NOTES ON THE DIFFERENTIATION OF QUASI-CONVEX FUNCTIONS

F. Reese Harvey and H. Blaine Lawson, Jr.*

ABSTRACT

This expository paper presents elementary proofs of four basic results concerning derivatives of quasi-convex functions. They are combined into a fifth theorem which is simple to apply and adequate in many cases. Along the way we establish the equivalence of the basic lemmas of Jensen and Slodkowski.

1. Introduction.

Let $u(x)$ be a real-valued function on an open set $X \subset \mathbf{R}^n$. Then $u(x)$ is said to be *quasi-convex* if the function $u(x) + \frac{\lambda}{2}|x|^2$ is convex for some $\lambda \geq 0$. There are four basic results concerning the differentiability of such functions. To state some of them we need the following concept. Let $\mathrm{Sym}^2(\mathbf{R}^n)$ denote the set of $n \times n$ symmetric matrices.

Definition 1.1. A point $x \in X$ is called an **upper contact point for** u if there exists $(p, A) \in \mathbf{R}^n \times \mathrm{Sym}^2(\mathbf{R}^n)$ such that

$$u(y) \;\leq\; u(x) + \langle p, y - x\rangle + \tfrac{1}{2}\langle A(y-x), y-x\rangle \qquad \forall\, y \text{ near } x. \tag{1.1}$$

In this case, (p, A) is called an **upper contact jet for** u **at** x.

The first result is the differentiability at upper contact points.

Lemma 1.2. (D at UCP). *Suppose u is quasi-convex. If x is an upper contact point for u, then u is differentiable at x. Moreover, if (p, A) is any upper contact jet for u at x, then $p = D_x$ is unique.*

Another even more standard result is called partial continuity of the gradient, or first derivative.

Lemma 1.3. (PC of FD). *Suppose u is quasi-convex and $x_j \to x$. If u is differentiable at each x_j and at x, then $D_{x_j}u \to D_x u$.*

The next two results concern the second-order contact of quasi-convex functions and are of a deeper nature.

*Partially supported by the N.S.F.

THEOREM 1.4. (Alexandrov). *A locally quasi-convex function is twice differentiable almost everywhere.*

For the next result we need two variations of the notion of an upper contact jet. First, we say that (p, A) is a **strict** upper contact jet for $u \in \mathrm{USC}(X)$ at $x_0 \in X$ if the upper contact inequality (1.1) is strict for $y \neq x_0$. An understanding of the strict upper contact jets will be adequate for our discussion since (p, A) is an upper contact jet if and only if $(p, A + \epsilon I)$ is a strict upper contact jet for all $\epsilon > 0$. Second, we need a notion of upper contact point and jet, which requires the inequality (3.1) to hold globally.

Definition 1.5. Given $u \in \mathrm{USC}(X)$ and $A \in \mathrm{Sym}^2(\mathbf{R}^n)$, a point x is called a **global upper contact point of type** A **on** X if for some $p \in \mathbf{R}^n$

$$u(y) \;\leq\; u(x) + \langle p, y - x\rangle + \tfrac{1}{2}\langle A(y - x), y - x\rangle \quad \forall\, y \in X. \tag{1.2}$$

Let $\mathcal{C}(u, X, A)$ denote the set of all global upper contact points of type A on X for the function u.

Remark 1.6. Note that if u is quasi-convex, then by (D at UCP) each point $x \in \mathcal{C}(u, X, A)$ is a point of differentiability and the only p in (1.2) is $p = D_x u$.

THEOREM 1.7. (Jensen-Slodkowski). *Suppose that u is a quasi-convex function possessing a strict upper contact jet (p, A) at x. Let B_ρ denote the ball of radius ρ about x. Then there exists $\bar\rho > 0$ such that the measure*

$$|\mathcal{C}(u, B_\rho, A)| \;>\; 0 \qquad \forall\, 0 < \rho \leq \bar\rho. \tag{1.3}$$

This result follows in a straightforward/elementary manner (see Section 4) from Slodkowski's Lemma 4.1 below, which in turn is proved in Sections 5–7.

On the other hand, Slodkowski's Lemma 4.1 and Jensen's Lemma 9.1 below are equivalent special cases of Theorem 1.7 (see Section 9 for a proof of this equivalence).

The four results above yield the following useful theorem concerning the upper contact jets of a quasi-convex function. The order two part of this theorem can be considered a "partial upper semi-continuity of the second derivative" (PUSC of SD).

THEOREM 1.8. (Upper Contact Jets). *Suppose u is quasi-convex with an upper contact jet (p_0, A_0) at a point x. Then*

(D at UCP) *u is differentiable at x and $D_x u = p_0$.*

Suppose E is a set of full measure in a neighborhood of x. Then there exists a sequence $\{x_j\} \subset E$ with $x_j \to x$ such that u is twice differentiable at each x_j and

(PC of FD) $\quad D_{x_j} u \;\to\; D_x u = p_0,$

(PUSC of SD) $\quad D^2_{x_j} u \;\to\; A \leq A_0.$

Proof. By Alexandrov's Theorem, the set of points $x \in E$ where u is twice differentiable, is a set of full measure. In order to apply the Jensen-Slodkowski Lemma we replace (p_0, A_0) by the strict upper contact jet $(p_0, A_0 + \epsilon I)$. Now choose a sequence $\epsilon_j \to 0$, and pick

a point $x_j \in B_{\epsilon_j}(x_0)$ such that: (1) $x_j \in E$, (2) u is twice differentiable at x_j, and (3) x_j is a global upper contact point of type $A_0 + \epsilon_j I$ on $B_{\epsilon_j}(x_0)$ for u. By the basic differential calculus fact Lemma 3.3, $D^2_{x_j} u \leq A_0 + \epsilon_j I$. Since u is λ-quasi-convex, we have $D^2_{x_j} u + \lambda I \geq 0$. Thus,

$$-\lambda I \ \leq \ D^2_{x_j} u \ \leq \ A_0 + \epsilon_j I. \tag{1.4}$$

By compactness there is a subsequence such that $D^2_{x_j} u \to A \leq A_0$. ∎

Theorem 1.8 can be stated succinctly in terms of the subset $J^+(u) \subset J^2(X) \equiv X \times \mathbf{R} \times \mathbf{R}^n \times \mathrm{Sym}^2(\mathbf{R}^n)$ of upper contact jets for u, and another subset depending on E. Define $J(u, E) \subset J^2(X)$ to be the subset of tuples $(x, u(x), D_x u, D^2_x u + P)$ such that $x \in E$, u is twice differentiable at x, and $P \geq 0$. Then Theorem 1.8 condenses to:

$$\text{If } u \text{ is quasi}-\text{convex and } E \text{ has full measure, then } \ J^+(u) \ \subset \ \overline{J(u, E)}. \tag{1.5}$$

We will deduce the four results from the special case where u is convex, and for Lemma 1.7 we will reduce to the special case where $A = \lambda I$, i.e., Slodkowski's Lemma 4.1.

2. Convex Functions – The Subdifferential.

In this section we shall assume that u is convex and prove some of the basic properties using the following standard concept involving lower contact. If a convex function u is defined on a convex open set $X \subset \mathbf{R}^n$, the **subdifferential** ∂u **of** u is defined to be the set

$$\partial u \ \equiv \ \{(x, p) \in X \times \mathbf{R}^n : \ u(x) + \langle p, y - x \rangle \leq u(y) \ \ \forall\, y \in X\}. \tag{2.1}$$

Geometrically this means that the graph of the affine function $u(x) + \langle p, y - x \rangle$ lies below the graph of u and the graphs touch above the point x, that is, the hyperplane is a supporting hyperplane for the convex set $\{y \geq u(x)\}$ (or, in the language of Definition 1.5, that x is a lower contact point of type $A = 0$ on X).

The fibre of ∂u over $x \in X$ is denoted $\partial u(x)$. Note that

$$u \ \text{ is convex} \qquad \Longleftrightarrow \qquad \partial u(x) \ \neq \ \emptyset \ \text{ for each } \ x \in X. \tag{2.2}$$

For most of the purposes of this section, this can be taken as the definition of convexity.

The useful inequalities

$$\langle p, y - x \rangle \ \leq \ u(y) - u(x) \ \leq \ \langle q, y - x \rangle \quad \text{ for all } \quad p \in \partial u(x), q \in \partial u(y) \tag{2.3}$$

follow immediately from the definition (2.1) of the set ∂u. These inequalities, stated geometrically, say that (with coordinates intrinsic to the affine line determined by x and y) if $x \leq y$, then $\bar{p} \leq s \leq \bar{q}$ where s is the slope of the chord above $[x, y]$ and $\bar{p}$ and $\bar{q}$ are the slopes of the supporting lines above x and y. Note that (2.3) implies that ∂u is monotone (non-decreasing). That is,

$$0 \ \leq \ \langle q - p, y - x \rangle \quad \text{ for all } \quad p \in \partial u(x), q \in \partial u(y). \tag{2.4}$$

For each compact set $K \subset X$

$$\partial u \cap (K \times \mathbf{R}^n) \text{ is a nonempty compact set with convex fibres.} \tag{2.5}$$

Proof. Let $K_\delta = \{x \in X : \operatorname{dist}(x, K) \le \delta\}$ and choose δ small enough so that $K_\delta \subset X$. Since $|u|$ is bounded on K_δ, the left hand inequality in (2.3) gives the upper bound

$$\langle p, y - x\rangle \;\le\; 2\|u\|_{K_\delta} \qquad \forall\, p \in \partial u(x),\ x \in K, \text{ and } y \in K_\delta.$$

Choosing $y = x + \delta \frac{p}{|p|}$ gives $\langle p, y - x\rangle = \delta|p| \le 2\|u\|_{K_\delta}$, and so $|p|$ is bounded on K. Since $\partial u \cap (K \times \mathbf{R}^n)$ is closed and vertically bounded, the compactness assertion follows.

Now fix x and note that

$$\begin{aligned} \partial u(x) &= \{p : u(x) + \langle p, y - x\rangle \le u(y) \ \ \forall\, y \in X\} \\ &= \bigcap_{y \in X} \{p : L_y(p) \le u(y)\} \end{aligned}$$

is an intersection of affine half-spaces and therefore convex. ∎

Combining (2.5) with the inequalities (2.3) easily yields two important facts. The first is that u is Lipschitz on K.

Lemma 2.1.

$$|u(y) - u(x)| \;\le\; C|y - x| \tag{2.6}$$

where the Lipschitz constant C is the supremum of $|p|$ taken over $p \in \partial u(x), x \in K$.

The second is the following.

Lemma 2.2. *u is differentable at x $\iff$ $\partial u(x) = \{p\}$ is a singleton, in which case*

$$p \;=\; D_x u \;=\; \lim_{\substack{y \to x \\ q \in \partial u(y)}} q. \tag{2.7}$$

Proof. If $\partial u(x) = \{p\}$ is a singleton, then by the compactness in (2.5)

$$p \;=\; \lim_{\substack{y \to x \\ q \in \partial u(y)}} q. \tag{2.8}$$

With $p \in \partial u(x)$, $q \in \partial u(y)$ and $e = (y - x)/|y - x|$, the inequalities (2.3) can be rewritten as

$$0 \;\le\; \frac{u(y) - u(x) - \langle p, y - x\rangle}{|y - x|} \;\le\; \langle q - p, e\rangle. \tag{2.9}$$

Combining (2.8) and (2.9) shows that if $\partial u(x)$ is a singleton, then u is differentiable at x with $D_x u = p$.

Suppose now that u is differentiable at x and $p \in \partial u(x)$. Then $\langle p, y - x\rangle \le u(y) - u(x)$. Hence, for each $e \in \mathbf{R}^n$, $|e| = 1$, we have $t\langle p, e\rangle \le u(x + te) - u(x)$ for t small. This proves that $\langle p, e\rangle \le \langle D_x u, e\rangle$ for all $|e| = 1$, and hence $p = D_x u$. Since $\partial u(x) \ne \emptyset$, there always exists $p \in \partial u(x)$, and hence $D_x u \in \partial u(x)$. ∎

Corollary 2.3.

$$u \text{ is differentiable everywhere} \iff \partial u \text{ is single valued} \iff u \text{ is } C^1.$$

Remark 2.4. (Critical Points). Note that by (2.1)

$$x \text{ is a minimum point for } u \qquad \Longleftrightarrow \qquad 0 \in \partial u(x)$$

We say that x is a *critical point* for u if u is differentiable at x and $D_x u = 0$. By Lemma 2.2

$$x \text{ is a critical point for } u \qquad \Longleftrightarrow \qquad \partial u(x) \;=\; \{0\}. \tag{2.10}$$

In particular,

$$\text{If } x \text{ is a critical point for } u, \text{ then } x \text{ is a minimum point for } u. \tag{2.11}$$

Finally, in the proof of Alexandrov's Theorem it will be helpful to use additivity of the subdifferential for the sum of a convex function u and a quadratic polynomial φ with $D^2\varphi \geq 0$.

Lemma 2.5. *Suppose $\varphi(y) \equiv c + \langle q, y\rangle + \frac{1}{2}\langle Py, y\rangle$ with $P \geq 0$. Then*

$$\partial(u+\varphi)(x) \;=\; \partial u(x) + \partial\varphi(x) \;=\; \partial u(x) + q + Px.$$

Proof. Of course, $\partial\varphi(x) = \{q+Px\}$. Moreover, it follows directly from the definition of the subdifferential that $\partial u(x) + \partial\varphi(x) \subset \partial(u+\varphi)(x)$. It remains the prove that $\partial(u+\varphi)(x) \subseteq \partial u(x) + q + Px$. Suppose $p+q+Px \in \partial(u+\varphi)(x)$. We want to show that $p \in \partial u(x)$. Our assumption is that $u(y) + \varphi(y) \geq u(x) + \varphi(x) + \langle p+q+Px, y-x\rangle$ which is equivalent to

$$u(y) \;\geq\; u(x) + \langle p, y-x\rangle - \tfrac{1}{2}\langle P(y-x), y-x\rangle \;\equiv\; \psi(y).$$

This means that the epigraph of u is contained in the epigraph of ψ in $\mathbf{R}^{n+1}$. This remains true if one applies the dilation ρ_t of $\mathbf{R}^{n+1}$ by $t > 0$ centered at the point $(x, u(x))$. By convexity $\mathrm{epi}(u) \subset \rho_t(\mathrm{epi}(u))$ for all $t \geq 1$. That is, $\mathrm{epi}(u) \subset \rho_t(\mathrm{epi}(u)) \subset \rho_t(\mathrm{epi}(\psi))$. As $t \to \infty$, the dilations $\rho_t(\mathrm{epi}(\psi))$ decrease down to the half-space $H \equiv \{y : u(x) + \langle p, y-x\rangle \geq 0\}$, proving that $\mathrm{epi}(u) \subseteq H$ or that $p \in \partial u(x)$. ∎

3. Proof of Lemmas 1.2 and 1.3.

These two results follow easily from the convex case, enabling us to assume that u is convex in the proofs.

Proof of Lemma 1.2 (D at UCP). By (2.2) there exists $\bar{p} \in \mathbf{R}^n$ with $u(x) + \langle \bar{p}, y-x\rangle \leq u(y)$, $\forall y \in X$. Subtracting the affine function of y on the left from u, we can assume $0 \leq u(y)$ and $u(x) = 0$. Now if (p, A) is an upper contact jet for u at x, then

$$0 \;\leq\; u(y) \;\leq\; \langle p, y-x\rangle + \tfrac{1}{2}\langle A(y-x), y-x\rangle \tag{3.1}$$

which implies that u is differentiable at x with $D_x u = p$. This proves Lemma 1.2. ∎

Remark 3.1. This argument proves more. First note that by (2.1) if $0 \leq u(y)$ and $u(x) = 0$, then $0 \in \partial u(x)$. Now because of (2.10) the inequalities (3.1) imply that $p = 0$ and $A \geq 0$. Therefore,

$$\text{If } (p, A) \text{ is an upper contact jet for a convex function } u, \quad \text{then } A \geq 0. \tag{3.2}$$

For the converse, namely:

If each upper contact jet (p, A) of an u.s.c. function u satisfies $A \geq 0$, then u is convex,

the reader is referred to [HL_1].

Proof of Lemma 1.3 (PC of FD). By Lemma 2.2

$$\text{If } u \text{ is differentiable at } x, \text{ then } \lim_{\substack{y \to x \\ q \in \partial u(y)}} q = D_x u. \tag{3.3}$$

This is, in fact, a stronger version of Lemma 1.3. ∎

4. The Reduction of the Jensen-Slodkowski Lemma to the Slodkowski Lemma.

The Jensen-Slodkowski Lemma 1.7 contains the following as a special case.

Lemma 4.1. (Slodkowski). *Suppose that u is a convex function with a strict upper contact jet $(0, \lambda I)$ at a point x. Then there exists $\bar{\rho} > 0$ such that the measure*

$$|\mathcal{C}(u, B_\rho, \lambda I)| \; > \; 0 \qquad \forall\, 0 < \rho \leq \bar{\rho}.$$

The following trivial lemma is all that is needed for the reduction. First note that a degree-2 polynomial $\varphi(y)$ satisfies

$$\varphi(y) \;=\; \varphi(x) + \langle D_x\varphi, y - x\rangle + \tfrac{1}{2}\langle (D_x^2\varphi)(y-x), y-x\rangle \qquad \forall\, x, y \in \mathbf{R}^n.$$

and $D_x^2\varphi$ is independent of x.

Lemma 4.2. *Suppose φ is a degree-2 polynomial. Set $B \equiv D_x^2\varphi$.*

(1) If (p, A) is an upper contact jet for u at x on X, then $(p + D_x\varphi, A + B)$ is an upper contact jet for $w \equiv u + \varphi$ at x on X.

(2) (p, A) is strict for $u \;\Rightarrow\; (p + D_x\varphi, A + B)$ is strict for $w \equiv u + \varphi$.

(3) $\mathcal{C}(u + \varphi, X, A + B) \;=\; \mathcal{C}(u, X, A)$

Proof. For any point $x \in X$

$$\begin{aligned} & u(y) \;\leq\; u(x) + \langle p, y - x\rangle + \tfrac{1}{2}\langle A(y-x), y-x\rangle \quad \forall\, y \in X \\ \Longleftrightarrow \qquad & w(y) \;\leq\; w(x) + \langle p + D_x\varphi, y - x\rangle + \tfrac{1}{2}\langle (A+B)(y-x), y-x\rangle \quad \forall\, y \in X \end{aligned}$$

■

We now claim the following.

The special case Lemma 4.1 implies the full Jensen – Slodkowski Lemma. (4.3)

Proof. Suppose (p, A) is a strict upper contact jet for u at x. Take $\varphi(y) \equiv -\langle p, y - x\rangle - \frac{1}{2}\langle A(y-x), y-x\rangle + \frac{\lambda}{2}|y-x|^2$ and apply Lemma 4.2. Then $(0, \lambda I)$ is a strict upper contact jet for $w \equiv u + \varphi$ at x on $X \equiv B_{\bar{\rho}}(x)$ for some $\bar{\rho} > 0$. Moreover, $\mathcal{C}(w, X, \lambda I) = \mathcal{C}(u, X, A)$. Finally take λ sufficiently large so that $w \equiv u + \varphi$ is convex. (If u is α-quasi-convex and $A \leq \beta I$, take $\lambda \geq \alpha + \beta$.) Now the Slodkowski Lemma 4.1 can be applied to w. ■

In the next section we establish the elementary convex geometric fact needed to prove Slodkowski's Lemma.

5. The Convex Hull of Two Open Paraboloids of the Same Radius.

Now we begin the proof of Slodkowski's Lemma 4.1, which is completed in Section 8. Our proof is an adaptation of his proof, in which we use paraboloids of radius r in place of balls of radius r. It is important that these paraboloids be *open*. Each such paraboloid is determined by its vertex $(v, \varphi(v)) \in \mathbf{R}^n \times \mathbf{R}$, and by definition, is the open epigraph $\mathrm{epi}(\varphi)$ of the quadratic function

$$\varphi(y) \;\equiv\; \varphi(v) + \frac{1}{2r}|y - v|^2.$$

Given two such open paraboloids $\mathrm{epi}(\varphi_1)$ and $\mathrm{epi}(\varphi_2)$ with vertices $(v_1, \varphi(v_1))$ and $(v_2, \varphi(v_2))$ respectively, we compute the convex hull $\mathrm{ch}(\mathrm{epi}(\varphi_1) \cup \mathrm{epi}(\varphi_2))$ of the union of these two open sets. We shall emphasize what is needed in the application.

Lemma 5.1. *There is an open vertical slab* $\mathrm{SLAB} \subset \mathbf{R}^{n+1}$ *written as the intersection* $\mathrm{SLAB} = \mathcal{H}_1 \cap \mathcal{H}_2$ *of two parallel vertical open half-spaces with the following property. Let* $\mathrm{ch} \equiv \mathrm{ch}(\mathrm{epi}(\varphi_1) \cup \mathrm{epi}(\varphi_2))$. *Then*

$$\mathrm{graph}(\varphi_1) \cap \mathrm{ch} \;\subset\; \mathcal{H}_1 \qquad \text{and} \qquad \mathrm{graph}(\varphi_2) \cap \mathrm{ch} \;\subset\; \mathcal{H}_2 \tag{5.1}$$

Moreover, the width of SLAB *is* $|v_1 - v_2|$.

Proof. Set $e \equiv \frac{v_2 - v_1}{|v_2 - v_1|}$ and let $m \equiv \frac{\varphi(v_2) - \varphi_1(v_1)}{|v_2 - v_1|}$ denote the slope of the line segment from the first vertex to the second. Define $\mathcal{H}_1$ to be the open half-space whose boundary hyperplane $\partial\mathcal{H}_1$ has interior normal $(e, 0)$ and passes through $(v_1 + rme, 0)$. Similarly, define $\mathcal{H}_2$ to have interior normal $(-e, 0)$ and boundary $\partial\mathcal{H}_2$ passing through $(v_2 + rme, 0)$. Then $\mathrm{SLAB} \equiv \mathcal{H}_1 \cap \mathcal{H}_2$ clearly has width $|v_2 - v_1|$. It remains to prove (5.1).

This can be seen by determining the pairs of points

$$z_1 \;\equiv\; (y_1, \varphi_1(y_1)) \in \mathrm{graph}(\varphi_1) \qquad \text{and} \qquad z_2 \;\equiv\; (y_2, \varphi_2(y_2)) \in \mathrm{graph}(\varphi_2)$$

which have a common tangent plane H. Equating normals $(D_{y_1}\varphi_1, -1)$ and $(D_{y_2}\varphi_2, -1)$ yields $y_1 - v_1 = y_2 - v_2$. Thus $y_1 = v_1 + w$ and $y_2 = v_2 + w$ for some $w \in \mathbf{R}^n$.

Now $N \equiv (\frac{w}{r}, -1)$ is normal to H. Hence, $z_1, z_2 \in H$ implies that $\frac{1}{r}\langle y_1, w\rangle - \varphi_1(y_1) = \frac{1}{r}\langle y_2, w\rangle - \varphi_2(y_2)$. Therefore $\langle v_2 - v_1, w\rangle = \langle y_2 - y_1, w\rangle = r(\varphi_2(y_2) - \varphi_1(y_1))$. However,

$$\varphi_2(y_2) - \varphi_1(y_1) = \varphi_2(v_2) + \frac{1}{2r}|y_2 - v_2|^2 - \varphi_1(v_1) - \frac{1}{2r}|y_1 - v_1|^2 = \varphi_2(v_2) - \varphi_1(v_1)$$

proving that $\langle e, w\rangle = rm$. Let $\mathbf{R}^{n-1}$ denote $e^\perp$ in $\mathbf{R}^n$. This proves that there exists $\bar{w} \in \mathbf{R}^{n-1}$ with

$$z_1 \;=\; (v_1 + rme + \bar{w}, \varphi_1(v_1 + rme + \bar{w})) \qquad \text{and} \qquad z_2 \;=\; (v_2 + rme + \bar{w}, \varphi_1(v_2 + rme + \bar{w})).$$

The mapping $\bar{w} \to z_1$ with $\bar{w} \in \mathbf{R}^{n-1}$ parameterizes $\partial\mathcal{H}_1 \cap \mathrm{graph}(\varphi_1)$, and similarly $\bar{w} \to z_2$ parameterizes $\partial\mathcal{H}_2 \cap \mathrm{graph}(\varphi_2)$. ∎

Remark 5.2. Consider the closure C of $\mathrm{ch}(\mathrm{epi}(\varphi_1) \cup \mathrm{epi}(\varphi_2))$. The points in $\mathrm{epi}(\varphi_1) \cup \mathrm{epi}(\varphi_2) \sim \mathrm{SLAB}$ are extreme points of C. For each $\bar{w} \in \mathbf{R}^{n-1}$ as above, the associated hyperplane H supports C and intersects C along the line segment from z_1 to z_2.

6. Upper Semi-Continuous Functions – Radius-r Upper Contact Points.

Analogous to the fact that the subdifferential is basic for understanding convex functions, is the fact that upper contact quadratics of radius r are basic for understanding general upper semi-continuous functions. In this section u is any upper semi-continuous function on X. Let φ be a quadratic function, and note that its second derivative $D_x^2\varphi$ is independent of the point x. We say that φ has **radius** r if $D_x^2\varphi = \frac{1}{r}I$ where $0 < r < \infty$. In this case φ has a unique minimum point v which we call the **vertex point**. We say the graph of φ has its **vertex** at $(v, \varphi(v))$. The radius r and vertex v determine φ up to its **height** $\varphi(v) = c$, that is

$$\varphi(y) \;=\; c + \frac{1}{2r}|y - v|^2 \tag{6.1}$$

Now assume that X is a compact set in $\mathbf{R}^n$ and u is an upper semi-continuous function on X. For large c the graph of φ lies above the graph of u. The smallest such c is

$$\begin{aligned} \widehat{c} \;&\equiv\; \inf\left\{c : u(y) \le c + \tfrac{1}{2r}|y - v|^2 \;\forall\, y \in X\right\} \\ &=\; \sup_{y\in X}\left(u(y) - \tfrac{1}{2r}|y - v|^2\right). \end{aligned} \tag{6.2}$$

Since u is upper semi-continuous and X is compact, this supremum $\widehat{c}$ is attained at some point $x \in X$. Thus with the height $c = \widehat{c}$ defined by (6.2), the polynomial $\varphi(y)$ given by (6.1) satisfies

$$\begin{aligned} &\text{(a)}\;\; u(y) \;\le\; \varphi(y) \quad \text{for all } y \in X, \text{ and} \\ &\text{(b)}\;\; u(x) \;= \varphi(x) \quad \text{for some } x \in X. \end{aligned} \tag{6.3}$$

The contact set $\{u = \varphi\}$ is compact since X is compact. It will be denoted by

$$\mathcal{C}(u, X, \tfrac{1}{r}I, v). \tag{6.4}$$

If v belongs to the interior of X and the radius r is small, these contact points x, where graph(φ) touches graph(u), should occur in the interior of X. The basic estimate proving this is given in the following lemma. Let $\mathrm{Osc}_X(u) \equiv \sup_X u - \inf_X u$ denote the oscillation of u on X. Note that $\mathrm{Osc}_X(u) < \infty$ if and only if u is bounded below, since u is upper semi-continuous. Finite oscillation must be assumed in order for the estimate to have content.

Lemma 6.1. *If $x \in \mathcal{C}(u, X, \frac{1}{r}I, v)$, then*

$$|x - v| \;\le\; \sqrt{2r\mathrm{Osc}_X(u)}.$$

Proof. Since the graph of φ lies above the graph of u and $\varphi(x) = u(x)$, the change in φ from v to x, which equals $\varphi(x) - \varphi(v) = \frac{1}{2r}|x - v|^2$, is $\le$ the change $u(x) - u(v)$, which is $\le \mathrm{Osc}_X(u)$. ∎

This result can be put in a more useful form, guaranteeing lots of upper contact points for any upper semi-continuous function which is bounded below.

Lemma 6.1′. *Set* $\delta \equiv \sqrt{2r\mathrm{Osc}_X(u)}$ *and* $X_\delta \equiv \{y \in X : \mathrm{dist}(y, \partial X) > \delta\}$. *For any point* $v \in X_\delta$ *the contact set* $\mathcal{C}(u, X, \frac{1}{r}I, v)$ *is a non-empty compact subset of the open set* X_δ. *In fact, it is contained in the closed ball* $B_\delta(v)$ *about* v *of radius* δ.

The Upper Vertex Map

Now suppose $X \subset \mathbf{R}^n$ is open and

$$u(y) \;\leq\; u(x) + \langle p, y - x\rangle + \frac{1}{2r}|y - x|^2 \quad \forall\, y \in X. \tag{6.5}$$

By Definition 1.5 $x \in \mathcal{C}(u, X, \frac{1}{r}I)$, that is, x is a global upper contact point of type $\frac{1}{r}I$ (radius r) on X for u. We also say that $(p, \frac{1}{r}I)$ is an upper contact jet for u at x with the upper contact inequality holding on all of X.

Set

$$\varphi(y) \;\equiv\; u(x) + \langle p, y - x\rangle + \frac{1}{2r}|y - x|^2. \tag{6.6}$$

Then φ has radius r, and the vertex point v for φ is given by

$$v \;=\; x - rp, \tag{6.7}$$

since $0 = D_v\varphi = p + \frac{1}{r}(v - x)$. Furthermore, if $x \in \mathrm{Diff}^1(u)$, the set of points where u is differentiable, then the p satisfying (6.5) is unique and equal to $D_x u$.

Definition 6.3. (The Upper Vertex Map). The map

$$V : \mathcal{C}(X, u, \tfrac{1}{r}I) \cap \mathrm{Diff}^1(u) \;\longrightarrow\; \mathbf{R}^n$$

defined by $V(x) \equiv x - rD_x u$, i.e., $V \equiv I - rDu$, will be called the **vertex map for** u.

This map has the property that

$$v \;=\; V(x) \qquad \Rightarrow \qquad |v - x| \;\leq\; \sqrt{2r\mathrm{Osc}_X(u)} \tag{6.8}$$

by the basic estimate Lemma 6.1.

Convex Functions

The constructions of the previous section apply to a convex function u. By (D at UCP) we have $\mathcal{C}(u, X, \frac{1}{r}I) \subset \mathrm{Diff}^1(u)$.

Therefore, the vertex map $V \equiv I - rDu$ is a well defined map

$$V : \mathcal{C}(u, X, \tfrac{1}{r}I) \;\longrightarrow\; \mathbf{R}^n. \tag{6.9}$$

7. The Vertex Map is a Contraction.

Proposition 7.1. *Given a convex function u defined on an open convex set $X \subset \mathbf{R}^n$, the vertex map $V : \mathcal{C}(u, X, \frac{1}{r}I)) \to \mathbf{R}^n$ is a contraction.*

Remark 7.2. If u is smooth, then the Jacobian J of V is $I - rD^2u$ and $0 \le J \le I$ on $\mathcal{C}(u, X, \frac{1}{r}I))$ is obvious.

Proof. Given $x_1, x_2 \in \mathcal{C}(u, X, \frac{1}{r}I))$ we must show that

$$|V(x_2) - V(x_1)| \ \le \ |x_2 - x_1|. \tag{7.1}$$

Let $\varphi_1(y)$ denote the quadratic of radius r whose graph lies above graph(u) and touches at $(x_1, u(x_1))$, i.e., $\varphi_1(x_1) = u(x_1)$, and define φ_2 similarly. Let $v_1 \equiv V(x_1)$ and $v_2 \equiv V(x_2)$ denote the vertex points.

Since u is convex and $\varphi_k \ge u$, $k = 1, 2$, we have

$$\text{ch}\,(\text{epi}(\varphi_1) \cup \text{epi}(\varphi_2)) \ \subset \ \text{epi}(u).$$

Now $(x_1, \varphi_1(x_1)) = (x_1, u(x_1)) \notin \text{epi}(u)$ (recall that epi(u) is the *open* epigraph). Hence, $(x_1, \varphi_1(x_1)) \notin \text{ch}\,(\text{epi}(\varphi_1) \cup \text{epi}(\varphi_2))$, and therefore $(x_1, \varphi_1(x_1)) \notin \mathcal{H}_1$ by Lemma 5.1. Similarly, $(x_2, \varphi_1(x_2)) \notin \mathcal{H}_2$. We conclude that these points lie on opposite sides of SLAB and so $|x_2 - x_1| \ge \text{width(SLAB)} = |v_2 - v_1|$. ∎

8. Completion of the Proof of Slodkowski's Lemma.

The fact that the vertex map is a contraction combined with a standard perturbation argument is all that is needed to prove Slodkowski's Lemma 4.1.

We may assume that $x = 0$ and $u(x) = 0$. Then the assumption on the convex function u which occurs in Slodkowski's Lemma is:

$$(0, \lambda I) \ \text{ is a strict upper contact jet for } u \text{ at } x. \tag{8.1}$$

With $\lambda = 1/R$ we claim that this is equivalent to the existence of $\bar{\rho} > 0$ such that:

$$0 \ \le \ u(y) \ < \ \frac{1}{2R}|y|^2 \qquad \text{for } 0 < |y| \le \bar{\rho}. \tag{8.1}'$$

Assuming (8.1), the function u is differentiable at $x_0 = 0$ and $D_{x_0}u = 0$ by (D at UCP). The convexity of u then implies that $0 \le u(y)$ (cf.(2.5)), showing that (8.1) $\Rightarrow$ (8.1)′. That (8.1)′ $\Rightarrow$ (8.1) is clear.

Now we apply Lemma 6.1′ to u with $X \equiv \overline{B}_\rho$ where $\rho \le \bar{\rho}$ and $r \le R$. Since $u(0) = 0$ and $u(y) \ge 0$ on $\overline{B}_\rho$, we have $\text{Osc}_{\text{X}} = \sup_{\overline{\text{B}}_\rho} \text{u}$, which we denote by $M(\rho)$ so that $\delta(\rho) \equiv \sqrt{2rM(\rho)}$. Now $X(\delta) = B_{\rho-\delta}$. According to Lemma 6.1′, if $v \in B_{\rho-\delta}$, then $\mathcal{C}(u, B_\rho, \frac{1}{r}I, v)$ is a non-empty subset of $B_{\rho-\delta}$. Choose such an upper contact point x. Then, as noted above, since u is convex, the vertex map V is defined at x. Thus for all

$v \in B_{\rho-\delta}$, we have that $v = V(x)$ for some $x \in \mathcal{C}(u, \overline{B}_\rho, \frac{1}{r}I)$, i.e., $\overline{B}_{\rho-\delta} \subset V(\mathcal{C}(u, \overline{B}_\rho, \frac{1}{r}I))$. Since V is a contraction, this proves that:

$$|\overline{B}_{\rho-\delta}| \;\leq\; |V(\mathcal{C}(u, \overline{B}_\rho, \tfrac{1}{r}I))| \;\leq\; |\mathcal{C}(u, \overline{B}_\rho, \tfrac{1}{r}I)|. \tag{8.2}$$

Since $M(\rho) = \sup_{\partial B_\rho} u$ by the maximum principle, (8.1)$'$ implies that $M(\rho) < \rho^2/(2R)$, and hence, $\delta(\rho) < \rho\sqrt{\frac{r}{R}}$. This proves that

$$\rho - \delta(\rho) \;>\; \rho\left(1 - \sqrt{\tfrac{r}{R}}\right). \tag{8.3}$$

Taking $r = R$, this completes the proof of Lemma 4.1, since $\rho - \delta(\rho) > 0$. ∎

Our proof of this lemma, which uses paraboloids instead of spheres, also provides an analogue of Slodkowski's Lemma 3.4 which estimates the density of the contact set, using (8.3).

Proposition 8.1. *If u is a convex function on $B_{\overline{\rho}}$ satisfying $0 \leq u(y) < \frac{|y|^2}{2R}$ for $y \neq 0$. Then for $0 < r < R$,*

$$|B_\rho|\left(1 - \sqrt{\tfrac{r}{R}}\right)^n \;\leq\; \left|\mathcal{C}\left(u, \overline{B}_\rho, \tfrac{1}{r}I\right)\right| \qquad \forall\, 0 < \rho \leq \overline{\rho}.$$

9. The Equivalence of Slodkowski's Lemma and Jensen's Lemma.

The results of this section are not needed elsewhere in these notes, but they might have some historical interest.

Another special case of the general Slodkowski-Jensen Lemma 1.7 is Jensen's Lemma.

Lemma 9.1. (Jensen). *Suppose that w is a quasi-convex function with the strict upper contact jet $(0,0)$ at x (equivalently, w has a strict local maximum at x). Then there exists $\bar{\rho} > 0$ such that*

$$|\mathcal{C}(w, B_\rho, 0)| \;>\; 0 \qquad \forall\, 0 < \rho \leq \bar{\rho}.$$

In a very strong sense the Slodkowski Lemma is equivalent to Jensen$'$s Lemma. (9.1)

The precise statement (9.1) is embedded in the following proof.

Proof. Set

$$u(y) \;\equiv\; w(y) + \frac{\lambda}{2}|y - x|^2. \tag{9.2}$$

Then by definition

$$u \text{ is convex} \qquad \Longleftrightarrow \qquad w \text{ is } \lambda\text{-quasi-convex.}$$

By Lemma 4.2 parts (1) and (2)

$$\begin{aligned} &(0, \lambda I) \;\text{ is a strict upper contact jet for } u \text{ at } x \qquad \Longleftrightarrow \\ &(0,0) \;\text{ is a strict upper contact jet for } w \text{ at } x, \end{aligned} \tag{9.3}$$

while part (3) state that

$$\mathcal{C}(u, B_\rho, \lambda I) \;=\; \mathcal{C}(w, B_\rho, 0). \tag{9.4}$$

Thus by (9.3) the hypotheses of Slodkowski and Jensen are equivalent, while by (9.4) the conclusions of Slodkowski and Jensen are identical (not just equivalent).

10. The Proof of Alexandrov's Theorem.

We will evoke two local results about Lipschitz maps $G : \mathbf{R}^n \to \mathbf{R}^n$ which can be found many places (e.g. [F]). Otherwise the proof is elementary and complete. It combines elements of the proofs in [CIL] and [AA] with the Legendre transform. It is also worth noting that while upper contact quadratics of radius r were key to the proof of Slodkowski's Lemma, lower contact points of radius r are the key to proving Alexandrov's Theorem.

Rademacher's Theorem. *The derivative of G exists almost everywhere.*

It is convenient to lable the variables as $x = G(y)$. We will say that y *is a critical point for* G if G is differentiable at y and D_yG is singular. The image of the set of all critical points under the mapping G is the *set of critical values of* G.

The Lipschitz Version of Sard's Theorem. *The set of critical values of G has measure zero.*

Since convex functions are locally Lipschitz (see Lemma 2.1), the scalar version of Rademacher's Theorem implies that

$$\text{A convex function is differentiable almost everywhere.} \tag{10.1}$$

This fact will also be used in the proof.

Now we begin the proof of Alexandrov's Theorem. Given a convex function u on a convex open set X with upper bound N and $r > 0$, we will show that

$$f(x) \;\equiv\; ru(x) + \tfrac{1}{2}|x|^2$$

is twice differentiable a.e. on $X_\delta \equiv \{x \in X : \text{dist}(x, \partial X) > \delta\}$ where $\delta \equiv 4\sqrt{r|u|_\infty}$ and $|u|_\infty \equiv \sup_X |u|$. First note that by Lemma 2.5

$$\partial f(x) \;=\; x + r\partial u(x), \quad \text{i.e.,} \quad \partial f \;=\; I + r\partial u. \tag{10.2}$$

That is, with $x \in X$,

$$\text{If } y \text{ and } p \text{ are related by } y = x + rp, \quad \text{then} \quad (x, y) \in \partial f \quad \Longleftrightarrow \quad (x, p) \in \partial u. \tag{10.3}$$

Lemma 10.1. *The multi-valued map ∂f is expansive. That is, if $(x_1, y_1) \in \partial f$ and $(x_2, y_2) \in \partial f$, then*

$$|x_1 - x_2| \;\le |y_1 - y_2|.$$

Proof. Note that $|y_1-y_2||x_1-x_2| \geq \langle y_1-y_2, x_1-x_2\rangle = |x_1-x_2|^2 + r\langle p_1-p_2, x_1-x_2\rangle$ with p_1 and p_2 defined by (10.3) so that $p_1 \in \partial u(x_1)$ and $p_2 \in \partial u(x_2)$. By the monotonicity (2.4) of ∂u this expression is $\geq |x_1-x_2|^2$. ∎

Because of this inequality, if $(x_1, y), (x_2, y) \in \partial f$, then $x_1 = x_2$. Thus the inverse of $F \equiv \partial f$ is single-valued. We denote this single-valued mapping, which is defined on the set

$$Y \;\equiv\; \mathrm{Im}F \;\equiv\; \text{the projection of } \partial f \text{ onto the second factor of } \mathbf{R}^n \times \mathbf{R}^n,$$

by $x = G(y)$. Now Lemma 10.1 states that

$$|G(y_1) - G(y_2)| \;\leq\; |y_1 - y_2| \qquad \text{for} \;\; y_1, y_2 \in Y. \tag{10.4}$$

That is, *G is 1-Lipschitz* or *contractive*.

Lemma 10.2.

$$X_\delta \;\subset\; \mathrm{Dom(G)} \;\equiv\; Y \quad \text{where} \;\; \delta = 2\sqrt{r|u|_\infty}. \tag{10.5}$$

Proof. Given $y \in X_\delta$ pick a minimum point x for the function $ru(z) + \frac{1}{2}|z-y|^2$ on the ball $B_\delta(y) \equiv \{x : |x-y| \leq \delta\} \subset X$. This is also a minimum for the same function on all of X. To see this consider any $\bar{x} \in X$ with $|y-\bar{x}| > \delta$. Then $u(\bar{x}) - u(y) + \frac{1}{2r}|\bar{x}-y|^2 > -2|u|_\infty + \frac{\delta^2}{2r} = 0$ by the definition of δ. This implies that the value of $ru(z) + \frac{1}{2}|z-y|^2$ at $\bar{x}$ is greater than the value at y, and therefore greater than the value at x.

The fact that x is a minimum point of $ru(z) + \frac{1}{2}|z-y|^2$ on X implies that $0 \in \partial(ru(z) + \frac{1}{2}|z-y|^2)(x)$. By Lemma 2.5 there exists $p \in \partial u(x)$ with $0 = rp + (x-y)$. By (10.3) above, $y \in \partial f(x)$. ∎

The Legendre Transform

The map G is the inverse of the multi-valued map $F = \partial f$. This map G also has a scalar potential g, with $G = \partial G$, which is classically called the Legendre transform of f. Let $Y = \mathrm{Im}F$ as above an note that with $(x, y) \in X \times Y$

$$\begin{aligned}
y \in \partial f(x) \qquad &\Longleftrightarrow \qquad f(x) + \langle y, z-x\rangle \leq f(z) \quad \forall\, z \in X \\
&\Longleftrightarrow \qquad f(x) - \langle y, x\rangle \leq f(z) - \langle y, z\rangle \quad \forall\, z \in X \\
&\Longleftrightarrow \qquad f(z) - \langle y, z\rangle \;\; \text{has a minimum point at} \;\; z = x.
\end{aligned}$$

Define $-g(y)$ to be the minimum value, so that:

$$f(x) + g(y) \;=\; \langle x, y\rangle \qquad \forall\,(x, y) \in \partial f. \tag{10.6}$$

Now

$$\begin{aligned}
x \in \partial g(y) \qquad &\Longleftrightarrow \qquad g(y) + \langle x, w-y\rangle \leq g(w) \quad \forall\, w \in Y \\
&\Longleftrightarrow \qquad g(y) - \langle x, y\rangle \leq g(w) - \langle x, w\rangle \quad \forall\, w \in Y \\
&\Longleftrightarrow \qquad g(w) - \langle x, w\rangle \;\; \text{has a minimum point at} \;\; w = y.
\end{aligned}$$

The minimum value is $g(y) - \langle x, y\rangle$, which equals $f(x)$. This is the proof of the classical fact that the Legendre transform is an involution.

Summarizing, if f and g correspond under the Legendre transform, then

$$y \in \partial f \qquad \Longleftrightarrow \qquad x \in \partial g(y). \tag{10.7}$$

That is, the multi-valued maps $F = \partial f$ and $G = \partial g$ are inverses of each other.

Note that g convex, since $g(y)$ is the supremum of the family $\langle y, x\rangle - f(x)$ of affine functions of y.

The Legendre transform g of f enjoys nicer properties than the convex function $f(x) \equiv ru(x) + \frac{1}{2}|x|^2$. We state more than required.

Lemma 10.3. *The Legendre transform g of f is a convex C^1 function on X_δ with derivative $Dg = G$. If G is differentiable at y with first derivative $D_y G = B$, then g is twice differentiable at y with second derivative $D_y^2 g = B$.*

Proof. Since $\partial g = G$ and G is single valued, we have $g \in C^1$ by Corollary 2.3. This enables us to apply the standard Mean Value Theorem. Now assume that G is differentiable at y_0 with $D_{y_0} G = B$. We can assume $y_0 = 0$, $g(y_0) = 0$, and $G(y_0) \equiv D_{y_0} g = 0$, so that

$$G(y) - By \;=\; o(|y|). \tag{10.8}$$

By the Mean Value Theorem applied to the function $\phi(y) \equiv g(y) - \frac{1}{2}\langle By, y\rangle$, there exists $\xi \in [0, y]$ such that

$$g(y) - \tfrac{1}{2}\langle By, y\rangle \;=\; \phi(y) \;=\; \langle D_\xi \phi, y\rangle \;=\; \langle D_\xi g - B\xi, y\rangle. \tag{10.9}$$

By (10.8)

$$D_\xi g - B\xi \;=\; G(\xi) - B(\xi) \;=\; o(|\xi|) \;=\; o(|y|).$$

(This last equality is because $|\xi| \le |y|$.) Therefore, the right hand side of (10.9) is $o(|y|^2)$. This proves $\phi(y) = o(|y|^2)$ and hence g is twice differentiable at 0 with $D_0^2 g = B$. ∎

Now we are ready to prove the analogous lemma for f.

Lemma 10.4. *Suppose that G is differentiable at $y_0 \in X_\delta$ and let $B \equiv D_{y_0} G$ denote the derivative. Assume that $x_0 = G(y_0)$ is not a critical value of G. Further assume that the convex function f is differentiable at x_0, and hence $D_{x_0} f = y_0$. Then the function f is twice differentiable at x_0 with second derivative $D_{x_0}^2 f = B^{-1}$.*

Proof. We can assume that $x_0 = 0$ and that $f(0) = D_0 f = 0$, by modifying f by an affine function. Since f is differentiable at 0, the subdifferential $\partial f(0) = \{y_0\} = \{D_0 f\} = \{0\}$ by Lemma 2.2. Since $y_0 = 0$ is not a critical point of G, the derivative $D_0 G \equiv B$ is invertible. Let $A \equiv B^{-1}$. We must show that:

$$f(x) - \tfrac{1}{2}\langle Ax, x\rangle \;=\; o\left(|x|^2\right). \tag{10.10}$$

For $(x, y) \in \partial f$ the identity $f(x) + g(y) \;=\; \langle x, y\rangle$ can be written as

$$f(x) - \tfrac{1}{2}\langle Ax, x\rangle \;=\; \tfrac{1}{2}\langle By, y\rangle - g(y) + \tfrac{1}{2}\langle y - Ax, x\rangle + \tfrac{1}{2}\langle x - By, y\rangle. \tag{10.11}$$

We have by (10.8) that

$$x - By \;=\; G(y) - By \;=\; o(|y|). \tag{10.8$'$}$$

Since $D^2_{y_0} g = B$, we have $g(y) - \frac{1}{2}\langle By, y\rangle = o(|y|^2)$. The remaining term of the RHS of (10.11) is also $o(|y|^2)$ since $|x| \le |y|$ and

$$y - Ax \;=\; -A(x - By) \;=\; o(\|A\||y|) \;=\; o(|y|). \tag{10.12}$$

This proves that, with $x = G(y)$,

$$f(x) - \tfrac{1}{2}\langle Ax, x\rangle \;=\; o\left(|y|^2\right). \tag{10.13}$$

Finally for $(x, y) \in \partial f$, $|y| = |ABy| \le \|A\|(|x - By| + |x|)$. By (10.8)$'$ we have $|x - By| = o(|y|)$, so this proves $|y| = O(|x|)$. ∎

Remark . Note that $|y| = O(|x|)$ combined with (10.12) yields $y - Ax = o(|x|)$. This is the statement that

$$\lim_{\substack{x \to x_0 \\ y \in \partial f(x)}} \frac{y - y_0 - A(x - x_0)}{|x - x_0|} \;=\; 0, \tag{10.14}$$

which says that the multi-valued function ∂f is differentiable at x_0 with derivative A.

Let D denote the set of points in $G(X_\delta)$ where the convex function f is differentiable. Let N denote the set of points in X_δ where G is not differentiable. Let C denote the set of critical points for G in X_δ (where G is differentiable but the derivative of G is singular). Lemma 10.4 applies to each point $x_0 = G(y_0) \in D$ with $y_0 \notin N \cup C$.

Lemma 10.5. *The set $D - G(N \cup C)$ has full measure in $G(X_\delta)$.*

Proof. As noted in (10.1) D has full measure. By Rademacher's Theorem N has measure zero. Since G is contractive, $G(N)$ also has measure zero. Finally, by the Lipschitz version of Sard's Theorem $G(C)$ has measure zero. ∎

This proves that f is twice differentiable on a set of full measure in $G(X_\delta)$. Since $X_{2\delta} \subset G(X_\delta)$, the proof of Alexandrov's Theorem is complete. ∎

Appendix A. A Quasi-Convexity Characterization of $C^{1,1}$.

It is interesting that the condition that a function be $C^{1,1}$ is directly related to quasi-convexity, in fact it is equivalent to the function being simultaneously quasi-convex and quasi-concave. This was probably first observed by Hiriart-Urruty and Plazanet in [HP]. An alternate proof appeared in [E]. For the benefit of the reader we include a proof here.

THEOREM A.1.

$$u \text{ is } \lambda - C^{1,1} \qquad \Longleftrightarrow \qquad \text{both} \quad \pm u \text{ are } \lambda - \text{quasi} - \text{convex}$$

Proof. ($\Rightarrow$) Suppose that u is λ-C^{1-1}, i.e., $u \in C^1$ and $|D_x u - D_y u| \leq \lambda |x-y|$ for all x, y. Set $f \equiv u + \frac{\lambda}{2}|x|^2$. Then

$$D_x f - D_y f \;=\; \lambda(x-y) + D_x u - D_y u,$$

and hence

$$\begin{aligned} \langle D_x f - D_y f, x-y\rangle &= \lambda|x-y|^2 + \langle D_x u - D_y u, x-y\rangle \\ &\geq \lambda|x-y|^2 - |D_x u - D_y u||x-y| \\ &= (\lambda|x-y| - |D_x u - D_y u|)|x-y| \;\geq\; 0. \end{aligned}$$

This form of monotonicity of Df is one of the standard definitions of f being convex. The same proof works for $-u$

($\Leftarrow$) We state the converse as a proposition.

Proposition A.2. *If u and $-u$ are λ-quasi-convex, then $u \in C^1$ and*

$$|D_x u - D_y u| \;\leq\; \lambda|x-y|, \quad \text{i.e., } u \text{ is } \lambda - C^{1,1}.$$

Proof. We first show that this is true if $u \in C^\infty$. Note that $\pm u$ are λ-quasi-convex $\Longleftrightarrow$ $D_x^2 u + \lambda I \geq 0$ and $-D_x^2 u + \lambda I \geq 0$ for all x $\Longleftrightarrow$ $-\lambda I \leq D_x^2 u \leq \lambda I$ for all x. By the Mean Value Theorem, $D_x u - D_y u = (D_\xi^2 u)(x-y)$ for some $\xi \in [x, y]$, and hence $|D_x u - D_y u| \leq \lambda |x-y|$.

In general, since the graph of $u + \frac{\lambda}{2}|x|^2$ has a supporting hyperplane from below and the graph of $u - \frac{\lambda}{2}|x|^2$ has a supporting hyperplane from above, at every point, the function u is differentiable everywhere. By partial continuity of the first derivative for quasi-convex functions (Lemma 1.3), we have $u \in C^1$.

Now standard convolution $u^\epsilon \equiv u * \varphi_\epsilon$ works just fine to complete the proof since $\pm u^\epsilon$ is λ-quasi-convex by the next lemma, and the fact that $u \in C^1 \;\Rightarrow\; Du^\epsilon \to Du$ locally uniformly.

Lemma A.3. *u is λ-quasi-convex $\;\Rightarrow\; u^\epsilon \equiv u * \varphi_\epsilon$ is λ-quasi-convex.*

Proof. Suppose u is λ-quasi-convex, i.e., $f \equiv u + \frac{\lambda}{2}|x|^2$ is convex. Standard convolution of f with an approximate identity φ_ϵ based on φ (i.e., $\varphi_\epsilon(x) \equiv \frac{1}{\epsilon^n}\varphi(\frac{x}{\epsilon})$) yields $f^\epsilon \equiv f * \varphi_\epsilon$ smooth and convex. Note that $(|x|^2 * \varphi_\epsilon) = |x|^2 + \langle a, x\rangle + c$ preserves $|x|^2$ modulo an affine

function, since $\int |x+\epsilon y|^2\varphi(y)\,dy = |x|^2 + \epsilon\langle a,x\rangle + C\epsilon^2$ where $\langle a,x\rangle = 2\int\langle x,y\rangle\varphi(y)\,dy$ and $C = \int |y|^2\varphi(y)\,dy$. Therefore, $D^2 f^\epsilon = D^2 u^\epsilon + \lambda I$, proving that each u^ϵ is λ-quasi-convex.■

References.

[AA] G. Alberti, and L. Ambrosio, *A geometrical approach to monotone functions in* $\mathbf{R}^n$, Math. Z. **230** (1999), no. 2, 259-316.

[Al] A. D. Alexandrov, *Almost everywhere existence of the second differential of a convex function and properties of convex surfaces connected with it (in Russian)*, Lenningrad State Univ. Ann. Math. **37** (1939), 3-35.

[CIL] M. G. Crandall, H. Ishii and P. L. Lions *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. (N. S.) **27** (1992), 1-67.

[E] A. Eberhard, *Prox-regularity and subjets*, in: A. Rubinov (Ed.), Optimization and Related Topics, Applied Optimization Volumes, Kluwer Academic Publishers, Dordrecht, 2001, pp. 237313.

[F] H. Federer, Geometric Measure Theory, Springer-Verlag, Berlin-Heidelberg, 1969.

[HL_1] F. R. Harvey and H. B. Lawson, Jr, *Dirichlet duality and the non-linear Dirichlet problem*, Comm. on Pure and Applied Math. **62** (2009), 396-443. ArXiv:math.0710.3991

[HL_2] F. R. Harvey and H. B. Lawson, Jr., *The AE-Theorem and addition theorems for quasi-convex functions*. ArXiv:1309:1770.

[HP] J.-B. Hiriart-Urruty, Ph. Plazanet, *Moreaus Theorem revisited*, Analyse Nonlinéaire (Perpignan, 1987), Ann. Inst. H. Poincaré **6** (1989), 325338.

[S] Z. Slodkowski, *The Bremermann-Dirichlet problem for q-plurisubharmonic functions*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **11** (1984), 303-326.

[J] R. Jensen, *The maximum principle for viscosity solutions of fully nonlinear second order partial differential equations*, Arch. Ratl. Mech. Anal. **101** (1988), 1-27.