sense of Definition 5.2. First observe that the expression for $\nabla_t\tilde{p}_i(t,\mu)$ in Lemma 5.6 implies the bound

$$
2\bar{\kappa}_C \ge \|\operatorname{diag}(\nabla_t\tilde{p}_i(g_i(\xi,\bar{x}),\mu)))\|_\infty \quad \forall\,(\xi,\bar{x},\mu)\in\Xi\times X\times\mathbb{R}_{++}.
\tag{5.14}
$$

Since this bound is independent of μ, we can use it in conjunction with the representation (5.13) to provide a Lipschitz constant for $\tilde{f}$ analogous to (5.4).

LEMMA 5.9 (Smoothed CM Integrands are LL Integrands). *Let $\tilde f : \Xi \times X \to \mathbb{R}$ be as in Theorem 5.8. Then, for every $\mu \in \mathbb{R}_{++}$, $\tilde f(\cdot,\cdot,\mu)$ is an LL integrand on $\Xi\times X$, where, for all $\bar x \in X$, one may take $\kappa_{\tilde f_\mu}(\cdot,\bar x) := \bar\kappa_{\mathsf{q}}[\hat\kappa_{\mathsf{c}}(\cdot,\bar x) + 2\bar\kappa_C\hat\kappa_g(\cdot,\bar x)]$.*

We also have the following elementary bounds for the functions p_i, $\tilde p_i$, $\nabla_t\tilde p_i$, and η_i.

LEMMA 5.10. *For $i = 1,\dots,m$, let $\nabla_t\tilde p_i$ and η_i be as in Lemma 5.6, and set*

$$\gamma_i(t) := \begin{cases} |t-t_2|, & \text{if } r_i = 2 \text{ and } t \neq t_2,\\ +\infty, & \text{if } r_i = 2 \text{ and } t = t_2,\\ \min\{|t-t_{ij}| \mid j\in\{2,\dots,r_i\},\ t\neq t_{ij}\}, & \text{if } r_i \ge 3 \text{ and } t\neq t_{ij}\ j=1,\dots,r_i,\\ \min\{|t_{i\bar j}-t_{i(\bar j-1)}|, |t_{i\bar j}-t_{i(\bar j+1)}|\}, & \text{if } r_i\ge 3 \text{ and } t=t_{i\bar j},\ \bar j\in\{2,\dots,r_i\}. \end{cases}$$

Then, for each $i = 1,\dots,m$,

$$\bar\kappa_C(|\bar t - t| + \mu\omega) \ge |\, p_i(\bar t) - \tilde p_i(t,\mu)\,| \quad \forall\, \bar t, t\in\mathbb{R}, \quad \text{and} \tag{5.15}$$
$$\mathsf{b}_i(t,\mu) := (\bar r+1)\bar\kappa_C\varphi(-\mu^{-1}\gamma_i(t)) \ \ge\ |\,\nabla_t\tilde p_i(t,\mu) - \eta_i(t)\,|, \tag{5.16}$$

where ω is defined in (5.5), $\bar\kappa_C$ is defined in (5.3) and $\bar r := \max\{r_1,\dots,r_m\}$. Moreover, for $i = 1,\dots,m$, b_i is continuous on $\mathbb{R}\times(0,+\infty)$ when $r_i = 2$ and is continuous on $(\mathbb{R}\setminus\{t_{i2},\dots,t_{ir_i}\})\times(0,+\infty)$ when $r_i \ge 3$. In addition, for all $(t,\mu)\in\mathbb{R}\times(0,\infty)$, $0 \le \mathsf{b}(t,\mu) \le \frac{1}{2}(\bar r+1)\bar\kappa_C$, and $\mathsf{b}(t,\cdot)$ is non-decreasing on $(0,+\infty)$ with $\lim_{\mu\uparrow\infty}\mathsf{b}(t,\mu) = \frac12(\bar r+1)\bar\kappa_C$ and $\lim_{\mu\downarrow 0}\mathsf{b}(t,\mu) = 0$.

Proof. The bound (5.15) is given in the proof of [6, Lemma 4.1]. Next, fix $i \in \{1,\dots,m\}$. Since i is fixed, we suppress it in the proof to follow. Let $t\in\mathbb{R}$ and let k denote some integer in $\{2,\dots,r\}$. One of the following five mutually exclusive cases must hold: (i) $r = 2$ and $t \ne t_2$, (ii) $r = 2$ and $t = t_2$, (iii) $r \ge 3$ and ($t < t_2$ or $t > t_r$), (iv) $r \ge 3$ and $t = t_k$, and (v) $r \ge 3$ and $t_k < t < t_{k+1}$ with $2 \le k \le r-1$. Each of the five cases is addressed separately. In each case, we make free use of the properties of the function φ as described in (5.5)-(5.7).

<u>(i) $r = 2$ and $t \neq t_2$:</u>

$$|\,\nabla_t\tilde p_i(t,\mu) - \eta_i(t)\,| \le r\bar\kappa_C\varphi(\mu^{-1}(|t-t_2|)) = r\bar\kappa_C\varphi(-\mu^{-1}\gamma(t)).$$

<u>(ii) $k = 2$ and $t = t_2$:</u>

$$|\,\nabla_t\tilde p_i(t,\mu) - \eta_i(t)\,| = 0 = \varphi(-\mu^{-1}\gamma(t)).$$

<u>(iii) $r \ge 3$ and ($t < t_2$ or $t > t_r$):</u>

$$\begin{aligned} (t<t_2):\ |\,\nabla_t\tilde p_i(t,\mu) - \eta_i(t)\,| &\le r\bar\kappa_C\varphi(\mu^{-1}(t-t_2)) = r\bar\kappa_C\varphi(-\mu^{-1}\gamma(t)).\\ (t>t_r):\ |\,\nabla_t\tilde p_i(t,\mu) - \eta_i(t)\,| &\le r\bar\kappa_C\left(1-\varphi(\mu^{-1}(t-t_r))\right) = r\bar\kappa_C\varphi(\mu^{-1}(t_r-t))\\ &= r\bar\kappa_C\varphi(-\mu^{-1}\gamma(t)). \end{aligned}$$

(iv) $r \geq 3$ and $t = t_k$:

$$
\begin{aligned}
|\,\nabla_t \tilde{p}_i(t,\mu) - \eta_i(t)\,| \leq{}& (k-2)\bar{\kappa}_C\left(1-\varphi(\mu^{-1}(t-t_{k-1}))\right) \\
&+\left|a_{k-1}\Big(\varphi(\mu^{-1}(t-t_{k-1}))-\frac{1}{2}\Big)+a_k\Big(\frac{1}{2}-\varphi(\mu^{-1}(t-t_{k+1}))\Big)-\frac{1}{2}(a_{(k-1)}+a_k)\right| \\
&+(r-k)\bar{\kappa}_C\varphi(\mu^{-1}(t-t_{k+1})) \\
&\leq (k-1)\bar{\kappa}_C\left(1-\varphi(\mu^{-1}(t-t_{k-1}))\right)+(r-k+1)\bar{\kappa}_C\varphi(\mu^{-1}(t-t_{k+1})) \\
&= (k-1)\bar{\kappa}_C\varphi(\mu^{-1}(t_{k-1}-t))+(r-k+1)\bar{\kappa}_C\varphi(\mu^{-1}(t-t_{k+1})) \\
&\leq r\bar{\kappa}_C\varphi(-\mu^{-1}\gamma(t)).
\end{aligned}
$$

(v) $r \geq 3$ and $t_k < t < t_{k+1}$ with $2 \leq k \leq r-1$:

$$
\begin{aligned}
|\,\nabla_t \tilde{p}_i(t,\mu) - \eta_i(t)\,| \leq{}& (k-1)\bar{\kappa}_C\left(1-\varphi(\mu^{-1}(t-t_{k-1}))\right)+(r-k)\bar{\kappa}_C\varphi(\mu^{-1}(t-t_{k+1})) \\
&+\left|a_k\Big(\varphi(\mu^{-1}(t-t_k))-\varphi(\mu^{-1}(t-t_{(k+1)}))\Big)-a_k\right| \\
&\leq k\bar{\kappa}_C\left(1-\varphi(\mu^{-1}(t-t_k))\right)+(r-k+1)\bar{\kappa}_C\varphi(\mu^{-1}(t-t_{k+1})) \\
&= k\bar{\kappa}_C\varphi(\mu^{-1}(t_k-t))+(r-k+1)\bar{\kappa}_C\varphi(\mu^{-1}(t-t_{k+1})) \\
&\leq (r+1)\bar{\kappa}_C\varphi(-\mu^{-1}\gamma(t)).
\end{aligned}
$$

The bound (5.16) follows. The properties stated for the function b follow from its definition. ∎

THEOREM 5.11 (Measurable Smoothing Functions for CM Integrands). *Let f be a CM integrand and let $\tilde{f} : \Xi \times U \times \mathbb{R}_{++} \to \mathbb{R}$ be as given in Theorem 5.8. Then $\tilde{f}$ is a measurable smoothing function for f on $\Xi \times U \times \mathbb{R}_{++}$. Moreover, the functions $F(x) := \mathbb{E}[f(\xi, x)]$ and $\widetilde{F}(x,\mu) := \mathbb{E}[\tilde{f}(\xi, x, \mu)]$ are well defined on U and $U \times \mathbb{R}_{++}$, respectively, with $\widetilde{F}$ a smoothing function for F on U.*

Proof. By Lemma 5.4, we need only establish (i) and (ii) in Definition 4.2 to show that $\tilde{f}$ is a measurable smoothing function for f. First note that the bound (5.15) in Lemma 5.10 shows that (4.1) in part (i) in Definition 4.2 is satisfied. The bound (4.2) is also satisfied since, by Lemma 5.9,

$$|\tilde{f}(\xi,x,\mu)| \leq |\tilde{f}(\xi,\bar{x},\mu)| + \bar{\kappa}_{\mathsf{q}}[\hat{\kappa}_{\mathsf{c}}(\xi,\bar{x}) + 2\bar{\kappa}_C\hat{\kappa}_g(\xi,\bar{x})] \quad \forall\,(\xi,x,\mu) \in \Xi \times \mathbb{B}_\epsilon(\bar{x}) \times (0,\bar{\mu}]$$

(see Definition 3.2 and Lemma 5.4 for the definition of the terms in this bound). Hence Definition 5.2(i) is satisfied.

By (5.13), for all $\mu > 0$, the gradient $\nabla_x \tilde{f}(\xi, x, \mu)$ exists, and $\nabla_x \tilde{f}(\xi, \cdot, \mu)$ is continuous on U for all $\xi \in \Xi$. Also, by (5.13), Definition 5.1 and (5.14) (or, more simply Lemma 5.9),

$$|\nabla_x \tilde{f}(\xi,x,\mu)| \leq \bar{\kappa}_q[\hat{\kappa}_{\mathsf{c}}(\xi,\bar{x}) + 2\bar{\kappa}_C\hat{\kappa}_g(\xi,\bar{x})] \quad \forall\;(\xi,x,\mu) \in \Xi \times \mathbb{B}_{\epsilon(\bar{x})}(\bar{x}) \times (0,\bar{\mu}],$$

which establishes a bound stronger than (4.2) in Definition 5.2(ii) since it is independent of μ.

The final statement of the theorem follows from Lemma 4.4. ∎

5.2. Gradient sub-consistency of CM integrands. We now examine conditions under which the smoothing (5.12) of CM integrands satisfy the gradient sub-consistency property (4.10). Our approach is to develop conditions under which Theorem 4.7 can be applied. The key condition in this regard is the uniform subgradient approximation property (4.9). This property is equivalent to saying that there exists $\bar{\nu} \in (0,1)$ such that for all $\nu \in (0,\bar{\nu})$ there exist $\delta(\nu,\bar{x}) > 0$ and $\Xi(\nu,\bar{x}) \in \mathcal{M}$ with $\rho(\Xi(\nu,\bar{x})) \geq 1-\nu$ for which

$$\operatorname{dist}\left(\nabla_x \tilde{f}(\xi,x,\mu) \mid \partial_x f(\xi,\bar{x})\right) \leq \nu \ \forall\, (x,\mu) \in [(\bar{x},0)+\delta(\nu,\bar{x})(\mathbb{B}\times(0,1))] \ \text{ a.e. } \xi \in \Xi(\nu,\bar{x}). \tag{5.17}$$

To establish this condition we use Theorem 5.8 to derive a bound on the distance to $\partial_x f(\xi,\bar{x})$ in terms of the distances to the subdifferentials $\partial_t p_i(g_i(\xi,\bar{x}))$. For this we require the following Lipschitz hypothesis on the Jacobians $\nabla \mathsf{q}$, $\nabla_x g$ and $\nabla_x c$: for all $\bar{x} \in U$, there is a $\bar{\delta}(\bar{x}) > 0$ for which there exist $k_g(\bar{x}) > 0$ and $k_c(\bar{x}) > 0$ such that, for all $\xi \in \Xi$ and $x \in \mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x})$,

$$\begin{aligned} \|\nabla \mathsf{q}(y) - \nabla \mathsf{q}(\bar{y})\| &\leq \bar{\kappa}_q \|y-\bar{y}\| \\ \|\nabla_x g(\xi,x) - \nabla_x g(\xi,\bar{x})\| &\leq k_g(\bar{x})\|x-\bar{x}\|, \text{ and} \\ \|\nabla_x c(\xi,x) - \nabla_x c(\xi,\bar{x})\| &\leq k_c(\bar{x})\|x-\bar{x}\| \quad \forall\, \xi \in \Xi,\ x \in \mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x}), \end{aligned} \tag{5.18}$$

where $\bar{\kappa}_q$ is the Lipschitz constant for $\nabla \mathsf{q}$ given in Definition 5.1. The Lipschitz continuity of $\nabla_x c$ and $\nabla_x g$ on $\mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x})$ uniformly in ξ on Ξ implies that these functions are bounded on $\mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x})$ uniformly in ξ on Ξ. Denote these bounds by $\kappa_c(\bar{x})$ and $\kappa_g(\bar{x})$, respectively. we also assume that $\nabla \mathsf{q}$ is bounded by κ_{q}. Taken together, we have

$$\|\nabla \mathsf{q}\| \leq \kappa_{\mathsf{q}},\ \|\nabla_x c(\xi,x)\| \leq \kappa_c(\bar{x}) \ \text{ and } \ \|\nabla_x g(\xi,x)\| \leq \kappa_g(\bar{x}) \quad \forall\, (\xi,x) \in \Xi \times \mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x}). \tag{5.19}$$

Lemma 5.12. *Let f be a CM integrand as given in Definition 5.2 and let $\tilde{f}$ be the smoothing function for f as in Theorem 5.8 for which (5.18) and (5.19) hold, and set*

$$\begin{aligned} &K_1(\bar{x}) := [\kappa_q(k_c(\bar{x}) + 2\sqrt{m}k_g(\bar{x})\bar{\kappa}_C) + \bar{\kappa}_q(\kappa_c(\bar{x}) + \bar{\kappa}_C\kappa_g(\bar{x}))(\kappa_c(\bar{x}) + \sqrt{m}\kappa_g(\bar{x})\bar{\kappa}_C)], \\ &K_2(\bar{x}) := \sqrt{m}\omega\bar{\kappa}_q\bar{\kappa}_C(\kappa_c(\bar{x}) + \sqrt{m}\kappa_g(\bar{x})\bar{\kappa}_C), \quad \textit{and} \\ &K_3(\bar{x}) := \sqrt{m}\kappa_q\kappa_g(\bar{x}). \end{aligned}$$

If $\bar{x} \in U$ is such that $\operatorname{rank}\nabla_x g(\xi,x) = m$ *for all $x \in \mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x})$ for almost all $\xi \in \Xi$, then*

$$\begin{aligned} \operatorname{dist}\left(\nabla_x \tilde{f}(\xi,x,\mu) \,|\, \partial_x f(\xi,\bar{x})\right) \leq\, & K_1(\bar{x})\|x-\bar{x}\| + K_2(\bar{x})\mu \\ & + K_3(\bar{x}) \max_{i=1,\dots,m} \operatorname{dist}\left(\nabla_t \tilde{p}_i(g_i(\xi,x),\mu) \,|\, \partial_t p_i(g_i(\xi,\bar{x}))\right) \end{aligned} \tag{5.20}$$

for all $(\xi,x) \in \Xi \times \mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x})$ and $\mu > 0$.

Proof. Let $x \in \mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x})$ and set $Y = \operatorname{diag}(y)$ and $Z := \operatorname{diag}(z)$ where $y_i := \nabla_t \tilde{p}_i(g_i(\xi,x),\mu))$ and $z_i \in \partial_t p_i(g_i(\xi,\bar{x}))$, $i = 1,\dots,m$. Then, by Theorem 5.8, for almost all $\xi \in \Xi$,

$$\tilde{g} := \nabla_x \tilde{f}(\xi,x,\mu) = (\nabla_x \mathsf{c}(\xi,x) + Y\nabla_x g(\xi,x))^T \nabla \mathsf{q}(\mathsf{c}(\xi,x) + \tilde{C}(g(\xi,x)))$$

and

$$g := (\nabla_x \mathsf{c}(\xi,\bar{x}) + Z\nabla_x g(\xi,\bar{x}))^T \nabla \mathsf{q}(\mathsf{c}(\xi,\bar{x}) + C(g(\xi,\bar{x}))) \in \partial_x f(\xi,\bar{x}).$$

By using the bound (5.15), the constants defined in (5.18) and (5.19), and the fact that $\|\cdot\|_2 \leq \sqrt{m}\,\|\cdot\|_\infty$ on $\mathbb{R}^m$, we have

(5.21)
$$
\begin{aligned}
\|\tilde{g} - g\| \leq{}& [(k_c(\bar{x}) + 2\sqrt{m}k_g(\bar{x})\bar{\kappa}_C)\,\|x - \bar{x}\| + \kappa_g(\bar{x})\,\|Y - Z\|]\kappa_q \\
&+ [\kappa_c(\bar{x}) + \sqrt{m}\kappa_g(\bar{x})\bar{\kappa}_C]\bar{\kappa}_q\left[\kappa_c(\bar{x})\,\|x - \bar{x}\| + \left\|\tilde{C}(g(\xi, x)) - C(g(\xi, \bar{x}))\right\|\right] \\
\leq{}&[(k_c(\bar{x}) + 2\sqrt{m}k_g(\bar{x})\bar{\kappa}_C)\,\|x - \bar{x}\| + \kappa_g(\bar{x})\,\|Y - Z\|]\kappa_q \\
&+ [\kappa_c(\bar{x}) + \sqrt{m}\kappa_g(\bar{x})\bar{\kappa}_C]\bar{\kappa}_q\left[\kappa_c(\bar{x})\,\|x - \bar{x}\| + \bar{\kappa}_C(\kappa_g(\bar{x})\,\|x - \bar{x}\| + \sqrt{m}\omega\mu)\right] \\
\leq{}&[\kappa_q(k_c(\bar{x})+2\sqrt{m}k_g(\bar{x})\bar{\kappa}_C)+\bar{\kappa}_q(\kappa_c(\bar{x})+\bar{\kappa}_C\kappa_g(\bar{x}))(\kappa_c(\bar{x})+\sqrt{m}\kappa_g(\bar{x})\bar{\kappa}_C)]\,\|x-\bar{x}\| \\
&+\sqrt{m}\omega\bar{\kappa}_q\bar{\kappa}_C(\kappa_c(\bar{x})+\sqrt{m}\kappa_g(\bar{x})\bar{\kappa}_C)\mu+\sqrt{m}\kappa_q\kappa_g(\bar{x})\max_{i=1,\ldots,m}|\nabla_t\tilde{p}_i(g_i(\xi, x), \mu)-z_i|,
\end{aligned}
$$

or equivalently,

$$
\|\tilde{g} - g\| \leq K_1(\bar{x})\,\|x - \bar{x}\| + K_2(\bar{x})\mu + K_3(\bar{x})\max_{i=1,\ldots,m}|\nabla_t\tilde{p}_i(g_i(\xi, x), \mu) - z_i|,
$$

for almost all $\xi \in \Xi$, which proves the lemma. ∎

Lemma 5.12 shows that if we can obtain a bound on the distances to the subdifferentials $\partial_t p_i(g_i(\xi, \bar{x}))$ similar to the bound in (5.17), then we can choose $\hat{\delta}(\bar{x})$ and μ small enough to ensure that (5.17) also holds.

LEMMA 5.13. *Let f and $\tilde{f}$ satisfy the hypotheses of Lemma 5.12. Set*

$$
\bar{\tau} := \bar{\kappa}_C(\bar{r} + 1)/2, \quad \bar{\varepsilon} := \frac{1}{4}\min\left\{|t_{ij} - t_{i(j+1)}| \mid i = 1, \cdots, m,\ j = 2, \cdots, r_i - 1\right\}
$$

and, for every $\varepsilon \in (0, \bar{\varepsilon}]$ and $x \in U$, define

$$
\bar{\Xi}_\varepsilon(x) := \left\{ \xi \in \Xi \;\middle|\; \begin{aligned} &\exists\, i \in \{1, \cdots, m\},\ g_i(\xi, x) \in \bigcup_{j=2}^{r_i}(t_{ij} + [-\varepsilon, \varepsilon]) \\ &\qquad\qquad \text{but } g_i(\xi, x) \notin \{t_{i1}, \cdots, t_{ir_i}\} \end{aligned} \right\}.
$$

Let $\bar{x} \in U$ and consider the following assumption:

$$
\text{for any } \tau \in (0, \bar{\tau}),\ \exists\, \tilde{\varepsilon}(\tau, \bar{x}) \in (0, \bar{\epsilon}),\ \text{ s.t. } \rho(\bar{\Xi}_\varepsilon(\bar{x})) \leq \tau, \quad \forall\, \varepsilon \in (0, \tilde{\varepsilon}(\tau, \bar{x})). \tag{5.22}
$$

If $\bar{x} \in U$ is such that (5.22) holds, then, for all $i \in \{1, \ldots, m\}$, $\tau \in (0, \bar{\tau})$ and $\varepsilon \in (0, \tilde{\varepsilon}(\tau, \bar{x}))$,

$$
\operatorname{dist}\left(\nabla_t\tilde{p}_i(g_i(\xi, x), \mu) \mid \partial_t p_i(g_i(\xi, \bar{x}))\right) \leq \tau \tag{5.23}
$$

for all $\xi \in \bar{\Xi}^c_\varepsilon(\bar{x}) := \Xi \setminus \bar{\Xi}_\varepsilon(\bar{x})$ whenever $(x, \mu) \in \mathcal{B}_{\tilde{\delta}(\varepsilon,\tau,\bar{x})}(\bar{x}) \times (0, \tilde{\mu}(\varepsilon, \tau, \bar{x}))$, where $\tilde{\delta}(\varepsilon, \tau, \bar{x}) := \min\{\bar{\delta}(\bar{x}), \varepsilon/(2\kappa_g(\bar{x}))\}$ and

$$
\tilde{\mu}(\epsilon, \tau, \bar{x}) := \frac{\varepsilon}{-2\varphi^{-1}_{\min}(\frac{\tau}{(\bar{r}+1)\bar{\kappa}_C})}
$$

with $\bar{\delta}(\bar{x})$ and $\kappa_g(\bar{x})$ as given in Lemma 5.12 and (5.19), respectively.

Proof. Let $\bar{\delta}(\bar{x})$, $\bar{\kappa}_q$, κ_q, $k_g(\bar{x})$, $\kappa_g(\bar{x})$, $k_c(\bar{x})$ and $\kappa_c(\bar{x})$ be as in Lemma 5.12 and it's proof. Observe that, for every $\varepsilon \in (0, \bar{\varepsilon}]$ and $x \in U$,

$$
\bar{\Xi}^c_\varepsilon(x) = \left\{ \xi \in \Xi \;\middle|\; \begin{aligned} &\forall i \in \{1, \cdots, m\}, g_i(\xi, x) \in [t_{ij} + \varepsilon, t_{i(j+1)} - \varepsilon], \\ &\qquad\qquad \text{or } g_i(\xi, x) \in \{t_{i1}, \cdots, t_{ir_i}\} \end{aligned} \right\},
$$

and note that these sets are measurable.

By (5.22), for any $\tau \in (0, \bar{\tau})$, we have $\rho(\bar{\Xi}_\varepsilon(\bar{x})) \le \tau$ for all $\varepsilon \in (0, \tilde{\varepsilon}(\tau, \bar{x})]$. Let $\varepsilon \in (0, \tilde{\varepsilon}(\tau, \bar{x}))$. Then, for all $(\xi, x) \in \Xi \times \mathcal{B}_{\tilde{\delta}(\varepsilon,\tau,\bar{x})}(\bar{x})$, we have for $i = 1, \cdots, m$, $\|g_i(\xi, x) - g_i(\xi, \bar{x})\| \le \frac{\varepsilon}{2}$.
Let $\varepsilon \in (0, \tilde{\varepsilon}(\tau, \bar{x})]$ and $\xi \in \bar{\Xi}^c_\varepsilon(\bar{x})$. We consider two cases, both of which make use of the following two elementary facts without reference:

(a) If $t < t_1 < t_2$, then
$$|\varphi(\mu^{-1}(t - t_1)) - \varphi(\mu^{-1}(t - t_2))| \le \varphi(\mu^{-1}(t - t_2)) = \varphi(-\mu^{-1}|t - t_2|).$$

(b) If $t > t_1 > t_2$, then
$$|\varphi(\mu^{-1}(t - t_1)) - \varphi(\mu^{-1}(t - t_2))| \le 1 - \varphi(\mu^{-1}(t - t_1)) = \varphi(-\mu^{-1}|t - t_1|).$$

Case 1: $(g_i(\xi, \bar{x}) \in [t_{i\bar{j}} + \varepsilon, t_{i(\bar{j}+1)} - \varepsilon]$ for some $\bar{j} \in \{1, \cdots, r_i\})$ Let $x \in \mathcal{B}_{\tilde{\delta}(\varepsilon,\tau,\bar{x})}(\bar{x})$ and $\mu > 0$. Then $g_i(\xi, x) \in [t_{i\bar{j}} + \frac{\varepsilon}{2}, t_{i(\bar{j}+1)} - \frac{\varepsilon}{2}]$ for all $x \in \mathcal{B}_{\tilde{\delta}(\varepsilon,\tau,\bar{x})}(\bar{x})$, in which case $\nabla_t p_i(g_i(\xi, \bar{x})) = \nabla_t p_i(g_i(\xi, x)) = \eta_i(g_i(\xi, x))$. By Lemma 5.10, we have

$$|\nabla_t \tilde{p}_i(g_i(\xi, x), \mu) - \nabla_t p_i(g_i(\xi, \bar{x}))| \le (r_i + 1)\bar{\kappa}_C \varphi(-\mu^{-1}\gamma_i(g_i(\xi, x))) \le (\bar{r} + 1)\bar{\kappa}_C \varphi(-\tfrac{\varepsilon}{2\mu}).$$

Since $\tau/(\bar{\kappa}_C(\bar{r}+1)) \le 1/2$ (so that $\varphi^{-1}_{\min}(\frac{\tau}{(\bar{r}+1)\bar{\kappa}_C}) < 0$ by (5.6)), we have the inequality $|\nabla_t \tilde{p}_i(g_i(\xi, x), \mu) - \nabla_t p_i(g_i(\xi, \bar{x}))| \le \tau$ whenever $0 < \mu \le \tilde{\mu}(\epsilon, \tau, \bar{x})$. Hence, for any $(x, \mu) \in \mathcal{B}_{\tilde{\delta}(\varepsilon,\bar{x})}(\bar{x}) \times (0, \tilde{\mu}(\varepsilon, \tau, \bar{x}))$, we have

$$|\nabla_t \tilde{p}_i(g_i(\xi, x), \mu) - \nabla_t p_i(g_i(\xi, \bar{x}))| \le \tau.$$

Case 2: $(g_i(\xi, \bar{x}) = t_{i\bar{j}}$ for some $\bar{j} \in \{2, \cdots, r_i\})$ In this case $\partial_t p_i(g_i(\xi, \bar{x})) = [a_{i(\bar{j}-1)}, a_{i\bar{j}}]$. Clearly,

$$\tilde{\eta}(g_i(\xi, x), \mu) := a_{i(\bar{j}-1)}\left(1 - \varphi(\mu^{-1}(g_i(\xi, x) - t_{i\bar{j}}))\right) + a_{i\bar{j}}\varphi(\mu^{-1}(g_i(\xi, x) - t_{i\bar{j}})) \in \partial_t p_i(g_i(\xi, \bar{x})),$$

and so

$$\operatorname{dist}\left(\nabla_t \tilde{p}_i(g_i(\xi, x), \mu) \mid \partial_t p_i(g_i(\xi, \bar{x}))\right) \le |\nabla_t \tilde{p}_i(g_i(\xi, x), \mu) - \tilde{\eta}_i(g_i(\xi, \bar{x}), \mu)|.$$

If $r_i = 2$, then (5.8) tells us that $\nabla_t \tilde{p}_i(g_i(\xi, x), \mu) = \tilde{\eta}_i(g_i(\xi, \bar{x}), \mu)$, so that

$$|\nabla_t \tilde{p}_i(g_i(\xi, x), \mu) - \tilde{\eta}_i(g_i(\xi, \bar{x}))| = 0 \le (r_2 + 1)\bar{\kappa}_C \varphi(-\frac{\varepsilon}{2\mu}).$$

If $r_i \ge 3$ and $2 \ne \bar{j} \ne r_i$, the expression in (5.8) for $\nabla_t \tilde{p}_i(g_i(\xi, x), \mu)$ tells us that

$$\begin{aligned}
& |\nabla_t \tilde{p}_i(g_i(\xi, x), \mu) - \tilde{\eta}_i(g_i(\xi, \bar{x}), \mu)| \\
\le\ & \textstyle\sum_{j=1}^{(\bar{j}-2)} \bar{\kappa}_C(1 - \varphi(\mu^{-1}(g_i(\xi, x) - t_{ij}))) \\
& \qquad + |a_{i(\bar{j}-1)}(\varphi(\mu^{-1}(g_i(\xi, x) - t_{i(\bar{j}-1)})) - \varphi(\mu^{-1}(g_i(\xi, x) - t_{i\bar{j}}))) \\
& \qquad + a_{i\bar{j}}(\varphi(\mu^{-1}(g_i(\xi, x) - t_{i\bar{j}})) - \varphi(\mu^{-1}(g_i(\xi, x) - t_{i(\bar{j}+1)}))) - \tilde{\eta}_i(g_i(\xi, \bar{x}), \mu)| \\
& + \textstyle\sum_{j=\bar{j}+1}^{r_i} \bar{\kappa}_C \varphi(\mu^{-1}(g_i(\xi, x) - t_{ij})) \\
\le\ & (\bar{j} - 2)\bar{\kappa}_C \varphi(-\tfrac{\varepsilon}{2\mu}) + |a_{i(\bar{j}-1)}|(1 - \varphi(\mu^{-1}(g_i(\xi, x) - t_{i(\bar{j}-1)})))| \\
& \qquad + |a_{i\bar{j}}|\varphi(\mu^{-1}(g_i(\xi, x) - t_{i(\bar{j}+1)})) + (r_i - \bar{j})\bar{\kappa}_C \varphi(-\tfrac{\varepsilon}{2\mu}) \\
\le\ & \bar{r}\bar{\kappa}_C \varphi(-\tfrac{\varepsilon}{2\mu}).
\end{aligned}$$

If $r_i \geq 3$ and $\bar{j} = 2$ or $\bar{j} = r_i$,

$$|\nabla_t \tilde{p}_i(g_i(\xi, x), \mu) - \tilde{\eta}_i(g_i(\xi, \bar{x}), \mu)| \leq (\bar{r} - 1)\bar{\kappa}_C \varphi(-\frac{\varepsilon}{2\mu}).$$

Hence, we always have

$$\operatorname{dist}(\nabla_t \tilde{p}_i(g_i(\xi, x), \mu) \mid \partial_t p_i(g_i(\xi, \bar{x}))) \leq (\bar{r} + 1)\bar{\kappa}_C \varphi(-\frac{\varepsilon}{2\mu}),$$

and so, as in Case 1, whenever $0 < \mu \leq \tilde{\mu}(\epsilon, \tau, \bar{x})$, we have

$$\operatorname{dist}(\nabla_g \tilde{p}_i(g_i(\xi, x), \mu) \mid \partial_g p_i(g_i(\xi, \bar{x}))) \leq \tau.$$

The result follows by combining these two cases. ∎

Remark 5.14. One can strengthen the hypothesis (5.22) to

$$(5.24) \quad \exists\, \tau > 0 \text{ s.t. } \forall\, \tau \in (0, \bar{\tau})\ \exists\, \tilde{\varepsilon}(\tau, \bar{x}) \in (0, \bar{\epsilon}) \text{ s.t. } \rho(\widehat{\Xi}_\varepsilon(\bar{x})) \leq \tau \quad \forall\, \varepsilon \in (0, \tilde{\varepsilon}(\tau, \bar{x})).$$

Then Lemma 5.13 still holds. However, under (5.24), we have that $\rho(\widetilde{\Xi}(\bar{x})) = 1$, where

$$\widetilde{\Xi}(\bar{x}) = \left\{ \xi \in \Xi \;\middle|\; \forall\, i \in \{1, \ldots, m\}\ \xi \notin (g_i(\cdot, x))^{-1}(\{t_{i2}, \ldots, t_{ir_i}\}) \right\}$$

is defined in Proposition 5.5. Consequently, (5.24) implies that $f(\xi, \cdot)$ is continuously differentiable and subdifferentially regular at $\bar{x}$ for almost all $\xi \in \Xi$.

We now combine Lemmas 5.12 and 5.13 to establish conditions under which (5.17) is satisfied.

THEOREM 5.15 (Uniform Subgradient Approximation for CM Integrands). *Let f be a CM integrand as given in Definition 5.2 and let $\tilde{f}$ be the smoothing function for f given in Theorem 5.8. Suppose that the basic assumptions of Lemmas 5.12 and 5.13 are satisfied so that their conclusions hold. Then $\tilde{f}$ satisfies the uniform subgradient approximation property at $\bar{x}$. That is, there exists $\bar{\nu} \in (0, 1)$ such that, for all $\nu \in (0, \bar{\nu})$, there exists $\delta(\nu, \bar{x}) > 0$ and $\Xi(\nu, \bar{x}) \in \mathcal{M}$ with $\rho(\Xi(\nu, \bar{x})) \leq 1 - \nu$ for which (5.17) is satisfied.*

Proof. Let $\bar{x} \in U$ be such that $\operatorname{rank}\nabla_x g(\xi, \bar{x}) = m$ and let $\nu \in (0, \bar{\nu})$. Let $\bar{\delta}(\bar{x})$, $K_1(\bar{x})$, $K_3(\bar{x})$ and $K_3(\bar{x})$ be as given by Lemma 5.12 so that (5.20) is satisfied for all $x \in \mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x})$ and $\mu > 0$. Set

$$\delta_1(\nu, \bar{x}) := \min\{\bar{\delta}(\bar{x}), \nu/(3K_1(\bar{x}))\}, \;\; \mu_1(\nu, \bar{x}) := \nu/(3K_2(\bar{x})), \;\; \text{and} \;\; \tau := \nu/(3K_3(\bar{x}){+}1).$$

Let $\tilde{\varepsilon}(\tau, \bar{x})$ be as in (5.22). Take $\varepsilon \in (0, \tilde{\varepsilon}(\tau, \bar{x}))$, and set $\Xi(\nu, \bar{x})$ equal to $\bar{\Xi}^c_\epsilon(\bar{x})$. Observe that $\rho(\Xi(\nu, \bar{x})) \geq 1 - \tau \geq 1 - \nu$ by construction. Set $\delta_2(\nu, \bar{x}) := \min\{\delta_1(\nu, \bar{x}), \tilde{\delta}(\varepsilon, \tau, \bar{x})\}$ and $\mu_2(\nu, \bar{x}) := \min\{\mu_1(\nu, \bar{x}), \tilde{\mu}(\varepsilon, \tau, \bar{x})\}$ where $\tilde{\delta}(\varepsilon, \tau, \bar{x})$ and $\tilde{\mu}(\varepsilon, \tau, \bar{x})$ are given in (5.23). Then, by (5.20) and the definitions given above,

$$\begin{aligned}
\operatorname{dist}\Big(\nabla_x \tilde{f}(\xi, x, \mu) \,|\, \partial_x f(\xi, \bar{x})\Big) \leq\, & K_1(\bar{x})\|x - \bar{x}\| + K_2(\bar{x})\mu \\
& + K_3(\bar{x}) \max_{i=1,\ldots,m} \operatorname{dist}(\nabla_t \tilde{p}_i(g_i(\xi, x), \mu) \,|\, \partial_t p_i(g_i(\xi, \bar{x}))) \\
\leq\, & \frac{\nu}{3} + \frac{\nu}{3} + \frac{\nu}{3} = \nu
\end{aligned}$$

for all $x \in \mathcal{B}_{\delta_2(\nu, \bar{x})}(\bar{x})$ and $\mu \in (0, \mu_2(\nu, \bar{x}))$. ∎

We apply Theorem 4.7 to obtain the gradient sub-consistency of smoothed CM integrands.

THEOREM 5.16 (Gradient Sub-Consistency of Smoothed CM Integrands). *Let f be a CM integrand as given in Definition 5.2 and let $\tilde{f}$ be the smoothing function for f given in Theorem 5.8. Suppose that the basic assumptions of Lemmas 5.12 and 5.13 are satisfied so that their conclusions hold. We further assume that $f(\xi,\cdot)$ is subdifferentially regular $\bar{x}$ for almost all $\xi \in \Xi$ or $-f(\xi,\cdot)$ is subdifferentially regular at $\bar{x}$ for almost all $\xi \in \Xi$. Then $\widetilde{F}(x,\mu) := \mathbb{E}[\tilde{f}(\xi,x,\mu)]$ satisfies the gradient sub-consistency property (4.10) at $\bar{x}$, i.e.,*

$$\mathrm{co}\left\{\operatorname*{Limsup}_{x\to\bar{x},\mu\downarrow 0} \nabla \widetilde{F}(x,\mu)\right\} \subseteq \partial F(\bar{x}) = \mathbb{E}\left[\mathrm{co}\left\{\operatorname*{Limsup}_{x\to\bar{x},\mu\downarrow 0} \nabla_x \tilde{f}(\xi,x,\mu)\right\}\right].$$

Proof. The result follows once it is shown that the hypotheses of Theorem 4.7 are satisfied. Theorem 5.15 tells us that the uniform subgradient approximation property is satisfied. Lemma 5.4 shows that every CM integrand is an LL integrand, so, the subdifferential regularity requirement implies that the hypotheses of Corollary 3.6 are satisfied. Hence, all hypotheses of Theorem 4.7 are satisfied at $\bar{x}$. ∎

5.3. Subgradient approximation via smoothing without regulartity. Theorem 5.16 uses the subdifferential regularity of $f(\xi,x)$ or $-f(\xi,x)$ for almost all ξ to obtain the gradient sub-consistency property of the smoothing approximation $\widetilde{F}$. However, gradient sub-consistency is often stronger than what is required in some applications. In this section it is shown that a useful subgradient approximation result can be obtained without assumptions on subdifferential regularity.

Let $\tilde{f}$ be the measurable smoothing function introduced in (5.12). We show that if $(\xi,\bar{x}) \in \Xi \times X$ is such that $\mathrm{rank}\nabla_x g(\xi,\bar{x}) = m$, then the limit $\lim_{\mu\downarrow 0} \nabla_x \tilde{f}(\xi,\bar{x},\mu)$ exits, and we provide an explicit formula for this limit. For $i = 1,\ldots,m$, define the functions

$$\mathsf{z}_i(\xi,\bar{x}) := \eta_i(g_i(\xi,\bar{x}))\nabla_x g_i(\xi,\bar{x})$$

$$\mathsf{h}_i^1(\xi,\bar{x})(v) := \begin{cases} a_{ij}\nabla_x g_i(\xi,\bar{x})^T v, & t_{ij} < g_i(\xi,\bar{x}) < t_{i(j+1)},\ j = 1,\ldots,r_i \\ a_{ij}\nabla_x g_i(\xi,\bar{x})^T v, & \langle \nabla_x g_i(\xi,\bar{x}), v\rangle \geq 0,\ g_i(\xi,\bar{x}) = t_{ij},\ j = 2,\ldots,r_i \\ a_{i(j-1)}\nabla_x g_i(\xi,\bar{x})^T v, & \langle \nabla_x g_i(\xi,\bar{x}), v\rangle < 0,\ g_i(\xi,\bar{x}) = t_{ij},\ j = 2,\ldots,r_i, \end{cases}$$

$$\mathsf{h}_i^2(\xi,\bar{x})(v) := \begin{cases} a_{ij}\nabla_x g_i(\xi,\bar{x})^T v, & t_{ij} < g_i(\xi,\bar{x}) < t_{i(j+1)},\ j = 1,\ldots,r_i \\ a_{i(j-1)}\nabla_x g_i(\xi,\bar{x})^T v, & \langle \nabla_x g_i(\xi,\bar{x}), v\rangle \geq 0,\ g_i(\xi,\bar{x}) = t_{ij},\ j = 2,\ldots,r_i \\ a_{ij}\nabla_x g_i(\xi,\bar{x})^T v, & \langle \nabla_x g_i(\xi,\bar{x}), v\rangle < 0,\ g_i(\xi,\bar{x}) = t_{ij},\ j = 2,\ldots,r_i, \end{cases}$$

where the functions η_i are defined in (5.9). Note that

$$\langle \mathsf{z}_i(\xi,\bar{x}), v\rangle = \frac{1}{2}(\mathsf{h}_i^1(\xi,\bar{x})(v) + \mathsf{h}_i^2(\xi,\bar{x})(v)), \tag{5.25}$$

and, by Lemma 5.6,

$$\mathsf{z}(\xi,\bar{x}) = \mathrm{diag}(\eta_i(g_i(\xi,\bar{x})))\nabla_x g(\xi,x). \tag{5.26}$$

LEMMA 5.17. *Consider the CM integrand f and its smoothing function $\tilde{f}$ defined in (5.12). Assume that $\mathrm{rank}\nabla_x g(\xi,\bar{x}) = m$ for a fixed $(\xi,\bar{x}) \in \Xi \times X$. Then the* *following limits exist as given with* $u(\xi,\bar{x}) \in \partial_x f(\xi,\bar{x})$*: for all* $v \in \mathbb{R}^n$,

$$u(\xi,\bar{x}) := \lim_{\mu\downarrow 0} \nabla_x \tilde{f}(\xi,\bar{x},\mu) = (\nabla_x \mathsf{c}(\xi,\bar{x}) + (\mathsf{z}_1(\xi,\bar{x}),\ldots,\mathsf{z}_m(\xi,\bar{x}))^T \nabla \mathsf{q}(\mathsf{c}(\xi,\bar{x}) + C(g(\xi,\bar{x}))) \tag{5.27}$$

$$\ell_1(\xi,\bar{x};v) := \lim_{t\downarrow 0} \frac{f(\xi,\bar{x}+2tv) - f(\xi,\bar{x}+tv)}{t} = \nabla \mathsf{q}(\mathsf{c}(\xi,\bar{x}) + C(g(\xi,\bar{x})))^T(\nabla_x \mathsf{c}(\xi,\bar{x})v + (\mathsf{h}_1^1(\xi,\bar{x})(v),\cdots,\mathsf{h}_m^1(\xi,\bar{x})(v))^T) \tag{5.28}$$

and

$$\ell_2(\xi,\bar{x};v) := \lim_{t\downarrow 0} \frac{f(\xi,\bar{x}-tv) - f(\xi,\bar{x}-2tv)}{t} = \nabla \mathsf{q}(\mathsf{c}(\xi,\bar{x}) + C(g(\xi,\bar{x})))^T(\nabla_x \mathsf{c}(\xi,\bar{x})v + (\mathsf{h}_1^2(\xi,\bar{x})(v),\cdots,\mathsf{h}_m^2(\xi,\bar{x})(v))^T). \tag{5.29}$$

Moreover, by (5.25)*, we have*

$$\langle u(\xi,\bar{x}), v\rangle = \frac{1}{2}(\ell_1(\xi,\bar{x};v) + \ell_2(\xi,\bar{x};v)) \quad \forall\ v \in \mathbb{R}^n. \tag{5.30}$$

Proof. By combining (5.13) and (5.9), we find that the right hand side of (5.27) is an element of $\partial_x f(\xi,\bar{x})$. Moreover, by (5.9) in Lemma 5.6 and (5.13) in Theorem 5.8, the limit $u(\xi,\bar{x})$ exists as given in (5.27). Since (5.30) follows from (5.28) and (5.29), it remains only to establish the limits $\ell_1(\xi,\bar{x};v)$ and $\ell_2(\xi,\bar{x};v)$ exist as given.

First consider the nonsmooth functions $h_i(\xi,x) := p_i(g_i(\xi,x))$, $i = 1,\cdots,m$. For each ξ, the functions $h_i(\xi,\cdot)$ are convex-composite functions [4]. Hence, by [4, Section 2],

$$\partial_x h_i(\xi,\bar{x}) = \partial_x(p_i \circ g_i)(\xi,\bar{x})) = \partial_t p_i(g_i(\xi,\bar{x}))\nabla_x g_i(\xi,\bar{x}) \tag{5.31}$$

and

$$\nabla_x(\tilde{p}_i \circ g_i)(\xi,\bar{x}),\mu) = \nabla_t \tilde{p}_i(g_i(\xi,\bar{x}),\mu)\nabla_x g_i(\xi,\bar{x}).$$

By combining (5.2), (5.26) and (5.31), we have that $\mathsf{z}_i(\xi,x) \in \partial_x(p_i \circ g_i)(\xi,x)$ for all $(\xi,x) \in \Xi \times X$. Since, for each $x \in X$, $\mathsf{z}_i(\xi,x)$ is defined by η_i which is the limit of measurable functions in ξ from (5.9), $\mathsf{z}_i(\xi,x)$ is measurable in ξ. In addition, by [4, Section 2], each of the mappings $x \mapsto h_i(\xi,x) = p_i(g_i(\xi,x))$ is Clarke regular. Since $g(\xi,x)$ is smooth, $\lim_{x'\to\bar{x}} \nabla_x g_i(\xi,x') = \nabla_x g_i(\xi,\bar{x})$. Combining (5.2) and Lemma 3.8, for any $x \in X$ and direction $v \in \mathbb{R}^n$, we have

$$\mathsf{h}_i^1(\xi,\bar{x})(v) = \lim_{t\downarrow 0} \frac{h_i(\xi,x+2tv) - h_i(\xi,x+tv)}{t} = \max_{\nu\in\partial_x h_i(\xi,x)} \langle \nu, v\rangle$$

and

$$\mathsf{h}_i^2(\xi,\bar{x})(v) = \lim_{t\downarrow 0} \frac{h_i(\xi,x-tv) - h_i(\xi,x-2tv)}{t} = \min_{\nu\in\partial_x h_i(\xi,x)} \langle \nu, v\rangle.$$

Note that, for every $t > 0$, the mean value theorem tells us that there exists $w_t \in \mathbb{R}^m$ on the line segment connecting the two vectors $\mathsf{c}(\xi,\bar{x}-tv) + C(g(\xi,\bar{x}-tv))$

and $\mathsf{c}(\xi,\bar{x}-2tv)+C(g(\xi,\bar{x}-2tv))$ such that

$$\begin{aligned}\ell_2(\xi,\bar{x};v)&=\lim_{t\downarrow 0}\frac{\mathsf{q}(\mathsf{c}(\xi,\bar{x}-tv)+C(g(\xi,\bar{x}-tv)))-\mathsf{q}(\mathsf{c}(\xi,\bar{x}-2tv)+C(g(\xi,\bar{x}-2tv)))}{t}\\&=\lim_{t\downarrow 0}\nabla\mathsf{q}(w_t)^T\frac{(\mathsf{c}(\xi,\bar{x}-tv)+C(g(\xi,\bar{x}-tv)))-(\mathsf{c}(\xi,\bar{x}-2tv)+C(g(\xi,\bar{x}-2tv)))}{t}\\&=\lim_{t\downarrow 0}\nabla\mathsf{q}(w_t)^T\left(\frac{\mathsf{c}(\xi,\bar{x}-tv)-\mathsf{c}(\xi,\bar{x}-2tv)}{t}+\frac{C(g(\xi,\bar{x}-tv))-C(g(\xi,\bar{x}-2tv))}{t}\right)\\&=\nabla\mathsf{q}(\mathsf{c}(\xi,\bar{x})+C(g(\xi,\bar{x})))^T(\nabla\mathsf{c}(\xi,\bar{x})^Tv+(\mathsf{h}_1^2(\xi,\bar{x})(v),\cdots,\mathsf{h}_m^2(\xi,\bar{x})(v))^T),\end{aligned}$$

and similarly,

$$\ell_2(\xi,\bar{x};v)=\nabla\mathsf{q}(\mathsf{c}(\xi,\bar{x})+C(g(\xi,\bar{x})))^T(\nabla\mathsf{c}(\xi,\bar{x})^Tv+(\mathsf{h}_1^1(\xi,\bar{x})(v),\cdots,\mathsf{h}_m^1(\xi,\bar{x})(v))^T).$$

This establishes (5.28) and (5.29) which combined imply (5.30). ∎

THEOREM 5.18 (Subgradient Approximation by Smoothing). *Consider the CM integrand f and its smoothing function $\tilde{f}$ defined in (5.12), and suppose the hypotheses of Lemma 5.12 hold. Set $F(x):=\mathbb{E}[f(\xi,x)]$ and $\widetilde{F}(x,\mu):=\mathbb{E}[\tilde{f}(\xi,x,\mu)]$ for all $x\in X$. Then, for almost all $\xi\in\Xi$,*

(5.32)
$$\operatorname{dist}\left(\nabla_x\tilde{f}(\xi,\bar{x},\mu)\,|\,\partial_x f(\xi,\bar{x})\right)\le K_2(\bar{x})\mu+K_3(\bar{x})(\bar{r}+1)\bar{\kappa}_C\max_{i=1,\ldots,m}\varphi(-\mu^{-1}\gamma_i(g_i(\xi,\bar{x}))).$$

Moreover, $\widetilde{F}(\cdot,\mu)$ is differentiable at $\bar{x}$ for all $\mu>0$ with $\nabla_x\widetilde{F}(\bar{x},\mu)=\mathbb{E}[\nabla_x\tilde{f}(\xi,\bar{x},\mu)]$, the function u in (5.27) is well defined, and,

$$\lim_{\mu\downarrow 0}\nabla_x\widetilde{F}(\bar{x},\mu)=\lim_{\mu\downarrow 0}\mathbb{E}[\nabla_x\tilde{f}(\xi,\bar{x},\mu)]=\mathbb{E}[u(\xi,\bar{x})]\in\partial\mathbb{E}[f(\xi,\bar{x})]=\partial F(\bar{x}).$$

Proof. Combining (5.16) and (5.20), we have (5.32).

By Lemma 5.4, f is an LL integrand. By Lemma 5.17, the function u in (5.27) is well defined a.e. in Ξ and measurable. By Theorem 5.11, $\widetilde{F}(\cdot,\mu)$ is differentiable at $\bar{x}$ for all $\mu>0$ with $\nabla_x\widetilde{F}(\bar{x},\mu)=\mathbb{E}[\nabla\tilde{f}(\xi,\bar{x},\mu)]$. By Theorem 5.11, (5.27) and the Dominated Convergence Theorem,

$$\lim_{\mu\downarrow 0}\nabla_x\widetilde{F}(\bar{x},\mu)=\lim_{\mu\downarrow 0}\mathbb{E}[\nabla_x\tilde{f}(\xi,\bar{x},\mu)]=\mathbb{E}[u(\xi,\bar{x})].$$

Since $\partial\mathbb{E}[f(\xi,\bar{x})]=\partial F(\bar{x})$, it remains only to show that $\mathbb{E}[u(\xi,\bar{x})]\in\partial F(\bar{x})$.

Let $v\in\mathbb{R}^n$. By [14, Theorem 2.7], each of the mappings $\nabla\tilde{f}(\xi,x,\mu)$ is measurable in ξ for each $(x,\mu)\in X\times\mathbb{R}_{++}$. By Lemma 5.17 and [14, Proposition 2.7], $u(\xi,\bar{x})\in\partial_x f(\xi,\bar{x})$ is measurable. Moreover, by (5.30) in Lemma 5.17,

$$\ell_1(\xi,\bar{x};v)-\langle u(\xi,\bar{x}),v\rangle=-(\ell_2(\xi,\bar{x};v)-\langle u(\xi,\bar{x}),v\rangle)\quad\text{a.e. }\xi\in\Xi,$$

with both limits existing and measurable. Consequently,

$$\mathbb{E}[\ell_1(\xi,\bar{x};v)]-\mathbb{E}[\langle u(\xi,\bar{x}),v\rangle]=-(\mathbb{E}[\ell_2(\xi,\bar{x};v)]-\mathbb{E}[\langle u(\xi,\bar{x}),v\rangle]).\tag{5.33}$$

Since f is an LL integrand,

$$\max\left\{\frac{|f(\xi,\bar{x}+2tv)-f(\xi,\bar{x}+tv)|}{t},\ \frac{|f(\xi,\bar{x}-tv)-f(\xi,\bar{x}-2tv)|}{t}\right\}\le\kappa_f(\xi,\bar{x})\,\|v\|\,.$$

Therefore, by the Dominated Convergence Theorem,

$$\mathbb{E}[\ell_1(\xi,\bar{x};v)] = \lim_{t\downarrow 0} \mathbb{E}\left[\frac{f(\xi,\bar{x}+2tv)-f(\xi,\bar{x}+tv)}{t}\right] \quad \text{and}$$
$$\mathbb{E}[\ell_1(\xi,\bar{x};v)] = \lim_{t\downarrow 0} \mathbb{E}\left[\frac{f(\xi,\bar{x}-tv)-f(\xi,\bar{x}-2tv)}{t}\right],$$

which tells us that

$$\begin{aligned}\widehat{d}F(\bar{x})(v) &= \limsup_{t\downarrow 0, z\to\bar{x}} \mathbb{E}\left[\frac{f(\xi,z+tv)-f(\xi,z)}{t}\right]\\ &\geq \lim_{t\downarrow 0} \mathbb{E}\left[\frac{f(\xi,\bar{x}+2tv)-f(\xi,\bar{x}+tv)}{t}\right] = \mathbb{E}[\ell_1(\xi,\bar{x};v)]\end{aligned}$$

and

$$\begin{aligned}\widehat{d}F(\bar{x})(v) &= \limsup_{t\downarrow 0, z\to\bar{x}} \mathbb{E}\left[\frac{f(\xi,z+tv)-f(\xi,z)}{t}\right]\\ &\geq \lim_{t\downarrow 0} \mathbb{E}\left[\frac{f(\xi,\bar{x}-tv)-f(\xi,\bar{x}-2tv)}{t}\right] = \mathbb{E}[\ell_2(\xi,\bar{x};v)].\end{aligned}$$

Hence, by (5.33),

$$\begin{aligned}\widehat{d}F(\bar{x})(v) - \langle \mathbb{E}[u(\xi,\bar{x})]\,,\, v\rangle &\geq \mathbb{E}[\ell_1(\xi,\bar{x};v)] - \langle \mathbb{E}[u(\xi,\bar{x})]\,,\, v\rangle\\ &= -(\mathbb{E}[\ell_2(\xi,\bar{x};v)] - \langle \mathbb{E}[u(\xi,\bar{x})]\,,\, v\rangle)\end{aligned}$$

and

$$\widehat{d}F(\bar{x})(v) - \langle \mathbb{E}[u(\xi,\bar{x})]\,,\, v\rangle \geq \mathbb{E}[\ell_2(\xi,\bar{x};v)] - \langle \mathbb{E}[u(\xi,\bar{x})]\,,\, v\rangle\,,$$

and so

$$\widehat{d}F(\bar{x})(v) - \langle \mathbb{E}[u(\xi,\bar{x})]\,,\, v\rangle \geq |\, \mathbb{E}[\ell_2(\xi,\bar{x};v)] - \langle \mathbb{E}[u(\xi,\bar{x})]\,,\, v\rangle \,| \geq 0.$$

Since v was chosen arbitrarily, Definition 2.2 tells us that $\mathbb{E}[u(\xi,\bar{x})] \in \partial F(\bar{x})$. ∎

COROLLARY 5.19. *Let the assumptions of Theorem 5.18 hold. For $\mu > 0$ and $\bar{x} \in U$, set*

$$K_4(\bar{x}) := [K_1(\bar{x})\mu + 2\bar{r}K_3(\bar{x})\kappa_g(\bar{x})].$$

Then

$$\left\|\nabla\tilde{f}(\xi,x,\mu) - \nabla\tilde{f}(\xi,\bar{x},\mu)\right\| \leq \frac{K_4(\bar{x})}{\mu}\,\|x-\bar{x}\| \quad \forall\ \xi\in\Xi \ \ \textit{and} \ \ x\in\mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x}) \tag{5.34}$$

and

$$\operatorname{dist}\Big(\mathbb{E}[\nabla\tilde{f}(\xi,x,\mu)]\,|\,\partial F(\bar{x})\Big) \leq \frac{K_4(\bar{x})}{\mu}\,\|x-\bar{x}\| + \operatorname{dist}\Big(\nabla_x\widetilde{F}(\bar{x},\mu)\,|\,\partial F(\bar{x})\Big)\ \forall\ x\in\mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x}). \tag{5.35}$$

Moreover, we have the following gradient sub-consistency property at $\bar{x}$ for any $\gamma \in (0,1)$:

$$\operatorname*{Limsup}_{x\to\bar{x},\,\mu=O(\|x-\bar{x}\|^\gamma)} \mathbb{E}[\nabla\tilde{f}(\xi,x,\mu)] \in \partial F(\bar{x}). \tag{5.36}$$

Proof. The proof of (5.34) follows the pattern of proof given for (5.21) to first establish that

$$\|\nabla \tilde f(\xi,x,\mu) - \nabla \tilde f(\xi,\bar x,\mu)\| \le K_1(\bar x)\|x-\bar x\| + K_3(\bar x) \max_{i=1,\dots,m} |\nabla_t \tilde p_i(g_i(\xi,x,\mu) - \nabla_t \tilde p_i(g_i(\xi,\bar x,\mu)|.$$

Then use (5.11) to obtain the bound (5.34).

To see (5.35), note that

$$\begin{aligned}
\operatorname{dist}\Big(\mathbb{E}[\nabla \tilde f(\xi,x,\mu)] \,|\, \partial F(\bar x)\Big) &\le \|\mathbb{E}[\nabla \tilde f(\xi,x,\mu)] - \mathbb{E}[\nabla \tilde f(\xi,\bar x,\mu)]\| \\
&\quad + \operatorname{dist}\Big(\nabla_x \mathbb{E}[\tilde f(\xi,\bar x,\mu)] \,|\, \partial F(\bar x)\Big) \\
&\le \frac{K_4(\bar x)}{\mu}\|x-\bar x\| + \operatorname{dist}\Big(\nabla_x \mathbb{E}[\tilde f(\xi,\bar x,\mu)] \,|\, \partial F(\bar x)\Big) \\
&\le K_4(\bar x)\frac{\|x-\bar x\|}{\mu} + \operatorname{dist}\Big(\nabla_x \widetilde F(\bar x,\mu) \,|\, \partial F(\bar x)\Big).
\end{aligned}$$

Hence, (5.36) follows from Theorem 5.18. ∎

6. Conclusion. We provide a framework for the study of smoothing functions for non-smooth random integrands with the primary focus being the study of the gradient consistency property and the approximation of Clarke subgradients of expectation functions. For the large class of measurable CM functions, we show the gradient sub-consistency property when the integrand or its negative is subdifferentially regular for almost all $\xi \in \Xi$ (Theorems 5.15-5.16). Moreover, when this subdifferential regularity hypothesis fails, we show that for any $x \in \mathbb{R}^n$,

$$\lim_{\mu\downarrow 0} \mathbb{E}[\nabla \tilde f(\xi,x,\mu)] \in \partial \mathbb{E}[f(\xi,x)] \quad \text{and} \quad \operatorname*{Limsup}_{x\to\bar x,\, \mu=O(\|x-\bar x\|^\gamma)} \mathbb{E}[\nabla \tilde f(\xi,x,\mu)] \in \partial \mathbb{E}[f(\xi,x)].$$

Consequently, we can approximate an element of the Clarke subgradient of the expectation function using gradients of a smoothing function for the non-smooth integrand (Theorem 5.18 and Corollary 5.19). Measurable CM functions arise in several important applications, e.g.

$$\mathbb{E}[\|\min(x,\varphi(\xi,x))\|^2] \quad \text{and} \quad \mathbb{E}[(\max(a(\xi)^T x,0) - b(\xi))^2] + \lambda \sum_{i=1}^m \log(1+|d_i^T x|)).$$

The first example comes from stochastic nonlinear complementarity problems with a continuously differentiable function $\varphi : \Xi \times R^n \to R^n$ [10, 11], and the second is from optimal statistical learning problems with $a(\xi) \in R^n, b(\xi) \in R$ and $d_i \in R^n$ [1, 3]. Our goal is to apply these approximation techniques in cases where the inclusion $\partial\mathbb{E}[f(\xi,x)] \subseteq \mathbb{E}[\partial f(\xi,x)]$ is insufficient for guiding both numerical optimization and optimality assessment.

REFERENCES

[1] M. Ahn, J.S. Pang and J. Xin, *Difference-of-convex learning I: directional stationarity, optimality, and sparsity,* Manuscript, University of Southern California, Los Angeles, California, U. S. A., 2016.

[2] R.J. Aumann, *Integrals of set-valued functions,* J. Math. Anal. Appl., **12** (1965), pp. 1–12.

[3] R. Blundell and J.L. Powell, Censored regression quantiles with endogenous regression, J. Econometrics, **141** (2007), pp. 65-83.

[4] J.V. Burke, *Second order necessary and sufficient conditions for convex composite NDO,* Math. Program., **38** (1987), pp. 287–302.
[5] J.V. Burke and T. Hoheisel, *Epi-convergent smoothing with applications to convex composite functions,* SIAM J. Optim., **23** (2013), pp. 1457–1479.
[6] J.V. Burke, T. Hoheisel and C. Kanzow, *Gradient consistency for integral-convolution smoothing functions,* Set-Valued Var. Anal., **21** (2013), pp. 359 – 376.
[7] J.V. Burke and D.R. Luke, *Variational analysis applied to the problem of optimal phase retieval,* SIAM J. Control Optim., **42** (2003), pp. 576 – 595.
[8] C. Chen and O.L. Mangasarian, *A class of smoothing functions for nonlinear and mixed complementarity problems*, Comput. Optim. Appl., **5** (1996), pp. 97–138.
[9] X. Chen, *Smoothing methods for nonsmooth, nonconvex minimization,* Math. Program., **134** (2012), pp. 71–99.
[10] X. Chen and M. Fukushima, *Expected residual minimization method for stochastic linear complementarity problems,* Math. Oper. Res., **30** (2005), pp. 1022–1038.
[11] X. Chen, R. Wets and Y. Zhang, *Stochastic variational inequalities: residual minimization smoothing sample average approximations,* SIAM J. Optim., **22** (2012), pp. 649–673.
[12] F.H. Clarke, *Optimization and Nonsmooth Analysis*, volume 5 of *Classics in Applied Mathematics*, SIAM, Philadelphia, 1990.
[13] F.H. Clarke, Yu. S. Ledyaev, R. J. Stern, and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*, Springer Verlag, New York, 1998.
[14] G.B. Folland, *Real Analysis, 2nd edition*, Wiley, New York, 1999.
[15] W. Hildenbrand, *Core and Equilibria of a Large Economy*, Princeton University Press, Princeton, New Jersey, 1974.
[16] A. Nemirovski, A. Juditsky, G. Lan and A. Shapiro, *Robust stochastic approximation approach to stochastic programming,* SIAM J. Optim., **19** (2009), pp. 1574–1906.
[17] R.T. Rockafellar, *Conjugate Duality and Optimization*, SIAM, Philadelphia, 1974.
[18] R.T. Rockafellar, *Integral functionals, normal integrands and measurable selections,* In Nonlinear Operators in the Calculus of Variations, number 543 in Lecture Notes in Mathematics, Springer-Verlag, New York, (1976), pp. 157–207.
[19] R.T. Rockafellar and R. Wets, *Variational Analysis*, Springer, New York, 1998.
[20] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, San Francisco, 3rd edition, 1976.
[21] A. Shapiro, D. Dentcheva and A. Ruszczynski, *Lectures on stochastic programming: modeling and theory*, SIAM, Philadelphia, 2009.
[22] R. Wets, *Stochastic Programming,* In: Nemhauser, G.L., et al. (eds.) Handbooks in OR & MS, **1** (1989), pp. 573–629.
[23] H. Xu and J.J. Ye, *Necessary optimality conditions for two-stage stochastic programs with equilibaium constraints,* SIAM J. Optim., **20** (2010), pp. 1685–1715.
[24] H. Xu and D. Zhang, *Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications,* Math. Program., **119** (2009), pp. 371–401.