Guided proofs of Rademacher and Alexandrov Theorems

Theorem 1 (Rademacher)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a locally Lipschitz continuous function. Then f is differentiable $\mathscr{L}^d$-almost everywhere.*

Sketch of the proof of Theorem 1.

1. Use Riesz representation theorem (from Functional Analysis) to prove that there exists a *weak gradient* of f, namely there exists an L^∞_{loc}-function $\tilde{\nabla} f = (\tilde{\partial}_1 f, \dots, \tilde{\partial}_d f) : \mathbb{R}^d \to \mathbb{R}^d$ such that
$$\int_{\mathbb{R}^d} f \; \partial_i g \; dx = -\int_{\mathbb{R}^d} \tilde{\partial}_i f \; g \; dx$$
for all $g \in C^\infty_c(\mathbb{R}^d, \mathbb{R})$ and for all $i = 1, \dots, d$. Here with $\partial_i g$ we denote the classical derivative of g with respect to the coordinate x_i.

   To do this, use that $\partial_i g$ is the limit of incremental ratios.

2. Show that, for each $x_0 \in \mathbb{R}^d$, and each $r_j \downarrow 0$, the sequence of functions
$$f_{r_j}(y) = \frac{f(x_0 + r_j y) - f(x_0)}{r_j}$$
admits a uniformly convergent subsequence to a function f_0 in $\overline{B(0,1)}$.

3. Let $x_0 \in \mathbb{R}^d$ be a Lebesgue point for $\tilde{\nabla} f$. That is, a point such that
$$\frac{1}{\mathscr{L}^d(B_r(x_0))} \int_{B_r(x_0)} |\tilde{\nabla} f - \tilde{\nabla} f(x_0)| \; dx \to 0 \quad \text{as } r \to 0.$$
Show that the weak gradients $\tilde{\nabla} f_{r_j}$ converge to $\tilde{\nabla} f(x_0)$ in $L^1(B(0,1), \mathbb{R}^d)$ as $r_j \downarrow 0$.

4. Show that $\nabla f_0 \equiv \tilde{\nabla} f(x_0)$ in $B(0,1)$ and deduce that f is differentiable at x_0, its classical gradient being exactly $\tilde{\nabla} f(x_0)$. Conclude by the Lebesgue differentiation theorem.

$\square$

Theorem 2 (Alexandrov)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then f is twice differentiable $\mathscr{L}^d$-almost everywhere.*

To prove Theorem 2 we first need a couple of lemmas. The first is about a mean-value property of convex functions. The second regards uniformly convex (λ-convex) functions.

Lemma 3 (Non-smooth mean value theorem)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then, for every $x, y \in \mathbb{R}^d$, there exist z in the closed segment with endpoints x, y, and $\xi \in \partial f(z)$, such that*

$$f(y) - f(x) = \langle \xi, y - x \rangle.$$

Hint: Approximate f by convolution with smooth convex functions. Then use the stability of subdifferentials under uniform convergence.

Definition 4 (λ-convexity)**.** *A function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be λ-convex, for some $\lambda \in \mathbb{R}$, if $f - (\lambda/2)|x|^2$ is convex.*

Lemma 5 (Properties of λ-convex functions)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be λ-convex for some $\lambda \geq 0$ (in particular, f is convex). Then the following hold.*

i) For every $x, y \in \mathbb{R}^d$ and every $t \in [0, 1]$:

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) - \frac{\lambda}{2}t(1-t)|y - x|^2.$$

ii) The subdifferential of f at x can be written as

$$\partial f(x) = \left\{ \xi \in \mathbb{R}^d : f(y) \geq f(x) + \langle \xi, y - x \rangle + \frac{\lambda}{2}|y - x|^2 \text{ for every } y \in \mathbb{R}^d \right\}.$$

iii) For every $x, y \in \mathbb{R}^d$ and every $\xi \in \partial f(x), \eta \in \partial f(y)$ we have

$$\langle \eta - \xi, y - x \rangle \geq \lambda |y - x|^2.$$

iv) If $\lambda > 0$, then for every $\xi \in \mathbb{R}^d$, there exists a unique point $x =: \Psi(\xi) \in \mathbb{R}^d$ such that $\xi \in \partial f(x)$. Moreover, $\Psi : \mathbb{R}^d \to \mathbb{R}^d$ is λ^{-1}-Lipschitz.

Hint: Points i),ii) and iii) are direct consequences of the definition of λ-convexity and the corresponding properties of convex functions. Regarding point iv), to find x such that $\xi \in \partial f(x)$ it is enough to minimize the function $y \mapsto f(y) - \langle \xi, y \rangle$. To prove that such $x =: \Psi(\xi)$ is unique and that Ψ is λ^{-1}-Lipschitz, use point iii).

Sketch of the proof of Theorem 2.

1. First of all observe that we may assume without loss of generality that f is λ-convex, for some $\lambda > 0$. Then, thanks to point iv) in Lemma 5 it is well-defined the "inverse" of the subdifferential $\Psi : \mathbb{R}^d \to \mathbb{R}^d$ as the λ^{-1}-Lipschitz map which associates each point $\xi \in \mathbb{R}^d$ with the unique $\Psi(\xi) = x \in \mathbb{R}^d$ for which $\xi \in \partial f(x)$.

2. Define the set

$$\Sigma := \left\{ \xi \in \mathbb{R}^d : \text{either } \Psi \text{ is not differentiable at } \xi \text{ or it is but } \det D\Psi(\xi) = 0 \right\}.$$

   Use Theorem 1 and the area formula to prove that

$$\mathscr{L}^d(\Psi(\Sigma)) = 0.$$

   Define the set

$$\Omega := \left\{ x \in \mathbb{R}^d : f \text{ is differentiable at } x \right\} \setminus \Psi(\Sigma).$$

   Observe that by Theorem 1 again (remember that convex functions are locally Lipschitz),

$$\mathscr{L}^d(\mathbb{R}^d \setminus \Omega) = 0.$$

3. Given $x \in \Omega$, take $p = \nabla f(x)$, and notice that $\Psi(p) = x$, Ψ is differentiable at p and $\det D\Psi(p) \neq 0$. Hence it is well-defined the matrix

$$S(x) := D\Psi(p)^{-1}.$$

   Notice that since Ψ is λ^{-1}-convex, $|S(x)| \leq \lambda$. Prove that

$$\lim_{\substack{y \to x \\ \eta \in \partial f(y)}} \frac{|\eta - \nabla f(x) - S(x)(y - x)|}{|y - x|} = 0. \tag{1}$$

4. Show that the function

$$g(y) := f(y) - f(x) - \langle \nabla f(x), y - x \rangle - \frac{1}{2} \langle S(x)(y - x), (y - x) \rangle$$

   is convex. Notice that (1) can be rephrased as

$$\lim_{\substack{y \to x \\ \eta \in \partial g(y)}} \frac{|\eta|}{|y - x|} = 0.$$

   Finally, use Lemma 3 on g to conclude that

$$\lim_{y \to x} \frac{\left| f(y) - f(x) - \langle \nabla f(x), y - x \rangle - \frac{1}{2} \langle S(x)(y - x), (y - x) \rangle \right|}{|y - x|^2} = 0.$$

$\square$