Mathematical Programming (2020) 181:229–264
https://doi.org/10.1007/s10107-019-01441-9

FULL LENGTH PAPER

Series B

The subdifferential of measurable composite max integrands and smoothing approximation

James V. Burke[1] · **Xiaojun Chen**[2] · **Hailin Sun**[3]

Received: 24 March 2018 / Accepted: 12 October 2019 / Published online: 24 October 2019

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 of the expectation equals the expectation of their Clarke subdifferential. In particular, this holds for convex integrands. However, little is known about the calculation of Clarke subgradients for the expectation of 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*. This class contains non-regular integrands from stochastic complementarity problems as well as stochastic optimization problems arising in statistical learning.

Keywords Stochastic optimization · Clarke subgradient · Smoothing · Non-regular integrands

Mathematics Subject Classification 90C15

This work is support by NSF Grant No. DMS-1908890, Hong Kong Research Grant Council Grant 153000/17P, and NSFC Grant Nos. 11871276 and 11571178.

✉ Xiaojun Chen
maxjchen@polyu.edu.hk

James V. Burke
jvburke@uw.edu

Hailin Sun
hlsun@njnu.edu.cn

1 Department of Applied Mathematics, University of Washington, Seattle, USA

2 Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China

3 Jiangsu Key Lab for NSLSCS, School of Mathematical Sciences, Nanjing Normal University, Nanjing 210023, China

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

$$\min_{x \in X} \quad F(x) := \mathbb{E}[f(\xi, x)], \tag{1}$$

where $\xi : \Omega \to \Xi$ is a random variable on the probability space $(\Omega, \mathcal{F}, \varsigma)$ with ς absolutely continuous with respect to Lebesgue measure, $f : \Xi \times X \to \mathbb{R}$ 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) if it satisfies

$$0 \in \partial F(x) + \mathcal{N}_X(x), \tag{2}$$

where ∂ denotes the Clarke subdifferential (see Appendix Definition 9 (i)) and $\mathcal{N}_X(x)$ is the normal cone to X at $x \in X$. Condition (2) is a first-order necessary condition for optimality of problem (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 [29,31,33] employs the first-order necessary condition

$$0 \in \mathbb{E}[\partial_x f(\xi, x)] + \mathcal{N}_X(x), \tag{3}$$

where

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

is the Aumann (set-valued) expectation of $\partial_x f(\xi, x)$ with respect to ξ defined in [2]. Points x satisfying (3) are called *weak stationary points* for problem (1) [21,34]. In some cases, elements of $\mathbb{E}[\partial_x f(\xi, x)]$ can be computed [26]. However, as the name implies, condition (3) is much weaker than condition (2). In particular, it is always the case that

$$\partial F(x) \subseteq \mathbb{E}[\partial_x f(\xi, x)] = \operatorname{co} \mathbb{E}[\partial_x f(\xi, x)], \tag{4}$$

where "co" denotes the convex hull. In (4), the inclusion is given by Clarke [14, Theorem 2.7.2] and equivalence follows from either Aumann's Convexity Theorem [2,23] or Lyapunov's Convexity Theorem [19,30]. Moreover, since the Clarke subdifferential is the closed convex hull of the Mordukhovich subdifferential (M-subdifferential) ∂^M ((see Appendix Definition 9) [27, Definition 8.3]), the subdifferential inclusion

$$\partial^M F(x) \subseteq \partial F(x) \subseteq \mathbb{E}[\partial_x f(\xi, x)] = \mathbb{E}\left[\operatorname{co} \partial_x^M f(\xi, x)\right] = \mathbb{E}\left[\partial_x^M f(\xi, x)\right]$$

holds where the final equality follows from [2, Theorem 3] (see [20, Lemma 6.18] for connections to the M-subdifferential).

Consequently, if x satisfies (2), then x satisfies (3), but the converse is not in general true. On the other hand, [14, Theorem 2.7.2] tells us that if $f(\xi, \cdot)$ is Clarke regular [14, Definition 2.3.4] on X for almost all $\xi \in \Xi$, then equality holds in (4). Unfortunately, in many applications of interest, Clarke regularity fails to hold, and the set $\mathbb{E}[\partial_x f(\xi, x)]$ is much larger than the set $\partial\mathbb{E}[f(\xi, x)]$. For example, this occurs in stochastic nonlinear complementarity problems [11,13] and optimal statistical learning problems [1,3]. In such cases, condition (3) is too weak for assessing optimality. By way of illustration, consider $f(\xi, x) = \xi|x|$ with $\xi \sim N(0, 1)$ and $x \in \mathbb{R}$. Then $\mathbb{E}[f(\xi, x)] = \mathbb{E}[\xi|x|] \equiv 0$ for $x \in \mathbb{R}$, but $\mathbb{E}[\partial f(\xi, 0)] = \sqrt{2/\pi}[-1, 1]$.

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 [11,13] and optimization problems in statistical learning [1,3].

The paper is organized as follows. In Sect. 2 and the Appendix, we recall some basic definitions and properties from variational analysis, the theory of measurable multifunctions, and the study of the variational properties of the expectation function. In Sect. 3, we define measurable smoothing functions and give a few fundamental properties. In particular, we introduce the notions of gradient consistency and sub-consistency. In Sect. 4, 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 approximated via smoothing in the absence of regularity.

2 The subdifferential properties of $F(x) := \mathbb{E}[f(\xi, x)]$

In this section, we study the relationship between the variational properties of f and $F(x) = \mathbb{E}[f(\xi, x)]$. 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. In particular, we study those properties of f that give access to the desired variational properties of F. For example, it has already been mentioned 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 [14, Hypothesis 2.7.1]).

2.1 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. In particular, this implies that ρ is non-atomic. Let $\hat{\rho}$ denote the induced product measure on $\mathbb{R}^\ell \times \mathbb{R}^n$. We consider the following class of functions.

Definition 1 (*Carathéodory Mappings*) [27, 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 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

$$|f(\xi, x_1) - f(\xi, x_2)| \leq \kappa_f(\xi, \bar{x})\|x_1 - x_2\| \quad \forall\, x_1, x_2 \in \mathbb{B}_{\epsilon(\bar{x})}(\bar{x}) \text{ and a.e. } \xi \in \Xi,$$

where a.e. denotes "almost every" and $\mathbb{B}_\epsilon(\bar{x}) := \{x \mid \|x - \bar{x}\| \leq \epsilon\} \subseteq U$.

Here and throughout, $L_n^m(\mathbb{R}^\ell, \mathcal{M}, \rho)$ denotes the Banach space of m-integrable functions $\kappa : \mathbb{R}^\ell \to \mathbb{R}^n$ defined on the measure space $(\mathbb{R}^\ell, \mathcal{M}, \rho)$, where the norm is given by $\|\kappa\|_m := [\int_{\mathbb{R}^\ell}[\sum_{j=1}^n |\kappa_j|^m]d\rho]^{1/m}$.

In what follows we often have functions of two variables, $h(u, v)$, but need to discuss the variational objects for this function with respect to only one of the two variables. For this we use $\nabla_v h(u, v)$, $\partial_v h(u, v)$ and $d_v h(u, v)(d)$ to denote the derivative, the Clarke subdifferential, and the subderivative of $h(u, v)$ in the direction d, respectively, in v for fixed u.

Lemma 1 (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 Appendix Definition 10) for a.e. $\xi \in \Xi$ with $\operatorname{lip}_x f(\xi, \bar{x}) \leq \kappa_f(\xi, \bar{x})$ 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 $\widehat{d}_x f(\xi, x)(v)$ (see Appendix Definition 8) 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 $\bar{x} \in U$*

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

for all $x \in \mathbb{B}_{\epsilon(\bar{x})}(\bar{x})$, where $\|\kappa_f(\cdot, \bar{x})\|_2 := \sqrt{\int_\Xi \kappa_f^2(\xi, \bar{x})d\rho}$.

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

(b) This follows immediately from the inequality

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

(c) This follows from the well known fact that the limsup of measurable functions is measurable, e.g. [17, Theorem 2.7].

(d) This follows immediately from Proposition 2 of the Appendix since $\kappa_f(\cdot, \bar{x}) \in L_1^2(\mathbb{R}^\ell, \mathcal{M}, \rho)$.

(e) This is an immediate consequence of [14, 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 [27, 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 11.

2.2 Subdifferential regularity

If f is an LL integrand on $\Xi \times U$, then, by Lemma 1(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 1 (The subdifferential of F) *[14, 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*

$$\partial F(x) \subseteq \mathbb{E}[\partial_x f(\xi, x)] \quad \forall x \in U. \tag{5}$$

If, in addition, $\bar{x} \in U$ is such that $f(\xi, \cdot)$ is subdifferentially regular (see Appendix Definition 9 (ii)) in x at $\bar{x}$ for a.e. $\xi \in \Xi$, then F is subdifferentially regular at $\bar{x}$ and equality holds in (5).

Remark 1 In [14, 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 1 *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 a.e. $\xi \in \Xi$ or $-f(\xi, \cdot)$ is subdifferentially regular at $\bar{x} \in U$ for a.e. $\xi \in \Xi$, then equality holds in (5).*

Proof If $f(\xi, \cdot)$ is subdifferentially regular in x at $\bar{x} \in U$ for a.e. $\xi \in \Xi$, then the result follows from Theorem 1. If $-f(\xi, \cdot)$ is subdifferentially regular in x at $\bar{x}$ for a.e. $\xi \in \Xi$, then, by [14, Proposition 2.3.1] and Theorem 1,

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

Note that, in opposition to Theorem 1, 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 1 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, observe that

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

Nonetheless, by Corollary 1, $\partial F(0) = \mathbb{E}[\partial f(\xi, 0)]$. This can also be verified by direct computation.

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

Lemma 2 *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*

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

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

$$
\begin{aligned}
\ell_1(\xi, x; v) &= d_x h(\xi, x)(v) = \sup\{\langle g\,,\ v\rangle \mid g \in \partial_x h(\xi, x)\} \qquad \textit{and} \\
\ell_2(\xi, x; v) &= -d_x h(\xi, x)(-v) = \inf\{\langle g\,,\ v\rangle \mid g \in \partial_x h(\xi, x)\}.
\end{aligned} \tag{6}
$$

Proof Strict continuity (Appendix Definition 10) tells us that

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

so that $d_x h(\xi, x)(v)$ is finite for all $v \in \mathbb{R}^n$. Therefore, by (55), the limit $\ell_1(\xi, x; v)$ exists and

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

The first equvalence in (6) now follows from (56). The second equivalence follows from the first by replacing v by $-v$.

3 Smoothing functions for $F(x) := \mathbb{E}[f(\xi, x)]$

3.1 Measurable smoothing functions

Definition 3 (*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 [27, Definition 5.41], i.e., $\lim_{\mu\downarrow 0, x\to\bar{x}} \widetilde{F}(x, \mu) = F(\bar{x}) \quad \forall \bar{x} \in U$, 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 (*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 $\tilde{f}(\cdot, x, \mu) \in L_1^1(\mathbb{R}^\ell, \mathcal{M}, \rho)$ 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 a.e. $\xi \in \Xi$ in the sense of [27, Definition 5.41], i.e.,

$$\lim_{\mu\downarrow 0, x\to\bar{x}} \tilde{f}(\xi, x, \mu) = f(\xi, \bar{x}) \quad \forall \bar{x} \in U \text{ and a.e. } \xi \in \Xi, \tag{7}$$

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

$$|\tilde{f}(\xi, x, \mu)| \le \kappa_f(\xi, \bar{x}, \bar{\mu}) \quad \forall (\xi, x, \mu) \in \Xi \times V \times (0, \bar{\mu}]\,. \tag{8}$$

(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

$$\left\|\nabla_x \tilde{f}(\xi, x, \mu)\right\| \leq \hat{\kappa}_f(\xi, \bar{x}, \bar{\mu}), \quad \forall\, (\xi, x, \mu) \in \Xi \times V \times (0, \bar{\mu}]\,. \tag{9}$$

Remark 2 Just as in Lemma 1, Lemma 11 can be applied to show that (9) implies that

$$\int_E \nabla_x \tilde{f}(\xi, x, \mu) d\rho \in \left\|\hat{\kappa}_f(\cdot, \bar{x}, \bar{\mu})\right\|_2 \rho(E)\mathbb{B} \quad \forall\, (x, \mu) \in V \times (0, \bar{\mu}] \tag{10}$$

for all $E \in \mathcal{M}$. Note that if we assume that $\tilde{f}(\xi, \cdot, \cdot)$ is Lipschitz continuous, then (8) holds. The conditions in (8) and (9) are added to the usual notion of smoothing function in Definition 3 to facilitate the application of the Dominated Convergence Theorem when needed.

Lemma 3 (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*

$$\nabla_x \widetilde{F}(x, \mu) = \mathbb{E}[\nabla_x \tilde{f}(\xi, x, \mu)] \quad \forall\, (x, \mu) \in U \times \mathbb{R}_{++}. \tag{11}$$

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 (7), (8) and the Dominated Convergence Theorem, for all $x \in U$,

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

which establishes (i) in Definition 3.

Next let $(\bar{x}, \bar{\mu}) \in U \times \mathbb{R}_{++}$ and $d \in \mathbb{R}^n$ with $d \neq 0$. By (9) 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

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

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 (11).

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 (11), (9) and the Dominated Convergence Theorem,

$$\lim_{x \to \bar{x}} \nabla_x \widetilde{F}(x, \bar{\mu}) = \lim_{x \to \bar{x}} \mathbb{E}[\nabla_x \tilde{f}(\xi, x, \bar{\mu})]$$
$$= \mathbb{E}\left[\lim_{x \to \bar{x}} \nabla_x \tilde{f}(\xi, x, \bar{\mu})\right] = \mathbb{E}[\nabla_x \tilde{f}(\xi, \bar{x}, \bar{\mu})] = \nabla_x \widetilde{F}(\bar{x}, \bar{\mu}).$$

3.2 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 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

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

where the limit superior is taken in the multi-valued sense (57).

When $\widetilde{F} \equiv F$, the definition reduces to that of the Clarke subdifferential for the finite-dimensional case (Appendix Definition 9).

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 2 (Gradient consistency and subgradient approximation) *Let $U \subseteq \mathbb{R}^n$, $\bar{x} \in U$, and $f : \Xi \times U \to \mathbb{R}$ be as in Corollary 1, 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 a.e. $\xi \in \Xi$, $\tilde{f}(\xi, \cdot, \cdot)$ is gradient consistent at $\bar{x}$, i.e.*

$$\operatorname{co}\left\{\operatorname*{Limsup}_{x \to \bar{x}, \mu \downarrow 0} \nabla_x \tilde{f}(\xi, x, \mu)\right\} = \partial_x f(\xi, \bar{x}) \quad a.e.\ \xi \in \Xi, \tag{12}$$

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

$$\partial F(\bar{x}) = \mathbb{E}\left[\operatorname{co}\left\{\operatorname*{Limsup}_{x \to \bar{x}, \mu \downarrow 0} \nabla_x \tilde{f}(\xi, x, \mu)\right\}\right] = \operatorname{co} \mathbb{E}\left[\operatorname*{Limsup}_{x \to \bar{x}, \mu \downarrow 0} \nabla_x \tilde{f}(\xi, x, \mu)\right]. \tag{13}$$

Proof The fact that $\widetilde{F}$ is a smoothing function for F comes from Theorem 3. Therefore, the result is an immediate consequence of Corollary 1 and the Lyapunov convexity theorem [30].

The B-subdifferential, denote ∂^B, is defined in Definition 9 part (iii) of the "Appendix". In particular, we know that for a fixed $\xi \in \Xi$,

$$\partial_x^B f(\xi, x) \subseteq \partial_x^M f(\xi, x) \subseteq \partial_x f(\xi, x)$$

and

$$\mathrm{co}\, \partial_x^B f(\xi, x) = \mathrm{co}\, \partial_x^M f(\xi, x) = \partial_x f(\xi, x).$$

Moreover, by [2, Theorems 1 and 3]

$$\mathbb{E}[\partial_x^B f(\xi, x)] = \mathbb{E}[\partial_x^M f(\xi, x)] = \mathbb{E}[\partial_x f(\xi, x)].$$

If we replace (12) by

$$\left\{ \operatorname*{Limsup}_{x \to \bar{x}, \mu \downarrow 0} \nabla_x \tilde{f}(\xi, x, \mu) \right\} = \partial_x^B f(\xi, \bar{x}) \quad \text{a.e. } \xi \in \Xi, \tag{14}$$

then we have the subdifferential inclusion

$$\partial F(\bar{x}) = \mathbb{E}\left[\mathrm{co} \left\{ \operatorname*{Limsup}_{x \to \bar{x}, \mu \downarrow 0} \nabla_x \tilde{f}(\xi, x, \mu) \right\} \right] = \mathbb{E}[\mathrm{co}\, \partial_x^B f(\xi, x)] = \mathbb{E}[\partial_x^B f(\xi, x)]. \tag{15}$$

Obviously, condition (14) implies condition (12). The above argument also holds if we replace ∂^B in (14)–(15) by ∂^M. However, the converse is not true. For example, consider $f(\xi, x) = -\xi |x|$, where ξ follows uniform distribution over [0.5, 1.5]. It is clear that both $-f(\xi, x)$ and $-\mathbb{E}[f(\xi, x)] = |x|$ are subdifferentially regular. Taking $\tilde{f}(\xi, x, \mu) := -\xi \sqrt{x^2 + \mu}$, we have

$$\operatorname*{Limsup}_{x \to 0, \mu \downarrow 0} \nabla_x \tilde{f}(\xi, x, \mu) = \operatorname*{Limsup}_{x \to 0, \mu \downarrow 0} \frac{\xi x}{\sqrt{x^2 + \mu}} = \partial_x f(\xi, 0) = [-\xi, \xi] \quad \text{a.e. } \xi \in \Xi,$$

and

$$\partial \mathbb{E}[f(\xi, 0)] = \mathbb{E}[\partial_x f(\xi, 0)] = [-1, 1].$$

However,

$$\partial_x^B f(\xi, 0) = \{-\xi, \xi\}, \;\; \partial^B \mathbb{E}[f(\xi, 0)] = \{-1, 1\}, \;\; \mathbb{E}[\partial_x^B f(\xi, 0)] = [-1, 1]$$
$$\partial_x^M f(\xi, 0) = \{-\xi, \xi\}, \;\; \partial^M \mathbb{E}[f(\xi, 0)] = \{-1, 1\}, \;\; \mathbb{E}[\partial_x^M f(\xi, 0)] = [-1, 1].$$

Thus (14) does not hold for either ∂^B or ∂^M. This simple example tells us that even for a function $f(\xi, x)$ for which $-f(x, \xi)$ is subdifferentially regular in x for all $\xi \in \Xi$, it may not be the case that

$$\partial^B \mathbb{E}[f(\xi, x)] = \mathbb{E}[\partial_x^B f(\xi, x)], \quad \partial^M \mathbb{E}[f(\xi, x)] = \mathbb{E}[\partial_x^M f(\xi, x)]$$

and the gradient sub-consistency. Therefore, in this paper we consider the smoothing approximation for the Clarke subdifferential.

The pointwise condition (12) does not imply the gradient consistency of $\widetilde{F}$ at $\bar{x}$. To obtain such a result from (13) we also need to know that

$$\operatorname{co}\mathbb{E}\left[\operatorname*{Limsup}_{x\to\bar{x},\mu\downarrow 0}\nabla_x\tilde{f}(\xi,x,\mu)\right]=\operatorname{co}\left\{\operatorname*{Limsup}_{x\to\bar{x},\mu\downarrow 0}\mathbb{E}[\nabla_x\tilde{f}(\xi,x,\mu)]\right\}. \qquad (16)$$

The equivalence (16) is nontrivial requiring stronger hypotheses.

Since $\partial_x f(\xi,x)$ is compact-, convex-valued in x, the left-hand side of (12) is contained in the right-hand side if and only if for a.e. $\xi\in\Xi$ 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 3 (Gradient sub-consistency) *Let $U\subseteq\mathbb{R}^n$ and $f:\Xi\times U\to\mathbb{R}$ be as in Corollary 1, 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 $\rho(\Xi(\nu,\bar{x}))\geq 1-\nu$ satisfying for a.e. $\xi\in\Xi(\nu,\bar{x})$*

$$\nabla_x\tilde{f}(\xi,x,\mu)\in\partial_x f(\xi,\bar{x})+\nu\mathbb{B}\quad\forall\,(x,\mu)\in[(\bar{x},0)+\delta(\nu,\bar{x})(\mathbb{B}\times(0,1))], \qquad (17)$$

then

$$\operatorname{co}\left\{\operatorname*{Limsup}_{x\to\bar{x},\mu\downarrow 0}\nabla\widetilde{F}(x,\mu)\right\}\subseteq\partial F(\bar{x})=\operatorname{co}\mathbb{E}\left[\operatorname*{Limsup}_{x\to\bar{x},\mu\downarrow 0}\nabla_x\tilde{f}(\xi,x,\mu)\right]. \qquad (18)$$

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 1(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

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

Then

$$\begin{aligned}g&=\lim_{k\to\infty}\nabla_x\widetilde{F}(x^k,\mu_k)\\ &=\lim_{k\to\infty}\int_{\xi\in\Xi(\nu,\bar{x})}\nabla_x\tilde{f}(\xi,x^k,\mu_k)d\xi+\int_{\xi\in\Xi/\Xi(\nu,\bar{x})}\nabla_x\tilde{f}(\xi,x^k,\mu_k)d\xi\end{aligned}$$

$$
\begin{aligned}
&= \lim_{k\to\infty} \int_{\xi\in\Xi} q_k(\xi)d\xi - \int_{\xi\in\Xi\setminus\Xi(\nu,\bar{x})} s(\xi)d\xi + \int_{\xi\in\Xi\setminus\Xi(\nu,\bar{x})} \nabla_x \tilde{f}(\xi, x^k, \mu_k)d\xi \\
&\in \mathbb{E}[\partial_x f(\xi,\bar{x})] + \nu\mathbb{B} + \nu\left\|\kappa_f(\cdot,\bar{x})\right\|_2 \mathbb{B} + \nu\left\|\hat{\kappa}_f(\cdot,\bar{x},\bar{\mu})\right\|_2 \mathbb{B} \\
&= \partial F(\bar{x}) + \nu(1 + \left\|\kappa_f(\cdot,\bar{x})\right\|_2 + \left\|\hat{\kappa}_f(\cdot,\bar{x},\bar{\mu})\right\|_2)\mathbb{B},
\end{aligned}
$$

where the second equation follows from Theorem 3, the inclusion follows from (17), Theorem 1(f), and (10), and the final equation follows from Corollary 1. Since $\nu \in (0, \bar{\nu})$ was chosen arbitrarily, this proves the inclusion in (18) since $\partial F(\bar{x})$ is closed. The equivalence in (18) follows from Theorem 2 since (17) implies (12).

In what follows, we refer to (18) as the *gradient sub-consistency property* for the smoothing function $\widetilde{F}$ at $\bar{x}$, and we refer to (17) as the *uniform subgradient approximation property* for the measurable smoothing function $\tilde{f}$ at $\bar{x}$.

4 Composite max (CM) integrands

In this section we introduce the class of CM integrands, and smoothing functions for these integrands, that 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 appeared in the work of Chen and Mangasarian [8] and were later expanded by Chen [9] to a broader class of non-smooth 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 [5,6,8,9], 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 6 (*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 $x \in U$, the gradient $\nabla_x g(\xi, x)$ is locally L^2 bounded in x 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

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

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

We now define CM integrands.

Definition 7 (*Composite max (CM) Integrands*) A CM integrand on $\Xi \times X$ is any mapping of the form

$$f(\xi, x) := \mathsf{q}(\mathsf{c}(\xi, x) + C(g(\xi, x))) \tag{19}$$

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

(i) $C : \mathbb{R}^m \to \mathbb{R}^m$ 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,
(ii) 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
(iii) the mapping $\mathsf{q} : \mathbb{R}^m \to \mathbb{R}$ is continuously differentiable with Lipschitz continuous derivative on the set

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

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

Remark 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); the conditional value at risk (CVaR) [24]; and the difference of two Clarke regular functions where nonsmoothness occurs due the presence of compositions with piecewise convex functions. Ralph and Xu [22] discussed Aumann's integral of piecewise random set-valued mappings which include some special CM integrands. The censored regression problem in statistics and machine learning has many important applications [3] and takes the form

$$\min_x \mathbb{E}[(\max(a(\xi)^T x, 0) - b(\xi))^2].$$

The function $f(\xi, x) := (\max(a(\xi)^T x, 0) - b(\xi))^2$ is an example of a CM integrand. In this case, $m = 1$, and

$$c(\xi, x) := -b(\xi),\ \ C(y) := \max(y, 0) - b(\xi),\ \ g(\xi, x) := a(\xi)^T x,\ \ q(z) := z^2.$$

Following [6, Section 4], we assume with no loss of generality that for each $i = 1, \ldots, m$, there is a positive integer r_i and scalar pairs $(a_{ij}, b_{ij}),\ i = 1, \ldots, m,\ j = 1, \ldots, r_i$ such that $p_i(t) := \max\left\{a_{ij}t + b_{ij} \mid j = 1, \ldots, r_i\right\}$, where $a_{i1} < a_{i2} < \cdots < a_{i(r_i-1)} < a_{ir_i}$. Again with no loss of generality, we assume that the scalar pairs $(a_{ij}, b_{ij}),\ i = 1, \ldots, m,\ j = 1, \ldots, r_i$ 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 $j = 1, \ldots, r_i - 1$, $a_{ij}t_{i(j+1)} + b_{ij} = a_{i(j+1)}t_{i(j+1)} + b_{i(j+1)}$ and

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

This representation for the functions p_i gives

$$\partial p_i(t) = \begin{cases} a_{ij}, & t_{ij} < t < t_{i(j+1)},\ j = 1, \ldots, r_i \\ [a_{i(j-1)}, a_{ij}], & t = t_{ij},\ j = 2, \ldots, r_i. \end{cases} \tag{20}$$

It is easily shown that the functions p_i and C (19) are globally Lipschitz continuous with common Lipschitz constant

$$\bar{\kappa}_C := \max\{|a_{ij}| \mid i = 1, \ldots, m,\ j = 1 \ldots, r_i\}. \tag{21}$$

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 4 (CM Integrands are LL Integrands) *Let $f : \Xi \times X \to \mathbb{R}$ be an CM integrand as in (19). Then f is an LL integrand on $\Xi \times X$, where, for all $\bar{x} \in X$, one may take*

$$\kappa_f(\cdot, \bar{x}) := \bar{\kappa}_q[\hat{\kappa}_{\mathsf{c}}(\cdot, \bar{x}) + \bar{\kappa}_C \hat{\kappa}_g(\cdot, \bar{x})], \tag{22}$$

where $\bar{\kappa}_q$ and $\bar{\kappa}_C$ are defined in Definition 7 and (21), respectively, and $\hat{\kappa}_{\mathsf{c}}$ and $\hat{\kappa}_g$ are given in Definition 6.

Since the functions p_i are continuously differentiable on the open set $\mathbb{R}\backslash\{t_{i2}, \ldots, t_{ir_i}\}$ 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 1 *Let $f : \Xi \times X \to \mathbb{R}$ be a CM integrand as in (19), and, for each $i = 1, \ldots, m$, set $q_i(\xi, x) := p_i(g_i(\xi, x))$. Given $(\xi, x) \in \Xi \times U$ set*

$$\widetilde{U}_i(\xi) := \left\{ x \in U \;\middle|\; x \notin (g_i(\xi, \cdot))^{-1}(\{t_{i2}, \ldots, t_{ir_i}\}) \right\},\ i = 1, \ldots, m,$$
$$\widetilde{\Xi}_i(x) := \left\{ \xi \in \Xi \;\middle|\; \xi \notin (g_i(\cdot, x))^{-1}(\{t_{i2}, \ldots, t_{ir_i}\}) \right\},\ i = 1, \ldots, m,$$
$$\widetilde{U}(\xi) := \bigcap_{i=1}^{m} \widetilde{U}_i(\xi) \ \ \text{and}\ \ \widetilde{\Xi}(x) := \bigcap_{i=1}^{m} \widetilde{\Xi}_i(x).$$

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

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

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

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

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 a.e.* $\xi \in \Xi$.

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 with the amenable derivatives of $\mathsf{c}(\xi, \cdot)$ and $g(\xi, \cdot)$ (e.g. see [28, 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 [27, 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 evident.

4.1 Smoothing CM integrands

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

$$\int_{\mathbb{R}} \beta(t)\, dt = 1,\ \ \beta(t) = \beta(-t)\ \text{ and }\ \omega := \int_{\mathbb{R}} |t|\beta(t)\, dt < \infty. \tag{23}$$

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

$$\phi(0) = \frac{1}{2},\ \ 1 - \phi(x) = \phi(-x),\ \ \lim_{x\to\infty} \phi(x) = 1\ \text{ and }\ \lim_{x\to-\infty} \phi(x) = 0. \tag{24}$$

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

$$\phi^{-1}_{\min}(\alpha) := \inf\left\{\zeta \;\middle|\; \zeta \in \phi^{-1}(\alpha)\right\} \leq \phi^{-1}_{\max}(\alpha) := \sup\left\{\zeta \;\middle|\; \zeta \in \phi^{-1}(\alpha)\right\} \tag{25}$$

with both $\phi^{-1}_{\min}(\alpha)$ and $\phi^{-1}_{\max}(\alpha)$ finite. Finally, we note that since β is a non-negative, piecewise continuous density function, 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., $|\phi(t_1) - \phi(t_2)| \leq \beta_{\max}|t_1 - t_2|$.

Lemma 5 [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 non-negative, symmetric, piecewise continuous density satisfying* (23). *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

$$\begin{aligned}\nabla_t \tilde{p}_i(t,\mu) = \; & a_{i1}\Big(1-\phi\Big(\frac{t-t_{i2}}{\mu}\Big)\Big)\\ & +\sum_{j=2}^{r_i-1} a_{ij}\Big(\phi\Big(\frac{t-t_{ij}}{\mu}\Big)-\phi\Big(\frac{t-t_{i(j+1)}}{\mu}\Big)\Big)+a_{ir_i}\phi\Big(\frac{t-t_{ir_i}}{\mu}\Big),\end{aligned} \tag{26}$$

$$\eta_i(t) := \lim_{\mu\downarrow 0}\nabla_t \tilde{p}_i(t,\mu) = \begin{cases} a_{ij}, & t_{ij} < t < t_{i(j+1)},\ j=1,\ldots,r_i \\ \frac{1}{2}(a_{i(j-1)}+a_{ij}), & t = t_{ij},\ j=2,\ldots,r_i \end{cases} \tag{27}$$

is an element of $\partial p_i(t)$, *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*

$$|\tilde{p}_i(\hat{t},\hat{\mu})-\tilde{p}_i(\bar{t},\bar{\mu})| \le \bar{\kappa}_C[\,|\hat{t}-\bar{t}|+|\hat{\mu}-\bar{\mu}|\,] \;\; \textit{and} \tag{28}$$

$$|\nabla_t\tilde{p}_i(\hat{t},\hat{\mu})-\nabla_t\tilde{p}_i(\bar{t},\bar{\mu})| \le \frac{2r_i}{\hat{\mu}}[\,|\hat{t}-\bar{t}|+(1-\hat{\mu}/\bar{\mu})\max_j|\bar{t}-t_{ij}|\,]. \tag{29}$$

Remark 4 The bounds (28) and (29) do not appear in [6], but are straightforward to verify directly from the definitions and (26).

Theorem 4 (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*

$$\tilde{f}(\xi,x,\mu) := q(c(\xi,x)+\tilde{C}(g(\xi,x),\mu)), \tag{30}$$

where $\tilde{C}(y,\mu) := [\tilde{p}_1(y_1,\mu),\tilde{p}_2(y_2,\mu),\ldots,\tilde{p}_m(y_m,\mu)]^T$ *with each* $\tilde{p}_i$ *is as given in Lemma 5, 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* $\mu>0$,

$$\begin{aligned}\nabla_x\tilde{f}(\xi,\bar{x},\mu) = (\nabla_x c(\xi,\bar{x}) + \operatorname{diag}(\nabla_t\tilde{p}_i(g_i(\xi,\bar{x}),\mu))\nabla_x g(\xi,\bar{x}))^T\nabla q(c(\xi,\bar{x})\\ +\tilde{C}(g(\xi,\bar{x})))\end{aligned} \tag{31}$$

and

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

where

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

We now proceed to show that the function $\tilde{f}$ defined in (30) is a measurable smoothing function for f in the sense of Definition 7. First observe that the expression for $\nabla_t \tilde{p}_i(t,\mu)$ in Lemma 5 implies the bound

$$2\bar{\kappa}_C \geq \|\mathrm{diag}(\nabla_t \tilde{p}_i(g_i(\xi,\bar{x}),\mu)))\|_\infty \quad \forall\,(\xi,\bar{x},\mu) \in \Xi \times X \times \mathbb{R}_{++}. \tag{32}$$

Since this bound is independent of μ, we can use it in conjunction with the representation (31) to provide a Lipschitz constant for $\tilde{f}$ analogous to (22).

Lemma 6 (Smoothed CM integrands are LL integrands) *Let $\tilde{f} : \Xi \times X \to \mathbb{R}$ be as in Theorem 4. 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 $\kappa_{\tilde{f}_\mu}(\cdot,\bar{x}) := \bar{\kappa}_q[\hat{\kappa}_c(\cdot,\bar{x}) + 2\bar{\kappa}_C\hat{\kappa}_g(\cdot,\bar{x})]$.*

We also have the following bounds for the functions p_i, $\tilde{p}_i$, $\nabla_t \tilde{p}_i$, and η_i.

Lemma 7 *For $i = 1,\dots,m$, let $\nabla_t \tilde{p}_i$ and η_i be as in Lemma 5, and set*

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

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

$$\bar{\kappa}_C(|\bar{t} - t| + \mu\omega) \geq |\, p_i(\bar{t}) - \tilde{p}_i(t,\mu)\,| \quad \forall\,\bar{t}, t \in \mathbb{R}, \quad \text{and} \tag{33}$$

$$b_i(t,\mu) := (\bar{r}+1)\bar{\kappa}_C\phi(-\mu^{-1}\gamma_i(t)) \;\geq\; |\,\nabla_t \tilde{p}_i(t,\mu) - \eta_i(t)\,|, \tag{34}$$

where ω is defined in (23), $\bar{\kappa}_C$ is defined in (21) 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 $(\mathbb{R}\backslash\{t_{i2},\dots,t_{ir_i}\}) \times (0,+\infty)$ when $r_i \geq 3$. In addition, for all $(t,\mu) \in \mathbb{R} \times (0,\infty)$, $0 \leq b(t,\mu) \leq \frac{1}{2}(\bar{r}+1)\bar{\kappa}_C$, and $b(t,\cdot)$ is non-decreasing on $(0,+\infty)$ with $\lim_{\mu\uparrow\infty} b(t,\mu) = \frac{1}{2}(\bar{r}+1)\bar{\kappa}_C$ and $\lim_{\mu\downarrow 0} b(t,\mu) = 0$.

Proof The bound (33) 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 $t \neq t_2$, (ii) $r = 2$ and $t = t_2$, (iii) $r \geq 3$ and ($t < t_2$ or $t > t_r$), (iv) $r \geq 3$ and $t = t_k$, and (v) $r \geq 3$ and $t_k < t < t_{k+1}$ with $2 \leq k \leq r-1$. 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 (23)–(25).

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

$$|\,\nabla_t \tilde{p}_i(t,\mu) - \eta_i(t)\,| \leq r\bar{\kappa}_C\phi(\mu^{-1}(|t - t_2|)) = r\bar{\kappa}_C\phi(-\mu^{-1}\gamma(t)).$$

(ii) $k = 2$ and $t = t_2$:

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

(iii) $r \geq 3$ and ($t < t_2$ or $t > t_r$) :

$$\begin{aligned}
(t < t_2): |\nabla_t \tilde{p}_i(t,\mu) - \eta_i(t)| &\leq r\bar{\kappa}_C \phi(\mu^{-1}(t - t_2)) &&= r\bar{\kappa}_C \phi(-\mu^{-1}\gamma(t)).\\
(t > t_r): |\nabla_t \tilde{p}_i(t,\mu) - \eta_i(t)| &\leq r\bar{\kappa}_C \left(1 - \phi(\mu^{-1}(t - t_r))\right) &&= r\bar{\kappa}_C \phi(\mu^{-1}(t_r - t))\\
& &&= r\bar{\kappa}_C \phi(-\mu^{-1}\gamma(t)).
\end{aligned}$$

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

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

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

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

The bound (34) follows. The properties stated for the function b follow from its definition.

Theorem 5 (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 4. 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 4, we need only establish (i) and (ii) in Definition 4 to show that $\tilde{f}$ is a measurable smoothing function for f. First note that the bound (33) in Lemma 7 shows that (7) in part (i) in Definition 4(i) is satisfied. The bound (8) is also satisfied since, by Lemma 6,

$$|\tilde{f}(\xi, x, \mu)| \leq |\tilde{f}(\xi, \bar{x}, \mu)| + \bar{\kappa}_{\mathsf{q}}[\hat{\kappa}_{\mathsf{c}}(\xi, \bar{x}) + 2\bar{\kappa}_C \hat{\kappa}_g(\xi, \bar{x})]$$

$\forall\, (\xi, x, \mu) \in \Xi \times \mathbb{B}_{\epsilon}(\bar{x}) \times (0, \bar{\mu}]$ (see Definition 2 and Lemma 4 for the definition of the terms in this bound). Hence Definition 7(i) is satisfied.

By (31), 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 (31), Definition 6 and (32) (or, more simply Lemma 6),

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

which establishes a bound stronger than (9) in Definition 7(ii) since it is independent of μ.

The final statement of the theorem follows from Lemma 3.

4.2 Gradient sub-consistency of CM integrands

We now examine conditions under which the smoothing (30) of CM integrands satisfy the gradient sub-consistency property (18). Our approach is to develop conditions under which Theorem 3 can be applied. The key condition in this regard is the uniform subgradient approximation property (17). 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 $\rho(\Xi(\nu, \bar{x})) \geq 1 - \nu$ satisfying, for a.e. $\xi \in \Xi(\nu, \bar{x})$,

$$\operatorname{dist}\left(\nabla_x \tilde{f}(\xi, x, \mu) \mid \partial_x f(\xi, \bar{x})\right) \leq \nu \;\forall\, (x, \mu) \in [(\bar{x}, 0) + \delta(\nu, \bar{x})(\mathbb{B} \times (0, 1))]. \tag{35}$$

To establish this condition, we use Theorem 4 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 $x \in \mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x})$,

$$\begin{aligned} \|\nabla \mathsf{q}(y) - \nabla \mathsf{q}(\bar{y})\| &\leq \bar{\kappa}_q \|y - \bar{y}\| \\ \|\nabla_x g(\xi, x) - \nabla_x g(\xi, \bar{x})\| &\leq k_g(\bar{x}) \|x - \bar{x}\|, \text{ and} \\ \|\nabla_x c(\xi, x) - \nabla_x c(\xi, \bar{x})\| &\leq k_c(\bar{x}) \|x - \bar{x}\| \quad \forall\, \xi \in \Xi,\ x \in \mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x}), \end{aligned} \tag{36}$$

where $\bar{\kappa}_q$ is the Lipschitz constant for $\nabla \mathsf{q}$ given in Definition 6. 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

$$\begin{aligned} &\|\nabla \mathsf{q}\| \leq \kappa_{\mathsf{q}}, \ \|\nabla_x c(\xi, x)\| \leq \kappa_c(\bar{x}) \ \text{ and} \\ &\|\nabla_x g(\xi, x)\| \leq \kappa_g(\bar{x}) \quad \forall\, (\xi, x) \in \Xi \times \mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x}). \end{aligned} \tag{37}$$

Lemma 8 *Let f be a CM integrand as given in Definition 7 and let $\tilde{f}$ be the smoothing function for f as in Theorem 4 for which (36) and (37) 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 a.e. $\xi \in \Xi$, then*

$$\begin{aligned} &\operatorname{dist}\left(\nabla_x \tilde{f}(\xi, x, \mu) \mid \partial_x f(\xi, \bar{x})\right) \\ &\quad \leq K_1(\bar{x})\|x - \bar{x}\| + K_2(\bar{x})\mu \\ &\qquad + K_3(\bar{x}) \max_{i=1,\dots,m} \operatorname{dist}\left(\nabla_t \tilde{p}_i(g_i(\xi, x), \mu) \mid \partial p_i(g_i(\xi, \bar{x}))\right) \end{aligned} \tag{38}$$

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 $z_i \in \partial p_i(g_i(\xi, \bar{x}))$, $i = 1, \dots, m$. Then, by Theorem 4, for a.e. $\xi \in \Xi$,

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

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 (33), the constants defined in (36) and (37), and the fact that $\|\cdot\|_2 \leq \sqrt{m}\,\|\cdot\|_\infty$ on $\mathbb{R}^m$, we have

$$\begin{aligned} \|\tilde{g} - g\| &\leq [(k_c(\bar{x}) + 2\sqrt{m}k_g(\bar{x})\bar{\kappa}_C)\,\|x - \bar{x}\| + \kappa_g(\bar{x})\,\|Y - Z\|]\kappa_q \\ &\quad + [\kappa_c(\bar{x}) + \sqrt{m}\kappa_g(\bar{x})\bar{\kappa}_C]\bar{\kappa}_q\left[\kappa_c(\bar{x})\,\|x - \bar{x}\| + \left\|\tilde{C}(g(\xi, x)) - C(g(\xi, \bar{x}))\right\|\right] \\ &\leq [(k_c(\bar{x}) + 2\sqrt{m}k_g(\bar{x})\bar{\kappa}_C)\,\|x - \bar{x}\| + \kappa_g(\bar{x})\,\|Y - Z\|]\kappa_q \\ &\quad + [\kappa_c(\bar{x}) + \sqrt{m}\kappa_g(\bar{x})\bar{\kappa}_C]\bar{\kappa}_q\,[\kappa_c(\bar{x})\,\|x - \bar{x}\| \\ &\quad + \bar{\kappa}_C(\kappa_g(\bar{x})\,\|x - \bar{x}\| + \sqrt{m}\omega\mu)] \\ &\leq [\kappa_q(k_c(\bar{x}) + 2\sqrt{m}k_g(\bar{x})\bar{\kappa}_C) + \bar{\kappa}_q(\kappa_c(\bar{x}) + \bar{\kappa}_C\kappa_g(\bar{x}))(\kappa_c(\bar{x}) \\ &\quad + \sqrt{m}\kappa_g(\bar{x})\bar{\kappa}_C)]\,\|x - \bar{x}\| \end{aligned}$$

$$+\sqrt{m}\omega\bar{\kappa}_q\bar{\kappa}_C(\kappa_c(\bar{x})+\sqrt{m}\kappa_g(\bar{x})\bar{\kappa}_C)\mu$$
$$+\sqrt{m}\kappa_q\kappa_g(\bar{x})\max_{i=1,\ldots,m}|\nabla_t\tilde{p}_i(g_i(\xi,x),\mu)-z_i|, \tag{39}$$

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 a.e. $\xi\in\Xi$, which proves the lemma.

Lemma 8 shows that if we can obtain a bound on the distances to the subdifferentials $\partial p_i(g_i(\xi,\bar{x}))$ similar to the bound in (35), then we can choose $\hat{\delta}(\bar{x})$ and μ small enough to ensure that (35) also holds.

Lemma 9 *Let f and $\tilde{f}$ satisfy the hypotheses of Lemma 8. Set*

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

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

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

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

$$\textit{for any } \tau\in(0,\bar{\tau}),\;\exists\,\tilde{\varepsilon}(\tau,\bar{x})\in(0,\bar{\epsilon}),\quad \text{s.t. } \rho(\bar{\Xi}_\varepsilon(\bar{x}))\leq\tau,\;\forall\,\varepsilon\in(0,\tilde{\varepsilon}(\tau,\bar{x})). \tag{40}$$

If $\bar{x}\in U$ is such that (40) holds, then, for all $i\in\{1,\ldots,m\}$, $\tau\in(0,\bar{\tau})$ and $\varepsilon\in(0,\tilde{\varepsilon}(\tau,\bar{x}))$,

$$\operatorname{dist}\left(\nabla_t\tilde{p}_i(g_i(\xi,x),\mu)\;|\;\partial p_i(g_i(\xi,\bar{x}))\right)\leq\tau \tag{41}$$

for all $\xi\in\bar{\Xi}^c_\varepsilon(\bar{x}):=\Xi\backslash\bar{\Xi}_\varepsilon(\bar{x})$ 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

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

with $\bar{\delta}(\bar{x})$ and $\kappa_g(\bar{x})$ as given in Lemma 8 and (37), 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 8 and it's proof. Observe that, for every $\varepsilon \in (0, \bar{\varepsilon}]$ and $x \in U$,

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

and note that these sets are measurable.

By (40), for any $\tau \in (0, \bar{\tau})$, we have $\rho(\bar{\Xi}_\varepsilon(\bar{x})) \leq \tau$ 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 $i = 1, \cdots, m$, $\|g_i(\xi, x) - g_i(\xi, \bar{x})\| \leq \frac{\varepsilon}{2}$.

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

$$|\phi(\mu^{-1}(t - t_1)) - \phi(\mu^{-1}(t - t_2))| \leq \phi(\mu^{-1}(t - t_2)) = \phi(-\mu^{-1}|t - t_2|).$$

(b) If $t > t_1 > t_2$, then

$$|\phi(\mu^{-1}(t - t_1)) - \phi(\mu^{-1}(t - t_2))| \leq 1 - \phi(\mu^{-1}(t - t_1)) = \phi(-\mu^{-1}|t - t_1|).$$

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 $\nabla p_i(g_i(\xi, \bar{x})) = \nabla p_i(g_i(\xi, x)) = \eta_i(g_i(\xi, x))$. By Lemma 7, we have

$$|\nabla_t \tilde{p}_i(g_i(\xi, x), \mu) - \nabla p_i(g_i(\xi, \bar{x}))| \leq (r_i + 1)\bar{\kappa}_C \phi\left(\frac{-1}{\mu}\gamma_i(g_i(\xi, x))\right) \leq (\bar{r} + 1)\bar{\kappa}_C \phi\left(\frac{-\varepsilon}{2\mu}\right).$$

Since $\tau/(\bar{\kappa}_C(\bar{r} + 1)) \leq 1/2$ (so that $\phi^{-1}_{\min}(\frac{\tau}{(\bar{r}+1)\bar{\kappa}_C}) < 0$ by (24)), we have the inequality $|\nabla_t \tilde{p}_i(g_i(\xi, x), \mu) - \nabla p_i(g_i(\xi, \bar{x}))| \leq \tau$ whenever $0 < \mu \leq \tilde{\mu}(\epsilon, \tau, \bar{x})$. 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

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

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

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

and so

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

If $r_i = 2$, then (26) tells us that $\nabla_t \tilde{p}_i(g_i(\xi, x), \mu) = \tilde{\eta}_i(g_i(\xi, \bar{x}), \mu)$, so that

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

If $r_i \geq 3$ and $2 \neq \bar{j} \neq r_i$, the expression in (26) for $\nabla_t \tilde{p}_i(g_i(\xi, x), \mu)$ tells us that

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

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

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

Hence, we always have

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

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

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

The result follows by combining these two cases.

Remark 5 One can strengthen the hypothesis (40) to

$$\exists\, \tau > 0 \text{ s.t. } \forall\, \tau \in (0, \bar{\tau})\ \exists\, \tilde{\varepsilon}(\tau, \bar{x}) \in (0, \bar{\epsilon}) \text{ s.t. } \rho(\widehat{\varXi}_\varepsilon(\bar{x})) \leq \tau \quad \forall\, \varepsilon \in (0, \tilde{\varepsilon}(\tau, \bar{x})). \tag{42}$$

Then Lemma 9 still holds. However, under (42), we have that $\rho(\widetilde{\varXi}(\bar{x})) = 1$, where

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

is defined in Proposition 1. Consequently, (42) implies that $f(\xi, \cdot)$ is continuously differentiable and subdifferentially regular at $\bar{x}$ for a.e. $\xi \in \Xi$.

We now combine Lemmas 8 and 9 to establish conditions under which (35) is satisfied.

Theorem 6 (Uniform subgradient approximation) *Let f be a CM integrand as given in Definition 7 and let $\tilde{f}$ be the smoothing function for f given in Theorem 4. Suppose that the basic assumptions of Lemmas 8 and 9 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 $\rho(\Xi(\nu, \bar{x})) \le 1 - \nu$ for which (35) 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 $\bar{\delta}(\bar{x})$, $K_1(\bar{x})$, $K_2(\bar{x})$ and $K_3(\bar{x})$ be as given by Lemma 8 so that (38) is satisfied for all $x \in \mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x})$ and $\mu > 0$. Set

$$\begin{aligned} &\delta_1(\nu, \bar{x}) := \min\{\bar{\delta}(\bar{x}), \nu/(3K_1(\bar{x}))\}, \quad \mu_1(\nu, \bar{x}) := \nu/(3K_2(\bar{x})), \quad \text{and} \\ &\qquad \tau := \nu/(3K_3(\bar{x}) + 1). \end{aligned}$$

Let $\tilde{\varepsilon}(\tau, \bar{x})$ be as in (40). 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 $\rho(\Xi(\nu, \bar{x})) \ge 1 - \tau \ge 1 - \nu$ 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 (41). Then, by (38) and the definitions given above,

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

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

We can now apply Theorem 3 to obtain the gradient sub-consistency of smoothed CM integrands.

Theorem 7 (Gradient sub-consistency of smoothed CM integrands) *Let f be a CM integrand as given in Definition 7 and let $\tilde{f}$ be the smoothing function for f given in Theorem 4. Suppose that the basic assumptions of Lemmas 8 and 9 are satisfied so that their conclusions hold. We further assume that $f(\xi, \cdot)$ is subdifferentially regular $\bar{x}$ for a.e. $\xi \in \Xi$ or $-f(\xi, \cdot)$ is subdifferentially regular at $\bar{x}$ for a.e. $\xi \in \Xi$. Then $\widetilde{F}(x, \mu) := \mathbb{E}[\tilde{f}(\xi, x, \mu)]$ satisfies the gradient sub-consistency property (18) at $\bar{x}$, i.e.,*

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

Proof The result follows once it is shown that the hypotheses of Theorem 3 are satisfied. Theorem 6 tells us that the uniform subgradient approximation property is satisfied. Lemma 4 shows that every CM integrand is an LL integrand, so, the subdifferential regularity requirement implies that the hypotheses of Corollary 1 are satisfied. Hence, all hypotheses of Theorem 3 are satisfied at $\bar{x}$.

4.3 Subgradient approximation via smoothing without regulartity

Theorem 7 uses the subdifferential regularity of $f(\xi, x)$ or $-f(\xi, x)$ for a.e. 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 (30). We show that if $(\xi, \bar{x}) \in \Xi \times X$ is such that $\operatorname{rank} \nabla_x g(\xi, \bar{x}) = m$, 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

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

where the functions η_i are defined in (27). Note that

$$
\langle \mathsf{z}_i(\xi, \bar{x}), v \rangle = \frac{1}{2}(\mathsf{h}_i^1(\xi, \bar{x})(v) + \mathsf{h}_i^2(\xi, \bar{x})(v)), \tag{43}
$$

and, by Lemma 5,

$$
\mathsf{z}(\xi, \bar{x}) = \operatorname{diag}(\eta_i(g_i(\xi, \bar{x}))) \nabla_x g(\xi, x). \tag{44}
$$

Lemma 10 *Consider the CM integrand f and its smoothing function $\tilde{f}$ defined in* (30)*. Assume that* $\operatorname{rank} \nabla_x g(\xi, \bar{x}) = m$ *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 $v \in \mathbb{R}^n$,*

$$
\begin{aligned}
u(\xi, \bar{x}) &:= \lim_{\mu \downarrow 0} \nabla_x \tilde{f}(\xi, \bar{x}, \mu) \\
&= (\nabla_x c(\xi, \bar{x}) + (\mathsf{z}_1(\xi, \bar{x}), \ldots, \mathsf{z}_m(\xi, \bar{x}))^T \nabla q(c(\xi, \bar{x}) + C(g(\xi, \bar{x})))
\end{aligned}
\tag{45}
$$

$$\begin{aligned}\ell_1(\xi,\bar{x};v) &:= \lim_{t\downarrow 0}\frac{f(\xi,\bar{x}+2tv)-f(\xi,\bar{x}+tv)}{t}\\ &= \nabla q(c(\xi,\bar{x})+C(g(\xi,\bar{x})))^T(\nabla_x c(\xi,\bar{x})v)\\ &\quad+(h_1^1(\xi,\bar{x})(v),\cdots,h_m^1(\xi,\bar{x})(v))^T)\end{aligned}\tag{46}$$

and

$$\begin{aligned}\ell_2(\xi,\bar{x};v) &:= \lim_{t\downarrow 0}\frac{f(\xi,\bar{x}-tv)-f(\xi,\bar{x}-2tv)}{t}\\ &= \nabla q(c(\xi,\bar{x})+C(g(\xi,\bar{x})))^T(\nabla_x c(\xi,\bar{x})v)\\ &\quad+(h_1^2(\xi,\bar{x})(v),\cdots,h_m^2(\xi,\bar{x})(v))^T).\end{aligned}\tag{47}$$

Moreover, by (43)*, we have*

$$\langle u(\xi,\bar{x}),v\rangle=\frac{1}{2}(\ell_1(\xi,\bar{x};v)+\ell_2(\xi,\bar{x};v))\quad\forall\, v\in\mathbb{R}^n.\tag{48}$$

Proof By combining (31) and (27), we find that the right hand side of (45) is an element of $\partial_x f(\xi,\bar{x})$. Moreover, by (27) in Lemma 5 and (31) in Theorem 4, the limit $u(\xi,\bar{x})$ exists as given in (45). Since (48) follows from (46) and (47), 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],

$$\partial_x h_i(\xi,\bar{x})=\partial_x(p_i\circ g_i)(\xi,\bar{x}))=\partial p_i(g_i(\xi,\bar{x}))\nabla_x g_i(\xi,\bar{x})\tag{49}$$

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 (20), (44) and (49), 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 (27), $\mathsf{z}_i(\xi,x)$ 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 (20) and Lemma 2, for any $x\in X$ and direction $v\in\mathbb{R}^n$, we have

$$\begin{aligned}\mathsf{h}_i^1(\xi,\bar{x})(v)&=\lim_{t\downarrow 0}\frac{h_i(\xi,x+2tv)-h_i(\xi,x+tv)}{t}=\max_{\nu\in\partial_x h_i(\xi,x)}\langle\nu,v\rangle\quad\text{and}\\ \mathsf{h}_i^2(\xi,\bar{x})(v)&=\lim_{t\downarrow 0}\frac{h_i(\xi,x-tv)-h_i(\xi,x-2tv)}{t}=\min_{\nu\in\partial_x h_i(\xi,x)}\langle\nu,v\rangle.\end{aligned}$$

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

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

and similarly,

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

This establishes (46) and (47) which combined imply (48).

Theorem 8 (Subgradient approximation by smoothing) *Consider the CM integrand f and its smoothing function $\tilde{f}$ defined in (30), and suppose the hypotheses of Lemma 8 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 a.e. $\xi \in \Xi$,*

$$\begin{aligned}&\operatorname{dist}\left(\nabla_x \tilde{f}(\xi, \bar{x}, \mu) \mid \partial_x f(\xi, \bar{x})\right) \\ &\quad \le K_2(\bar{x})\mu + K_3(\bar{x})(\bar{r}+1)\bar{\kappa}_C \max_{i=1,\dots,m} \phi\left(\frac{-1}{\mu}\gamma_i(g_i(\xi, \bar{x}))\right).\end{aligned} \tag{50}$$

Moreover, $\widetilde{F}(\cdot, \mu)$ is differentiable at $\bar{x}$ for all $\mu > 0$ with $\nabla_x \widetilde{F}(\bar{x}, \mu)$ equal $= \mathbb{E}[\nabla_x \tilde{f}(\xi, \bar{x}, \mu)]$, the function u in (45) 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 (34) and (38), we have (50).

By Lemma 4, f is an LL integrand. By Lemma 10, the function u in (45) is well defined a.e. in Ξ and measurable. By Theorem 5, $\widetilde{F}(\cdot, \mu)$ 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, (45) 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 [17, 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 10 and [17, Proposition 2.7], $u(\xi, \bar{x}) \in$

$\partial_x f(\xi, \bar{x})$ is measurable. Moreover, by (48) in Lemma 10,

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

with both limits existing and measurable. Consequently,

$$\mathbb{E}[\ell_1(\xi, \bar{x}; v)] - \mathbb{E}[\langle u(\xi, \bar{x}), v\rangle] = -(\mathbb{E}[\ell_2(\xi, \bar{x}; v)] - \mathbb{E}[\langle u(\xi, \bar{x}), v\rangle]). \tag{51}$$

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,

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

which tells us that

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

and

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

Hence, by (51),

$$\begin{aligned}&\widehat{d}F(\bar{x})(v) - \langle \mathbb{E}[u(\xi, \bar{x})],\, v\rangle \ge \mathbb{E}[\ell_1(\xi, \bar{x}; v)] - \langle \mathbb{E}[u(\xi, \bar{x})],\, v\rangle \\ &\qquad\qquad = -(\mathbb{E}[\ell_2(\xi, \bar{x}; v)] - \langle \mathbb{E}[u(\xi, \bar{x})],\, v\rangle) \\ \text{and } &\widehat{d}F(\bar{x})(v) - \langle \mathbb{E}[u(\xi, \bar{x})],\, v\rangle \ge \mathbb{E}[\ell_2(\xi, \bar{x}; v)] - \langle \mathbb{E}[u(\xi, \bar{x})],\, v\rangle, \quad \text{and so} \\ &\widehat{d}F(\bar{x})(v) - \langle \mathbb{E}[u(\xi, \bar{x})],\, v\rangle \ge |\, \mathbb{E}[\ell_2(\xi, \bar{x}; v)] - \langle \mathbb{E}[u(\xi, \bar{x})],\, v\rangle \,| \ge 0.\end{aligned}$$

Since v was chosen arbitrarily, Appendix Definition 9 tells us that $\mathbb{E}[u(\xi, \bar{x})] \in \partial F(\bar{x})$.

Corollary 2 *Let the assumptions of Theorem 8 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*

$$\left\|\nabla \tilde{f}(\xi, x, \mu) - \nabla \tilde{f}(\xi, \bar{x}, \mu)\right\| \le \frac{K_4(\bar{x})}{\mu}\|x - \bar{x}\| \quad \forall\, \xi \in \Xi \text{ and } x \in \mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x}) \tag{52}$$

and

$$\operatorname{dist}\left(\mathbb{E}[\nabla \tilde{f}(\xi, x, \mu)] \mid \partial F(\bar{x})\right) \le \frac{K_4(\bar{x})}{\mu}\|x - \bar{x}\| + \operatorname{dist}\left(\nabla_x \widetilde{F}(\bar{x}, \mu) \mid \partial F(\bar{x})\right) \ \forall\, x \in \mathcal{B}_{\bar{\delta}(\bar{x})}(\bar{x}). \tag{53}$$

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

$$\operatorname*{Limsup}_{x \to \bar{x},\, \mu = O(\|x - \bar{x}\|^{\gamma})} \mathbb{E}[\nabla \tilde{f}(\xi, x, \mu)] \in \partial F(\bar{x}). \tag{54}$$

Proof The proof of (52) follows the pattern of proof given for (39) to first establish that

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

Then use (29) to obtain the bound (52).

To see (53), note that

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

Hence, (54) follows from Theorem 8.

One of stopping criteria in smoothing algorithms is to require that the smoothing gradient $\nabla \widetilde{F}(x^k, \mu)$ is sufficiently small. However, in keeping with our program, we prefer a stopping criteria based on the integrand. Such a criteria is provided by Theorem 8 where it is shown that the expectation $\mathbb{E}[u(\xi, \bar{x})]$, with u is defined in (45), provides an arguably better estimate of proximity to stationarity in the CM function setting. This expectation is computable and satisfies $u(\xi, \bar{x}) \in \partial_x f(\xi, \bar{x})$ a.e. ξ.

5 Conclusion

In this paper, 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 a.e. $\xi \in \Xi$ (Theorems 6, 7). Moreover, when this subdifferential regularity hypothesis fails, we show that for any $x \in \mathbb{R}^n$,

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

Consequently, we can approximate an element of the Clarke subdifferential of the expectation function using gradients of a smoothing function for the non-smooth integrand (Theorem 8 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 comes from stochastic nonlinear complementarity problems with a continuously differentiable function $\varphi : \Xi \times R^n \to R^n$ [11,13], 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]. Other interesting application might be stochastic programs with the P-matrix linear complementarity constraints. The P-matrix linear complementarity constraints can be rewritten as piecewise linear constraints [10,22,32] and approximated by continuously differentiable constraint functions using a smoothing approximation. 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.

Acknowledgements We would like to thank Associate Editor and two referees for their helpful comments.

6 Appendix: Background

6.1 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: $\operatorname{epi} f := \{(x, \mu) \mid f(x) \le \mu\}$. 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 8 (*Subderivatives*) [27, 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 9 (*Subgradients, subdifferentials and subdifferential regularity*) Consider a locally Lipschitz 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) [27, Theorem 8.49] The vector v is a *Clarke subgradient* of f at u_* if 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) [27, Corollary 8.19] $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be *subdifferentially regular* (or Clarke regular) at $u_* \in \operatorname{dom} f$ with $\partial f(u_*) \neq \emptyset$ if

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

(iii) [14, Definition 2.6.1] [12] The vector v is a *B-subgradient* of f at u_* if

$$v = \lim_{u^k \to u_*} \nabla f(u^k), \quad \text{where } f \text{ is differentiable at } u^k.$$

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

(iv) [27, Definition 8.3] The vector v is an *M-subgradient* of f at u_* if there are sequences $u^k \to u_*$ and $v^k \to v$ with

$$\liminf_{u \to u^k} \frac{f(u) - f(u^k) - \left\langle v^k\,,\, u - u^k\right\rangle}{\|u - u^k\|} \geq 0.$$

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

Remark 6 In [27], the notion of subdifferential regularity is defined in [27, 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 10 (*Strict continuity and strict differentiability*) Let $H : D \to \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, and $h : \mathbb{R}^n \to \overline{\mathbb{R}}$.

(i) [Strict Continuity [27, Definition 9.1]] We say that H is strictly continuous at $\bar{x} \in \operatorname{int}(D)$ if

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

(ii) [Strict Differentiability [27, Definition 9.17]] We say that h is strictly differentiable at a point $\bar{x} \in \operatorname{dom} h$ if h is differentiable at $\bar{x}$ and

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

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 $\partial h(x) = \{\nabla h(x)\}$ for all $x \in U$ ([27, 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 8 and Definition 9(ii) combine to tell us that

$$df(u_*)(w) = \widehat{d} f(u_*)(w) = \lim_{\tau\downarrow 0} \frac{f(u_* + \tau w) - f(u_*)}{\tau} \quad \forall\, w \in \mathbb{R}^n, \tag{55}$$

wherever f is strictly continuous and subdifferentially regular at u_*. Moreover, in this case, [27, Theorem 8.30] tells us that

$$df(x)(v) = \sup\{\langle g \, , \, v\rangle \mid g \in \partial f(x)\} . \tag{56}$$

Remark 7 (Subdifferentials of Compositions) If $g : X \subset \mathbb{R}^n \to \overline{\mathbb{R}}$ is given as the composition of two functions $f : Y \subset \mathbb{R}^m \to \overline{\mathbb{R}}$ 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 9 (Strict differentiability and the subdifferential) [27, Theorem 9.18] [14, Proposition 2.2.4] *Let $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ with $\bar{x} \in \operatorname{dom} h$. Then h is strictly differentiable at $\bar{x}$ if and only if h is strictly continuous at $\bar{x}$ and $\partial h(\bar{x}) = \{\nabla h(\bar{x})\}$.*

6.2 Measurable multi-functions

We now review some of the properties of measurable multi-functions used in this paper [2,15,18,27]. For more information on this topic, we refer the interested reader to [27, Chapter 14] and [25].

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 $S : \mathbb{R}^k \rightrightarrows \mathbb{R}^s$. The *outer limit* of S at $\bar{x} \in \mathbb{R}^k$ *relative to* $X \subseteq \mathbb{R}^k$ is

$$\operatorname*{Limsup}_{x \to_X \bar{x}} S(x) := \left\{ v \in \mathbb{R}^s \mid \exists \{x^k\} \to_X \bar{x}, \{v^k\} \to v \in \mathbb{R}^s : v^k \in S(x^k) \quad \forall k \in \mathbb{N} \right\} \tag{57}$$

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

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

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 Ξ be a nonempty subset of $\mathbb{R}^\ell$ and let $\mathcal{A}$ be a σ-field of subsets of Ξ, called the *measurable* subsets of Ξ 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 $(\Xi, \mathcal{A}, \rho)$. A multi-function $\Psi : \Xi \rightrightarrows \mathbb{R}^n$ 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 $\Xi \times \mathbb{R}^n$ 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 Ξ obtained as the projections of measurable subsets $\{G\}$ of $\Xi \times \mathbb{R}^n$:

$$\{G\} \in \mathcal{A} \otimes \mathcal{B}^n \quad \Longrightarrow \quad \left\{ \xi \in \Xi \;\middle|\; \exists\, v \in \mathbb{R}^n \text{ with } (\xi, v) \in \{G\} \right\} \in \mathcal{A}.$$

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 [27, Theorem 14.8].

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

Theorem 10 (Measurable selections) [27, Corollary 14.6] *A closed-valued measurable map $\Psi : \Xi \rightrightarrows \mathbb{R}^n$ always admits a measurable selection.*

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

$$\|v\|_\infty \leq a(\xi) \tag{58}$$

for all pairs $(\xi, v) \in \Xi \times \mathbb{R}^n$ satisfying $v \in \Psi(\xi)$. Here and elsewhere we interpret vector inequalities as element-wise inequalities. Let $1 \leq p \leq \infty$. When $\Xi = \mathbb{R}^\ell$, 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

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

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

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

If the function a in (58) is such that $\|a(\xi)\|_p$ is integrable with respect to the measure ρ on the measure space $(\Xi, \mathcal{A}, \rho)$, then the multi-function Ψ is said to be L^p-bounded, where $\|\cdot\|_p$ denotes the p-norm of vectors.

Proposition 2 [7, Proposition 2.2] and [16, Corollary IV.8.4](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 $L^2_m(\mathbb{R}^\ell, \mathcal{M}^n, \lambda_n)$.*

We now develop some properties of integrals of multi-valued mappings. Given a measurable multi-function $\Psi : \Xi \rightrightarrows \mathbb{R}^n$, we define the integral of Ψ over Ξ with respect to the measure ρ by

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

The next theorem, due to Hildenbrand [18], is a restatement of Theorems 3 and 4 of Aumann [2] for multi-functions on the non-atomic measure space $(\Xi, \mathcal{A}, \rho)$. These results are central to the theory of integrals of multi-valued functions.

Theorem 11 (Integrals of multi-functions) [18, Theorem 4 and Proposition 7] *The following properties hold for integrably bounded multi-functions $\Psi : \Xi \rightrightarrows \mathbb{R}^n$ on non-atomic measure spaces $(\Xi, \mathcal{A}, \rho)$.*

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

(b) *If Ψ is closed valued (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 $\Psi : \Xi \rightrightarrows \mathbb{R}^n$ of the form

$$\Psi(\xi) := \kappa(\xi)\mathbb{B}, \tag{59}$$

where $\mathbb{B} := \{x \mid \|x\|_2 \le 1\}$ is the closed unit ball in $\mathbb{R}^n$ and $\kappa : \Xi \to \mathbb{R}_+$ is measurable.

Lemma 11 (Tubes) *Let $\Psi : \Xi \rightrightarrows \mathbb{R}^n$ be a measurable tube as in (59) with $\kappa \in L_1^2(\Xi, \mathcal{A}, \rho)$ non-negative a.e. on Ξ. 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 (59) is obviously closed valued and measurable. Therefore, Theorem 10 tells us that $\mathcal{S}(\Psi)$ is non-empty. Let $E \in \mathcal{A}$ and $s \in \mathcal{S}(\Psi)$. Then

$$\left|\int_E s(\xi) d\rho\right| \le \int_E |s(\xi)| d\rho \le \int_E \kappa(\xi) d\rho,$$

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

References

1. Ahn, M., Pang, J.S., Xin, J.: Difference-of-convex learning I: directional stationarity, optimality, and sparsity. SIAM J. Optim. **27**, 1637–1665 (2017)
2. Aumann, R.J.: Integrals of set-valued functions. J. Math. Anal. Appl. **12**, 1–12 (1965)
3. Blundell, R., Powell, J.L.: Censored regression quantiles with endogenous regression. J. Econom. **141**, 65–83 (2007)
4. Burke, J.V.: Second order necessary and sufficient conditions for convex composite NDO. Math. Program. **38**, 287–302 (1987)
5. Burke, J.V., Hoheisel, T.: Epi-convergent smoothing with applications to convex composite functions. SIAM J. Optim. **23**, 1457–1479 (2013)
6. Burke, J.V., Hoheisel, T., Kanzow, C.: Gradient consistency for integral-convolution smoothing functions. Set-Valued Var. Anal. **21**, 359–376 (2013)
7. Burke, J.V., Luke, D.R.: Variational analysis applied to the problem of optimal phase retieval. SIAM J. Control Optim. **42**, 576–595 (2003)
8. Chen, C., Mangasarian, O.L.: A class of smoothing functions for nonlinear and mixed complementarity problems. Comput. Optim. Appl. **5**, 97–138 (1996)
9. Chen, X.: Smoothing methods for nonsmooth, nonconvex minimization. Math. Program. **134**, 71–99 (2012)
10. Chen, X., Fukushima, M.: A smoothing method for a mathematical program with P-matrix linear complementarity constraints. Comput. Optim. Appl. **27**, 223–246 (2004)
11. Chen, X., Fukushima, M.: Expected residual minimization method for stochastic linear complementarity problems. Math. Oper. Res. **30**, 1022–1038 (2005)
12. Chen, X., Qi, L., Sun, D.: Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities. Math. Comput. **67**, 519–540 (1998)
13. Chen, X., Wets, R., Zhang, Y.: Stochastic variational inequalities: residual minimization smoothing sample average approximations. SIAM J. Optim. **22**, 649–673 (2012)
14. Clarke, F.H.: Optimization and Nonsmooth Analysis, Volume 5 of Classics in Applied Mathematics. SIAM, Philadelphia (1990)

15. Clarke, F.H., Ledyaev, Y.S., Stern, R.J., Wolenski, P.R.: Nonsmooth Analysis and Control Theory. Springer, New York (1998)
16. Dunford, N., Schwartz, J.T.: Linear Operators, Part1: General Theory. Wiley, Hoboken (1988)
17. Folland, G.B.: Real Analysis, 2nd edn. Wiley, New York (1999)
18. Hildenbrand, W.: Core and Equilibria of a Large Economy. Princeton University Press, Princeton (1974)
19. Lyapunov, A.: Sur les fonctions-vecteur complétement additives. Bull. Acad. Sci. USSR Ser. Math. **4**, 465–478 (1940)
20. Mordukhvich, B.: Variational Analysis and Generalized Differentiation II. Springer, Berlin (2006)
21. Nemirovski, A., Juditsky, A., Lan, G., Shapiro, A.: Robust stochastic approximation approach to stochastic programming. SIAM J. Optim. **19**, 1574–1906 (2009)
22. Ralph, D., Xu, H.: Convergence of stationary points of sample average two stage stochastic programs: a generalized equation approach. Math. Oper. Res. **36**, 568–592 (2011)
23. Richter, H.: Verallgemeinerung eines in der statistik benötigten satzes der masstheorie. Math. Ann. **150**, 85–90 (1963)
24. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk **2**, 21–41 (2000)
25. Rockafellar, R.T.: Integral functionals, normal integrands and measurable selections. In: Nonlinear Operators in the Calculus of Variations, Volume 543 in Lecture Notes in Mathematics, pp. 157–207. Springer, New York (1976)
26. Rockafellar, R.T., Wets, R.: On the interchange of subdifferentiation and conditional expectation for convex functions. Stochastics **7**, 173–182 (1982)
27. Rockafellar, R.T., Wets, R.: Variational Analysis. Springer, New York (1998)
28. Rudin, W.: Principles of Mathematical Analysis, 3rd edn. McGraw-Hill, San Francisco (1976)
29. Shapiro, A., Dentcheva, D., Ruszczynski, A.: Lectures on Stochastic Programming: Modeling and Theory. SIAM, Philadelphia (2009)
30. Tardella, F.: A new proof of the Lyapunov convexity theorem. SIAM J. Control Optim. **28**, 478–481 (1990)
31. Wets, R.: Stochastic programming. In: Nemhauser, G.L., et al. (eds.) Handbooks in OR & MS, vol. 1, pp. 573–629 (1989)
32. Xu, H.: An implicit programming approach for a class of stochastic mathematical programs with complementarity constraints. SIAM J. Optim. **16**, 670–696 (2006)
33. Xu, H., Ye, J.J.: Necessary optimality conditions for two-stage stochastic programs with equilibaium constraints. SIAM J. Optim. **20**, 1685–1715 (2010)
34. Xu, H., Zhang, D.: Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications. Math. Program. **119**, 371–401 (2009)