On generalized gradients and optimization–highlights

November 2008

Locally Lipschitz functions:

Def. 2.1 $f : \mathbb{R}^n \to (-\infty, +\infty]$ is *Lipschitz near* $x_0 \in \text{int dom } f$ if

$$\exists_{K \geq 0} \forall_{x,x' \in B_\delta(x_0)} \; |f(x) - f(x')| \leq K|x - x'|$$

where $\delta > 0$ is so small that $B_\delta(x_0) \subset \text{dom } f$.

If f is Lipschitz near every point in $\mathbb{R}^n$, then f is called *locally Lipschitz* (= LL).

From now on: only functions into $\mathbb{R}$, except when they are convex.

Example 2.2 (i) If $f : \mathbb{R}^n \to (-\infty, +\infty]$ is convex, then it is LL at every $x_0 \in \text{int dom } f$.

Example 2.2 (ii) If $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable at $x_0 \in \mathbb{R}^n$, then it is LL at x_0.

Generalized gradients:

Def. 2.3 (Clarke) Let $f : \mathbb{R}^n \to \mathbb{R}$, $x_0 \in \mathbb{R}^n$. The *generalized directional derivative* of f at x_0 in the direction $d \in \mathbb{R}^n$ is defined by

$$f^o(x_0; d) := \limsup_{\substack{h \to 0 \\ \lambda \downarrow 0}} \frac{f(x_0 + h + \lambda d) - f(x_0 + h)}{\lambda} :=$$

$$:= \lim_{\delta \downarrow 0} \sup_{h \in B_\delta(0), \lambda \in (0,\delta)} \frac{f(x_0 + h + \lambda d) - f(x_0 + h)}{\lambda}.$$

Also, the *generalized gradient* of f at x_0 is defined by

$$\bar{\partial} f(x_0) := \{\xi \in \mathbb{R}^n : f^o(x_0; d) \geq \xi^t d \forall_{d \in \mathbb{R}^n}\}.$$

Big difference with convex analysis: Def. 2.3 has difference quotients

$$\frac{f(x_0 + h + \lambda d) - f(x_0 + h)}{\lambda}$$

with a *variable* “base” $x_0 + h$!

Ex. 2.4 a. For $f(x) := |x|$ on $\mathbb{R}$ one has $\bar{\partial} f(0) = [-1, 1]$.

b. For $f(x) := -|x|$ on $\mathbb{R}$ one also has $\bar{\partial} f(0) = [-1, 1]$.

In general, $\bar{\partial} f(x_0)$ could be empty. However, this is not so for LL functions:

Thm. 2.5 Let $f : \mathbb{R}^n \to \mathbb{R}$ be LL near $x_0 \in \mathbb{R}^n$ with LL constant K. Then

(i) $\bar{\partial} f(x_0) \neq \emptyset$.

(ii) $|\xi| \leq K$ for all $\xi \in \bar{\partial} f(x_0)$.

(iii) $\forall_{d \in \mathbb{R}^n}$ $f^0(x_0; d) = \sup_{\xi \in \bar{\partial} f(x_0)} \xi^t d$.

Lemma 2.6 is key to Thm. 2.5: generalized directional derivative functions are *automatically* convex.

Lemma 2.6 In Thm. 2.5 $p : d \mapsto f^o(x_0; d)$ is positively homogeneous and subadditive (whence convex). Also, $|p(d)| \leq K|d|$ $\forall_{d \in \mathbb{R}^n}$.

Note: $\bar{\partial} f(x_0) := \partial p(0)$! This enables use of convex analysis.

Generalized gradients in two classical situations:

Prop. 2.7 If $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable at $x_0 \in \mathbb{R}^n$, then

$$\bar{\partial} f(x_0) = \{\nabla f(x_0)\},$$

Prop. 2.8 If $f : \mathbb{R}^n \to \mathbb{R}$ be convex, then for every $x_0 \in \mathbb{R}^n$

$$\bar{\partial} f(x_0) = \partial f(x_0),$$

Example 2.4.a illustrates Prop. 2.8.

Generalized gradient calculus:

Calculus rules for generalized gradients retain the "difficult" parts of the Moreau-Rockafellar theorem and the Dubovitskii-Milyutin theorem:

Thm. 2.9 Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be LL near $x_0 \in \mathbb{R}^n$. Then

$$\forall_{\alpha>0}\ \bar{\partial}(\alpha f)(x_0) = \alpha\bar{\partial} f(x_0),$$
$$\bar{\partial}(f+g)(x_0) \subset \bar{\partial} f(x_0) + \bar{\partial} g(x_0).$$

Thm. 2.10 Let $f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$ be LL near $x_0 \in \mathbb{R}^n$. Define $f : \mathbb{R}^n \to \mathbb{R}$ by

$$f(x) := \max_{1 \le i \le m} f_i(x)$$

and let $I(x_0) := \{i : f_i(x_0) = f(x_0)\}$. Then

$$\bar{\partial} f(x_0) \subset \operatorname{co}\left(\cup_{i \in I(x_0)} \bar{\partial} f_i(x_0)\right).$$

KKT theorems for generalized gradients:

Let $S \subset \mathbb{R}^n$ be closed; let $\text{dist}_S(x) := \inf_{x' \in S} |x - x'|$.

Thm. 3.1 ("small" KKT) Let $f : \mathbb{R}^n \to \mathbb{R}$ be LL near $\bar{x} \in S$. If $\bar{x}$ is a local optimal solution of

$$(P) \quad \inf_{x \in S} f(x).$$

then

$$\exists_{\bar{\eta} \in \mathbb{R}^n} \; 0 \in \bar{\partial} f(\bar{x}) + \bar{\eta}$$

and

$$\bar{\eta} \in \cup_{t>0} t \, \bar{\partial} \text{dist}_S(\bar{x}) \quad \textbf{(NCP)}.$$

Proof uses that $\bar{x}$ is also a local optimal solution of an auxiliary problem

$$(P') \quad \inf_{x \in B_{\epsilon/2}(\bar{x})} [f(x) + K \text{dist}_S(x)]$$

with $K :=$ LL-constant of f at $\bar{x}$. After this, result follows by Thm. 2.9, because

$$0 \in \bar{\partial}(f + K\text{dist}_S)(\bar{x}) \subset \bar{\partial} f(\bar{x}) + K \bar{\partial} \text{dist}_S(\bar{x}).$$

Compared to "small" KKT thm. in [OSC], above proof uses "penalty term" $K\text{dist}_S$ instead of χ_S.

Rem. 3.2 NCP stands for *normal cone property*: $\cup_{t>0} t \, \bar{\partial} \text{dist}_S(\bar{x})$ is called the *normal cone* to S at $\bar{x}$. If S is additionally *convex*, then this agrees with [OSC]. Reason: then dist_S is convex, so $\bar{\partial} \text{dist}_S(\bar{x}) =$

$\partial \mathrm{dist}_S(\bar{x})$. Hence (NCP) implies $\forall_{x\in S}\, \bar{\eta}^t(x-\bar{x}) \leq 0$, i.e., obtuse angle property.

Thm. 3.3 (KKT – no equality constraints) Let $f, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ be LL near $\bar{x} \in S$. If $\bar{x}$ is locally optimal for

$$(P) \quad \inf_{x\in S}\{f(x) : g_1(x) \leq 0, \ldots, g_m(x) \leq 0\}$$

then $\exists_{\bar{u}_0\in\{0,1\}, \bar{u}\in\mathbb{R}^m_+, \bar{\eta}\in\mathbb{R}^n, (\bar{u}_0,\bar{u})\neq(0,0)}$ with

$$\forall_i \bar{u}_i g_i(\bar{x}) = 0 \;\; \textbf{(CS)}$$

$$0 \in \bar{u}_0 \bar{\partial} f(\bar{x}) + \sum_{i\in I(\bar{x})} \bar{u}_i \bar{\partial} g_i(\bar{x}) + \bar{\eta} \;\; \textbf{(LI)},$$

$$\bar{\eta} \in \cup_{t>0}\, t\, \bar{\partial}\mathrm{dist}_S(\bar{x}) \;\; \textbf{(NCP)}.$$

Here $I(\bar{x}) := \{i : 1 \leq i \leq m, g_i(\bar{x}) = 0\}$. The proof is similar to that of the "convex" counterpart in [OSC] by using that $\bar{x}$ is also a local optimal solution of the auxiliary problem

$$(P') \quad \inf_{x\in S} \max\left[f(x) - f(\bar{x}), \max_{1\leq i\leq m} g_i(x)\right]$$

and applying "small" KKT Theorem 3.1.

Thm. 3.5 (KKT – general) Let $f, g_1, \ldots, g_m$, $h_1, \ldots, h_p : \mathbb{R}^n \to \mathbb{R}$ be LL near $\bar{x} \in S$. If $\bar{x}$ is locally optimal for

$$(P) \quad \inf_{x\in S}\{f(x) : \forall_i g_i(x) \le 0, \forall_j h_j(x) = 0\}$$

then $\exists_{\bar{u}_0\in\{0,1\},\bar{u}\in\mathbb{R}^m_+,\bar{v}\in\mathbb{R}^p,\bar{\eta}\in\mathbb{R}^n,(\bar{u}_0,\bar{u},\bar{v})\neq(0,0,0)}$ with **(CS)** and **(NCP)** holding and the following **(LI)**:

$$0 \in \bar{u}_0\bar{\partial}f(\bar{x}) + \sum_{i\in I(\bar{x})} \bar{u}_i\bar{\partial}g_i(\bar{x}) + \sum_{j=1}^{m} \bar{v}_j\bar{\partial}h_j(\bar{x}) + \bar{\eta},$$

Proof differs from "convex" counterpart in [OSC], by complications with equality constraints. It uses:

Ekeland's theorem Let $F : S \to \mathbb{R}$ be l.s.c. and bounded below. Let $\epsilon > 0$ and let $x_0 \in S$ be such that

$$F(x_0) \le \inf_{x\in S} F(x) + \epsilon.$$

Then there exists $\tilde{x}_\epsilon \in S$ such that $|x_0 - \tilde{x}_\epsilon| \le \sqrt{\epsilon}$ and

$$F(\tilde{x}) \le F(x) + \sqrt{\epsilon}|x - \tilde{x}_\epsilon| \text{ for all } x \in S.$$

In words: an ϵ-almost minimizer of F is $\sqrt{\epsilon}$-close to an *exact* minimizer of a "$\sqrt{\epsilon}$-penalization" of F.

Application gives: $\forall_{\epsilon>0}\exists_{\tilde{x}_\epsilon\in S\cap B_{\sqrt{\epsilon}}(\bar{x})}$ with $\tilde{x}_\epsilon$ a local optimal solution of

$$(P_\epsilon) \quad \inf_{x\in S} \left[F_\epsilon(x) + \sqrt{\epsilon}|x - \tilde{x}_\epsilon|\right]$$

where

$$F_\epsilon(x) := \max \left[f(x) - f(\bar{x}) + \epsilon, \max_i g_i(x), \max_j |h_j(x)| \right].$$

By applying "small" KKT theorem, then letting $\epsilon \to 0$ and using certain limit arguments, proof of Thm. 3.5 now follows from "small" KKT Theorem 3.1.

Corollary (KKT, smooth case) Let $f, g_1, \ldots, g_m$, $h_1, \ldots, h_p : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable in $\bar{x} \in \mathbb{R}^n$. If $\bar{x}$ is locally optimal for (P) with $S := \mathbb{R}^n$ then $\exists_{\bar{u}_0 \in \{0,1\}, \bar{u} \in \mathbb{R}^m_+, \bar{v} \in \mathbb{R}^p, (\bar{u}_0, \bar{u}, \bar{v}) \neq (0,0,0)}$ with **(CS)** and

$$0 = \bar{u}_0 \nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \nabla g_i(\bar{x}) + \sum_{j=1}^{m} \bar{v}_j \nabla h_j(\bar{x}).$$