Ec 181 KC Border
Convex Analysis and Economic Theory AY 2019–2020

Topic 18: Differentiability

18.1 Differentiable functions

In this section I want to introduce progressively stronger notions of derivatives and differentiability for functions between real vector spaces. The real line is considered to be a one-dimensional vector space. Recall that a derivative is some kind of limit of line segments joining points on the graph of a function. The simplest way to take such a limit is along a line segment containing x. We start by recalling that the notation

$$\lim_{x \to x_0} f(x) = y$$

means that for every neighborhood U of y, there is some neighborhood V of x_0 such that if $x \neq x_0$ and $x \in V$, then $f(x) \in U$. The reason we restrict attention to $x \neq x_0$ is so that when $x - x_0$ is a number or belongs to a normed vector space, we may divide by $|x - x_0|$ or $\|x - x_0\|$.

18.1.1 Definition (One-sided directional derivative) *Let A be a subset of the vector space X, let Y be a topological vector space, and let $f\colon A \to Y$.*

*We say that f has the **one-sided directional derivative** $f'(x;v)$ at x in the direction v, if $f'(x;v)$ is a vector in Y satisfying*

$$f'(x;v) = \lim_{\lambda \downarrow 0} \frac{f(x + \lambda v) - f(x)}{\lambda}.$$

In order for this definition to make sense, we implicitly require that there is some $\varepsilon > 0$ such that $0 \leqslant \lambda \leqslant \varepsilon$ implies that $x + \lambda v$ belongs to A, so that $f(x + \lambda v)$ is defined.

For the case $Y = \boldsymbol{R}$, we also permit f to assume one of the extended values $\pm\infty$, and also permit $f'(x;v)$ to assume one of the values $\pm\infty$.

Note that in the definition of $f'(x,v)$, the neighborhoods are taken in Y and in $\boldsymbol{R}$, so a topology is needed on Y, but none is necessary on X. Also note that $x + \lambda v$ need not belong to A for $\lambda < 0$. Considering $\lambda = 0$ implies $x \in A$. The next lemma shows that the set of v for which a one-sided directional derivative exists is a cone, and that $f'(x;v)$ is positively homogeneous in v on this cone.

18.1.2 Lemma *The one-sided directional derivative is positively homogeneous of degree one. That is, if $f'(x;v)$ exists, then*

$$f'(x;\alpha v) = \alpha f'(x;v) \quad \text{for } \alpha \geqslant 0.$$

Proof: This follows from $\frac{f(x+\lambda\alpha v)-f(x)}{\lambda} = \alpha\frac{f(x+\beta v)-f(x)}{\beta}$, where $\beta = \lambda\alpha$, and letting $\lambda, \beta \downarrow 0$. ∎

18.1.3 Definition *If $f'(x;v) = -f'(x;-v)$, then we denote the common value by $D_v f(x)$, that is,*

$$D_v f(x) = \lim_{\lambda\to 0}\frac{f(x+\lambda v)-f(x)}{\lambda},$$

and we say that f has ***directional derivative*** *$D_v f(x)$ at x in the direction v.*

It follows from Lemma 18.1.2 that if $D_v(x)$ exists, then $D_{\alpha v}(x) = \alpha D_v(x)$ for all α. In $\boldsymbol{R}^{\mathrm{n}}$, the i^{th} **partial derivative** of f at x, if it exists, is the directional derivative in the direction e^i, the i^{th} unit coordinate vector.

Note that this definition still uses no topology on X. This generality may seem like a good thing, but it has the side effect that since it does not depend on the topology of X, it cannot guarantee the continuity of f at x in the normed case. That is, f may have directional derivatives in all nonzero directions at x, yet not be continuous at x. Moreover, we may not be able to express directional derivatives as a linear combination of partial derivatives.

18.1.4 Example (Directional derivatives w/o continuity or linearity) Let $f\colon \boldsymbol{R}^2 \to \boldsymbol{R}$ via

$$f(x,y) = \begin{cases} \dfrac{xy}{x^2+y} & y \neq -x^2 \\ 0 & y = -x^2. \end{cases}$$

Observe that f has directional derivatives at $(0,0)$ in every direction $v = (x,y)$, as

$$\frac{f(\lambda x,\lambda y)-f(0,0)}{\lambda} = \frac{\left(\dfrac{\lambda^2 xy}{\lambda^2x^2+\lambda y}\right)}{\lambda} = \frac{xy}{\lambda x^2+y}.$$

If $y \neq 0$, then the limit of this expression is x as $\lambda \to 0$, and if $y = 0$, the limit is 0. Thus the directional derivative exists for every direction $(x,y) \neq (0,0)$, but it is not continuous at the x-axis.

But f is not continuous at $(0,0)$. For instance, for $\varepsilon > 0$,

$$f(\varepsilon, -\varepsilon^2-\varepsilon^4) = \frac{-\varepsilon(\varepsilon^2+\varepsilon^4)}{\varepsilon^2-\varepsilon^2-\varepsilon^4} = \frac{1}{\varepsilon}+\varepsilon \to \infty \text{ as } \varepsilon \to 0.$$

Note too that the mapping $v \mapsto D_v f(0)$ is not linear. □

18.1.5 Definition (The Gâteaux derivative) *Let X and Y be normed vector spaces. If $D_v f(x)$ exists for all $v \in X$ and the mapping $T\colon v \mapsto D_v f(x)$ from X to Y is a continuous linear mapping, then T is called the* ***Gâteaux derivative*** *or* ***Gâteaux differential*** *of f at x,*[1] *and we say that f is* ***Gâteaux differentiable*** *at x.*

[1] This terminology disagrees with Luenberger [14, p. 171], who does not require linearity. It is however, the terminology used by Aubin [3, Definition 1, p. 111] , Aubin and Ekeland [4, p. 33], and Ekeland and Temam [6, Definition 5.2, p. 23].

This notion uses the topology on X to define continuity of the linear mapping, but Gâteaux differentiability of f is still not strong enough to imply continuity of f, even in two dimensions. The next example may be found, for instance, in Aubin and Ekeland [4, p. 18].

18.1.6 Example (Gâteaux differentiability does not imply continuity) Define $f\colon \boldsymbol{R}^2 \to \boldsymbol{R}$ by

$$f(x,y) = \begin{cases} \dfrac{y}{x}(x^2+y^2) & x \neq 0 \\ 0 & x = 0. \end{cases}$$

Then for $x \neq 0$,

$$\frac{f(\lambda x, \lambda y) - f(0,0)}{\lambda} = \frac{\left(\dfrac{\lambda y}{\lambda x}\lambda^2(x^2+y^2)\right)}{\lambda} = \frac{\lambda y}{x}(x^2+y^2) \to 0.$$

Thus $D_v f(0) = 0$ for any v, so f has a Gâteaux derivative at the origin, namely the zero linear map.

But f is not continuous at the origin. For consider $v(\varepsilon) = (\varepsilon^4, \varepsilon)$. The $v(\varepsilon) \to 0$ as $\varepsilon \to 0$, but

$$f\big(v(\varepsilon)\big) = \frac{\varepsilon}{\varepsilon^4}(\varepsilon^8 + \varepsilon^2) = \varepsilon^5 + 1/\varepsilon.$$

Thus $f\big(v(\varepsilon)\big) \to \infty$ as $\varepsilon \downarrow 0$, and $f\big(v(\varepsilon)\big) \to -\infty$ as $\varepsilon \uparrow 0$, so $\lim_{\varepsilon\to 0} f\big(v(\varepsilon)\big)$ does not exist. □

A stronger notion of derivative has proven useful. Gâteaux differentiability requires that chords have a limiting slope along straight lines approaching x. The stronger requirement is that chords have a limiting slope along arbitrary approaches to x. The definition quite naturally applies to functions between any normed vector spaces, not just Euclidean spaces, so we shall work as abstractly as possible. Dieudonné [5] claims that this makes everything clearer, but I know some who may disagree.

18.1.7 Definition (The differential or Fréchet derivative) *Let X and Y be normed real vector spaces. Let U be an open set in X and let $f\colon U \to Y$. The Gâteaux derivative is called the* ***differential*** *at x (also known as a* ***Fréchet derivative****, a* ***total derivative****, or simply a* ***derivative****) if it satisfies*

$$\lim_{v\to 0} \frac{\|f(x+v) - f(x) - D_v f(x)\|}{\|v\|} = 0. \tag{D}$$

The differential is usually denoted $Df(x)$, and it is a function from X into Y. Its value at a point v in X is denoted $Df(x)(v)$ rather than $D_v f(x)$. The double parentheses are only slightly awkward, and you will get used to them after a while.

When f has a differential at x, we say that f is ***differentiable*** *at x, or occasionally for emphasis that f is* ***Fréchet differentiable*** *at x.*

Actually my definition is a bit nonstandard. I started out with directional derivatives and said that if the mapping was linear and satisfied (D), then it was the differential. That is, I defined the differential in terms of directional derivatives. The usual approach is to say that f has a differential at x if there is some continuous linear mapping T that satisfies

$$\lim_{v\to 0} \frac{\|f(x+v)-f(x)-T(v)\|}{\|v\|} = 0. \tag{D$'$}$$

It is then customary to prove the following lemma.

18.1.8 Lemma *If T satisfies* (D$'$), *then $T(v) = f'(x;v)$. Consequently, T is unique, so*

$$Df(x)(v) = D_v f(x) = f'(x;v) = \lim_{\lambda\downarrow 0} \frac{f(x+\lambda v)-f(x)}{\lambda}.$$

Proof: Fix $v \neq 0$ and replace v by λv in (D$'$), and conclude

$$\lim_{\lambda\downarrow 0} \frac{\|f(x+\lambda v)-f(x)-T(\lambda v)\|}{\lambda\|v\|} = \lim_{\lambda\downarrow 0} \frac{1}{\|v\|} \left\| \frac{f(x+\lambda v)-f(x)}{\lambda} - T(v) \right\| = 0.$$

That is, $T(v) = \lim_{\lambda\downarrow 0} \frac{f(x+\lambda v)-f(x)}{\lambda} = f'(x;v)$. ∎

The continuity (equivalently boundedness) of $Df(x)(\cdot)$ implies the continuity of f.

18.1.9 Lemma (Differentiability implies Lipschitz continuity) *If f is differentiable at x, then f is continuous at x. Indeed, f is Lipschitz continuous at x. That is, there is $M \geqslant 0$ and $\delta > 0$ such that if $\|v\| < \delta$, then*

$$\Delta_v f(x) < M\,\|v\|.$$

Proof: Setting $\varepsilon = 1$ in the definition of differentiability, there is some $\delta > 0$ so that $\|v\| < \delta$ implies $\|\Delta_v f(x) - Df(x)(v)\| < \|v\|$, so by the triangle inequality,

$$\|\Delta_v f(x)\| < \|v\| + \|Df(x)(v)\| \leqslant \big(\|Df(x)\| + 1\big)\|v\|,$$

where of course $\|Df(x)\|$ is the operator norm of the linear transformation $Df(x)$. Thus f is continuous at x. ∎

Rewriting the definition

There are other useful ways to state this definition that I may use from time to time. Start by defining the **first difference function** $\Delta_v f$ of f at x by v, where $\Delta_v f \colon X \to Y$, by[2]

$$\Delta_v f(x) = f(x+v) - f(x).$$

[2] Loomis and Sternberg would write this as $\Delta f_x(v)$.

We can rewrite the definition of differentiability in terms of the first difference as follows: f is (Fréchet) differentiable at x if

$$(\forall \varepsilon > 0)\ (\exists \delta > 0)\ (\forall v)\ [\, 0 < \|v\| < \delta \implies \|\Delta_v f(x) - Df(x)(v)\| < \varepsilon \|v\|\,].$$

Another interpretation of the definition is this. Fix x and define the difference quotient function d_λ by

$$d_\lambda(v) = \frac{f(x+\lambda v) - f(x)}{\lambda}.$$

If f is differentiable at X, then d_λ converges uniformly on norm-bounded sets to the linear function $Df(x)$ as $\lambda \to 0$.

Further notes on the definition

When $X = Y = \boldsymbol{R}$, the differential we just defined is closely related to the derivative as usually defined for functions of one variable. The differential is the linear function $Df(x)\colon v \mapsto f'(x)v$, where $f'(x)$ is the numerical derivative defined earlier. Despite this difference, some authors (including Dieudonné [5], Luenberger [14], Marsden [15], and Spivak [19]) call the differential a derivative, but with modest care no serious confusion results. Loomis and Sternberg [13, pp. 158–159] argue that the term differential ought to be reserved for the linear transformation and derivative for its *skeleton* or matrix representation. But these guys are rather extreme in their views on notation and terminology—for instance, on page 157 they refer to the "barbarism of the classical notation for partial derivatives."

Also note that my definition of differentiability does not require that f be continuous anywhere but at x. In this, I believe I am following Loomis and Sternberg [13, p. 142]. Be aware that some authors, such as Dieudonné [5, p. 149] only define differentiability for functions continuous on an open set. As a result the function $f\colon \boldsymbol{R}^2 \to \boldsymbol{R}$ defined by

$$f(x,y) = \begin{cases} x^2 + y^2 & x = y \\ 0 & x \neq y \end{cases}$$

is differentiable at $(x, y) = (0, 0)$ under my definition, but not under Dieudonné's definition. By the way, Dieudonné does not require that the differential be a continuous linear transformation, he proves it using the continuity of f. Since we do not assume that f is continuous, we must make continuity of $Df(x)(\cdot)$ part of the definition (as do Loomis and Sternberg).

More variations on the definition

In Definition 18.1.7, I required that f be defined on an open set U in a normed space. Some authors, notably Graves [8], do not impose this restriction. Graves's definition runs like this.

Let X and Y be normed real vector spaces. Let A be a subset of X and let $f\colon A \to Y$. The **differential** at x is a linear transformation $T\colon X \to Y$ such that ****************** This differential is also denoted $Df(x)$. Note that if A is small enough, the differential may not be uniquely defined.

18.2 Differentiability of convex functions on $\boldsymbol{R}^{\mathrm{m}}$

The one-dimensional case has implications for the many dimensional case. The next results may be found in Fenchel [7, Theorems 33–34, pp. 86–87].

18.2.1 Theorem *Let f be a concave function on the open convex set C in $\boldsymbol{R}^{\mathrm{m}}$. For each direction v, $f'(x;v)$ is a lower semicontinuous function of x, and $\{x : f'(x;v)+f'(x;-v) < 0\}$ has Lebesgue measure zero. Thus $f'(x;v)+f'(x;-v) = 0$ almost everywhere, so f has a directional derivative in the direction v almost everywhere. Moreover, the directional derivative $Df(\cdot;v)$ is continuous on the set on the set on which it exists.*

Proof: Since f is concave, it is continuous by Theorem 6.3.4. Fix v and choose $\lambda_n \downarrow 0$. Then g_n defined by $g_n(x) = \frac{f(x+\lambda_n v)-f(x)}{\lambda_n}$ is continuous and by Lemma 15.1.2, $g_n(x) \uparrow f'(x;v)$ for each x. Thus Proposition 13.4.5 implies that $f'(x;v)$ is lower semicontinuous in x for any v.

Now $f'(x;v) + f'(x;-v) \leqslant 0$ by concavity, so let

$$A = \{x : f'(x;v) + f'(x;-v) < 0\}.$$

Note that since $f'(\cdot;v)$ and $f'(\cdot;-v)$ are lower semicontinuous, then A is a Borel subset of $\boldsymbol{R}^{\mathrm{m}}$. If $x \in A^c$, that is, if $f'(x;v) + f'(x;-v) = 0$, then $f'(x;v) = -f'(x;-v)$, so f has a directional derivative $D_v(x)$ in the direction v. And since $f'(\cdot;-v)$ is lower semicontinuous, the function $-f(\cdot;-v)$ is upper semicontinuous, $f'(\cdot;v)$ is actually continuous on A^c.

Thus we want to show that $A = \{x : f'(x;v) + f'(x;-v) < 0\}$ has Lebesgue measure zero.

If $v = 0$, then $f'(x;0) = -f'(x;-0) = 0$, so assume $v \neq 0$. Consider a line $L_y = \{y + \lambda v : \lambda \in \boldsymbol{R}\}$ parallel to v. By Theorem 6.1.4, $L_y \cap A = \{x \in L_y : f'(x;v) + f'(x;-v) < 0\}$ is countable, and hence of one-dimensional Lebesgue measure zero. Let M be the subspace orthogonal to v, so $M \times L = \boldsymbol{R}^{\mathrm{m}}$, where $L = L_0$ is the one-dimensional subspace spanned by v. Every $x \in \boldsymbol{R}^{\mathrm{m}}$ can be uniquely written as $x = (x_M, x_v)$, where $x_M \in M$ and $x_v \in L$. Then by Fubini's theorem,

$$\int \mathbf{1}_A(x)\, d\lambda^n(x) = \int_M \int_L \mathbf{1}_A(x_M, x_v)\, d\lambda(x_v)\, d\lambda^{n-1}(x_M) = \int_M 0\, d\lambda^{n-1}(x_M) = 0.$$

■

18.2.2 Lemma *Let f be a concave function on the open convex set $C \subset \boldsymbol{R}^{\mathrm{m}}$. If all n partial derivatives of f exist at x, then f has a Gâteaux derivative at x. That is, all the directional derivatives exist and the mapping $v \mapsto D_v f(x)$ is linear.*

Proof: The mapping $v \mapsto f'(x;v)$ is itself concave, and since f has an i^{th} partial derivative, there is $\delta_i > 0$ so that $v \mapsto f'(x;v)$ is linear on the segment $L_i = (-\delta_i e^i, \delta_i e^i)$. Indeed $\lambda e^i \mapsto \frac{\partial f(x)}{\partial x_i}\lambda$. So by Lemma 18.2.5 below, the mapping $v \mapsto f'(x;v)$ is linear on $\operatorname{co}\bigcup_{i=1}^m L_i$. This makes it the Gâteaux derivative of f at x. ∎

The next result may be found in Fenchel [7, Property 32, p. 86], or Hiriart-Urruty–Lemaréchal [10, Proposition 4.2.1, p. 114].

18.2.3 Lemma *Let f be a concave function on the open convex set $C \subset \boldsymbol{R}^{\mathrm{m}}$. If f has a Gâteaux derivative at x, then it is a Fréchet derivative.*

Proof: Let $v \mapsto f'(x;v)$ be the Gâteaux derivative of f. We need to show that for every $\varepsilon > 0$ there is some $\delta > 0$ such that

$$(\forall 0 < \lambda < \delta)\ \left(\forall v_{\|v\|=1}\right)\ [\,\|f(x+\lambda v) - f(x) - \lambda f'(x;v)\| \leqslant \varepsilon\lambda\,].$$

Fix $\varepsilon > 0$. By definition, $f'(x;v) = \lim_{\lambda\downarrow 0}\big(f(x+\lambda v) - f(x)\big)/\lambda$, so for each v, there is a $\delta_v > 0$ such that for $0 < \lambda \leqslant \delta_v$,

$$\left|\frac{f(x+\lambda v) - f(x)}{\lambda} - f'(x;v)\right| < \varepsilon,$$

or multiplying by λ,

$$|f(x+\lambda v) - f(x) - \lambda f'(x;v)| < \varepsilon\lambda,$$

By Lemma 15.1.4 and the homogeneity of $f'(x;\cdot)$, for $\lambda > 0$ we have

$$f(x) + \lambda f'(x;v) - f(x+\lambda v) \geqslant 0.$$

Combining these two inequalities, for $0 < \lambda \leqslant \delta_v$, we have

$$0 \leqslant \lambda f'(x;v) - f(x+\lambda v) + f(x) < \varepsilon\lambda. \tag{$\star$}$$

Once again consider the 2^n vectors $u^1, \dots, u^{2^n}$ with coordinates ± 1, and let $\delta = \min_j \delta_{u^j}$. Then ($\star$) holds with $v = u^j$ for any $0 < \lambda < \delta$.

Let $U = \operatorname{co}\{u^1, \dots, u^{2^n}\}$, which is a convex neighborhood of zero that includes all the vectors v with $\|v\| = 1$. Fixing λ, the function

$$h_\lambda(v) = \lambda f'(x;v) - f(x+\lambda v) + f(x)$$

is convex in v, and any v in U can be written as a convex combination $v = \sum_{j=1}^{2^n} \alpha_j u^j$, so for any $0 < \lambda \leqslant \delta$,

$$0 \leqslant \lambda f'(x;v) - f(x+\lambda v) + f(x) = h_\lambda(v) \leqslant \sum_{j=1}^{2^n} \alpha_j h_\lambda(u^j) \leqslant \max_j h_\lambda(u^j) < \varepsilon\lambda.$$

Since this is true for every vector v of norm one, we are finished. ∎

18.2.4 Theorem *Let f be a concave function on the open convex set $C \subset \boldsymbol{R}^{\mathrm{m}}$. Then f is differentiable almost everywhere on C.*

Proof: By Theorem 18.2.1 for each i, the i^{th} partial derivative exists for almost every x. Therefore all n partial derivatives exist for almost every x. The result now follows from Lemma 18.2.2 and 18.2.3. ∎

This lemma is used in the proof of Lemma 18.2.2, which is required for the proof pf Theorem 18.2.4.

18.2.5 Lemma *Let g be concave on C and let $x \in \operatorname{ri} C$. Let $v^1, \dots, v^m$ be linearly independent and assume that g is affine on each of the segments $L_i = \{x + \lambda v^i : |\lambda| \leqslant \delta_i\} \subset C$, $i = 1, \dots, m$. Then g is affine on $A = \operatorname{co} \bigcup_{i=1}^m L_i$.*

Proof: By hypothesis, for each $i = 1, \dots, m$, there is an α_i satisfying

$$g(x + \lambda v^i) = g(x) + \alpha_i \lambda \text{ on } L_i.$$

Define ℓ on the span of $v^1, \dots, v^m$ by $\ell(\lambda_1 v^1 + \cdots + \lambda_m v^m) = \alpha_1\lambda_1 + \cdots + \alpha_m\lambda_m$. Then ℓ is linear, so the function h on A defined by $h(y) = g(x) + \ell(y - x)$ is affine. Moreover h agrees with g on each segment L_i. In particular $g(x) - h(x) = 0$.

Now any point y in A can be written as a convex combination of points $x + \lambda_i v^i$ belonging to L_i. Since g is concave, for a convex combination $\sum_i \alpha_i(x + \lambda_i v^i)$ we have

$$g\Big(\sum_i \alpha_i(x + \lambda_i v^i)\Big) \geqslant \sum_i \alpha_i g(x + \lambda_i v^i) = \sum_i \alpha_i h(x + \lambda_i v^i) = h\Big(\sum_i \alpha_i(x + \lambda_i v^i)\Big),$$

where the final equality follows from the affinity of h. Therefore $g - h \geqslant 0$ on A. But $g - h$ is concave, x belongs to $\operatorname{ri} A$, and $(g - h)(x) = 0$. Therefore $g - h = 0$ on A. (To see this, let y belong to A. Since x in $\operatorname{ri} A$, for some $z \in A$ and $0 < \lambda < 1$, we may write $x = \lambda y + (1-\lambda)z$, so $0 = (g-h)(x) \geqslant \lambda(g-h)(y) + (1-\lambda)(g-h)(z) \geqslant 0$, which can only happen if $(g - h)(y) = (g - h)(z) = 0$.)

Thus g is the affine function h on A. ∎

18.2.6 Remark This result depends on the fact that x belongs to $\operatorname{ri} C$, and can fail otherwise. For instance, let $C = \boldsymbol{R}^2_+$ and $f(x, y) = xy$. Then f is linear (indeed zero) on the nonnegative x and y axes, which intersect at the origin, but f is not linear on the convex hull of the axes. Of course, the origin is not in the relative interior of $\boldsymbol{R}^2_+$.

The next fact may be found in Fenchel [7, Theorem 35, p. 87ff], or Katzner [12, Theorems B.5-1 and B.5-2].

18.2.7 Fact *If $f\colon C \subset \boldsymbol{R}^{\mathrm{m}} \to \boldsymbol{R}$ is twice differentiable, then the Hessian H_f is everywhere negative semidefinite if and only if f is concave. If H_f is everywhere negative definite, then f is strictly concave.*

************ There are many ways to see this. One way is to look at the second difference $\Delta^2_{v,w} f = f(x + w + v) - f(x + w) - \big(f(x + v) - f(x)\big)$. By ************

18.3 Differentiability and the single subgradient

The next result may be found in Rockafellar [16, Theorem 25.1, p. 242].

18.3.1 Theorem *Let f be a convex function on $\boldsymbol{R}^{\mathrm{m}}$. Then f is differentiable at x if and only if the subdifferential $\partial f(x)$ is a singleton, in which case the lone subgradient is in fact the differential of f at x.*

Let f be a concave function on $\boldsymbol{R}^{\mathrm{m}}$. Then f is differentiable at the point $x \in \operatorname{int}\operatorname{dom} f$ if and only if the superdifferential $\partial f(x)$ is a singleton, in which case $\partial f(x) = \{f'(x)\}$.

Proof: I'll prove the concave case. ($\Longrightarrow$) Suppose f is differentiable at the interior point x. Then for any v, $f'(x;v) = f'(x) \cdot v$. Moreover there is an $\varepsilon > 0$ such that for any v, $x + \varepsilon v \in \operatorname{dom} f$. Now the superdifferential $\partial f(x)$ is nonempty, since $f'(x) \in \partial f(x)$, so by Lemma 15.1.6, if $p \in \partial f(x)$, then

$$p \cdot \varepsilon v \geqslant f'(x;\varepsilon v) = f'(x) \cdot \varepsilon v.$$

But this also holds for $-v$, so

$$p \cdot v = f'(x) \cdot v.$$

Since this holds for all v, we have $p = f'(x)$.

($\Longleftarrow$) Suppose $\partial f(x) = \{p\}$. Since x is interior there is an $\alpha > 0$ such that $x + \alpha B \subset \operatorname{dom} f$, where B is the unit ball in $\boldsymbol{R}^{\mathrm{m}}$. Define the concave function $g\colon \alpha B \to \boldsymbol{R}$ by

$$g(v) = f(x+v) - f(x) - p \cdot v.$$

Note that f is differentiable at x if and only if g is differentiable at 0, in which case $g'(0) = f'(x) - p$.

Now the supergradient inequality asserts that $f(x) + p \cdot v \geqslant f(x+v)$, so $g \leqslant 0$. But $g(0) = 0$, that is, 0 maximizes g over αB, so by Lemma 14.1.8, $0 \in \partial g(0)$.

In fact, $\partial g(0) = \{0\}$. For if $q \in \partial g(0)$, we have

$$\begin{aligned} g(0) + q \cdot v &\geqslant g(v) \\ 0 + q \cdot v &\geqslant f(x+v) - f(x) - p \cdot v \\ f(x) + (p+q) \cdot v &\geqslant f(x+v), \end{aligned}$$

which implies $p + q \in \partial f(x)$, so $q = 0$.

By Lemma 15.1.7, the closure of $g'(x;\cdot)$ is the cost function of $\partial g(0) = \{0\}$, so $\operatorname{cl} g'(x;\cdot) = 0$. But this implies that $g'(x;\cdot)$ is itself closed, and so identically zero. But zero is a linear function, so by Lemma 18.2.3, g is differentiable at zero. ■

Why is it closed? See Rockafellar 23.4.

18.3.1 Hotelling's and Shephard's Lemmas

The following twin results in economics are attributed to Shephard [18] and Hotelling. They are simple consequences of general results in convex analysis.

18.3.2 Shephard's Lemma *If x is the unique minimizer of p over a set $A \subset \boldsymbol{R}^{\mathrm{m}}$, then the cost function c_A is differentiable at p and $c_A'(p) = x$.*

18.3.3 Hotelling's Lemma *If x is the unique maximizer of p over a set $A \subset \boldsymbol{R}^{\mathrm{m}}$, then the profit function π_A is differentiable at p and $\pi_A'(p) = x$.*

Proof of Shephard's Lemma: Theorem 15.2.1 asserts that the subdifferential of a cost function at the price vector p is the set of cost-minimizing vectors. Theorem 18.3.1 asserts that if the cost minimizer is unique, then the cost function is differentiable and its gradient is the cost minimizer. ■

18.4 Convex functions on finite dimensional spaces

In this section we gather several important properties of convex functions on finite dimensional spaces. For a more detailed account see the definitive volume by R. T. Rockafellar [16].

If f is differentiable at every point in U, the mapping $x \mapsto Df(x)$ from U into $L(X,Y)$ is itself a function from an open subset of a normed space to a normed space. If this mapping is differentiable at x, its differential, which belongs to $L\big(X, L(X,Y)\big)$, is called the **second differential** of f at x, denoted $D^2f(x)$. (Thus for each $v \in X$, $D^2f(x)(v)$ is an operator in $L(X,Y)$. Its value at $w \in X$ is $D^2f(x)(v)(w)$. For the case $X = \boldsymbol{R}^{\mathrm{m}}$ and $Y = \boldsymbol{R}$, we can identify $L\big(\boldsymbol{R}^{\mathrm{m}}, L(\boldsymbol{R}^{\mathrm{m}}, \boldsymbol{R})\big)$ with $\boldsymbol{R}^{n^2}$ and $D^2f(x)$ with the **Hessian matrix** $\left[\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right]$ of f at x via $D^2f(x)(v)(w) = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f(x)}{\partial x_i \partial x_j} v_i w_j$.)

We already know from Proposition 14.1.5 that the subdifferential is nonempty at interior points and that when it is a singleton, it consists of the Fréchet derivative (Theorem 18.3.1).[3]

Check the crossref.

The next result shows that the derivative a convex function defined on an open subset of a finite dimensional space exists except possibly on a set of Lebesgue measure zero.

18.4.1 Theorem *If C is an open convex subset of $\boldsymbol{R}^{\mathrm{m}}$ and $f\colon C \to \boldsymbol{R}$ is a convex function, then:*

a. *the set A of points where f is differentiable is a dense $\mathcal{G}_\delta$ subset of C,*

b. *its complement has Lebesgue measure zero, and*

[3] In infinite dimensional spaces, it is in general the Gâteaux derivative.

c. *the function* $x \mapsto Df(x)$ *from* A *to* $L(\boldsymbol{R}^{\mathrm{m}}, \boldsymbol{R})$ *is continuous.*

A proof of this can be found in R. T. Rockafellar [16, Theorem 25.5 and Corollary 25.5.1, p. 246] or W. Fenchel [7, Theorems 33 and 34, pp. 86–87]. The proof that a convex function is differentiable almost everywhere has two parts. The first part uses Lemma 6.1.4 (3) and Fubini's Theorem to show that a convex function has partial derivatives almost everywhere, and the second part shows that if all the partial derivative of a convex function (on a finite dimensional space) exist at a point, then it is in fact differentiable at that point.

Need a reference to Fubini's Theorem.

18.4.2 Theorem *Let* $f\colon C \to \boldsymbol{R}$ *be a twice differentiable real function on an open convex subset of* $\boldsymbol{R}^{\mathrm{m}}$*. Then* f *is convex if and only its Hessian matrix is positive semidefinite*[4] *everywhere in* C*.*

For a proof of the above result, see C. D. Aliprantis [2, Problem 1.1.2, p. 3] or J.-B. Hiriart-Urruty and C. Lemaréchal [9, Theorem 4.3.1, p. 190]. We can also say something about the almost everywhere existence of the second differential. Let us say that a correspondence $\varphi\colon \boldsymbol{R}^{\mathrm{n}} \twoheadrightarrow \boldsymbol{R}^{\mathrm{m}}$ is **differentiable** at x if there is a linear mapping $T\colon \boldsymbol{R}^{\mathrm{n}} \to \boldsymbol{R}^{\mathrm{m}}$ (that is, an $m \times n$ matrix) satisfying

$$y_v = y + T(v) + o(\|v\|), \quad \text{for all} y \in \varphi(x),\ y_v \in \varphi(x+v).$$

If φ is differentiable at x, then it is singleton-valued at x, and the linear mapping T is unique and is called the **derivative** of φ at x.

The following theorem is due to A. D. Alexandroff [1]. The formulation stated here is based on R. Howard [11, Theorems 6.1 and 7.1] and J.-B. Hiriart-Urruty and C. Lemaréchal [9, Theorem 4.3.4, p. 192]. See also the enlightening discussion in Sections I.5.1–2 (pp. 30–33) and IV.4.3 (pp. 190–193) of [9].

18.4.3 Theorem (Alexandroff's Theorem) *If* $f\colon \boldsymbol{R}^n \to \boldsymbol{R}$ *is a proper convex function, and* $\operatorname{dom} f$ *has nonempty interior, then there exists a subset* A *of the interior of* $\operatorname{dom} f$ *such that:*

1. *The set* $\operatorname{int}(\operatorname{dom} f) \setminus A$ *has Lebesgue measure zero.*
2. *Both* f *and* ∂f *are differentiable at every point of* A*.*
3. *The derivative* T *of* ∂f *at each point of* A *is symmetric, positive definite, and satisfies the "second order Taylor expansion formula"*

$$f(x+v) = f(x) + Df(x)(v) + \tfrac{1}{2}T(x)(v)\cdot v + o\Big(\|v\|^2\Big).$$

[4] Recall that an $n \times n$ symmetric matrix M is **positive semidefinite** if $x^{\mathrm{t}}Mx \geq 0$ holds for each $x \in \boldsymbol{R}^{\mathrm{m}}$. Equivalently, M is positive semidefinite if its eigenvalues are real and nonnegative.

References

[1] A. D. Alexandroff. 1939. Almost everywhere existence of the second differential of a convex function and surfaces connected with it. *Leningrad State University Annals, Mathematics Series* 6:3–35. In Russian.

[2] C. D. Aliprantis. 1996. *Problems in equilibrium theory*. Heidelberg & New York: Springer–Verlag.

[3] J.-P. Aubin. 1979. *Mathematical methods of game and economic theory*. Number 7 in Studies in mathematics and its applications. New York: North Holland.

[4] J.-P. Aubin and I. Ekeland. 1984. *Applied nonlinear analysis*. Pure and Applied Mathematics: A Wiley-Interscience Series of Texts, Monographs, and Tracts. Mineola, New York: John Wiley & Sons. Reprint of the 1984 edition by John Wiley and Sons.

[5] J. Dieudonné. 1969. *Foundations of modern analysis*. Number 10-I in Pure and Applied Mathematics. New York: Academic Press. Volume 1 of Treatise on Analysis.

[6] I. Ekeland and R. Temam. 1976. *Convex analysis and variational problems*. Number 1 in Studies in Mathematics and its Applications. Amsterdam: North Holland.

[7] W. Fenchel. 1953. Convex cones, sets, and functions. Lecture notes, Princeton University, Department of Mathematics. From notes taken by D. W. Blackett, Spring 1951.

[8] L. M. Graves. 1956. *Theory of functions of real variables*, 2d. ed. International Series in Pure and Applied Mathematics. New York: McGraw–Hill.

[9] J.-B. Hiriart-Urruty and C. Lemaréchal. 1993. *Convex analysis and minimization algorithms I*. Number 305 in Grundlehren der mathematischen Wissenschaften. Berlin: Springer–Verlag.

[10] ——— . 2001. *Fundamentals of convex analysis*. Grundlehren Text Editions. Berlin: Springer–Verlag.

[11] R. Howard. 1998. Alexandrov's theorem on the second derivatives of convex functions via Rademacher's theorem on the first derivatives of Lipschitz functions. Lecture notes, Department of Mathematics, University of South Carolina, Columbia, SC.
http://www.math.sc.edu/~howard/Notes/alex.pdf

[12] D. W. Katzner. 1970. *Static demand theory*. London: Macmillan.

[13] L. H. Loomis and S. Sternberg. 1968. *Advanced calculus*. Reading, Massachusetts: Addison–Wesley.

[14] D. G. Luenberger. 1969. *Optimization by vector space methods*. Series in Decision and Control. New York: Wiley.

[15] J. E. Marsden. 1974. *Elementary classical analysis*. San Francisco: W. H. Freeman and Company.

[16] R. T. Rockafellar. 1970. *Convex analysis*. Number 28 in Princeton Mathematical Series. Princeton: Princeton University Press.

[17] T. Saijo. 1983. Differentiability of the cost functions is equivalent to strict quasiconcavity of the production functions. *Economics Letters* 12(2):135–139. DOI: 10.1016/0165-1765(83)90124-6

[18] R. W. Shephard. 1981. *Cost and production functions*. Number 194 in Lecture Notes in Economics and Mathematical Systems. Berlin: Springer–Verlag. Reprint of the 1953 edition published by Princeton University Press.

[19] M. Spivak. 1965. *Calculus on manifolds*. Mathematics Monograph Series. New York: Benjamin.