SOME FACTS ABOUT CONVEX FUNCTIONS

SCOTT ARMSTRONG

Definition 1. *The* epigraph *of a function* $f : \mathbb{R}^n \to (-\infty, \infty]$ *is the set*

$$\operatorname{epi} f = \{(x, \mu) \in \mathbb{R}^{n+1} : \mu \geq f(x)\}.$$

A function $f : \mathbb{R}^n \to (-\infty, \infty]$ *is called* convex *if* $\operatorname{epi} f$ *is convex. This is equivalent to the condition that for all* $0 \leq \lambda \leq 1$ *we have*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

The domain *of a convex function* f *is the set*

$$\operatorname{dom} f = \{x \in \mathbb{R}^n : f(x) < \infty\}.$$

A sublinear function *is a convex function which satisfies the stronger condition*

$$f(\lambda x + \mu y) \leq \lambda f(x) + \mu f(y) \quad \text{for all } \lambda, \mu \geq 0.$$

A function $f : \mathbb{R}^n \to (-\infty, \infty]$ *is called* positively homogeneous *if for all* $\lambda \geq 0$ *and* $x \in \mathbb{R}^n$ *we have*

$$f(\lambda x) = \lambda f(x)$$

(we take the convention that $0 \cdot \infty = 0$*). If*

$$f(x + y) \leq f(x) + f(y)$$

for all $x, y \in \mathbb{R}^n$*, then* f *is said to be* subadditive.

Proposition 1. *A function* $f : \mathbb{R}^n \to (-\infty, \infty]$ *is sublinear if and only if it is both positively homogeneous and subadditive.*

Proof. A sublinear function is clearly subadditive. If f is sublinear then

$$f(\lambda x) \leq \lambda f(x) \quad \text{for all } \lambda \geq 0. \tag{0.1}$$

In particular, $f(0) \leq 0$. But since f is subadditive, $f(0) \leq f(0) + f(0)$, hence $f(0) = 0$. If $\lambda > 0$, then using (0.1) we see that

$$\lambda f(x) = \lambda f(\lambda^{-1} \lambda x) \leq f(\lambda x).$$

Hence f is positively homogeneous.

If f is positively homogeneous and subadditive then for all $\lambda, \mu \geq 0$ and $x, y \in \mathbb{R}^n$

$$f(\lambda x + \mu y) \leq f(\lambda x) + f(\mu y) = \lambda f(x) + \mu f(y).$$

□

Definition 2. *The* linearity space *of a sublinear function* f *is the set*

$$\operatorname{lin} f = \{x \in \mathbb{R}^n : f(x) = -f(-x)\}.$$

Proposition 2. *If f is a sublinear function, then* $\operatorname{lin} f$ *is the largest subspace of* $\mathbb{R}^n$ *on which f is linear.*

Proof. Let $x, y \in \operatorname{lin} f$. Then if $\lambda, \mu \geq 0$

$$\begin{aligned} f(\lambda x + \mu y) &\leq \lambda f(x) + \mu f(y) \\ &= -(\lambda f(-x) + \mu f(-y)) \\ &\leq -f(-\lambda x - \mu y). \end{aligned} \tag{0.2}$$

Also,

$$0 = f(0) \leq f(\lambda x + \mu y) + f(-\lambda x - \mu y),$$

and thus

$$f(\lambda x + \mu y) = -f(-\lambda x - \mu y) \tag{0.3}$$

for all $\lambda, \mu \geq 0$. Since $-\operatorname{lin} f = \operatorname{lin} f$, we conclude that (0.3) holds for all $\lambda, \mu \in \mathbb{R}$. That is, $\operatorname{lin} f$ is a linear subspace of $\mathbb{R}^n$. Since we have equality in (0.2), it follows that

$$f(\lambda x + \mu y) = \lambda f(x) + \mu f(y)$$

for all $\lambda, \mu \geq 0$. Using again the fact that $-\operatorname{lin} f = \operatorname{lin} f$, we see that f is linear on $\operatorname{lin} f$.

For any subspace E of $\mathbb{R}^n$ such that $f|_E$ is linear, it is clear that $E \subseteq \operatorname{lin} f$. □

Definition 3. *The* directional derivative *of a function $f : \mathbb{R}^n \to (-\infty, \infty]$ at a point $x \in \mathbb{R}^n$ in the direction $d \in \mathbb{R}^n$ is*

$$f'(x; d) = \lim_{h \downarrow 0} \frac{f(x + hd) - f(x)}{h}$$

when this limit exists ($-\infty$ and ∞ are allowed).

Proposition 3. *If $f : \mathbb{R}^n \to (-\infty, \infty]$ is convex, then the directional derivative $f'(x; \cdot)$ is well-defined and positively homogeneous for each $x \in \operatorname{dom} f$. Moreover, for any point $x \in \operatorname{int}(\operatorname{dom} f)$, $f'(x; \cdot)$ is everywhere finite and sublinear.*

Proof. 1. Let $x \in \operatorname{dom} f$ and $d \in \mathbb{R}^n$. We claim that for $0 < s \leq t$,

$$\frac{f(x + sd) - f(x)}{s} \leq \frac{f(x + td) - f(x)}{t}. \tag{0.4}$$

By convexity, we have that

$$\begin{aligned} t f(x + sd) &\leq t \left(\frac{t - s}{t} f(x) + \frac{s}{t} f(x + td) \right) \\ &= (t - s) f(x) + s f(x + td), \end{aligned}$$

and upon rearranging we obtain (0.4). It follows that $f'(x;d)$ exists. If $\lambda > 0$ then

$$\begin{aligned} f'(x;\lambda d) &= \lim_{h\downarrow 0} \frac{f(x+h\lambda d)-f(x)}{h} \\ &= \lambda \lim_{h\downarrow 0} \frac{f(x+h\lambda d)-f(x)}{h\lambda} = \lambda f'(x;d), \end{aligned}$$

so $f'(x;\cdot)$ is positively homogeneous.

2. Let $x \in \operatorname{int}\operatorname{dom} f$. We claim that for $t > 0$,

$$\frac{f(x-td)-f(x)}{-t} \leq \frac{f(x+td)-f(x)}{t}. \tag{0.5}$$

To see this, we use convexity again:

$$2f(x) \leq f(x+td) + f(x-td),$$

which is (0.5) after rearrangement. Thus for sufficiently small $t > 0$, we have that

$$\infty > \frac{f(x+tb)-f(x)}{t} \downarrow f'(x;b) > \frac{f(x-tb)-f(x)}{-t} > -\infty.$$

We again use convexity to deduce that for any directions $d_1, d_2 \in \mathbb{R}^n$,

$$\frac{f(x+t(d_1+d_2))-f(x)}{t} \leq \frac{f(x+2td_1)-f(x)}{2t} + \frac{f(x+2td_2)-f(x)}{2t}.$$

Letting $t \downarrow 0$ gives the subadditivity of $f'(x;\cdot)$. □

Definition 4. *We say that $w \in \mathbb{R}^n$ is a* subgradient *of f at a point $x \in \mathbb{R}^n$ if*

$$f(y) - f(x) \geq w \cdot (y - x)$$

for all $y \in \mathbb{R}^n$. We denote the set of subgradients by $\partial f(x)$, called the subdifferential *of f at x. It is easy to show that $\partial f(x)$ is a closed convex set.*

Proposition 4. *If $f : \mathbb{R}^n \to (-\infty, \infty]$ is convex and $x \in \operatorname{dom} f$, then $w \in \mathbb{R}^n$ is a subgradient of f at x if and only if*

$$w \cdot z \leq f'(x;z)$$

for all $z \in \mathbb{R}^n$.

Proof. Observe that

$$\begin{aligned} w \in \partial f(x) &\iff w \cdot (y-x) \leq f(y) - f(x) \quad \text{for all } y \in \mathbb{R}^n \\ &\iff w \cdot \left(\frac{y-x}{h}\right) \leq \frac{f(y)-f(x)}{h} \quad \text{for all } h > 0, y \in \mathbb{R}^n \\ &\iff w \cdot z \leq \frac{f(x+hz)-f(x)}{h} \quad \text{for all } h > 0, z \in \mathbb{R}^n \\ &\iff w \cdot z \leq f'(x;z) \quad \text{for all } z \in \mathbb{R}^n, \end{aligned}$$

where in the last line we used the fact that

$$\frac{f(x+hz)-f(x)}{h} \downarrow f'(x;z) \quad \text{as } h \downarrow 0.$$

□

We are interested in showing that for convex functions f and $x \in \operatorname{int}(\operatorname{dom} f)$, $\partial f(x)$ is nonempty. We will accomplish this by recursively constructing sublinear functions. First we need the following lemma.

Lemma 5. *Suppose that $p : \mathbb{R}^n \to (-\infty, \infty]$ is a sublinear function and that $x \in \operatorname{int}(\operatorname{dom} f)$. Then the function $q = p'(x; \cdot)$ satisfies the following conditions:*

(i) *$q(\lambda x) = \lambda p(x)$ for all $\lambda \in \mathbb{R}$;*
(ii) *$q \leq p$;*
(iii) *$\operatorname{lin} p + \operatorname{span}\{x\} \subseteq \operatorname{lin} q$.*

Proof. 1. Select $\lambda \in \mathbb{R}$. Using the fact that p is positively homogeneous and that $(1 + h\lambda) > 0$ for all h sufficiently small, we calculate

$$\begin{aligned} q(\lambda x) = p'(x; \lambda x) &= \lim_{h \downarrow 0} h^{-1}[p(x + h\lambda x) - p(x)] \\ &= \lim_{h \downarrow 0} h^{-1}[(1 + h\lambda)p(x) - p(x) \\ &= \lim_{h \downarrow 0} h^{-1}[(h\lambda)p(x)] = \lambda p(x). \end{aligned}$$

This proves (i).

2. Fix $y \in \mathbb{R}^n$. Then

$$\begin{aligned} q(y) = p'(x; y) &= \lim_{h \downarrow 0} h^{-1}[p(x + hy) - p(x)] \\ &= \lim_{h \downarrow 0} h^{-1}[p(hy)] = p(y). \end{aligned}$$

This is (ii).

3. Clearly $x \in \operatorname{lin} q$ by (i). If $y \in \operatorname{lin} p$, then

$$\begin{aligned} q(y) = p'(x; y) &= \lim_{h \downarrow 0} h^{-1}[p(x + hy) - p(x)] \\ &= \lim_{h \downarrow 0} h^{-1}[p(hy) = \lim_{h \downarrow 0} h^{-1}[hp(y)] = p(y). \end{aligned}$$

That is, if $y \in \operatorname{lin} p$ then $q(y) = p(y)$. But then

$$q(y) = p(y) = -p(-y) = -q(-y),$$

since $-y \in \operatorname{lin} p$. Therefore $\operatorname{lin} p \subseteq \operatorname{lin} q$. Since $\operatorname{lin} q$ is a subspace of $\mathbb{R}^n$, it follows that

$$\operatorname{lin} p + \operatorname{span}\{x\} \subseteq \operatorname{lin} q.$$

□

Theorem 6 (Max Formula)**.** *If the function $f : \mathbb{R}^n \to (-\infty, \infty]$ is convex then for every point $x \in \operatorname{int}(\operatorname{dom} f)$ and direction $d \in \mathbb{R}^n$*

$$f'(x; d) = \max_{w \in \partial f(x)} w \cdot d.$$

In particular, $\partial f(x) \neq \emptyset$.

Proof. Fix $d \in \mathbb{R}^n$. By Proposition 4, it suffices to demonstrate the existence of a $w \in \partial f(x)$ such that $w \cdot d = f'(x; d)$. Choose a basis $\{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ with $v_1 = d$ if $d \neq 0$. Define a sequence $p_0, \ldots, p_n$ of sublinear functions by

$$\begin{aligned} p_0 &= f'(x; \cdot), \\ p_k &= p'_{k-1}(v_k; \cdot), \qquad (k = 1, \ldots, n). \end{aligned}$$

By Proposition 3, p_k is everywhere finite and sublinear. By (iii) of Lemma 5,

$$\operatorname{lin} p_{k-1} + \operatorname{span}\{v_k\} \subseteq \operatorname{lin} p_k.$$

Hence $\operatorname{span}\{v_1, \ldots, v_n\} \subseteq \operatorname{lin} p_n$, so p_n is linear. Choose $w \in \mathbb{R}^n$ such that for all $z \in \mathbb{R}^n$, $p_n(z) = x \cdot z$. Part (ii) of Lemma 5 implies that for all $y \in \mathbb{R}^n$

$$\begin{aligned} w \cdot (y - x) = p_n(y - x) &\leq p_{n-1}(y - x) \leq \ldots \leq p_0(y - x) \\ &\leq f'(x; y - x) = \lim_{h \downarrow 0} h^{-1}[f(x + h(y - x)) - f(x)] \\ &\leq f(y) - f(x). \end{aligned}$$

Thus $w \in \partial f(x)$. If $d = 0$, then $p_n(0) = 0 = f'(x; 0) = f'(x; d)$. If $d \neq 0$, then by part (i) of Lemma 5

$$\begin{aligned} p_n(d) \leq p_0(d) &= p_0(v_1) = -p'_0(v_1; -v_1) \\ &= -p_1(-v_1) = -p_1(-d) \leq -p_n(-d) = p_n(d). \end{aligned}$$

Hence $p_n(d) = p_0(d) = f'(x; d)$. □

Proposition 7. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is convex and $x \in \mathbb{R}^n$ such that f is bounded above on some neighborhood of x. Then there is an open ball B centered at x such that f is Lipschitz on B.*

Proof. We may assume without loss of generality that $x = 0$, $f(0) = 0$, and $f \leq 1$ on $2B$, where $B = \{z : |z| < 1\}$. By convexity, for all $z \in 2B$,

$$-f(z) = -f(z) + 2f(0) \leq -f(z) + f(z) + f(-z) = f(-z) \leq 1,$$

so $|f| \leq 1$ on $2B$.

Fix $x, y \in B$. Set $\alpha = |y - x|$ and $w = y + \alpha^{-1}(y - x)$ so that

$$y = \frac{\alpha w}{\alpha + 1} + \frac{x}{\alpha + 1}.$$

Then by the convexity of f

$$\begin{aligned} f(y) - f(x) &\leq \frac{\alpha}{\alpha + 1} f(w) + \frac{1}{\alpha + 1} f(x) - f(x) \\ &= \frac{\alpha}{\alpha + 1}[f(w) - f(x)] \leq 2\alpha = 2|y - x|. \end{aligned}$$

Now we interchange x and y and repeat the argument to get

$$|f(y) - f(x)| \leq 2|y - x|.$$

□

Lemma 8. *Let* $\Delta = \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n : 0 \leq x_i, \sum x_i \leq 1\}$. *If a function* $f : \Delta \to \mathbb{R}$ *is convex, then* f *is locally Lipschitz continuous on* $\operatorname{int}\Delta$.

Proof. By the preceding result, it suffices to show that f is bounded above on Δ. Let $\{e_i\}$ be the standard basis in $\mathbb{R}^n$ and fix $x \in \Delta$. Then by convexity

$$\begin{aligned} f(x) &= f\left(\sum x_i e_i + (1 - \sum x_i)0\right) \leq \sum x_i f(e_i) + (1 - \sum x_i) f(0) \\ &\leq \max\{f(0), f(e_1), \ldots, f(e_n)\}. \end{aligned}$$

□

Theorem 9 (Continuity of Convex Functions)**.** *Let* $f : \mathbb{R}^n \to (-\infty, \infty]$ *be a convex function. Then* f *is locally Lipschitz continuous on* $\operatorname{int}(\operatorname{dom} f)$.

Proof. If $x \in \operatorname{int}(\operatorname{dom} f)$, then by translation and scaling we may assume without loss of generality that $x \in \operatorname{int}\Delta \subset \operatorname{int}(\operatorname{dom} f)$. The result follows by the previous lemma. □