Subdifferentiation and Smoothing of Nonsmooth Integral Functionals

James V Burke
Mathematics, University of Washington

Joint work with
Xiaojun Chen (Hong Kong Polytechnic)
and
Hailin Sun (Nanjing University)

SIAM Conference on Optimization
Vancouver, British Columbia

May 23, 2017
The Problem

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

\(X \subseteq \mathbb{R}^n\)  convex compact set with non-empty interior

\(\Xi \subseteq \mathbb{R}^\ell\)  Lebesgue measurable closed set with non-empty interior

\(f : \Xi \times X \to \bar{\mathbb{R}}\)  continuous in \(x \forall \xi \in \Xi\), measurable in \(\xi \forall x \in X\)

\(\mathbb{E}[\cdot]\)  expectation over \(\Xi\)
The Problem

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

$X \subseteq \mathbb{R}^n$ convex compact set with non-empty interior

$\Xi \subseteq \mathbb{R}^\ell$ Lebesgue measurable closed set with non-empty interior

$f : \Xi \times X \to \bar{\mathbb{R}}$ continuous in $x \forall \xi \in \Xi$, measurable in $\xi \forall x \in X$

$\mathbb{E}[\cdot]$ expectation over $\Xi$

Applications of interest include stochastic nonlinear complementarity problems, stochastic gap functions, and optimization problems in statistical learning, where $f(\xi, x)$ is often not Clarke regular in $x$ for almost all $\xi$. 
Clarke Stationary Points

\[ 0 \in \partial F(x) + \mathcal{N}_X(x), \quad \text{where} \quad F(x) = \mathbb{E}[f(\xi, x)] \]
Clarke Stationary Points

\[ 0 \in \partial F(x) + \mathcal{N}_x(x), \quad \text{where} \quad F(x) = \mathbb{E}[f(\xi, x)] \]

Issues:
- Can we estimate \( \partial F(x) \) using \( f(\xi, x) \)?
Clarke Stationary Points

\[ 0 \in \partial F(x) + \mathcal{N}_x(x), \quad \text{where} \quad F(x) = \mathbb{E}[f(\xi, x)] \]

Issues:

- Can we estimate \( \partial F(x) \) using \( f(\xi, x) \)?
- (Clarke '83) If \( f(\xi, x) \) is Clarke regular a.e. \( \xi \), then
  \[ \partial F(x) = \mathbb{E}[\partial f(\xi, x)]. \]
Clarke Stationary Points

\[ 0 \in \partial F(x) + N_X(x), \quad \text{where} \quad F(x) = \mathbb{E}[f(\xi, x)] \]

Issues:

- Can we estimate \( \partial F(x) \) using \( f(\xi, x) \)?
- (Clarke '83) If \( f(\xi, x) \) is Clarke regular a.e. \( \xi \), then
  \[ \partial F(x) = \mathbb{E}[\partial f(\xi, x)]. \]
- In general, we only have
  \[ \partial F(x) \subseteq \mathbb{E}[\partial_x f(\xi, x)]. \]
Clarke Stationary Points

\[ 0 \in \partial F(x) + \mathcal{N}_x(x), \quad \text{where} \quad F(x) = \mathbb{E}[f(\xi, x)] \]

Issues:

- Can we estimate \( \partial F(x) \) using \( f(\xi, x) \)?
- (Clarke '83) If \( f(\xi, x) \) is Clarke regular a.e. \( \xi \), then \( \partial F(x) = \mathbb{E}[\partial f(\xi, x)] \).
- In general, we only have \( \partial F(x) \subseteq \mathbb{E}[\partial_x f(\xi, x)] \).

Example: \( f(\xi, x) = \xi|x| \) with \( \xi \sim N(0, 1) \). Then

\[ \mathbb{E}[f(\xi, x)] = \mathbb{E}[\xi|x|] \equiv 0 \implies \partial F(x) = 0 \quad \forall \ x \in \mathbb{R} \]

but

\[ \mathbb{E}[\partial f(\xi, 0)] = \sqrt{\pi}/2 \ [-1, 1]. \]
We say \( f : \Xi \times X \to \mathbb{R} \) is a Carathéodory mapping on \( \Xi \times X \) if \( f(\xi, \cdot) \) is continuous on an open set \( U \) containing \( X \) for all \( \xi \in \Xi \), and \( f(\cdot, x) \) is measurable on \( \Xi \) for all \( x \in X \).
We say $f : \Xi \times X \to \mathbb{R}$ is a Carathéodory mapping on $\Xi \times X$ if $f(\xi, \cdot)$ is continuous on an open set $U$ containing $X$ for all $\xi \in \Xi$, and $f(\cdot, x)$ is measurable on $\Xi$ for all $x \in X$.

We say that $f : \Xi \times U \to \mathbb{R}$ is a locally Lipschitz integrand on $\Xi \times U$ if $f$ is a Carathéodory mapping on $\Xi \times U$ and $
abla \bar{x} \in U \exists \epsilon(\bar{x}) > 0$ and an integrable mapping $\kappa_f(\cdot, \bar{x}) \in L^2_{1}(\mathbb{R}^\ell, \mathcal{M}, \rho)$ such that

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

where $B_\epsilon(\bar{x}) := \{x \mid \|x - \bar{x}\| \leq \epsilon\} \subseteq U$. 

Basic Assumptions (Clarke ’83)
Basic Assumptions (Clarke ’83)

We say $f : \Xi \times X \to \mathbb{R}$ is a Carathéodory mapping on $\Xi \times X$ if $f(\xi, \cdot)$ is continuous on an open set $U$ containing $X$ for all $\xi \in \Xi$, and $f(\cdot, x)$ is measurable on $\Xi$ for all $x \in X$.

We say that $f : \Xi \times U \to \mathbb{R}$ is a locally Lipschitz integrand on $\Xi \times U$ if $f$ is a Carathéodory mapping on $\Xi \times U$ and
\( \forall \, \bar{x} \in U \ \exists \, \epsilon(\bar{x}) > 0 \) and an integrable mapping $\kappa_f(\cdot, \bar{x}) \in L^2(\mathbb{R}^\ell, \mathcal{M}, \rho)$ such that
\[
|f(\xi, x_1) - f(\xi, x_2)| \leq \kappa_f(\xi, \bar{x}) \|x_1 - x_2\| \quad \forall \, x_1, x_2 \in B_\epsilon(\bar{x}) \text{ a.e. } \xi \in \Xi,
\]
where $B_\epsilon(\bar{x}) := \{ x \mid \|x - \bar{x}\| \leq \epsilon \} \subseteq U$.

If $f : \Xi \times U \to \mathbb{R}$ is a LL integrand then $F(x) := \mathbb{E}[f(\xi, x)]$ is locally Lipschitz continuous on $U$ with local Lipschitz modulus $\kappa_F(\bar{x}) := \mathbb{E}[\kappa_f(\xi, \bar{x})]$. 
Approximation by Smoothing Functions

Let $F : U \to \mathbb{R}$, where $U \subseteq \mathbb{R}^n$ is open. We say that

$$
\tilde{F} : U \times \mathbb{R}_{++} \to \mathbb{R}
$$

is a smoothing function for $F$ on $U$ if

(i) $\tilde{F}(\cdot, \mu)$ converges continuously to $F$ on $U$, i.e.,

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

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

\( \tilde{f} : \Xi \times U \times \mathbb{R}_{++} \to \mathbb{R} \) is a measurable smoothing integrand for \( f : \Xi \times U \to \mathbb{R} \) with smoothing parameter \( \mu > 0 \) if,

\[ \forall \mu > 0, \, \tilde{f}(\cdot, \cdot, \mu) \text{ is a Carathéodory map and} \]

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

(ii) \[ \forall (\bar{x}, \bar{\mu}) \in U \times \mathbb{R}_{++} \quad \exists \text{ open } V \subseteq U \text{ with } \bar{x} \in U \text{ and} \]

\[ \kappa_f(\cdot, \bar{x}, \bar{\mu}), \, \kappa_f(\cdot, \bar{x}, \bar{\mu}) \in L^2_1(\Xi, M, \rho) \]

such that

\[ |\tilde{f}(\xi, x, \mu)| \leq \kappa_f(\xi, \bar{x}, \bar{\mu}) \quad \text{and} \quad \left\| \nabla_x \tilde{f}(\xi, x, \mu) \right\| \leq \hat{\kappa}_f(\xi, \bar{x}, \bar{\mu}) \]

\[ \forall (\xi, x, \mu) \in \Xi \times V \times (0, \bar{\mu}). \]
Gradient Consistence of Smoothing Functions

Let $U \subseteq \mathbb{R}^n$ be open and let $F : U \to \mathbb{R}$ have smoothing function $\tilde{F} : U \times \mathbb{R}_{++} \to \mathbb{R}$ on $U$.

We say that $\tilde{F}$ is \textit{gradient consistent} at $\bar{x} \in U$ if

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

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

If

$$\text{co} \left\{ \text{Limsup}_{x \to \bar{x}, \mu \downarrow 0} \nabla \tilde{F}(x, \mu) \right\} \subseteq \partial F(\bar{x}),$$

we say the $\tilde{F}$ is \textit{gradient sub-consistent} at $\bar{x} \in U$

Chen (2012), B-Hoheisel-Kanzow (2013), B-Hoheisel (2013-16)
Gradient Sub-Consistency

If \( \bar{x} \in U \) is such that

\[
\exists \bar{\nu} > 0 \text{ s.t. } \forall \nu \in (0, \bar{\nu}) \; \exists \delta(\nu, \bar{x}) > 0 \; \text{ and } \Xi(\nu, \bar{x}) \in \mathcal{M} \text{ with } \rho(\Xi(\nu, \bar{x})) \geq 1 - \nu
\]

for which

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

a.e. \( \xi \in \Xi(\nu, \bar{x}) \),

then

\[
\text{co} \left\\{ \text{Limsup} \nabla \tilde{F}(x, \mu) \bigg\}_x \rightarrow \bar{x}, \mu \downarrow 0 \right\} \subseteq \partial F(\bar{x}) = \mathbb{E} \left[ \text{co} \left\{ \text{Limsup} \nabla x \tilde{f}(\xi, x, \mu) \bigg\}_x \rightarrow \bar{x}, \mu \downarrow 0 \right\} \right].
\]
Gradient Sub-Consistency

If \( \bar{x} \in U \) is such that

uniform subgradient approximation property

\[
\begin{aligned}
\exists \bar{\nu} > 0 \text{ s.t. } \\
\forall \nu \in (0, \bar{\nu}) \quad &\exists \delta(\nu, \bar{x}) > 0 \quad \text{and} \\
\Xi(\nu, \bar{x}) \in \mathcal{M} \quad &\text{with } \rho(\Xi(\nu, \bar{x})) \geq 1 - \nu \\
\end{aligned}
\]

for which

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

then

\[
\begin{aligned}
\co \left\{ \limsup_{x \rightarrow \bar{x}, \mu \downarrow 0} \nabla \tilde{F}(x, \mu) \right\} \subseteq \partial F(\bar{x}) = \mathbb{E} \left[ \co \left\{ \limsup_{x \rightarrow \bar{x}, \mu \downarrow 0} \nabla_x \tilde{f}(\xi, x, \mu) \right\} \right].
\end{aligned}
\]
Let $\Xi \times X \subseteq \mathbb{R}^\ell \times \mathbb{R}^n$ and let $U$ be an open set containing $X$. We say that the mapping $g : \Xi \times U \to \mathbb{R}^m$ is a measurable mapping with amenable derivative if the following two conditions are satisfied:

(i) Each component of $g$ is a Carathéodory mapping and, for all $\xi \in \Xi$, $g(\xi, \cdot)$ is continuously differentiable in $x$ on $U$;

(ii) For all $(\xi, x) \in \Xi \times U$, the gradient $\nabla_x g(\xi, x)$ is locally $L^2$ bounded in $x$ uniformly in $\xi$ 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

$$\forall \bar{x} \in X \exists \epsilon(\bar{x}) > 0 \text{ such that } \| \nabla_x g(\xi, x) \| \leq \hat{\kappa}_g(\xi, \bar{x}) \quad \forall x \in B_{\epsilon(\bar{x})}(\bar{x}).$$
Composite Max (CM) Integrands

A CM integrand on $\Xi \times X$ is a mapping of the form

$$f(\xi, x) := q(c(\xi, x) + C(g(\xi, x))) \quad (0.1)$$

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

1. $C : \mathbb{R}^m \to \mathbb{R}^m$ is of the form
   $$C(y) := [p_1(y_1), p_2(y_2), \ldots, p_m(y_m)]^T,$$
   where $p_i : \mathbb{R} \to \mathbb{R}$ ($i = 1, \ldots, m$) are finite piecewise linear convex with finitely many points of nondifferentiability,

2. the mappings $c$ and $g$ are measurable mappings with amenable derivatives and

3. the mapping $q : \mathbb{R}^m \to \mathbb{R}$ is continuously differentiable with Lipschitz continuous derivative.
Piecewise Linear Convex Functions on $\mathbb{R}$

For $i = 1, \ldots, m$, 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

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

where $a_{i1} < a_{i2} < \cdots < a_{i(r_i-1)} < a_{ir_i}$. 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

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

This representation for the functions $p_i$ gives

$$\partial p_i(t) = \begin{cases} 
  a_{ij}, & t_{ij} < t < t_{i(j+1)}, \quad j = 1, \ldots, r_i \\
  [a_{i(j-1)}, a_{ij}], & t = t_{ij}, \quad j = 2, \ldots, r_i.
\end{cases} \quad i = 1, \ldots, m.$$
Smoothing for CM Integrands

\(\beta : \mathbb{R} \to \mathbb{R}_+\) be a piecewise continuous \textit{density function} s.t.

\[
\beta(t) = \beta(-t) \quad \text{and} \quad \omega := \int_{\mathbb{R}} |t| \beta(t) \, dt < \infty.
\]

Denote the \textit{distribution function} for the density \(\beta\) by \(\varphi\), i.e.,

\[
\varphi : \mathbb{R} \to [0, 1] \text{ is given by } \varphi(x) = \int_{-\infty}^{x} \beta(t) \, dt.
\]

Since \(\beta\) is symmetric, \(\varphi\) is a non-decreasing continuous with

\[
\varphi(0) = \frac{1}{2}, \quad (1 - \varphi(x)) = \varphi(-x),
\]

\[
\lim_{x \to \infty} \varphi(x) = 1 \quad \text{and} \quad \lim_{x \to -\infty} \varphi(x) = 0.
\]
Smoothing the \( p_i \)

For \( i = 1, \ldots, m \), 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

\[
\nabla_t \tilde{p}_i(t, \mu) = a_{i1} \left( 1 - \varphi\left( \frac{t - t_{i2}}{\mu} \right) \right)
\]

\[+ \sum_{j=2}^{r_i-1} a_{ij} \left( \varphi\left( \frac{t - t_{ij}}{\mu} \right) - \varphi\left( \frac{t - t_{i(j+1)}}{\mu} \right) \right) + a_{ir_i} \varphi\left( \frac{t - t_{ir_i}}{\mu} \right),\]

\[
\eta_i(t) := \lim_{\mu \downarrow 0} \nabla_t \tilde{p}_i(t, \mu)
\]

\[
= \begin{cases}
  a_{ij} & t_{ij} < t < t_{i(j+1)}, \ j = 1, \ldots, r_i \\
  \frac{1}{2}(a_{i(j-1)} + a_{ij}) & t = t_{ij}, \ j = 2, \ldots, r_i
\end{cases}
\]

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

\[
\text{Limsup}_{t \to \bar{t}, \mu \downarrow 0} \nabla_t \tilde{p}_i(t, \mu) = \partial p_i(\bar{t}) \quad \forall \bar{t} \in \mathbb{R}.
\]
Let $f$ be a CM integrand. Then $\tilde{f} : \Xi \times U \times \mathbb{R}_{++} \to \mathbb{R}$ given by
$$\tilde{f}(\xi, x, \mu) := q(c(\xi, x) + \tilde{C}(g(\xi, x), \mu))$$
is a smoothing function for $f$, where
$$\tilde{C}(y, \mu) := [\tilde{p}_1(y_1, \mu), \tilde{p}_2(y_2, \mu), \ldots, \tilde{p}_m(y_m, \mu)]^T.$$

If $\text{rank} \nabla_x g(\xi, \bar{x}) = m$, then, for all $\mu > 0$,
$$\nabla_x \tilde{f}(\xi, \bar{x}, \mu) \text{ and } \partial_x f(\xi, \bar{x})$$
are given respectively by
$$(\nabla_x c(\xi, \bar{x}) + \text{diag}(\nabla_t \tilde{p}_i(g_i(\xi, \bar{x}), \mu))\nabla_x g(\xi, \bar{x}))^T \nabla q(c(\xi, \bar{x}) + \tilde{C}(g(\xi, \bar{x}))))$$
$$(\nabla_x c(\xi, \bar{x}) + \text{diag}(\partial_t p_i(g_i(\xi, \bar{x}), \mu))\nabla_x g(\xi, \bar{x}))^T \nabla q(c(\xi, \bar{x}) + C(g(\xi, \bar{x})))).$$

Moreover,
$$\limsup_{x \to \bar{x}, \mu \downarrow 0} \nabla_x \tilde{f}(\xi, x, \mu) \subseteq \partial_x f(\xi, \bar{x})$$
Gradient Sub-Consistency of Smoothed CM Integrands

\[ f(\xi, x) := q(c(\xi, x) + C(g(\xi, x), \mu)) \]
\[ \tilde{f}(\xi, x, \mu) := q(c(\xi, x) + \tilde{C}(g(\xi, x), \mu)) \]

If \( f(\xi, \cdot) \) is subdifferentially regular \( \bar{x} \) for almost all \( \xi \in \Xi \) or \( -f(\xi, \cdot) \) is subdifferentially regular at \( \bar{x} \) for almost all \( \xi \in \Xi \).

Then

\[ \tilde{F}(x) := \mathbb{E}[\tilde{f}(\xi, x)] \]

satisfies the gradient sub-consistency property i.e.,

\[
\text{co} \left\{ \limsup_{x \to \bar{x}, \mu \downarrow 0} \nabla \tilde{F}(x, \mu) \right\} \subseteq \partial F(\bar{x}) = \mathbb{E} \left[ \text{co} \left\{ \limsup_{x \to \bar{x}, \mu \downarrow 0} \nabla_x \tilde{f}(\xi, x, \mu) \right\} \right].
\]
What happens when Clarke regularity fails?

Consider the CM integrand $f$ and its smoothing function $\tilde{f}$:

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

Assume that $\text{rank}\nabla_x g(\xi, \bar{x}) = m$ for a fixed $(\xi, \bar{x}) \in \Xi \times X$.

Then the limit

$$u(\xi, \bar{x}) := \lim_{\mu \downarrow 0} \nabla_x f(\xi, \bar{x}, \mu)$$
$$= (\nabla_x c(\xi, \bar{x}) + (z_1(\xi, \bar{x}), \ldots, z_m(\xi, \bar{x}))^T \nabla q(c(\xi, \bar{x}) + C(g(\xi, \bar{x})))$$

exist as given with $u(\xi, \bar{x}) \in \partial_x f(\xi, \bar{x})$, where

$$z_i(\xi, \bar{x}) := \eta_i(g_i(\xi, \bar{x})) \nabla_x g_i(\xi, \bar{x}) \quad \text{with}$$

$$\eta_i(t) := \lim_{\mu \downarrow 0} \nabla_t \tilde{p}_i(t, \mu) \in \partial p_i(t).$$
Subgradient Approximation by Smoothing

\[ f(\xi, x) := q(c(\xi, x) + C(g(\xi, x), \mu)) \]
\[ \tilde{f}(\xi, x, \mu) := q(c(\xi, x) + \tilde{C}(g(\xi, x), \mu)) \]
\[ u(\xi, \bar{x}) := \lim_{\mu \downarrow 0} \nabla_x \tilde{f}(\xi, \bar{x}, \mu) \]

\[ F(x) := \mathbb{E}[f(\xi, x)] \quad \text{and} \quad \tilde{F}(x, \mu) := \mathbb{E}[\tilde{f}(\xi, x, \mu)] \quad \forall \ x \in X. \]

Then \( \tilde{F}(\cdot, \mu) \) is differentiable for all \( \mu > 0 \) with

\[ \nabla_x \tilde{F}(x, \mu) = \mathbb{E}[\nabla_x \tilde{f}(\xi, x, \mu)], \]

the function \( u \) is well defined, and, \( \forall x \in X \),

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

For $\mu > 0$ and $\bar{x} \in U$ there exits $K(\bar{x}) > 0$ and $\delta(\bar{x}) > 0$ s.t.

$$\left\| \nabla \tilde{f}(\xi, x, \mu) - \nabla \tilde{f}(\xi, \bar{x}, \mu) \right\| \leq \frac{K(\bar{x})}{\mu} \| x - \bar{x} \| \quad \forall \xi \in \Xi \text{ and } x \in B_{\delta(\bar{x})}(\bar{x})$$

and

$$\text{dist} \left( \mathbb{E}[\nabla \tilde{f}(\xi, x, \mu)] \mid \partial F(\bar{x}) \right) \leq \frac{K(\bar{x})}{\mu} \| x - \bar{x} \| + \text{dist} \left( \nabla_x \tilde{F}(\bar{x}, \mu) \mid \partial F(\bar{x}) \right)$$

$$\forall \ x \in B_{\delta(\bar{x})}(\bar{x}).$$

Moreover, for any $\gamma \in (0, 1)$:

$$\limsup_{x \to \bar{x}, \mu = O(\| x - \bar{x} \|^{\gamma})} \mathbb{E}[\nabla \tilde{f}(\xi, x, \mu)] \subset \partial F(\bar{x}).$$