Monday, January 7, 2008

1 Convex Functions on the Real Line

In what follows an *interval* $I \subset \mathbb{R}$ is any set of $\mathbb{R}$ such that if $x, y \in I$ and $x < y$, then $[x, y] \subset I$.

Definition 1 *Given an interval $I \subset \mathbb{R}$, a function $f : I \to \mathbb{R}$ is*

(i) convex *if*
$$f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$$
for all $x, y \in I$ and $\theta \in (0,1)$;

(ii) strictly convex *if the inequality in (i) is strict whenever $x \neq y$;*

(iii) concave *(respectively* strictly concave*) if $-f$ is convex (respectively strictly convex).*

Geometrically, the inequality in (i) means that if P, Q, and R are any three points on the graph of f with Q between P, and R, then Q is on or below the chord PR, or in terms of slopes
$$\text{slope } PQ \leq \text{slope } PR \leq \text{slope } QR, \tag{1}$$
with strict inequalities if f is strictly convex.

Example 2 *(i) The function $f : \mathbb{R} \to [0, \infty)$ defined by $f(x) := |x|^p$, $p > 0$, is convex if and only if $p \geq 1$, and is strictly convex if and only if $p > 1$. In particular, if $x, y \geq 0$, and $p \geq 1$, then by the convexity of f,*
$$\left(\frac{1}{2}x + \frac{1}{2}y\right)^p \leq \frac{1}{2}x^p + \frac{1}{2}y^p,$$
or equivalently,
$$(x+y)^p \leq 2^{p-1}(x^p + y^p).$$

(ii) The function $f : \mathbb{R} \to [0, \infty)$ defined by $f(x) := \sqrt{x^2+1}$ is strictly convex.

(iii) The function
$$f(x) := \log x \quad \text{if } x > 0$$
is strictly concave. In particular, if $x, y > 0$, $1 < p < \infty$, and q is its conjugate exponent, then by the concavity of f,
$$\log(xy) = \frac{1}{p}\log x^p + \frac{1}{q}\log y^q \leq \log\left(\frac{1}{p}x^p + \frac{1}{q}x^q\right),$$
or equivalently,
$$xy \leq \frac{1}{p}x^p + \frac{1}{q}y^q.$$
This is known as Young's inequality*. Note that, in view of the strict concavity of f, equality holds if and only if $x^p = y^q$.*

We say that $f : \mathbb{R} \to \mathbb{R}$ is *linear* if $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$ for all α, $\beta \in \mathbb{R}$ and all x, $y \in \mathbb{R}$. A linear function f takes the form $f(x) = mx$ for all $x \in \mathbb{R}$ and for some $m \in \mathbb{R}$. Given an interval $I \subset \mathbb{R}$, we say that $f : I \to \mathbb{R}$ is *affine* if it has the form $f(x) = mx + p$ for all $x \in I$ and for some m, $p \in \mathbb{R}$.

Exercise 3 *Given an interval $I \subset \mathbb{R}$, prove that $f : I \to \mathbb{R}$ is affine if and only if it is both convex and concave.*

1.1 Regularity of Convex Functions

Next we start looking at the regularity of convex functions.

Definition 4 *Let $E \subset \mathbb{R}$. A function $f : E \to \mathbb{R}$ is said to be*

(i) Lipschitz continuous *if there exists a constant $L > 0$ such that*

$$|f(x) - f(y)| \leq L|x - y|$$

for all x, $y \in E$;

(ii) locally Lipschitz continuous *if for every compact set $K \subset E$ if there exists a constant $L_K > 0$ such that*

$$|f(x) - f(y)| \leq L_K|x - y|$$

for all x, $y \in K$;

(iii) Hölder continuous *with exponent $0 < \alpha < 1$ if there exists a constant $L > 0$ such that*

$$|f(x) - f(y)| \leq L|x - y|^\alpha$$

for all x, $y \in E$.

Exercise 5 *The Weierstrass function*

$$f(s) := \sum_{n=1}^{\infty} \frac{1}{2^n} \sin 2^n s, \quad s \in \mathbb{R},$$

satisfies the estimate

$$|f(s) - f(t)| \leq C|s - t| \log \frac{1}{|s - t|}$$

*for all $s, t \in \mathbb{R}$, with $0 < |s - t| < 1$, and hence provides an example of a function that is Hölder continuous of any order $\alpha < 1$. Prove that f is not Lipschitz continuous, and actually it is nowhere differentiable (see [**?**]).*

In what follows I° denotes the interior of I.

Theorem 6 *Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be convex. Then $f : I^\circ \to \mathbb{R}$ is locally Lipschitz.*

Proof. **Step 1:** Let $[a,b] \subset I$. We begin by proving that f is bounded in $[a,b]$. Let $M = \max\{f(a), f(b)\}$. If $x \in [a,b]$ then we may write $x = \theta a + (1-\theta) b$ for some $\theta \in [0,1]$, and so by the convexity of f,

$$f(x) = f(\theta a + (1-\theta) b) \le \theta f(a) + (1-\theta) f(b) \le \theta M + (1-\theta) M = M,$$

which shows that f is bounded from above. To see that f is also bounded from below, write $x = \frac{a+b}{2} + t$. Then

$$\begin{aligned} f\left(\frac{a+b}{2}\right) &= f\left(\frac{1}{2}\left(\frac{a+b}{2}+t\right) + \frac{1}{2}\left(\frac{a+b}{2}-t\right)\right) \\ &\le \frac{1}{2} f\left(\frac{a+b}{2}+t\right) + \frac{1}{2} f\left(\frac{a+b}{2}-t\right). \end{aligned}$$

Hence

$$\begin{aligned} f(x) = f\left(\frac{a+b}{2}+t\right) &\ge 2f\left(\frac{a+b}{2}\right) - f\left(\frac{a+b}{2}-t\right) \\ &\ge 2f\left(\frac{a+b}{2}\right) - M =: m, \end{aligned}$$

which shows that f is also bounded from below.

Step 2: Let $\varepsilon > 0$ be so small that $a-\varepsilon$ and $b+\varepsilon$ belong to I, and let

$$M := \sup_{[a-\varepsilon, b+\varepsilon]} f, \quad m := \inf_{[a-\varepsilon, b+\varepsilon]} f. \tag{2}$$

If x, $y \in [a,b]$ and $x \ne y$, define

$$z := y + \varepsilon \frac{y-x}{|y-x|}.$$

Then $z \in [a-\varepsilon, b+\varepsilon]$ and $y = \theta z + (1-\theta) x$, where

$$\theta = \frac{|y-x|}{\varepsilon + |y-x|}.$$

Hence by the convexity of f,

$$f(y) = f(\theta z + (1-\theta) x) \le \theta f(z) + (1-\theta) f(x) = \theta (f(z) - f(x)) + f(x),$$

or, equivalently,

$$\begin{aligned} f(y) - f(x) \le \theta (f(z) - f(x)) \le \theta (M-m) &= \frac{|y-x|}{\varepsilon + |y-x|}(M-m) \\ &\le \frac{M-m}{\varepsilon} |y-x|. \end{aligned}$$

By interchanging the roles of x and y we obtain

$$|f(y) - f(x)| \le \frac{M-m}{\varepsilon} |y-x|, \tag{3}$$

which shows that f is locally Lipschitz in the interior of I. ■

Remark 7 *Note that a convex function may not be continuous at the boundary points of its domain, since it may have upward jumps there.*

Wednesday, January 16, 2008

Next we prove that f is actually differentiable except for at most a countable number of points. Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be a function. For $x \in I^{\circ}$ we define the *left* and *right* derivatives of f at x (whenever they exist)

$$f'_{-}(x) := \lim_{y \to x^{-}} \frac{f(y) - f(x)}{y - x}, \quad f'_{+}(x) := \lim_{y \to x^{+}} \frac{f(y) - f(x)}{y - x}.$$

We have the following result.

Theorem 8 *Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be convex (respectively strictly convex). Then $f'_{-}(x)$ and $f'_{+}(x)$ exist in $\mathbb{R}$ for all $x \in I^{\circ}$ and the functions f'_{-} and f'_{+} are increasing (respectively strictly increasing).*

Proof. Consider four points $w < x < y < z$ in I° with, P, Q, R, S the corresponding points on the graph of f. By (1),

$$\text{slope } PQ \le \text{slope } PR \le \text{slope } QR \le \text{slope } QS \le \text{slope } RS, \tag{4}$$

with strict inequalities if f is strictly convex. Since slope $PR \le$slope QR, we have that slope QR increases as $x \nearrow y$, while slope RS decreases as $z \searrow y$. Thus the left-hand side of the inequality

$$\frac{f(x) - f(y)}{x - y} \le \frac{f(z) - f(y)}{z - y}$$

increases as $x \nearrow y$ and the right-hand side decreases as $z \searrow y$. Hence we have proved that $f'_{-}(y)$ and $f'_{+}(y)$ exist in $\mathbb{R}$ and satisfy

$$f'_{-}(y) \le f'_{+}(y). \tag{5}$$

Moreover, by (4),

$$f'_{+}(w) \le \frac{f(x) - f(w)}{x - w} \le \frac{f(y) - f(x)}{y - x} \le f'_{-}(y), \tag{6}$$

with the second inequality strict if f is strictly convex. Hence also from (5),

$$f'_{-}(w) \le f'_{+}(w) \le f'_{-}(y) \le f'_{+}(y),$$

which shows that f'_{-} and f'_{+} are increasing (respectively strictly increasing) in in I°. ■

Exercise 9 *What can you conclude if one of the endpoints of I belongs to I?*

Corollary 10 *Let $I \subset \mathbb{R}$ be an open interval and let $f : I \to \mathbb{R}$ be convex. Then f is differentiable except at most on a countable set $E \subset I$. Moreover, $f' : I \setminus E \to \mathbb{R}$ is continuous.*

Proof. Fix any $w \in I$. Then for all $w < x < y$ in I, by (6) (with x and y in place of w and x) and the monotonicity of f'_+,

$$f'_+(w) \leq f'_+(x) \leq \frac{f(y) - f(x)}{y - x}.$$

Since f'_+ is increasing and f is continuous in I, letting $x \searrow w$ gives

$$f'_+(w) \leq \lim_{x \to w^+} f'_+(x) \leq \lim_{x \to w^+} \frac{f(y) - f(x)}{y - x} = \frac{f(y) - f(w)}{y - w}.$$

Letting $y \searrow w$ yields

$$\lim_{x \to w^+} f'_+(x) = f'_+(w). \tag{7}$$

Similarly, for all $x < y < w$ in I, by (6),

$$\frac{f(y) - f(x)}{y - x} \leq f'_+(y) \leq f'_-(w).$$

Since f'_+ is increasing and f is continuous in I, letting $y \nearrow w$ yields

$$\frac{f(w) - f(x)}{w - x} = \lim_{y \to w^-} \frac{f(y) - f(x)}{y - x} \leq \lim_{y \to w^-} f'_+(y) \leq f'_-(w).$$

Letting $x \nearrow w$ gives

$$\lim_{y \to w^-} f'_+(y) = f'_-(w). \tag{8}$$

It now follows from (7) and (8) that $f'_-(w) = f'_+(w)$ if and only if f'_+ is continuous at w. Thus the set E consists of the discontinuity points of the increasing function f'_+. This proves that E is countable. Since f'_+ is continuous on $I \setminus E$ and $f' = f'_+$ on $I \setminus E$ we have that $f' : I \setminus E \to \mathbb{R}$ is continuous. ■

Exercise 11 *What can you conclude if one of the endpoints of I belongs to I?*

1.2 Characterizations

The previous result leads to the first characterization of convex functions.

Theorem 12 *Let $I \subset \mathbb{R}$ be an open interval and let $f : I \to \mathbb{R}$ be a function. Then f is convex (respectively strictly convex) if and only if there exists an increasing (respectively strictly increasing) function $g : I \to \mathbb{R}$ such that*

$$f(y) - f(x) = \int_x^y g(t)\, dt$$

for all $x < y$ in I. In particular, if f is convex, then

$$f(y) - f(x) = \int_x^y f'_-(t)\, dt = \int_x^y f'_+(t)\, dt \tag{9}$$

for all $x < y$ in I.

Here the integral can be taken either in the sense of Riemann or of Lebesgue.

Proof. Assume first that f is convex. Fix any $x < y$ in I and consider any partition $P := \{x_0, \ldots, x_n\}$ of $[x, y]$, that is

$$x = x_0 < x_1 < \ldots < x_n = y.$$

By (6),

$$f'_+ (x_{i-1}) \le \frac{f (x_i) - f (x_{i-1})}{x_i - x_{i-1}} \le f'_- (x_i) \tag{10}$$

for all $i = 1, \ldots, n$, and since

$$f (y) - f (x) = f (x_n) - f (x_0) = \sum_{i=1}^{n} f (x_i) - f (x_{i-1}),$$

we have

$$\sum_{i=1}^{n} f'_+ (x_{i-1}) (x_i - x_{i-1}) \le f (y) - f (x) \le \sum_{i=1}^{n} f'_- (x_i) (x_i - x_{i-1}).$$

It follows in particular that

$$\underline{\int_x^y} f'_- (t)\ dt \le \underline{\int_x^y} f'_+ (t)\ dt \le f (y) - f (x) \le \overline{\int_x^y} f'_- (t)\ dt \le \overline{\int_x^y} f'_+ (t)\ dt,$$

where $\underline{\int_x^y}$ and $\overline{\int_x^y}$ are the lower and upper Riemann integrals. Since f'_- and f'_+ are increasing, they are Riemann integrable in $[x, y]$, and so (9) holds.

Conversely assume that there exists an increasing function $g : I \to \mathbb{R}$ such that

$$f (y) - f (x) = \int_x^y g (t)\ dt$$

for all $x < y$ in I. Then for all $x < y$ in I and $\theta \in (0, 1)$,

$$\begin{aligned}
&\theta f (x) + (1 - \theta) f (y) - f (\theta x + (1 - \theta) y) \\
&\quad = -\theta (f (\theta x + (1 - \theta) y) - f (x)) + (1 - \theta) (f (y) - f (\theta x + (1 - \theta) y)) \\
&\quad = -\theta \int_x^{\theta x + (1-\theta) y} g (t)\ dt + (1 - \theta) \int_{\theta x + (1-\theta) y}^{y} g (t)\ dt \\
&\quad \ge -\theta \int_x^{\theta x + (1-\theta) y} g (\theta x + (1 - \theta) y)\ dt + (1 - \theta) \int_{\theta x + (1-\theta) y}^{y} g (\theta x + (1 - \theta) y)\ dt \\
&\quad = g (\theta x + (1 - \theta) y) [-\theta (\theta x + (1 - \theta) y - x) + (1 - \theta) (y - \theta x - (1 - \theta) y)] \\
&\quad = 0,
\end{aligned}$$

where we have used the fact that g is increasing. This shows that f is convex. Finally, we observe that if g is increasing, then the previous inequality becomes strict, and thus we have that f is strictly convex. ■

Corollary 13 *Let $I \subset \mathbb{R}$ be an open interval and let $f : I \to \mathbb{R}$ be a differential function. Then f is convex (respectively strictly convex) if and only if f' is increasing (respectively strictly increasing).*

Proof. We have already proved in Theorem 8 that if f is convex (respectively strictly convex), then f' is increasing (respectively strictly increasing). Conversely, assume that f' is increasing (respectively strictly increasing). Then by the fundamental theorem of calculus

$$f(y) - f(x) = \int_x^y f'(t)\, dt$$

for all $x < y$ in I. The result now follows from the previous theorem. ■

Friday, January 18, 2008

We now provide another characterization of convex functions in terms of tangent lines.

Definition 14 *Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be a function. We say that f is* subdifferentiable *at $x_0 \in I$ if there exists $m \in \mathbb{R}$, such that*

$$f(x) \geq f(x_0) + m(x - x_0) \quad \text{for all } x \in I.$$

The element m is called a subgradient *of f at x_0, and the set of all subgradients at x_0 is called the* subdifferential *of f at x_0 and is denoted by $\partial f(x_0)$. If f is not subdifferentiable at x_0, then $\partial f(x_0) := \emptyset$.*

Remark 15 *Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be a function. It follows from the definition of subdifferentiability that f attains a minimum at some point $x_0 \in I$ if and only if $0 \in \partial f(x_0)$.*

Theorem 16 *Let $I \subset \mathbb{R}$ be an open interval and let $f : I \to \mathbb{R}$ be a function. Then f is convex if and only if it is subdifferentiable at every $x_0 \in I$.*

Proof. If f is convex and $x_0 \in I$, choose $m \in \left[f'_-(x_0), f'_+(x_0)\right]$. Then by (6),

$$m \leq f'_+(x_0) \leq \frac{f(x) - f(x_0)}{x - x_0} \quad \text{if } x > x_0,$$

while

$$\frac{f(x) - f(x_0)}{x - x_0} \leq f'_-(x_0) \leq m \quad \text{if } x < x_0.$$

Hence, $f(x) - f(x_0) \geq m(x - x_0)$ for all $x \in I$, or, equivalently

$$f(x) \geq f(x_0) + m(x - x_0) \quad \text{for all } x \in I.$$

This shows that $\left[f'_-(x_0), f'_+(x_0)\right] \subset \partial f(x_0)$.

Conversely, assume that f is subdifferentiable in I. Let $x, y \in I$ and $\theta \in (0, 1)$. If

$$x_0 = \theta x + (1 - \theta) y \in I,$$

let $m \in \partial f(x_0)$. Then

$$\begin{aligned} f(\theta x + (1-\theta) y) = f(x_0) &= \theta [f(x_0) + m(x - x_0)] + (1 - \theta)[f(x_0) + m(y - x_0)] \\ &\leq \theta f(x) + (1 - \theta) f(y), \end{aligned}$$

which shows that f is convex. ■

Exercise 17 *Extend the previous result to an arbitrary interval I.*

Corollary 18 *Let $I \subset \mathbb{R}$ be an open interval and let $f : I \to \mathbb{R}$ be convex. Then*

$$\partial f(x_0) = \left[f'_-(x_0), f'_+(x_0)\right].$$

In particular, f is differentiable at $x_0 \in I$ if and only if $\partial f(x_0)$ is a singleton. In this case, $\partial f(x_0) = \{f'(x_0)\}$.

Proof. We have already shown in the first part of the previous proof that

$$\left[f'_-(x_0), f'_+(x_0)\right] \subset \partial f(x_0).$$

To prove the opposite inclusion, let $m \in \partial f(x_0)$. Then

$$f(x) - f(x_0) \geq m(x - x_0) \quad \text{for all } x \in I.$$

For $x_1 < x_0 < x_2$, we have

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0} \leq m \leq \frac{f(x_2) - f(x_0)}{x_2 - x_0}.$$

Letting $x_2 \searrow x_0$ and $x_1 \nearrow x_0$ gives

$$f'_-(x_0) \leq m \leq f'_+(x_0).$$

Hence $\partial f(x_0) = \left[f'_-(x_0), f'_+(x_0)\right]$. ■

Exercise 19 *Extend the previous result to an arbitrary interval I.*

Corollary 20 *Let $I \subset \mathbb{R}$ be an open interval and let $f : I \to \mathbb{R}$ be convex. Then the multifunction ∂f is increasing, that is if $x_1 < x_2$ are in I, then $s_1 \leq s_2$ for all $s_1 \in \partial f(x_1)$ and $s_2 \in \partial f(x_2)$.*

Proof. By the previous corollary and (6) we have that

$$s_1 \leq f'_+(x_1) \leq f'_-(x_2) \leq s_2.$$

■

Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be convex. The *domain* of the subdifferential ∂f of f is defined as

$$\operatorname{dom} \partial f := \{x \in I : \partial f(x) \neq \emptyset\}.$$

By Theorem 16 and Exercise 17

$$I^\circ \subset \operatorname{dom} \partial f \subset I. \tag{11}$$

Note that the application

$$x \in \operatorname{dom} \partial f \mapsto \partial f(x)$$

is a set-valued function that is single valued whenever f is differentiable.

To study the second derivative of a convex function we recall the following definition of differentiability at points that are not necessarily interior points.

Definition 21 *Let $E \subset \mathbb{R}$ and let $x_0 \in E$ be an accumulation point of E. Given a function $g : E \to \mathbb{R}$, we say that g is* differentiable *at x_0 if there exists in $\mathbb{R}$ the limit*

$$\lim_{x \to x_0} \frac{g(x) - g(x_0)}{x - x_0}.$$

In this case the limit is called derivative *of f at x_0 and is denoted $g'(x_0)$ or $\frac{dg}{dx}(x_0)$.*

Theorem 22 *Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be convex. Then $f''(x)$ exists for $\mathcal{L}^1$ a.e. $x \in I$ and f'' is nonnegative and locally (Lebesgue) integrable.*

Proof. By Theorem 8, the function $f'_+ : I^\circ \to \mathbb{R}$ is increasing. By the Lebesgue differentiation theorem, it follows that $\left(f'_+\right)'(x)$ exists for $\mathcal{L}^1$ a.e. $x \in I$ and that $\left(f'_+\right)'$ is nonnegative and locally integrable. Since $f' = f'_+$ except on a countable set $E \subset I$, we have that $f' : I \setminus E \to \mathbb{R}$ is differentiable $\mathcal{L}^1$ a.e. in $I \setminus E$ and f'' is nonnegative and locally integrable in $I \setminus E$. ■

Corollary 23 *Let $I \subset \mathbb{R}$ be an open interval and let $f : I \to \mathbb{R}$ be a twice differential function. Then f is convex if and only if $f'' \geq 0$. Moreover, if $f'' > 0$, then f is strictly convex.*

Proof. Under the present hypotheses, we have that f' is increasing if and only if $f'' \geq 0$ and that f' is strictly increasing if $f'' > 0$. Hence the desired result follows from the previous corollary. ■

Example 24 *The function $f(x) = x^4$ is strictly convex but $f''(0) = 0$.*

Next we prove that for convex functions the existence of the second derivative at a point is equivalent to the validity of a second order Taylor's formula.

The proof will make use of the following result.

Corollary 25 (Mean Value Theorem) *Let $f : [a, b] \to \mathbb{R}$ be a convex and continuous function. Then there exists $c \in (a, b)$ such that*

$$\frac{f(b) - f(a)}{b - a} \in \partial f(c).$$

Proof. As in the proof of the mean value theorem, consider the function

$$g(x) := f(x) - f(a) - \frac{f(b) - f(a)}{b-a}(x-a), \quad x \in [a,b].$$

The function g is convex, continuous, and $g(a) = g(b)$. Hence it has a minimum at some point $c \in (a,b)$. It follows by Remark 15 that $0 \in \partial g(c)$. On the other hand, by direct calculations,

$$g'_-(x) = f'_-(x) - \frac{f(b) - f(a)}{b-a}, \quad g'_+(x) = f'_+(x) - \frac{f(b) - f(a)}{b-a},$$

and so by the previous corollary

$$0 \in \partial g(c) = \left[f'_-(c) - \frac{f(b) - f(a)}{b-a}, f'_+(c) - \frac{f(b) - f(a)}{b-a} \right],$$

which implies that

$$\frac{f(b) - f(a)}{b-a} \in \left[f'_-(c), f'_+(c) \right] = \partial f(c).$$

■

Theorem 26 *Let $I \subset \mathbb{R}$ be an open interval and let $f : I \to \mathbb{R}$ be a convex function differentiable at some $x_0 \in I$. Then $f''(x_0)$ exists if and only if there exists $\ell \in \mathbb{R}$ such that*

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{\ell}{2}(x - x_0)^2 + o\left((x - x_0)^2\right) \tag{12}$$

for all x near x_0. In this case, $\ell = f''(x_0)$.

Proof. Let $F \subset I$ be the set in which f is differentiable. Assume that $f''(x_0)$ exists and fix $\varepsilon > 0$. Then there exists $\delta > 0$ such that

$$\left| \frac{f'(x) - f'(x_0)}{x - x_0} - f''(x_0) \right| \le \varepsilon$$

for all $x \in F$ with $|x - x_0| \le \delta$, or, equivalently,

$$-\varepsilon |x - x_0| \le f'(x) - f'(x_0) - f''(x_0)(x - x_0) \le \varepsilon |x - x_0|$$

for all $x \in F$ with $|x - x_0| \le \delta$. Integrate between x_0 and x and use the fundamental theorem of calculus to obtain

$$-\frac{\varepsilon}{2}(x - x_0)^2 \le f(x) - f(x_0) - f'(x_0)(x - x_0) - \frac{1}{2} f''(x_0)(x - x_0)^2 \le \frac{\varepsilon}{2}(x - x_0)^2.$$

Conversely, assume that (12) holds and fix $\varepsilon > 0$. Fix $\theta \in (0,1)$ and apply (12) with $x = x_0 + \theta h$ and $x = x_0 + h$ and subtract the two identities to obtain

$$f(x_0 + h) - f(x_0 + \theta h) = (1 - \theta) f'(x_0) h + \frac{1}{2}\ell\left(1 - \theta^2\right) h^2 + o\left(h^2\right)$$

for all h sufficiently small, that is,

$$\frac{f(x_0+h)-f(x_0+\theta h)}{(1-\theta)h} = f'(x_0) + \frac{1}{2}\ell(1+\theta)h + o(h)$$

for all h sufficiently small. By the mean value theorem there exists c between x_0+h and $x_0+\theta h$ such that

$$s := \frac{f(x_0+h)-f(x_0+\theta h)}{(1-\theta)h} \in \partial f(c).$$

Hence $s = f'(x_0) + \frac{1}{2}\ell(1+\theta) + o(h) \in \partial f(c)$. Assume that $h>0$. By the monotonicity property of the subgradient for all h but a countable number

$$f'(x_0+\theta h) \le s \le f'(x_0+h), \tag{13}$$

that is

$$\begin{aligned} \frac{f'(x_0+\theta h)-f'(x_0)}{\theta h} &\le \frac{s-f'(x_0)}{\theta h} = \frac{1}{2}\ell\frac{1+\theta}{\theta} + o(1), \\ \frac{f'(x_0+h)-f'(x_0)}{h} &\ge \frac{s-f'(x_0)}{h} = \frac{1}{2}\ell(1+\theta) + o(1). \end{aligned}$$

If $h<0$ then the inequalities in (13) are reversed, but dividing by h we obtain the same last two inequalities. Letting $h \to 0$ we obtain

$$\begin{aligned} \limsup_{x\to x_0} \frac{f'(x)-f'(x_0)}{x} &= \limsup_{h\to 0} \frac{f'(x_0+\theta h)-f'(x_0)}{\theta h} \le \frac{1}{2}\ell\frac{1+\theta}{\theta}, \\ \liminf_{x\to x_0} \frac{f'(x)-f'(x_0)}{x} &= \liminf_{h\to 0} \frac{f'(x_0+h)-f'(x_0)}{h} \ge \frac{1}{2}\ell(1+\theta). \end{aligned}$$

Letting $\theta \to 1^-$ gives

$$\lim_{x\to x_0} \frac{f'(x)-f'(x_0)}{x} = \ell.$$

■

Remark 27 *(i) Note that for a (nonconvex) function the previous theorem is false. Indeed, take*

$$f(x) := \begin{cases} x^3 & \text{if } x \text{ is rational,} \\ 0 & \text{otherwise.} \end{cases}$$

Then $f'(0)=0$ and formula (12) holds with $\ell=0$, but f is discontinuous in $\mathbb{R}\setminus\{0\}$, and so it not differentiable in $\mathbb{R}\setminus\{0\}$.

(ii) Reasoning as in the last part of the proof, one can also show that (12) implies that f'_+ and f'_+ are differentiable at x_0 with

$$\left(f'_+\right)'(x_0) = \left(f'_-\right)'(x_0) = \ell.$$

1.3 Operations Preserving Convexity

We begin with a simple result.

Theorem 28 *Let $I \subset \mathbb{R}$ be an interval, let $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$ be convex, and let $\alpha \geq 0$. Then $f + g$ and αf are convex.*

Proof. Since f and g are convex,

$$f\left(\theta x + (1-\theta)\,y\right) \leq \theta f\left(x\right) + (1-\theta)\,f\left(y\right),$$
$$g\left(\theta x + (1-\theta)\,y\right) \leq \theta g\left(x\right) + (1-\theta)\,g\left(y\right)$$

for all x, $y \in I$ and $\theta \in (0,1)$. The result now follows by summing the two inequalities and by multiplying the second by $\alpha \geq 0$. ■

The product of convex functions is not convex, in general.

Example 29 *The functions* $f(x) = x$, $g(x) = -x$, $x \in \mathbb{R}$, *are linear, and so convex, but their product,* $(fg)(x) = -x^2$, $x \in \mathbb{R}$, *is not convex.*

However, we have the following.

Theorem 30 *Let* $I \subset \mathbb{R}$ *be an interval, let* $f : I \to [0, \infty)$ *and* $g : I \to [0, \infty)$ *be convex and increasing (respectively decreasing). Then* fg *is convex and increasing (respectively decreasing).*

Proof. If $x < y$ are in I, then

$$(f(x) - f(y))(g(y) - g(x)) \leq 0,$$

or, equivalently,

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

and so for any $\theta \in (0, 1)$,

$$\begin{aligned} & f(\theta x + (1-\theta) y) g(\theta x + (1-\theta) y) \\ & \quad \leq [\theta f(x) + (1-\theta) f(y)] [\theta g(x) + (1-\theta) g(y)] \\ & \quad = \theta^2 f(x) g(x) + \theta (1-\theta) [f(x) g(y) + f(y) g(x)] + (1-\theta)^2 f(y) g(y) \\ & \quad \leq \theta^2 f(x) g(x) + \theta (1-\theta) [f(x) g(x) + f(y) g(y)] + (1-\theta)^2 f(y) g(y) \\ & \quad = \theta (\theta + 1 - \theta) f(x) g(x) + (1-\theta)(\theta + 1 - \theta) f(y) g(y) \\ & \quad = \theta f(x) g(x) + (1-\theta) f(y) g(y). \end{aligned}$$

Note that we have used heavily the fact that f and g are nonnegative. ■

Theorem 31 *Let* I, $J \subset \mathbb{R}$ *be intervals, let* $f : I \to \mathbb{R}$ *be convex with* $f(I) \subset J$, *and let* $g : J \to \mathbb{R}$ *be convex and increasing. Then* $g \circ f : I \to \mathbb{R}$ *is convex.*

Proof. For all x, $y \in I$ and $\theta \in (0, 1)$, we have

$$\begin{aligned} (g \circ f)(\theta x + (1-\theta) y) &= g(f(\theta x + (1-\theta) y)) \\ &\overset{g \text{ increasing}}{\leq} g(\theta f(x) + (1-\theta) f(y)) \\ &\overset{g \text{ convex}}{\leq} \theta g(f(x)) + (1-\theta) g(f(y)). \end{aligned}$$

■

Theorem 32 *Let* $I \subset \mathbb{R}$ *be an interval, let* $f_\alpha : I \to \mathbb{R}$, $\alpha \in \Lambda$, *be an arbitrary family of convex functions, and let*

$$f(x) := \sup_{\alpha \in \Lambda} f_a(x), \quad x \in I.$$

If $J := \{x \in I : f(x) < \infty\}$ *is nonempty, then* J *is an interval and* $f : J \to \mathbb{R}$ *is convex.*

Proof. Since f_α is convex, for all x, $y \in I$ and $\theta \in (0,1)$,

$$\begin{aligned} f(\theta x + (1-\theta) y) &= \sup_{\alpha \in \Lambda} f_\alpha (\theta x + (1-\theta) y) \le \sup_{\alpha \in \Lambda} [\theta f_\alpha (x) + (1-\theta) f_\alpha (y)] \\ &\le \theta \sup_{\alpha \in \Lambda} f_\alpha (x) + (1-\theta) \sup_{\alpha \in \Lambda} f_\alpha (y) = \theta f(x) + (1-\theta) f(y), \end{aligned}$$

which shows simultaneously that if $f(x)$ and $f(y)$ are finite then f is finite in the interval of endpoints x and y, so that J is an interval, and that f is convex.
■

The infimum of convex functions is not convex, in general.

Example 33 *The functions $f(x) = x$, $g(x) = -x$, $x \in \mathbb{R}$, are linear, and so convex, but their minimum, $\min\{f, g\}(x) = -|x|$, $x \in \mathbb{R}$, is not convex.*

Theorem 34 *Let $I \subset \mathbb{R}$ be an interval, let $f_n : I \to \mathbb{R}$, $n \in \mathbb{N}$, be a sequence of convex functions such that for every $x \in I$ there exists in $\mathbb{R}$ the limit*

$$f(x) := \lim_{n\to\infty} f_n(x).$$

Then $f : I \to \mathbb{R}$ is convex and $\{f_n\}$ converges uniformly to f on any closed interval of I°.

Proof. Step 1: We prove that f is convex. Since f_n is convex, for all x, $y \in I$ and $\theta \in (0,1)$,

$$\begin{aligned} f(\theta x + (1-\theta) y) &= \lim_{n\to\infty} f_n (\theta x + (1-\theta) y) \le \lim_{n\to\infty} [\theta f_n (x) + (1-\theta) f_n (y)] \\ &= \theta \lim_{n\to\infty} f_n (x) + (1-\theta) \lim_{n\to\infty} f_n (y) = \theta f(x) + (1-\theta) f(y), \end{aligned}$$

which shows that f is convex.
Step 2: Let $a < b$ be two points in I°. As in the proof of Theorem 6 we have that

$$2 f_n\left(\frac{a+b}{2}\right) - \max\{f_n(a), f_n(b)\} \le f_n(x) \le \max\{f_n(a), f_n(b)\}$$

for all $x \in [a,b]$ and all $n \in \mathbb{N}$. Hence,

$$m \le f_n(x) \le M$$

for all $x \in [a,b]$ and all $n \in \mathbb{N}$, where

$$M := \sup_{n\in\mathbb{N}} \max\{f_n(a), f_n(b)\}, \quad m := 2 \inf_{n\in\mathbb{N}} f_n\left(\frac{a+b}{2}\right) - M.$$

Step 3: Let $[a,b] \subset I^\circ$. In view of the previous step (applied to a larger interval $[a-\varepsilon, b+\varepsilon] \subset I^\circ$ and of Theorem 6 (see in particular (2) and (3)), we may find a constant $L > 0$ depending on $[a,b]$ but independent of n such that

$$|f_n(y) - f_n(x)| \le L |y - x|,$$

for all $n \in \mathbb{N}$ and for all x, $y \in [a, b]$. Note that letting $n \to \infty$ in the previous inequality shows that the same condition holds for f. To prove uniform convergence, we could invoke the Ascoli–Arzelá theorem, or prove it directly. Fix $\varepsilon > 0$ and consider any partition $P := \{x_0, \ldots, x_m\}$ of $[a, b]$, that is

$$a = x_0 < x_1 < \ldots < x_m = b,$$

such that $|x_i - x_{i-1}| \le \frac{\varepsilon}{3L}$ for all $i = 1, \ldots, m$. Find $N \in \mathbb{N}$ such that

$$|f_n(x_i) - f(x_i)| \le \frac{\varepsilon}{3}$$

for all $i = 1, \ldots, m$ and for all $n \ge N$. If $x \in [a, b]$ let $i \in \{1, \ldots, m\}$ be such that $x \in [x_{i-1}, x_i]$. Then for $n \ge N$,

$$\begin{aligned}|f_n(x) - f(x)| &= |f_n(x) \pm f_n(x_i) \pm f(x_i) - f(x)| \\ &\le |f_n(x) - f_n(x_i)| + |f_n(x_i) - f(x_i)| + |f(x_i) - f(x)| \\ &\le \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.\end{aligned}$$

Hence

$$\sup_{x \in [a,b]} |f_n(x) - f(x)| \le \varepsilon$$

for all $n \ge N$, which proves uniform convergence. ■

1.4 Differences of Convex Functions

In this subsection we assume for simplicity that

$$I := [a, b].$$

The set of convex functions $u : [a, b] \to \mathbb{R}$ is not a vector space since the difference of convex functions is not convex in general. We now study the smallest vector space of functions $u : [a, b] \to \mathbb{R}$ that contains all convex functions. To avoid anomalies at the endpoints we will restrict this space slightly. Let $BC[a, b]$ be the class of functions $f : [a, b] \to \mathbb{R}$ that can be written as $f = g - h$, where $g : [a, b] \to \mathbb{R}$ and $h : [a, b] \to \mathbb{R}$ are two convex functions such that $g'_+(a)$, $g'_-(b)$, $h'_+(a)$, $h'_-(b)$ are all finite.

Exercise 35 *Prove that $BC[a, b]$ is a vector space and that all its elements are Lipschitz.*

Theorem 36 *A function $f : [a, b] \to \mathbb{R}$ belongs to the space $BC[a, b]$ if and only if there exists a function $g : [a, b] \to \mathbb{R}$ of pointwise bounded variation such that*

$$f(x) = f(a) + \int_a^x g(t)\, dt$$

for all $x \in [a, b]$.

Proof. If $f \in BC\,[a,b]$, then $f = g - h$, where $g : [a,b] \to \mathbb{R}$ and $h : [a,b] \to \mathbb{R}$ are two convex functions such that $g'_+(a)$, $g'_-(b)$, $h'_+(a)$, $h'_-(b)$ are all finite. By a slight variation of Theorem 12,

$$g(x) = g(a) + \int_a^x p(t)\ dt, \quad h(x) = h(a) + \int_a^x q(t)\ dt$$

for all $x \in [a,b]$ and for some increasing functions $p : [a,b] \to \mathbb{R}$ and $q : [a,b] \to \mathbb{R}$. Hence

$$f(x) = g(x) - h(x) = f(a) + \int_a^x [p(t) - q(t)]\ dt$$

for all $x \in [a,b]$.

Conversely assume that there exists a function $g : [a,b] \to \mathbb{R}$ of pointwise bounded variation such that

$$f(x) = f(a) + \int_a^x g(t)\ dt$$

for all $x \in [a,b]$. Since any function of pointwise bounded variation may be written as the difference of two increasing functions, we may write $g = p - q$, where $p : [a,b] \to \mathbb{R}$ and $q : [a,b] \to \mathbb{R}$ are increasing. Hence

$$\begin{aligned} f(x) &= f(a) + \int_a^x p(t)\ dt - \int_a^x q(t)\ dt \\ &=: g(x) - h(x) \end{aligned}$$

for all $x \in [a,b]$, which shows that f is the difference of two convex functions. Moreover,

$$p(a) = \frac{\int_a^x p(a)\ dt}{x-a} \le \frac{g(x) - g(a)}{x-a} = \frac{\int_a^x p(t)\ dt}{x-a} \le \frac{\int_a^x p(x)\ dt}{x-a} \le p(x),$$

and so

$$-\infty < p(a) \le g'_-(a) := \lim_{x \to a^+} \frac{g(x) - g(a)}{x-a} \le \lim_{x \to a^+} p(x) < \infty.$$

Similarly $g'_-(b)$, $h'_+(a)$, $h'_-(b)$ are all finite. ■

We recall that:

Definition 37 *A function* $f : [a,b] \to \mathbb{R}$ *has* pointwise bounded variation *if*

$$\begin{aligned} \operatorname{Var} f := \sup \Bigg\{ & \sum_{i=1}^n |f(x_i) - f(x_{i-1})| : \\ & P := \{x_0, \ldots, x_n\} \ \textit{is a partition of}\ [a,b] \Bigg\} < \infty. \end{aligned}$$

It turns out that the space of functions of bounded variation is the smallest vector space that contains all monotone functions.

A similar characterization holds for functions in $BC[a,b]$.

Definition 38 *A function $f:[a,b]\to\mathbb{R}$, we define*

$$K(f) := \sup\left\{\sum_{i=1}^{n-1}\left|\square_{i+1}f-\square_i f\right| : \right.$$
$$\left. P=\{x_0,\ldots,x_n\} \text{ is a partition of } [a,b]\right\},$$

where

$$\square_i f := \frac{f(x_i)-f(x_{i-1})}{x_i-x_{i-1}}.$$

To highlight the dependence of $K(f)$ we will write $K_{[a,b]}(f)$. Also, given a partition $P=\{x_0,\ldots,x_n\}$, we write

$$K(f,P) := \sum_{i=1}^{n-1}\left|\square_{i+1}f-\square_i f\right|,$$

so that

$$K(f) = \sup_{P \text{ partition of } [a,b]} K(f,P).$$

Theorem 39 *If $f:[a,b]\to\mathbb{R}$ is convex and $f'_+(a)$, $f'_-(b)$ are finite, then*

$$K(f) = f'_-(b) - f'_+(a).$$

Proof. By (10) for any partition $P = \{x_0, \dots, x_n\}$,

$$f'_+ (x_{i-1}) \le \Box_i f \le f'_- (x_i),$$

and so

$$0 \le f'_+ (x_i) - f'_- (x_i) \le \Box_{i+1} f - \Box_i f \le f'_- (x_{i+1}) - f'_+ (x_{i-1}).$$

Hence

$$\begin{aligned}\sum_{i=1}^{n-1} |\Box_{i+1} f - \Box_i f| &= \sum_{i=1}^{n-1} \Box_{i+1} f - \Box_i f = \Box_n f - \Box_1 f \\ &= \frac{f(b) - f(x_{i-1})}{b - x_{i-1}} - \frac{f(x_1) - f(a)}{x_1 - a} \le f'_- (b) - f'_+ (a),\end{aligned}$$

which shows that $K(f) \le f'_- (b) - f'_+ (a)$.

To prove the converse inequality, find $a < x_1 < x_2 < b$ such that

$$\left| f'_- (b) - \frac{f(x_2) - f(b)}{x_2 - b} \right| \le \frac{\varepsilon}{2}, \quad \left| f'_+ (a) - \frac{f(x_1) - f(a)}{x_1 - a} \right| \le \frac{\varepsilon}{2}.$$

Then

$$\begin{aligned}f'_- (b) - f'_+ (a) - \varepsilon &\le \frac{f(x_2) - f(b)}{x_2 - b} - \frac{f(x_1) - f(a)}{x_1 - a} = \Box_3 f - \Box_1 f \pm \Box_2 f \\ &\le \sum_{i=1}^{2} |\Box_{i+1} f - \Box_i f| \le K(f).\end{aligned}$$

It suffices to let $\varepsilon \to 0^+$. ■

Remark 40 *If $f \in BC[a,b]$, then $f = g - h$, where $g : [a,b] \to \mathbb{R}$ and $h : [a,b] \to \mathbb{R}$ are two convex functions such that $g'_+ (a)$, $g'_- (b)$, $h'_+ (a)$, $h'_- (b)$ are all finite. Hence*

$$K(f) \le K(g) + K(h) = g'_- (b) - g'_+ (a) + h'_- (b) - h'_+ (a) < \infty.$$

Proving the opposite implication requires much more work.

Theorem 41 *Let $f : [a,b] \to \mathbb{R}$ be such that $K(f) < \infty$. Then f'_- exists in $(a,b]$, f'_+ exists in $[a,b)$, and f is Lipschitz.*

Lemma 42 *Let $n \in \mathbb{N}$ be greater than 2, let $a_0, \dots, a_n$ be real numbers, and let $b_1, \dots, b_n$ be positive numbers. Then*

$$\begin{aligned}&\sum_{i=1}^{n-1} \left| \frac{a_{i+1} - a_i}{b_{i+1}} - \frac{a_i - a_{i-1}}{b_i} \right| \\ &\quad \ge \frac{1}{b_n} \left(\sum_{i=1}^{n} b_i \right) \left| \frac{a_n - a_0}{\sum_{i=1}^{n} b_i} - \frac{a_{n-1} - a_0}{\sum_{i=1}^{n-1} b_i} \right| + \sum_{i=1}^{n-2} \left| \frac{a_{i+1} - a_0}{\sum_{j=1}^{i+1} b_j} - \frac{a_i - a_0}{\sum_{j=1}^{i} b_j} \right|\end{aligned}$$

Proof. The proof is by induction on n. For the case $n = 3$,

$$\begin{aligned}
&\left|\frac{a_3 - a_2}{b_3} - \frac{a_2 - a_1}{b_2}\right| + \left|\frac{a_2 - a_1}{b_2} - \frac{a_1 - a_0}{b_1}\right| \\
&= \left|\frac{a_3 - a_2}{b_3} - \frac{a_2 - a_1}{b_2}\right| + \left(\frac{b_1}{b_1 + b_2} + \frac{b_2}{b_1 + b_2}\right)\left|\frac{a_2 - a_0}{b_2} - \frac{(a_1 - a_0)(b_1 + b_2)}{b_1 b_2}\right| \\
&= \left|\frac{a_3 - a_2}{b_3} - \frac{a_2 - a_1}{b_2}\right| + \frac{b_1}{b_2}\left|\frac{a_2 - a_0}{b_1 + b_2} - \frac{a_1 - a_0}{b_1}\right| + \left|\frac{a_2 - a_0}{b_1 + b_2} - \frac{a_1 - a_0}{b_1}\right| \\
&\geq \left|\frac{a_3 - a_2}{b_3} - \frac{a_2 - a_1}{b_2} + \frac{b_1}{b_2}\frac{a_2 - a_0}{b_1 + b_2} - \frac{a_1 - a_0}{b_2}\right| + \left|\frac{a_2 - a_0}{b_1 + b_2} - \frac{a_1 - a_0}{b_1}\right| \\
&= \left|\frac{a_3 - a_2 \pm a_0}{b_3} + \frac{-b_2}{b_2 (b_1 + b_2)}(a_2 - a_0)\right| + \left|\frac{a_2 - a_0}{b_1 + b_2} - \frac{a_1 - a_0}{b_1}\right| \\
&= \left|\frac{a_3 - a_0}{b_3} - \frac{b_1 + b_2 + b_3}{b_3 (b_1 + b_2)}(a_2 - a_0)\right| + \left|\frac{a_2 - a_0}{b_1 + b_2} - \frac{a_1 - a_0}{b_1}\right| \\
&= \frac{b_1 + b_2 + b_3}{b_3}\left|\frac{a_3 - a_0}{b_1 + b_2 + b_3} - \frac{a_2 - a_0}{b_1 + b_2}\right| + \left|\frac{a_2 - a_0}{b_1 + b_2} - \frac{a_1 - a_0}{b_1}\right|,
\end{aligned}$$

where we have used the triangle inequality.

Assume now that the result is true for n and let's prove it for $n+1$. By the induction hypothesis,

$$\begin{aligned}
&\sum_{i=1}^{n}\left|\frac{a_{i+1} - a_i}{b_{i+1}} - \frac{a_i - a_{i-1}}{b_i}\right| \\
&\quad = \left|\frac{a_{n+1} - a_n}{b_{n+1}} - \frac{a_n - a_{n-1}}{b_n}\right| + \sum_{i=1}^{n-1}\left|\frac{a_{i+1} - a_i}{b_{i+1}} - \frac{a_i - a_{i-1}}{b_i}\right| \\
&\quad \geq \left|\frac{a_{n+1} - a_n}{b_{n+1}} - \frac{a_n - a_{n-1}}{b_n}\right| + \frac{1}{b_n}\left(\sum_{i=1}^{n} b_i\right)\left|\frac{a_n - a_0}{\sum_{i=1}^{n} b_i} - \frac{a_{n-1} - a_0}{\sum_{i=1}^{n-1} b_i}\right| \\
&\quad + \sum_{i=1}^{n-2}\left|\frac{a_{i+1} - a_0}{\sum_{j=1}^{i+1} b_j} - \frac{a_i - a_0}{\sum_{j=1}^{i} b_j}\right| \\
&\quad = \left|\frac{a_{n+1} - a_n}{b_{n+1}} - \frac{a_n - a_{n-1}}{b_n}\right| + \frac{1}{b_n}\left(\sum_{i=1}^{n-1} b_i\right)\left|\frac{a_n - a_0}{\sum_{i=1}^{n} b_i} - \frac{a_{n-1} - a_0}{\sum_{i=1}^{n-1} b_i}\right| \\
&\quad + \sum_{i=1}^{n-1}\left|\frac{a_{i+1} - a_0}{\sum_{j=1}^{i+1} b_j} - \frac{a_i - a_0}{\sum_{j=1}^{i} b_j}\right| \\
&\quad =: I + II + III.
\end{aligned}$$

Now, again by the triangle inequality

$$
\begin{aligned}
&I + II \geq \\
&\left| \frac{a_{n+1} - a_n}{b_{n+1}} - \frac{a_n - a_{n-1}}{b_n} + \frac{1}{b_n} \left(\sum_{i=1}^{n-1} b_i \right) \frac{a_n - a_0}{\sum_{i=1}^{n} b_i} - \frac{1}{b_n} \left(\sum_{i=1}^{n-1} b_i \right) \frac{a_{n-1} - a_0}{\sum_{i=1}^{n-1} b_i} \right| \\
&= \left| \frac{a_{n+1} - a_0}{b_{n+1}} + \left(-\frac{1}{b_{n+1}} - \frac{1}{b_n} + \frac{\sum_{i=1}^{n-1} b_i \pm b_n}{b_n \sum_{i=1}^{n} b_i} \right) (a_n - a_0) \right| \\
&= \left| \frac{a_{n+1} - a_0}{b_{n+1}} - \frac{\sum_{i=1}^{n+1} b_i}{b_{n+1} \left(\sum_{i=1}^{n} b_i \right)} (a_n - a_0) \right| \\
&= \frac{1}{b_{n+1}} \left(\sum_{i=1}^{n+1} b_i \right) \left| \frac{a_{n+1} - a_0}{\sum_{i=1}^{n+1} b_i} - \frac{a_n - a_0}{\sum_{i=1}^{n} b_i} \right| .
\end{aligned}
$$

■

Lemma 43 *Let $n \in \mathbb{N}$ be greater than 2, let $a_0, \ldots, a_n$ be real numbers, and let*

$$c_0 \leq \cdots \leq c_n.$$

Then

$$\sum_{i=1}^{n-1} \left| \frac{a_{i+1} - a_i}{c_{i+1} - c_i} - \frac{a_i - a_{i-1}}{c_i - c_{i-1}} \right| \geq \sum_{i=1}^{n-1} \left| \frac{a_{i+1} - a_0}{c_{i+1} - c_0} - \frac{a_i - a_0}{c_i - c_0} \right|.$$

Proof. Define $b_i := c_i - c_{i-1}$ for $i = 1, \ldots, n$ and apply the previous lemma to obtain

$$\begin{aligned}
&\sum_{i=1}^{n-1} \left| \frac{a_{i+1} - a_i}{c_{i+1} - c_i} - \frac{a_i - a_{i-1}}{c_i - c_{i-1}} \right| \\
&\geq \frac{1}{(c_n - c_{n-1})} \left(\sum_{i=1}^{n} (c_i - c_{i-1}) \right) \left| \frac{a_n - a_0}{\sum_{i=1}^{n} (c_i - c_{i-1})} - \frac{a_{n-1} - a_0}{\sum_{i=1}^{n-1} (c_i - c_{i-1})} \right| \\
&+ \sum_{i=1}^{n-2} \left| \frac{a_{i+1} - a_0}{\sum_{j=1}^{i+1} (c_j - c_{j-1})} - \frac{a_i - a_0}{\sum_{j=1}^{i} (c_j - c_{j-1})} \right| \\
&= \frac{1}{(c_n - c_{n-1})} \left(c_n - c_{n-1} + \sum_{i=1}^{n-1} (c_i - c_{i-1}) \right) \left| \frac{a_n - a_0}{c_n - c_0} - \frac{a_{n-1} - a_0}{c_{n-1} - c_0} \right| \\
&+ \sum_{i=1}^{n-2} \left| \frac{a_{i+1} - a_0}{c_{i+1} - c_0} - \frac{a_i - a_0}{c_i - c_0} \right| \\
&\geq \left| \frac{a_n - a_0}{c_n - c_0} - \frac{a_{n-1} - a_0}{c_{n-1} - c_0} \right| + \sum_{i=1}^{n-2} \left| \frac{a_{i+1} - a_0}{c_{i+1} - c_0} - \frac{a_i - a_0}{c_i - c_0} \right|,
\end{aligned}$$

where we have used the fact that $\sum_{i=1}^{n-1} (c_i - c_{i-1}) \geq 0$. ■

Proof of Theorem 41. Step 1: We prove that for every $x_1 \in [a, b)$,

$$\limsup_{x \to x_1^+} \left| \frac{f(x) - f(x_1)}{x - x_1} \right| < \infty.$$

Fix $a \leq x_1 < x_2 < b$ and choose $x_2 < x_3 < b$. Let $a = x_0 \leq x_1 < x_2 < x_3 < x_4 = b$. Since

$$|\Box_2 f| - |\Box_4 f| = |\Box_2 f| - |\Box_4 f| \pm |\Box_3 f| \leq |\Box_3 f - \Box_2 f| + |\Box_4 f - \Box_3 f| \leq K(f),$$

we have

$$\left| \frac{f(x_2) - f(x_1)}{x_2 - x_1} \right| \leq K(f) + \left| \frac{f(x_4) - f(x_3)}{x_4 - x3} \right|. \tag{14}$$

Letting $x_2 \to x_1^+$ gives the desired result. Similarly, for any $x_1 \in (a, b]$,

$$\limsup_{x \to x_1^-} \left| \frac{f(x) - f(x_1)}{x - x_1} \right| < \infty.$$

Step 2: We prove that f'_+ exists in $[a,b)$. Assume by contradiction that there exists $x_0 \in [a,b)$ for which the limit

$$f'_+(x_0) := \lim_{x \to x_0^+} \frac{f(x) - f(x_0)}{x - x_0}$$

does not exists. Hence there exists an $\varepsilon_0 > 0$ with the property that for any $x \in (x_0, b)$ there exists $y \in (x_0, x)$ such that

$$\left| \frac{f(x) - f(x_0)}{x - x_0} - \frac{f(y) - f(x_0)}{y - x_0} \right| \geq \varepsilon_0.$$

By applying the previous inequality repeatedly for any $n \geq 2$ we may construct

$$c_0 = x_0 < c_1 \leq \ldots \leq c_n$$

such that

$$\sum_{i=1}^{n-1} \left| \frac{f(c_{i+1}) - f(x_0)}{c_{i+1} - x_0} - \frac{f(c_i) - f(x_0)}{c_i - x_0} \right| \geq (n-1)\,\varepsilon_0.$$

By the previous lemma we obtain

$$K(f) \geq \sum_{i=1}^{n-1} |\square_{i+1} f - \square_i f| = \sum_{i=1}^{n-1} \left| \frac{f(c_{i+1}) - f(c_i)}{c_{i+1} - c_i} - \frac{f(c_i) - f(c_{i-1})}{c_i - c_{i-1}} \right| \geq (n-1)\,\varepsilon_0,$$

which gives a contradiction as $n \to \infty$.
Step 3: We show that f is Lipschitz. Fix $a \leq x_1 < x_2 < b$ and choose $x_2 < x_3 < b$. By (14),

$$\left| \frac{f(x_2) - f(x_1)}{x_2 - x_1} \right| \leq K(f) + \left| \frac{f(b) - f(x_3)}{b - x_3} \right|.$$

Letting $x_3 \to b^-$ gives

$$\left| \frac{f(x_2) - f(x_1)}{x_2 - x_1} \right| \leq K(f) + \left| f'_-(b) \right|.$$

Thus we have proved that

$$|f(x_2) - f(x_1)| \leq \left(K(f) + \left| f'_-(b) \right| \right) |x_2 - x_1|$$

for all $x_1 < x_2$ in $[a,b)$. Since $f'_-(b) \in \mathbb{R}$, we have that f is continuous at b, and so the previous inequality holds also for $x_2 = b$. ■

We are now ready to prove the desired characterization.

Theorem 44 *A function $f : [a,b] \to \mathbb{R}$ belongs to $BC[a,b]$ if and only if $K(f) < \infty$.*

Proof. Step 1: Let $P = \{x_0, \ldots, x_n\}$ be any partition of $[a,b]$ and let $\bar{x} \in [a,b]$. We claim that the partition Q obtained by adding $\bar{x}$ to P is such that

$$K(f,P) \le K(f,Q).$$

To see this, let $i \in \{1, \ldots, n\}$ be such that $\bar{x} \in [x_{i-1}, x_i]$. If $\bar{x}$ coincides with one of the endpoints, then there is nothing to prove. Thus we can assume that $\bar{x} \in (x_{i-1}, x_i)$. Let

$$\Box_r f := \frac{f(x_i) - f(\bar{x})}{x_i - \bar{x}}, \quad \Box_l f := \frac{f(\bar{x}) - f(x_{i-1})}{\bar{x} - x_{i-1}}.$$

Then

$$\begin{aligned}\Box_i f = \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}} &= \frac{f(x_i) \pm f(\bar{x}) - f(x_{i-1})}{x_i - x_{i-1}} \\ &= \theta \Box_r f + (1-\theta)\Box_l f,\end{aligned}$$

where

$$\theta := \frac{x_i - \bar{x}}{x_i - x_{i-1}} \in (0,1).$$

Assume that $2 \le i \le n-1$ (the cases $i = 1$ and $i = 2$ are simpler). Then

$$\begin{aligned}|\Box_{i+1} f - \Box_i f| + |\Box_i f - \Box_{i-1} f| &= |\Box_{i+1} f - \Box_r f + (1-\theta)(\Box_r f - \Box_l f)| \\ &\quad + |\theta(\Box_r f - \Box_l f) + \Box_l f - \Box_{i-1} f| \\ &\le |\Box_{i+1} f - \Box_r f| + (1-\theta)|\Box_r f - \Box_l f| \\ &\quad + \theta|\Box_r f - \Box_l f| + |\Box_l f - \Box_{i-1} f| \\ &= |\Box_{i+1} f - \Box_r f| + |\Box_r f - \Box_l f| + |\Box_l f - \Box_{i-1} f|.\end{aligned}$$

Since the remaining terms in $K(f,P)$ and in $K(f,Q)$ are the same, this completes the proof of the claim.

Note that Step 1 implies in particular that the function $x \in [a,b] \mapsto K_{[a,x]}(f)$ is increasing and that there exists an increasing sequence $\{P_n\}$ of partitions of $[a,b]$ such that

$$\lim_{n\to\infty} K(f, P_n) = K(f).$$

■

Proof. Step 2: In view of Remark 40, it remains to show that if $K(f) < \infty$ then $f \in BC[a,b]$. Define the function

$$g(x) := \begin{cases} f'_-(x) - f'_+(a) & \text{if } a < x \leq b, \\ 0 & \text{if } x = a. \end{cases}$$

Fix $a < s < t \leq b$. Since by Step 1 of the previous theorem refinement of a partition increases $K(f,P)$, in view of the previous remark we can construct two increasing sequences $\{P_n\}$ and $\{Q_n\}$ of partitions of $[a,t]$ and of $[a,s]$, respectively, such that $Q_n \subset P_n$ for all $n \in \mathbb{N}$,

$$\lim_{n\to\infty} K(f,P_n) = K_{[a,t]}(f), \quad \lim_{n\to\infty} K(f,Q_n) = K_{[a,s]}(f), \tag{15}$$

and the size of the maximum interval of P_n goes to zero as $n \to \infty$. For every n write $Q_n := \left\{x_0^{(n)}, \ldots, x_{\ell_n}^{(n)}\right\}$ and $P_n := \left\{x_0^{(n)}, \ldots, x_{\ell_n}^{(n)}, x_{\ell_n+1}^{(n)}, \ldots, x_{m_n}^{(n)}\right\}$, where

$$x = x_0^{(n)} < x_1^{(n)} < \ldots < x_{\ell_n}^{(n)} = s < x_{\ell+1}^{(n)} < \ldots < x_{m_n}^{(n)} = t.$$

Then

$$\begin{aligned} K(f,P_n) - K(f,Q_n) &= \sum_{i=\ell_n}^{m_n-1} |\Box_{i+1} f - \Box_i f| \geq \left| \sum_{i=\ell_n}^{m_n-1} \Box_{i+1} f - \Box_i f \right| \\ &= |\Box_{m_n} f - \Box_{\ell_n} f| \\ &= \left| \frac{f(t) - f\left(x_{m_n-1}^{(n)}\right)}{t - x_{m_n-1}^{(n)}} - \frac{f(s) - f\left(x_{\ell_n-1}^{(n)}\right)}{s - x_{\ell_n-1}^{(n)}} \right|. \end{aligned}$$

Letting $n \to \infty$ and using (15) and the fact that $x_{m_n-1}^{(n)} \to t^-$ and $x_{\ell_n-1}^{(n)} \to s^-$, we obtain

$$K_{[a,t]}(f) - K_{[a,s]}(f) \geq \left| f'_-(t) - f'_-(s) \right| = |g(t) - g(s)|.$$

Note that this inequality continues to hold if $s = a$ since reasoning as before

$$K(f,P_n) \geq |\Box_{m_n} f - \Box_1 f| = \left| \frac{f(t) - f\left(x_{m_n-1}^{(n)}\right)}{t - x_{m_n-1}^{(n)}} - \frac{f(a) - f\left(x_1^{(n)}\right)}{a - x_1^{(n)}} \right|,$$

and so letting $n \to \infty$ and using (15) and the fact that $x_{m_n-1}^{(n)} \to t^-$ and $x_1^{(n)} \to a^+$, we obtain

$$K_{[a,t]}(f) - K_{[a,a]}(f) = K_{[a,t]}(f) - 0 \geq \left| f'_-(t) - f'_+(a) \right| = |g(t) - g(a)|.$$

Thus, we have proved that for all $a \leq s < t \leq b$,

$$K_{[a,t]}(f) - K_{[a,s]}(f) \geq \pm (g(t) - g(s)),$$

which implies that the functions $x \mapsto K_{[a,x]}(f) - g(x)$ and $x \mapsto K_{[a,x]}(f) + g(x)$ are increasing. It follows that the function

$$g(x) = \frac{1}{2}\left(K_{[a,x]}(f) + g(x)\right) - \frac{1}{2}\left(K_{[a,x]}(f) - g(x)\right), \quad x \in [a,b],$$

has pointwise bounded variation. Since f is Lipschitz continuous, it follows that f' exists for $\mathcal{L}^1$ a.e. $x \in [a,b]$ and that the fundamental theorem of calculus holds, so that

$$f(x) - f(a) = \int_a^x f'(t)\, dt = \int_a^x \left[g(t) + f'_+(a)\right] dt, \quad x \in [a,b].$$

It follows by Theorem 36 that $f \in BC[a,b]$. ■

1.5 Conjugate Functions

Let $g : [0,\infty) \to [0,\infty)$ be strictly increasing and continuous, with $g(0) = 0$ and $g(x) \to \infty$ as $x \to \infty$. Then g^{-1} exists and has the same property of g. Moreover, if we define the functions

$$f(x) := \int_0^x g(t)\, dt, \quad f^*(x) := \int_0^x g^{-1}(t)\, dt, \quad x \geq 0,$$

in view of Theorem 12 we have that f and f^* are convex. Looking at the picture we have that:

(A_1) **(Young inequality)** $xy \leq f(x) + f^*(y)$ for all x, $y \geq 0$;

(A_2) $xy = f(x) + f^*(y)$ if and only if $y = g(x) = f'(x)$;

(A_3) $(f^*)' = (f')^{-1}$;

(A_4) $f^{**} = f$;

(A_5) $f^*(y) = \sup_{x \geq 0} \{xy - f(x)\}$ for all $y \geq 0$.

The function f^* is called the conjugate of f. To extend this notion to general functions we will use (A_5) as a definition.

Definition 45 *Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be a function. The* conjugate *of f is the function defined by*

$$f^*(y) := \sup_{x \in I} \{xy - f(x)\},$$

with domain $I^* := \{y \in \mathbb{R} : f^*(y) < \infty\}$.

We will show that f^* is a convex and that it is closed.

Definition 46 *Let $J \subset \mathbb{R}$ be an interval and let $g : J \to \mathbb{R}$ be a convex function. Then g is* closed *if the set*

$$\{x \in J : \, g(x) \leq t\}$$

is a closed set for all $t \in \mathbb{R}$.

Exercise 47 *Let $J \subset \mathbb{R}$ be an interval and let $g : J \to \mathbb{R}$ be a convex function. Prove that g is closed if and only if it is continuous at each end point contained in J (if any) and such that $g(x) \to \infty$ as x approaches every (real) endpoint not contained in J (if any).*

Theorem 48 *Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be convex. Then I^* is an interval and $f^* : I^* \to \mathbb{R}$ is convex and closed.*

Proof. If $I = \{x_0\}$, then $f^*(y) = x_0 y - f(x_0)$, and so $I^* = \mathbb{R}$. If I does not consist of a single point, let $x_0 \in I^\circ$. By Corollary 18,

$$\partial f(x_0) = \left[f'_-(x_0), f'_+(x_0)\right].$$

Let $y \in \partial f(x_0)$. Then

$$f(x) \geq f(x_0) + y(x - x_0) \quad \text{for all } x \in I,$$

and so $xy - f(x) \leq x_0 y - f(x_0)$ for all $x \in I$. Hence,

$$f^*(y) = \sup_{x \in I} \{xy - f(x)\} \leq x_0 y - f(x_0) < \infty,$$

which shows that $y \in I^*$.

For every $x \in I$ define the affine function

$$g^x(y) := xy - f(x), \quad y \in \mathbb{R}.$$

Since g^x is convex and

$$f^*(y) = \sup_{x \in I} g^x(y), \quad y \in \mathbb{R},$$

it follows by Theorem 32 that I^* is an interval and that f^* is convex.

It remains to prove that f^* is closed. Fix $t \in \mathbb{R}$ and let $\{y_n\} \subset I^*$ be such that $y_n \to y$ and $f^*(y_n) \leq t$ for all $n \in \mathbb{N}$. We claim that $f^*(y) \leq t$. Since, $f^*(y_n) \leq t$ for all $n \in \mathbb{N}$, we have that

$$xy_n - f(x) \leq t$$

for all $x \in I$. Letting $n \to \infty$ gives

$$xy - f(x) \leq t$$

for all $x \in I$, and so $f^*(y) \leq t$. Hence the set $\{y \in I^* : f^*(y) \leq t\}$ is a closed set for all $t \in \mathbb{R}$. ■

Next we prove the appropriate versions of (A_1) and (A_2).

Theorem 49 (Young Inequality) *Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be convex. Then*

$$xy \leq f(x) + f^*(y)$$

for all $x \in I$ and all $y \in I^*$ and the equality holds if and only $y \in \partial f(x)$.

Proof. If $y \in I^*$, then

$$xy - f(x) \leq f^*(y)$$

for all $x \in I$, and so Young's inequality holds.

To prove the second statement of the proof, note that $y_0 \in \partial f(x_0)$ if and only if

$$f(x) \geq f(x_0) + y_0(x - x_0) \quad \text{for all } x \in I,$$

or equivalently

$$y_0 x_0 - f(x_0) \geq y_0 x - f(x) \quad \text{for all } x \in I.$$

In turn, this is equivalent to saying that

$$y_0 x_0 - f(x_0) \geq f^*(y_0).$$

It now follows by the Young inequality that $y_0 \in \partial f(x_0)$ if and only if $y_0 x_0 = f(x_0) + f^*(y_0)$. ■

To extend (A_3) we need some preliminary results.

Definition 50 *Given two nonempty sets X, Y, a* multifunction *or* correspondence *from X to Y is a map from X to the family of subsets of Y, namely*

$$\Gamma : X \to \mathcal{P}(Y).$$

The domain *of a multifunction is the set*

$$\operatorname{dom}\Gamma := \{x \in X : \Gamma(x) \neq \emptyset\}.$$

The graph *of a multifunction Γ is the set*

$$\operatorname{graph}\Gamma := \{(x, y) \in \operatorname{dom}\Gamma \times Y : y \in \Gamma(x)\}.$$

The inverse *of a multifunction Γ is the multifunction $\Gamma^{-1} : Y \to \mathcal{P}(X)$ defined by*

$$\Gamma^{-1}(y) := \{x \in X : y \in \Gamma(x)\}. \tag{16}$$

A multifunction $\Gamma : X \to \mathcal{P}(Y)$ is *univalued* on a set $E \subset X$ if $\Gamma(x)$ consists of at most one element for every $x \in E$. In this case the restriction of Γ to $E \cap \operatorname{dom}\Gamma$ may be identified with a function.

Definition 51 *A multifunction $\Gamma : \mathbb{R}^N \to \mathcal{P}(\mathbb{R}^N)$ is called* monotone *if*

$$(x_2 - x_1) \cdot (y_2 - y_1) \geq 0$$

for all (x_1, y_1), $(x_2, y_2) \in \operatorname{graph}\Gamma$. A monotone $\Gamma : \mathbb{R}^N \to \mathcal{P}(\mathbb{R}^N)$ is called maximal *if its graph is not a proper subset of the graph of a monotone multifunction.*

Note that Γ is a (maximal) monotone multifunction if and only if Γ^{-1} is a (maximal) monotone multifunction.

Theorem 52 *Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be convex. Then the multifunction ∂f is monotone.*[1]

Proof. Let (x_1, m_1), $(x_2, m_2) \in \operatorname{graph} \partial f$. Without loss of generality assume that $x_1 < x_2$. Then by Theorem 8 and Corollary 18,

$$f'_-(x_1) \leq m_1 \leq f'_+(x_1) \leq f'_-(x_2) \leq m_2 \leq f'_+(x_2),$$

which implies that $m_2 - m_1 \geq 0$. In turn, $(x_2 - x_1)(m_2 - m_1) \geq 0$, and so ∂f is monotone. ■

[1] Here we set $\partial f(x) := \varnothing$ if either $x \in I$ and f is not subdifferentiable at x or if x does not belong to I.

Theorem 53 *Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be closed and convex. Then the multifunction ∂f is maximal.*

Proof. To prove maximality, it suffices to show that if $(x_1, m_1) \notin \operatorname{graph} \partial f$, then there exists $(x_2, m_2) \in \operatorname{graph} \partial f$ such that

$$(x_2 - x_1)(y_2 - y_1) < 0.$$

Replacing f with the function

$$g(x) := f(x + x_1) - x m_1,$$

which is still convex and closed (see Theorem 30, and Exercise 47), without loss of generality we may assume that $x_1 = m_1 = 0$. Thus, we have to prove that if $(0,0) \notin \operatorname{graph} \partial f$, then there exists $(x, m) \in \operatorname{graph} \partial f$ such that $xm < 0$.

There are now three cases. If $x < 0$ for all $x \in \operatorname{dom} \partial f$, then by (11) the interval I is bounded from above. Let $b := \sup I \leq 0$. If $b \notin I$, then by Exercise 47,

$$\lim_{x \to b^-} f(x) = \infty$$

and so by (5) there must exists $x < 0$ such that $f'_+(x) > 0$ (since otherwise f would be bounded from above). Since $\left(x, f'_+(x)\right) \in \operatorname{graph} \partial f$ and $x f'_+(x) < 0$, the proof is concluded in the case in which $b \notin I$. If $b \in I$ and $f'_-(b) = \infty$, then (why?)

$$\lim_{y \to w^-} f'_+(y) = \infty,$$

and so we are back to the previous case. If $b \in I$ and $f'_-(b) < \infty$, then $b \in \operatorname{dom} \partial f$, so $b < 0$, and (why?)

$$\partial f(b) = \left[f'_-(b), \infty\right).$$

Hence we may find $m \in \partial f(b) \cap (0, \infty)$. Then $(b, m) \in \operatorname{graph} \partial f$ and $bm < 0$.

A similar argument holds if $x > 0$ for all $x \in \operatorname{dom} \partial f$.

Thus, it remains to consider the case in which $0 \in \operatorname{dom} \partial f$. Since $(0,0) \notin \operatorname{graph} \partial f$, it follows that $0 \notin \partial f(0)$, and so by Remark 15, $f(0)$ cannot be a minimum for f. Hence there exists $x_1 \in I$ such that $f(x_1) < f(0)$. If $x_1 < 0$, then by (9) we can find $x_2 \in [x_1, 0)$ such that $f'_+(x_1) > 0$, and so we are back to the previous case. Similarly, if $x_1 > 0$, then by (9) we can find $x_2 \in (0, x_1, 0]$ such that $f'_+(x_1) < 0$. ■

We are now ready to extend (A_3) and (A_4). Note that since $f^{**} = (f^*)^*$, the function f^{**} is convex and closed. Thus to recover (A_4) we need f to be convex and closed.

Theorem 54 *Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be convex. Then*

$$\operatorname{graph} (\partial f)^{-1} \subset \operatorname{graph} \partial f^*.$$

Moreover, if f is also closed, then

$$(\partial f)^{-1} = \partial f^{*}. \tag{17}$$

and $f^{**} = f$.

Proof. Fix $x \in I$. From the definition of f^{*},

$$f^{*}(y) - xy \geq -f(x) \quad \text{for all } y \in I^{*}.$$

Hence the function on the right-hand side is minimized (as a function of y) when there is equality in the Young inequality, that is when $y \in \partial f(x)$. Thus, if $y \in \partial f(x)$, then the function $h(z) := f^{*}(z) - xz$, $z \in I^{*}$, is minimized at $z = y$. In view of Remark 15, we have that $0 \in \partial h(y)$, which is equivalent to say that $x \in \partial f^{*}(y)$. Hence we have proved that if $y \in \partial f(x)$, then $x \in \partial f^{*}(y)$. In terms of graphs this means that if $(x, y) \in \operatorname{graph} \partial f$, then $(y, x) \in \operatorname{graph} \partial f^{*}$, in other words, $\operatorname{graph}(\partial f)^{-1} \subset \operatorname{graph} \partial f^{*}$.

If we assume that f is also closed, then by the previous theorem, ∂f is maximal monotone, and so is $(\partial f)^{-1}$. Since ∂f^{*} is also maximal monotone, it follows that $(\partial f)^{-1} = \partial f^{*}$.

It remains to prove that $f^{**} = f$. Applying (17) with f^{*} in place of f we have that

$$\partial f^{**} = (\partial f^{*})^{-1} = \left((\partial f)^{-1}\right)^{-1} = \partial f.$$

It follows by Corollary 18 that

$$(f^{**})'_{-} = f'_{-},$$

and so by (9),

$$f^{**}(y) - f^{**}(x) = \int_x^y (f^{**})'_{-}(t)\, dt = \int_x^y f'_{-}(t)\, dt = f(y) - f(x)$$

for all $x < y$ in I. To conclude the proof, we prove that $f(x_0) = f^{*}(x_0)$ at some point $x_0 \in I$. Choose $x_0 \in I$ and $y_0 \in I^{*}$ such that $y_0 \in \partial f(x_0)$. Then

$$(y_0, x_0) \in \operatorname{graph}(\partial f)^{-1} = \operatorname{graph} \partial f^{*}.$$

Applying the equality in the Young inequality first to f and then to f^{*}, we obtain

$$\begin{aligned} x_0 y_0 &= f(x_0) + f^{*}(y_0), \\ y_0 x_0 &= f^{*}(y_0) + f^{**}(x_0), \end{aligned}$$

which implies that $f(x_0) = f^{*}(x_0)$. ■

Monday, February 11, 2008

As a corollary of the previous theorem we can study the regularity of f^*.

Theorem 55 *Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be convex and closed. Then $f^* : I^* \to \mathbb{R}$ is differentiable in $(I^*)^\circ$ if and only if f is strictly convex in all intervals contained in*

$$\bigcup_{y \in (I^*)^\circ} \partial f^* (y) .$$

Proof. Assume that f is strictly convex in all intervals contained in

$$E := \bigcup_{y \in (I^*)^\circ} \partial f^* (y) ,$$

let $y_0 \in (I^*)^\circ$ and assume by contradiction that f^* is not differentiable at y_0. Then

$$(f^*)'_- (y_0) < (f^*)'_+ (y_0) .$$

Let $(f^*)'_- (y_0) \le x_1 < x_2 < (f^*)'_+ (y_0)$. Since x_1, $x_2 \in \partial f^* (y_0)$, it follows by the previous theorem that $y_0 \in \partial f (x_1) \cap \partial f (x_1)$. By the equality case in the Young inequality, we have

$$x_1 y_0 = f (x_1) + f^* (y_0) , \quad x_2 y_0 = f (x_2) + f^* (y_0) .$$

Hence, for any $\theta \in [0, 1]$ we have

$$\begin{aligned} \theta f (x_1) + (1 - \theta) f (x_2) + f^* (y_0) &= \theta (f (x_1) + f^* (y_0)) + (1 - \theta) (f (x_2) + f^* (y_0)) \\ &= \theta (x_1 y_0) + (1 - \theta) (x_2 y_0) \\ &= (\theta x_1 + (1 - \theta) x_2) y_0 \\ &\le f (\theta x_1 + (1 - \theta) x_2) + f^* (y_0) , \end{aligned}$$

where in the last inequality we have used the Young inequality. It follows that

$$\theta f (x_1) + (1 - \theta) f (x_2) = f (\theta x_1 + (1 - \theta) x_2) ,$$

which contradicts the strict convexity of f in the interval $[x_1, x_2] \subset E$.

Conversely, assume that $f^* : I^* \to \mathbb{R}$ is differentiable in $(I^*)^\circ$. By the previous theorem, $f = f^{**}$. Let J be any interval contained in E and assume by contradiction that f is not strictly convex in J. Then there exist $x_1 < x_2$ in J such that f is affine in $[x_1, x_2]$. Let $x := \frac{x_1 + x_2}{2} \in J \subset E$. By the definition of E, there exists $y \in (I^*)^\circ$ such that $x \in \partial (f^*) (y)$, and since f^* is differentiable in $(I^*)^\circ$, it follows that $(f^*)' (y) = x$. Using the facts that f is affine in $[x_1, x_2]$ and that $(f^*)' (y) = x$ it follows from the equality in the Young inequality that (recall that $f = f^{**}$) we have

$$0 = f (x) + f^* (y) - xy = \frac{1}{2} (f (x_1) + f^* (y) - x_1 y) + \frac{1}{2} (f (x_2) + f^* (y) - x_2 y) .$$

By the Young inequality, necessarily $f (x_1) + f^* (y) - x_1 y = 0$ and $f (x_2) + f^* (y) - x_2 y = 0$, which implies that x_1, $x_2 \in \partial (f^*) (y) = \{(f^*)' (y)\}$. This is a contradiction. ■

The next two exercises illustrate the previous function.

Exercise 56 *(i) Consider the function*

$$f(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \le 1, \\ |x| - \frac{1}{2} & \text{if } |x| > 1. \end{cases}$$

Note that f is not strictly convex. Prove that

$$f^*(y) = \begin{cases} \frac{1}{2}y^2 & \text{if } |y| \le 1, \\ \infty & \text{if } |y| > 1. \end{cases}$$

Note that f^* is differentiable in the interior of its domain. Why this does not contradict the previous theorem?

(ii) Consider the function

$$f(x) = \begin{cases} \frac{1}{2}(x+1)^2 & \text{if } x \le -1, \\ 0 & \text{if } -1 < x < 1, \\ \frac{1}{2}(x-1)^2 & \text{if } x \ge 1, \end{cases}$$

and prove that $f^*(y) = |y| + \frac{1}{2}y^2$. Why this does not contradict the previous theorem?

Concerning the second order derivative of f^{**} we have the following result.

Theorem 57 *Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be convex and closed. Let $x_0 \in I^\circ$ be such that f and f' are differentiable at x_0 and $f''(x_0) > 0$. Then $(f^*)''$ exists at $y_0 = f'(x_0)$ and*

$$(f^*)''(y_0) = \frac{1}{f''(x_0)}.$$

Proof. We claim that f^* is differentiable at y_0. Note that by Theorem 54,

$$\partial f^*(y_0) = (\partial f)^{-1}(y_0) = \{x \in \operatorname{dom} \partial f : y_0 \in \partial f(x)\},$$

and since $y_0 = f'(x_0)$, we have that $x_0 \in \partial f^*(y_0)$. If the convex set $\partial f^*(y_0)$ contains another element, say $x_0 + v$ for some $v \in \mathbb{R}$, then it contains the segment $x_0 + v[0,1]$. Again by Theorem 54 we have that $y_0 \in \partial f(x_0 + tv)$ for all all $t \in [0,1]$. Hence

$$f'(x_0 + tv) = y_0$$

for all but countably many $t \in [0,1]$, which contradicts the fact that $f''(x_0) > 0$. Thus the claim holds.

For all but countably many y near y_0 we have that $\partial f^*(y) = \{(f^*)'(y)\}$ and so again by Theorem 54, $y \in \partial f(x_y)$ for some $x_y \in \operatorname{dom} \partial f$. Hence

$$\frac{(f^*)'(y) - (f^*)'(y_0)}{y - y_0} = \frac{x_y - x_0}{y - y_0}.$$

If $x_y > x_0$ we have

$$\frac{x_y - x_0}{f'_-(x_y) - f'(x_0)} \leq \frac{x_y - x_0}{y - y_0} \leq \frac{x_y - x_0}{f'_+(x_y) - f'(x_0)}$$

Note that by the previous part of the proof, since f^* is differentiable at y_0, we have that $x_y \to x_0$ as $y \to y_0$ by (7) and (8). Hence the result follows by the fact that f'_+ is differentiable at x_0 by Theorem 26 and Remark 27. ∎

2 Convex Functions in $\mathbb{R}^N$

2.1 Affine Sets and Convex Sets

Definition 58 *Let V be a vector space. A set $E \subset V$ is said to be*

(i) affine *if for all $v_1, v_2 \in E$ and $\theta \in \mathbb{R}$,*

$$\theta v_1 + (1-\theta)\, v_2 \in E;$$

(ii) convex *if for all $v_1, v_2 \in E$ and $\theta \in (0,1)$,*

$$\theta v_1 + (1-\theta)\, v_2 \in E.$$

In a geometrical language, a set E is affine if whenever it contains two points, it also contains the line through these two points, while a set E is convex if whenever it contains two points, it also contains the segment joining these two points. Hence every affine set is convex, but not viceversa. The entire space V, Picture the empty set, a set consisting of a single point are both affine and convex sets. A segment is a convex set that is not affine. In $\mathbb{R}^2$ the interior of an ellipse or any regular polygon is a convex set.

The next proposition shows that an affine set can be regarded as a translation of a subspace.

Proposition 59 *A set $E \subset V$ with more that one point is affine if and only if it can be written as $v_0 + W$, where $v_0 \in E$ and W is a vector subspace of V.*

Proof. Suppose that $E \subset V$ can be written as $E = v_0 + W$, where $v_0 \in E$ and W is a vector subspace of V. If $v_1, v_2 \in E$ and $\theta \in \mathbb{R}$, then we may write $v_1 = v_0 + w_1, v_2 = v_0 + w_2$, where $w_1, w_2 \in W$. Hence

$$\begin{aligned}\theta v_1 + (1-\theta)\, v_2 &= \theta\,(v_0 + w_1) + (1-\theta)\,(v_0 + w_1)\\ &= v_0 + [\theta w_1 + (1-\theta)\, w_1] \in E,\end{aligned}$$

since $\theta w_1 + (1-\theta)\, w_1 \in W$.

Conversely, assume that E is affine and let v_0 be any vector in E. Define $W := -v_0 + E$. We claim that W is a subspace of V. To see this, let $w_1, w_2 \in W$ and $t \in \mathbb{R}$. Then we may write $w_1 = -v_0 + v_1, w_2 = -v_0 + v_2$, where $v_1, v_2 \in E$. Hence

$$w_1 + tw_2 = -v_0 + v_1 - tv_0 + tv_2 = -v_0 + t\left[2\left(\frac{1}{2}v_1 + \frac{1}{2}v_2\right) - v_0\right] + (1-t)\, v_1.$$

Since E is affine, $z := \frac{1}{2}v_1 + \frac{1}{2}v_2$ belongs to E, and in turn so does $v := 2z - 1v_0$, and so also $tv + (1-t)\, v_1$. Hence $w_1 + tw_2 = -v_0 + (tv + (1-t)\, v_1) \in W$, by definition of W. ■

2.2 Operations Preserving Convexity

Proposition 60 *If V, W are vector spaces and $E \subset V$, $F \subset W$ are affine (respectively convex), then $E \times F$ is affine (respectively convex).*

Proof. We give the proof only for convex sets. Let z_1, $z_2 \in E \times F$ and let $\theta \in (0,1)$. Then there exist v_1, $v_2 \in E$ and w_1, $w_2 \in F$ such that $z_1 = (v_1, w_1)$ and $z_2 = (v_2, w_2)$. Hence

$$\begin{aligned} \theta z_1 + (1-\theta) z_2 &= \theta (v_1, w_1) + (1-\theta)(v_2, w_2) \\ &= (\theta v_1 + (1-\theta) v_2, \theta w_1 + (1-\theta) w_2) \in E \times F, \end{aligned}$$

where we have used the convexity of E and F. ■

Given two vectors spaces V and W, an *affine transformation* $T : V \to W$ consists of a linear transformation followed by a translation, that is T can be written as

$$T(v) = L(v) + w_0, \quad v \in V,$$

where $L : V \to W$ is linear and $w_0 \in W$ is a fixed vector. Affine transformations preserve convexity.

Proposition 61 *Let V and W be two vector spaces, let $T : V \to W$ be an affine transformation and let $E \subset V$ be affine (respectively convex). Then $T(E)$ is affine (respectively convex).*

Proof. We give the proof only for convex sets. Let w_1, $w_2 \in T(E)$ and let $\theta \in (0,1)$. Then there exist v_1, $v_2 \in E$ such that $T(v_1) = w_1$ and $T(v_2) = w_2$. Hence

$$\begin{aligned} \theta w_1 + (1-\theta) w_2 &= \theta T(v_1) + (1-\theta) T(v_2) \\ &= \theta (L(v_1) + w_0) + (1-\theta)(L(v_2) + w_0) \\ &= L(\theta v_1 + (1-\theta) v_2) + w_0 = T(\theta v_1 + (1-\theta) v_2), \end{aligned}$$

where we have used the linearity of L. ■

Remark 62 *Some simple consequences of the previous proposition are the following:*

(i) ***(Projection)*** *If V, W are vector spaces and $E \subset V \times W$ is affine (respectively convex), then the projection of E into either V or W is affine (respectively convex). To see this, it suffices to consider the linear transformations*

$$\begin{aligned} &V \times W \to V, \qquad &&V \times W \to W, \\ &(v, w) \mapsto v, &&(v, w) \mapsto w. \end{aligned}$$

(ii) If V, W are vector spaces by considering the linear transformations

$$\begin{aligned} &V \times V \to V, \qquad &&V \to V, \\ &(v, w) \mapsto v + w, &&v \mapsto tv, \end{aligned}$$

where $t \in \mathbb{R}$, it follows that if E_1, $E_2 \subset V$ be affine (respectively convex), then the sets $E_1 + E_2$ and tE_1 are affine (respectively convex).

The arbitrary intersection of convex sets is still convex, but in general the union is not (the simplest example is the union of two disjoint closed segments on the real line).

Proposition 63 *Let V be a vector space and let $\{E_\alpha\}_{\alpha\in J}$ be an arbitrary family of affine (respectively convex) subsets of V. Then*

$$E_- := \bigcap_{\alpha\in J} E_\alpha$$

is affine (respectively convex). If $\{E_\alpha\}_{\alpha\in J}$ is totally ordered with respect to inclusion (that is, for α, $\beta \in J$, either $E_\alpha \subset E_\beta$ or $E_\beta \subset E_\alpha$), then

$$E_+ := \bigcup_{\alpha\in J} E_\alpha$$

is affine (respectively convex).

Proof. We give the proof only for convex sets. Let $v_1, v_2 \in E_-$ and $\theta \in [0,1]$. Then $v_1, v_2 \in E_\alpha$ for all $\alpha \in J$, and since E_α is convex, $\theta v_1 + (1-\theta)\, v_2 \in E_\alpha$ for all $\alpha \in J$. Hence $\theta v_1 + (1-\theta)\, v_2 \in E_-$.

Let $v_1, v_2 \in E_+$ and $\theta \in [0,1]$. Then $v_1 \in E_\alpha$ and $v_2 \in E_\beta$ for some α, $\beta \in J$. Then either $E_\alpha \subset E_\beta$ or $E_\beta \subset E_\alpha$, say, $E_\beta \subset E_\alpha$. It follows that $v_1, v_2 \in E_\alpha$, and since E_α is convex, $\theta v_1 + (1-\theta)\, v_2 \in E_\alpha$. Hence $\theta v_1 + (1-\theta)\, v_2 \in E_+$. ■

Remark 64 *In particular, if $C \subset \mathbb{R}^N$ is convex and H is any hyperplane then the intersection $C \cap H$ is convex. Thus, sections of convex sets are convex. This is useful for induction proofs on the dimension N.*

Corollary 65 *Let V be a vector space and let $\{E_n\}_{n\in\mathbb{N}}$ be a sequence of affine (respectively convex) subsets of V. Then*

$$\liminf_{n\to\infty} E_n := \bigcup_{k=1}^{\infty} \bigcap_{i=k} E_i$$

is affine (respectively convex).

Proof. We give the proof only for convex sets. By the previous proposition the sets

$$F_k := \bigcap_{i=k} E_i$$

are convex. Since the sequence $\{F_k\}$ is increasing, the set

$$\liminf_{n\to\infty} E_n = \bigcup_{k=1}^{\infty} F_k$$

is convex. ■

We consider next the case in which V is the Euclidean space $\mathbb{R}^N$.

Theorem 66 *If $C \subset \mathbb{R}^N$ is convex, then for every $\delta > 0$ the set*

$$C_\delta := \left\{x \in \mathbb{R}^N : \operatorname{dist}(x, C) < \delta\right\}$$

is convex.

Proof. Let $x_1,\, x_2 \in C_\delta$ and $\theta \in (0,1)$. By the definition of C_δ we may find $y_1,\, y_2 \in C$ such that

$$|x_1 - y_1| < \delta, \quad |x_2 - y_2| < \delta.$$

Since C is convex, $\theta y_1 + (1-\theta)\, y_2$ belongs to C, and so

$$\begin{aligned}|\theta x_1 + (1-\theta)\, x_2 - (\theta y_1 + (1-\theta)\, y_2)| &= |\theta\,(x_1 - y_1) + (1-\theta)\,(x_2 - y_2)| \\ &\le \theta\,|x_1 - y_1| + (1-\theta)\,|x_2 - y_2| < \delta,\end{aligned}$$

which implies that $\theta x_1 + (1-\theta)\, x_2 \in C_\delta$. ■

Remark 67 *The same proof works in a normed space. Note that taking a line in $\mathbb{R}^2$ shows that the previous result does not hold for affine sets.*

Corollary 68 *If $C \subset \mathbb{R}^N$ is convex, then so is its closure $\overline{C}$.*

Proof. First proof: Let $\delta_n \to 0^+$. Since

$$\overline{C} = \bigcap_{k=1}^{\infty} C_{\delta_n},$$

the result follows from the previous theorem and Proposition 63.
Second proof: Let $x_1,\, x_2 \in \overline{C}$ and $\theta \in (0,1)$. Then there exist two sequences $\{y_n\}$ and $\{z_n\}$ contained in C such that $y_n \to x_1$ and $z_n \to x_2$. Since C is convex, $\theta y_n + (1-\theta)\, z_n \in C$ for all n, and since

$$\lim_{n\to\infty} \theta y_n + (1-\theta)\, z_n = \theta x_1 + (1-\theta)\, x_2,$$

it follows that $\theta x_1 + (1-\theta)\, x_2 \in \overline{C}$. ■

Proposition 69 *Let $C \subset \mathbb{R}^N$ be a convex set. If $x_1 \in C^\circ$ and $x_2 \in \overline{C}$, then*

$$x := \theta x_1 + (1-\theta)\, x_2 \in C^\circ$$

for all $0 < \theta \le 1$. In particular, the interior of C is convex.

Proof. Since $x_1 \in C^\circ$ and $x_2 \in \overline{C}$, we may find $r > 0$ and a sequence $\{y_n\} \subset C$ such that $B(x_1, r) \subset C$ and $y_n \to x_2$. We claim that $B(x, \theta r) \subset C$. To see this, note that if $y \in B(x, \theta r)$, then $|y - x| < \theta r$, or equivalently,

$$\left|\frac{y}{\theta} - \frac{(1-\theta)}{\theta} x_2 - x_1\right| < r.$$

Since $y_n \to x_2$ we may find $n \in \mathbb{N}$ so large that

$$\left| \frac{y}{\theta} - \frac{(1-\theta)}{\theta} y_n - x_1 \right| < r.$$

Thus $\frac{y}{\theta} - \frac{(1-\theta)}{\theta} y_n \in B\left(x_1, r\right) \subset C$, which implies that

$$\frac{y}{\theta} - \frac{(1-\theta)}{\theta} y_n = \xi \in C.$$

In turn, $y = (1-\theta)\, y_n + \theta \xi \in C$, and the proof is complete. ■

Corollary 70 *If $C \subset \mathbb{R}^N$ is a convex set and C° is nonempty, then the closure of C° is the closure of C and the interior of $\overline{C}$ is C°.*

Proof. Since $C^\circ \subset C$, we have that $\overline{C^\circ} \subset \overline{C}$. Conversely, if $x_1 \in C^\circ$ and $x_2 \in \overline{C}$, then $\theta x_1 + (1-\theta)\, x_2 \in C^\circ$ $0 \leq \theta < 1$. Since $\lim_{\theta \to 0^+} \theta x_1 + (1-\theta)\, x_2 = x_2$, it follows that $x_2 \in \overline{C^\circ}$. Hence, $\overline{C^\circ} = \overline{C}$.

Next $C \subset \overline{C}$, and so $C^\circ \subset \left(\overline{C}\right)^\circ$. If $x \in \left(\overline{C}\right)^\circ \setminus C^\circ$, let $x_1 \in C^\circ$. Since $x \in \left(\overline{C}\right)^\circ$ there exists $B(x,r) \subset \overline{C}$. Pick a point $y \in B(x,r)$ on the segment $\overline{xx_1}$ on the opposite side of x_1, so that $x = \theta y + (1-\theta)\, x_1$ for some $\theta \in (0,1)$. In view of the previous theorem, we have that $x \in C^\circ$. This is a contradiction. Thus, $\left(\overline{C}\right)^\circ = C^\circ$. ■

2.3 Affine and Convex Hulls

Let V be a vector space. If $v_1, \ldots, v_n \in V$ and $\theta_1, \ldots, \theta_n \in \mathbb{R}$ with $\theta_1 + \ldots + \theta_n = 1$, then the vector $v := \theta_1 v_1 + \ldots + \theta_n v_n$ is called an *affine combination* of $v_1, \ldots, v_n \in V$. If in addition $\theta_i \in [0,1]$, then $v := \theta_1 v_1 + \ldots + \theta_n v_n$ is called a *convex combination* of $v_1, \ldots, v_n \in V$.

Proposition 71 *Let V be a vector space. A set $E \subset V$ is affine (respectively convex) if and only if every affine (respectively convex) combination of elements of E belongs to E.*

Proof. We give the proof only for convex sets. If a set contains every convex combination of its elements, then it is convex (take $n = 2$). Conversely, assume that E is convex. The proof is by induction on the number n of elements in the convex combination. If $n = 2$, this is just the definition that E is convex. Assume that the result is true for $n \in \mathbb{N}$ and let's prove it for $n+1$. Let $v_1, \ldots, v_{n+1} \in E$, $\theta_1, \ldots, \theta_{n+1} \in [0,1]$, with $\theta_1 + \ldots + \theta_{n+1} = 1$, and $v := \theta_1 v_1 + \ldots + \theta_{n+1} v_{n+1}$. If $\theta_{n+1} = 1$, then $v = v_{n+1} \in V$ and there is nothing to prove. Thus, assume that $\theta_{n+1} < 1$. Then $\theta_1 + \ldots + \theta_n = 1 - \theta_{n+1} > 0$, and so we may rewrite v as

$$v = (\theta_1 + \ldots + \theta_n) \left(\frac{\theta_1}{\theta_1 + \ldots + \theta_n} v_1 + \ldots + \frac{\theta_n}{\theta_1 + \ldots + \theta_n} v_n \right) + \theta_{n+1} v_{n+1}.$$

By the inductive hypothesis, the point

$$z := \frac{\theta_1}{\theta_1 + \ldots + \theta_n} v_1 + \ldots + \frac{\theta_n}{\theta_1 + \ldots + \theta_n} v_n$$

belongs to E, and since $v = (\theta_1 + \ldots + \theta_n)\, z + \theta_{n+1} v_{n+1}$, by the convexity of the set E we have that v belongs to E and the proof is complete. ■

Given any set $E \subset V$, the *convex hull* $\operatorname{co}(E)$ is the intersection of all convex sets that contain E.

Proposition 72 *Let V be a vector space and let $E \subset V$. Then*

$$\operatorname{co}(E) = \left\{ \sum_{i=1}^{n} \theta_i v_i : n \in \mathbb{N}, \sum_{i=1}^{n} \theta_i = 1, \theta_i \geq 0, v_i \in E, i = 1, \ldots, n \right\}. \tag{18}$$

Proof. If F is any convex set that contains E, then it must contain all convex combinations of elements of E, and so

$$F \supset \left\{ \sum_{i=1}^{n} \theta_i v_i : n \in \mathbb{N}, \sum_{i=1}^{n} \theta_i = 1, \theta_i \geq 0, v_i \in E, i = 1, \ldots, n \right\} =: G.$$

Since this holds for all convex sets containing E, it follows that

$$\operatorname{co}(E) \supset G.$$

To prove the opposite inclusion it suffices to show that G is convex and contains E. The latter assertion follows from the fact that if $u \in E$, then we can take $n = 1$ and $\theta_1 = 1$. To show that G is convex, let $0 \leq \theta \leq 1$ and let $u, v \in V$ be of the form

$$u = \sum_{i=1}^{n} \theta_i v_i, \quad v = \sum_{j=1}^{l} s_j w_j,$$

where $\sum_{i=1}^{n} \theta_i = \sum_{j=1}^{l} s_j = 1$, $\theta_i, s_j \geq 0$, $v_i, w_j \in E$, $i = 1, \ldots, n$, $j = 1, \ldots, l$. Note that without loss of generality, we may always assume that $n = l$. Indeed, if $n \neq l$, say $n > l$, then it suffices to set $s_{l+1}, \ldots, s_n := 0$ and $w_{l+1}, \ldots, w_n := w_1$. Then

$$\theta u + (1 - \theta) v = \sum_{i=1}^{n} \theta \theta_i v_i + \sum_{i=1}^{n} (1 - \theta) s_i w_i,$$

which is still a convex combination of elements of E, and so it belongs to G. ■

Note that without loss of generality, in (18) one may consider only positive coefficients θ_i. Carathéodory's theorem improves (18) in that it limits the number of terms in the convex combination to at most $N + 1$.

Theorem 73 (Carathéodory) *Let $E \subset \mathbb{R}^N$. Then*

$$\operatorname{co} E = \left\{ \sum_{i=1}^{N+1} \theta_i x_i : \sum_{i=1}^{N+1} \theta_i = 1, \theta_i \geq 0, x_i \in E, i = 1, \ldots, N + 1 \right\}.$$

Proof. Fix $x \in \operatorname{co} E$ and let

$$S := \{\ell \in \mathbb{N} : x \text{ is a convex combination of } \ell \text{ vectors of } E\}.$$

Note that by the previous proposition, S is nonempty. Let $k := \min S$. We claim that $k \leq N + 1$. Assume by contradiction that $k > N + 1$ and let

$$x = \sum_{i=1}^{k} \theta_i x_i,$$

where $\sum_{i=1}^{k}\theta_i = 1$, $\theta_i \in (0,1)$, $x_i \in E$, $i = 1,\dots,k$. Since $k-1 > N$, the $k-1$ vectors $x_2 - x_1, \dots, x_k - x_1$ are linearly dependent, and so we may find $s_2, \dots, s_k \in \mathbb{R}$ not all zero such that

$$\sum_{i=2}^{k} s_i \left(x_i - x_1\right) = 0.$$

Let $s_1 := -\sum_{i=2}^{k} s_i$. Then $\sum_{i=1}^{k} s_i x_i = 0$ and $\sum_{i=1}^{k} s_i = 0$. Since not all the s_i are zero, there must be positive ones. Define

$$c := \min\left\{\frac{\theta_i}{s_i} : s_i > 0,\ i = 1,\dots,k\right\}$$

and let m be such that $c = \frac{\theta_m}{s_m}$. Then $\theta_i - cs_i \geq 0$ for all $i = 1,\dots,k$ (if $s_i > 0$, then this follows from the definition of c, while if $s_i \leq 0$, then $-cs_i \geq 0$), $\theta_m - cs_m = 0$, and

$$\sum_{i=1}^{k}\left(\theta_i - cs_i\right) = \sum_{i=1}^{k}\theta_i - c\sum_{i=1}^{k} s_i = 1 - 0.$$

Since

$$x = \sum_{i=1}^{k}\theta_i x_i = \sum_{i=1}^{k}\theta_i x_i - 0 = \sum_{i=1}^{k}\left(\theta_i - cs_i\right) x_i,$$

we have written x as a convex combination of less than k elements ($\theta_m - cs_m = 0$), which contradicts the definition of k. ■

Exercise 74 *Prove that the convex hull of an open set $A \subset \mathbb{R}^N$ is open.*

Analogously, the *affine hull* $\text{aff}(E)$ is the intersection of all affine sets that contain E. Reasoning as in the proof of Proposition 72, it can be shown that

$$\text{aff}(E) = \left\{ \sum_{i=1}^{n} \theta_i v_i : n \in \mathbb{N}, \ \sum_{i=1}^{n} \theta_i = 1, \ \theta_i \in \mathbb{R}, \ v_i \in E, \ i = 1, \dots, n \right\}.$$

2.4 Relative Interior

The *relative interior of a set* $E \subset \mathbb{R}^N$ *with respect to* $\text{aff}(E)$, denoted by $\text{ri}_{\text{aff}}(E)$, is the set of points $x \in E$ such that $B(x,r) \cap \text{aff}(E) \subset E$ for some $r > 0$. A set $E \subset \mathbb{R}^N$ is *relatively open* if $\text{ri}_{\text{aff}}(E) = E$.

The *relative boundary of* E *with respect to* $\text{aff}(E)$, denoted by $\text{rb}_{\text{aff}}(E)$, is the set $\overline{E} \setminus \text{ri}_{\text{aff}}(E)$.

Exercise 75 *Let $C \subset \mathbb{R}^N$ be a nonempty convex set and let $T : \mathbb{R}^N \to \mathbb{R}^N$ be a bijective affine transformation. Prove that*

(i) $T\left(\overline{C}\right) = \overline{T(C)}$;

(ii) $\text{ri}_{\text{aff}}(T(C)) = T(\text{ri}_{\text{aff}}(C))$.

Remark 76 *The previous exercise shows that closures and relative interiors are preserved under bijective affine transformations. Hence if $\text{aff}(C)$ is the translation of a subspace of $\mathbb{R}^N$ of dimension m, there exists a bijective affine transformation $T : \mathbb{R}^N \to \mathbb{R}^N$ that carries $\text{aff}(C)$ onto the subspace*

$$\{y = (y_1, \dots, y_m, 0, \dots, 0) : y_1, \dots, y_m \in \mathbb{R}\}.$$

This subspace can be regarded as a copy of $\mathbb{R}^m$. Hence it is often possible to reduce a question about general convex sets to the case in which the convex set has the whole space as its affine hull. Note that in this case the relative interior is simply the interior of the set.

The *dimension of a convex set* $C \subset \mathbb{R}^N$ (with at least two elements) is the dimension of the translation of its affine hull thorough the origin. Note that if the dimension of C is m, then there exist $x_0, \dots, x_m \in C$ such that the m vectors $x_1 - x_0, \dots, x_m - x_0$ are linearly independent.

Proposition 77 *Let $C \subset \mathbb{R}^N$ be a nonempty convex set. Then $\text{ri}_{\text{aff}}(C)$ is convex and nonempty. Moreover, if $x \in \overline{C}$ and $x_0 \in \text{ri}_{\text{aff}}(C)$, then*

$$\theta x + (1 - \theta) x_0 \in \text{ri}_{\text{aff}}(C)$$

for all $0 \leq \theta < 1$.

Proof. If C is a singleton, then $\mathrm{ri}_{\mathrm{aff}}(C) = C$, and so there is nothing to prove. Thus assume that C has at least two elements. By Exercise 75 and the previous remark, without loss of generality, we may assume that the convex set has the whole space as its affine hull. Note that in this case the relative interior is simply the interior of the set and so it remains to show that C° is nonempty. Applying an affine transformation, if necessary, we may assume that the vectors

$$e_1 = (1, 0, \ldots, 0), \ldots, e_N = (0, \ldots, 0, 1)$$

belong to C. Since $0 \in C$, any point of the form

$$w = \sum_{i=1}^{N} \theta_i e_i = 0\left(1 - \sum_{i=1}^{N} \theta_i\right) + \sum_{i=1}^{N} \theta_i e_i,$$

where $\sum_{i=1}^{n} \theta_i \le 1$ and $\theta_i \ge 0$ for all $i = 1, \ldots, N$, belongs to C. Hence,

$$\left\{(x_1, \ldots, x_N) : \sum_{i=1}^{n} x_i \le 1,\, x_i \ge 0,\, i = 1, \ldots, N\right\} \subset C.$$

Since the interior of the set $\{(x_1, \ldots, x_N) : \sum_{i=1}^{n} x_i \le 1,\, x_i \ge 0,\, i = 1, \ldots, N\}$ is

$$\left\{(x_1, \ldots, x_N) : \sum_{i=1}^{n} x_i < 1,\, x_i > 0,\, i = 1, \ldots, N\right\},$$

which is clearly nonempty, we have proved that C° is nonempty. The second part of the statement follows from Proposition 69. ■

Example 78 *Note that if $C_1 \subset C_2$ are two nonempty convex sets, then in general one cannot conclude that*

$$\mathrm{ri}_{\mathrm{aff}}(C_1) \subset \mathrm{ri}_{\mathrm{aff}}(C_2).$$

Indeed, let C_2 be the closed unit cube in $\mathbb{R}^3$ and let C_1 be one of its faces. Then $\mathrm{ri}_{\mathrm{aff}}(C_2)$ *is the open unit cube, while* $\mathrm{ri}_{\mathrm{aff}}(C_1)$ *is the face without its four edges. Hence* $\mathrm{ri}_{\mathrm{aff}}(C_1)$ *and* $\mathrm{ri}_{\mathrm{aff}}(C_2)$ *are disjoint and nonempty.*

Proposition 79 *Let C_1, C_2 be two nonempty convex sets of $\mathbb{R}^N$. Then the following three conditions are equivalent:*

(i) $\overline{C_1} = \overline{C_2}$;

(ii) $\mathrm{ri}_{\mathrm{aff}}(C_1) = \mathrm{ri}_{\mathrm{aff}}(C_2)$;

(iii) $\mathrm{ri}_{\mathrm{aff}}(C_1) \subset C_2 \subset \overline{C_1}$.

Proof. We begin by observing that by Corollary 70 and remark 76 for any nonempty convex set $C \subset \mathbb{R}^N$,

$$\overline{C} = \overline{\mathrm{ri}_{\mathrm{aff}}(C)}, \quad \mathrm{ri}_{\mathrm{aff}}\left(\overline{C}\right) = \mathrm{ri}_{\mathrm{aff}}(C). \tag{19}$$

If (i) holds, then by (19),

$$\mathrm{ri}_{\mathrm{aff}}\left(C_1\right) = \mathrm{ri}_{\mathrm{aff}}\left(\overline{C_1}\right) = \mathrm{ri}_{\mathrm{aff}}\left(\overline{C_2}\right) = \mathrm{ri}_{\mathrm{aff}}\left(C_2\right),$$

that is, (ii) is true.

Similarly, if (ii) holds, then again by (19),

$$\overline{C_1} = \overline{\mathrm{ri}_{\mathrm{aff}}\left(C_1\right)} = \overline{\mathrm{ri}_{\mathrm{aff}}\left(C_2\right)} = \overline{C_2},$$

which is (i).

If (iii) holds, then by (19),

$$\overline{\mathrm{ri}_{\mathrm{aff}}\left(C_1\right)} \subset \overline{C_2} \subset \overline{C_1} = \overline{\mathrm{ri}_{\mathrm{aff}}\left(C_1\right)},$$

and (i) is satisfied.

Finally, if (i) holds, then so does (ii), and we deduce that

$$\mathrm{ri}_{\mathrm{aff}}\left(C_1\right) = \mathrm{ri}_{\mathrm{aff}}\left(C_2\right) \subset C_2 \subset \overline{C_2} = \overline{C_1},$$

which is (iii). ■

Exercise 80 *Let C_1, C_2 be two nonempty convex sets of $\mathbb{R}^N$ and let $t \in \mathbb{R}$. Prove that*

(i) $\mathrm{ri}_{\mathrm{aff}}\left(tC_1\right) = t\,\mathrm{ri}_{\mathrm{aff}}\left(C_1\right)$*;*

(ii) $\mathrm{ri}_{\mathrm{aff}}\left(C_1 + C_2\right) = \mathrm{ri}_{\mathrm{aff}}\left(C_1\right) + \mathrm{ri}_{\mathrm{aff}}\left(C_2\right)$*.*

2.5 Projection

Theorem 81 *Let $C \subset \mathbb{R}^N$ be a nonempty closed convex set. Then for every $x \in \mathbb{R}^N$ there exists a unique point $y \in C$ such that*

$$|x - y| \le |x - z| \quad \text{for all } z \in C. \tag{20}$$

Moreover,

$$(x - y) \cdot (z - y) \le 0 \quad \text{for all } z \in C. \tag{21}$$

Proof. Fix $x \in \mathbb{R}^N$. For $r > 0$ sufficiently large, the set $\overline{B(x,r)} \cap C$ is compact and nonempty. Hence the continuous function

$$z \in \mathbb{R}^N \mapsto |x - z|$$

attains a minimum on this set, say at $y \in \overline{B(x,r)} \cap C$. Hence

$$|x - y| \le |x - z| \quad \text{for all } z \in \overline{B(x,r)} \cap C.$$

If $z \in C \setminus \overline{B(x,r)}$, then $|x - z| > r \ge |x - y|$, and so we have shown (20).

To prove uniqueness, let $y_1 \in C$ be such that

$$|x - y_1| \le |x - z| \quad \text{for all } z \in C. \tag{22}$$

Then $y_2 := \frac{(y+y_1)}{2} \in C$ and

$$\left| x - \frac{(y + y_1)}{2} \right| < |x - y|,$$

unless $y_1 = y$.

To prove (21), note that by squaring both sides of (22) we get

$$|x - y|^2 \le |x - z|^2 \quad \text{for all } z \in C.$$

In particular, for any $w \in C$, taking $z := y + \theta (w - y) \in C$, $0 < \theta \le 1$, in the previous inequality yields

$$\begin{aligned} 0 &\ge (x - y) \cdot (x - y) - (x - z) \cdot (x - z) \\ &= (x - y) \cdot (x - y) - (x - y - \theta (w - y)) \cdot (x - y - \theta (w - y)) \\ &= 2\theta (x - y) \cdot (w - y) - \theta^2 (w - y) \cdot (w - y). \end{aligned}$$

Dividing by $2\theta > 0$ and letting $\theta \to 0^+$ gives

$$0 \ge (x - y) \cdot (w - y),$$

which is (21). ■

The point y is the *projection* of the point x onto C and is denoted $p_C(x)$.

Exercise 82 *Let $C \subset \mathbb{R}^N$ be a nonempty closed convex set.*

(i) Prove that if $x \in \mathbb{R}^N$ and if a point $y \in C$ satisfies (21), then

$$|x - y| \leq |x - z| \quad \textit{for all } z \in C,$$

that is, $y = p_C(x)$.

(ii) Prove that the mapping $p_C : \mathbb{R}^N \to C$ is Lipschitz continuous with Lipschitz constant less than or equal one, that is

$$|p_C(x) - p_C(x_1)| \leq |x - x_1|$$

for all x, $x_1 \in \mathbb{R}^N$.

The existence and uniqueness of a projection mapping actually characterizes convex sets.

Theorem 83 *Let $C \subset \mathbb{R}^N$ be a nonempty set with the property that to each point $x \in \mathbb{R}^N \setminus C$ there exists a unique nearest point in C. Then C is closed and convex.*

Proof. Step 1: Let $\{x_n\} \subset C$ be such that $x_n \to x$. We claim that $x \in C$. If not, then there exist a unique point $y \in C$ closest to x. This implies that

$$\operatorname{dist}(x, C) = |x - y| > 0,$$

which is a contradiction since $\operatorname{dist}(x, C) \leq |x - x_n| \to 0$.

Step 2: If C is not convex, there exist two points x, $y \in C$ such that the segment $\overline{xy}$ is not completely contained in C. Since C is closed (and so its complement is open), by changing endpoints, we can actually assume that $\overline{xy} \cap C = \{x, y\}$. Let $x_0 := \frac{x+y}{2}$ and construct a ball $B(x_0, r_0) \subset \mathbb{R}^N \setminus C$. Consider the family $\mathcal{F}$ of all open balls whose contained in $\mathbb{R}^N \setminus C$ and that contain $B\left(x_0, \frac{r_0}{2}\right)$ and let

$$R := \sup\left\{r : \, B(z, r) \in \mathcal{F} \text{ for some } z \in \mathbb{R}^N\right\}.$$

Note that $R < \infty$, since if the radii get too large the balls must contain x and y in their inside. In turn the set of centers is bounded. Hence by a simple compactness argument there exists a ball $B(z, R) \in \mathcal{F}$, (consider $B(z_n, r_n) \in \mathcal{F}$ such that $r_n \to R$ and find a subsequence such that $z_n \to z$. Then $B(z, R) \subset \mathbb{R}^N \setminus C$). By maximality, the ball $B(z, R)$ touches C in a unique point p. There are now two cases. If $\partial B(z, R) \cap \partial B\left(x_0, \frac{r_0}{2}\right)$ is nonempty, then let q be the unique point of intersection, while if the two boundaries don't intersect, take q to be z. For $\varepsilon > 0$ sufficiently small, the ball $\varepsilon(q - p) + B(z, R)$ contains $\partial B\left(x_0, \frac{r_0}{2}\right)$ and does not intersect C (otherwise $B(z, R)$ would touch C into two different points: we are moving away from p so if we touch we must touch in a different point). Hence we can enlarge the ball slightly to get a larger ball contained in $\mathbb{R}^N \setminus C$ and that contains $B\left(x_0, \frac{r_0}{2}\right)$. This contradicts the maximality of R. ■

2.6 Separating Theorems

In this section we prove some separation theorems for convex sets in $\mathbb{R}^N$. Their counterparts in the infinite-dimensional setting are the Hahn–Banach theorems, which are stated in the appendix.

Theorem 84 *Let C, $K \subset \mathbb{R}^N$ be nonempty disjoint convex sets, with C closed and K compact. Then there exist a vector $b \in \mathbb{R}^N \setminus \{0\}$ and two numbers $\alpha \in \mathbb{R}$ and $\varepsilon > 0$ such that*

$$b \cdot x \le \alpha - \varepsilon \quad \textit{for all } x \in C \textit{ and } b \cdot x \ge \alpha + \varepsilon \quad \textit{for all } x \in K.$$

Proof. Define
$$C_0 := C - K.$$
Then C_0 is closed. Indeed, if $z_n \in C_0$ is such that $z_n \to z$, then writing $z_n = x_n - y_n$, where $x_n \in C$ and $y_n \in K$, by compactness there exists a subsequence $\{y_{n_k}\}$ of $\{y_n\}$ that converge to some $y \in K$. In turn, $x_{n_k} = z_{n_k} + y_{n_k} \to z + y =: x \in C$, since C is closed, and thus $z = x - y \in C_0$.

Since C and K are disjoint, we have that 0 does not belong to C_0. Let $y_0 \in C_0$ be the projection of 0 onto C_0. Then $y_0 \neq 0$ (since $0 \notin C_0$) and by (21),

$$-y_0 \cdot (z - y_0) \le 0 \quad \text{for all } z \in C_0,$$

or equivalently
$$-y_0 \cdot z \le -|y_0|^2 < 0 \quad \text{for all } z \in C_0.$$
Define $b := 0 - y_0$. Then

$$b \cdot z \le -|b|^2 < 0 \quad \text{for all } z \in C_0.$$

Writing $z = x - y$, with $x \in C$ and $y \in K$, we get

$$b \cdot x - b \cdot y \le -|b|^2 < 0 \quad \text{for all } x \in C \text{ and } y \in K.$$

Taking first the supremum over all $x \in C$ and then over all $y \in K$, and using the fact that $\sup(-E) = -\inf E$, gives

$$\sup_{x \in C} (b \cdot x) - \inf_{y \in K} b \cdot y \le -|b|^2 < 0,$$

which gives
$$\sup_{x \in C} (b \cdot x) + |b|^2 \le \inf_{y \in K} b \cdot y.$$
This is the desired inequality. ■

Exercise 85 *Prove that a nonempty closed convex set $C \subset \mathbb{R}^N$ is given by the intersection of all the closed half-spaces that contain it.*

Exercise 86 *Let $x_1, \ldots, x_k \in \mathbb{R}^N$ be linearly independent vectors and let s_i, $s_i^{(n)} \in \mathbb{R}$, $i = 1, \ldots, k$, $n \in \mathbb{N}$. Prove that if*

$$\lim_{n\to\infty} \sum_{i=1}^{k} s_i^{(n)} x_i = \sum_{i=1}^{k} s_i x_i,$$

then $\lim_{n\to\infty} s_i^{(n)} = s_i$ for every $i = 1, \ldots, k$.

Exercise 87 *Let $C \subset \mathbb{R}^N$ be a nonempty convex set and let $x \in C$ and $x_0 \in \operatorname{ri}_{\text{aff}}(C)$.*

(i) Prove that $x_0 + t(x_0 - x) \in \operatorname{aff}(C)$ for all $t \in \mathbb{R}$.

(ii) Prove that the function $g : \mathbb{R} \to \operatorname{aff}(C)$, defined by $g(t) := x_0 + t(x_0 - x)$, $t \in \mathbb{R}$, is continuous.

(iii) Prove that $x_0 + t(x_0 - x) \in \operatorname{ri}_{\text{aff}}(C)$ for all t sufficiently small.

Theorem 88 *Let C_1, $C_2 \subset \mathbb{R}^N$ be nonempty convex sets. Then there exist a vector $b \in \mathbb{R}^N \setminus \{0\}$ and $\alpha \in \mathbb{R}$ such that*

$$b \cdot x \le \alpha \quad \text{for all } x \in C_1 \text{ and } b \cdot x \ge \alpha \quad \text{for all } x \in C_2,$$

and $C_1 \cup C_2$ is not contained in the hyperplane $\left\{x \in \mathbb{R}^N : b \cdot x = \alpha\right\}$ if and only if $\operatorname{ri}_{\text{aff}}(C_1) \cap \operatorname{ri}_{\text{aff}}(C_2) = \emptyset$.

Proof. Step 1: Assume that C_1, $C_2 \subset \mathbb{R}^N$ are nonempty convex sets with $\operatorname{ri}_{\text{aff}}(C_1) \cap \operatorname{ri}_{\text{aff}}(C_2) = \emptyset$. Define

$$C := C_1 - C_2.$$

By Exercise 80,

$$\operatorname{ri}_{\text{aff}}(C) = \operatorname{ri}_{\text{aff}}(C_1) - \operatorname{ri}_{\text{aff}}(C_2),$$

and so by hypothesis $0 \notin \operatorname{ri}_{\text{aff}}(C)$. To complete the proof in this case it suffices to prove that there exists $b \in \mathbb{R}^N \setminus \{0\}$ such that $b \cdot x \ge 0$ for all $x \in C$ with strict inequality for at least one element of C. There are two cases. If $0 \notin \overline{C}$ then it suffices to apply Theorem 84 to the closed set $\overline{C}$ and the compact set $\{0\}$. Thus assume that $0 \in \overline{C}$. Define the set

$$E = \bigcup_{t>0} t \operatorname{ri}_{\text{aff}}(C).$$

Then E is convex, $\operatorname{ri}_{\text{aff}}(C) \subset E \subset \operatorname{aff}(E)$, and $0 \notin E$. Since $\operatorname{ri}_{\text{aff}}(C)$ is nonempty by Proposition 77 and $\mathbb{R}^N$ is finite-dimensional, there is a finite maximal (with respect to inclusion) set of linearly independent vectors $x_1, \ldots, x_k \in$

$\mathrm{ri_{aff}}(C)$. Again by Proposition 77 we have that $\mathrm{ri_{aff}}(C)$ is convex, and so the vector

$$x_0 := \sum_{i=1}^{k} \frac{1}{k} x_i$$

belongs to $\mathrm{ri_{aff}}(C)$. We claim that $-x_0 \notin \overline{E}$. Indeed, assume by contradiction that $-x \in \overline{E}$ and find a sequence $\{w_n\} \subset E$ converging to $-x_0$. Then by definition of E we may write each w_n as $w_n = t_n \xi_n$, where $t_n > 0$ and $\xi_n \in \mathrm{ri_{aff}}(C)$. Since $x_1, \ldots, x_k$ form a maximal set of linearly independent vectors in $\mathrm{ri_{aff}}(C)$, each ξ_n can be written as their linear combination (and so can $w_n = t_n \xi_n$). Thus we may write

$$w_n = \sum_{i=1}^{k} s_i^{(n)} x_i$$

for some $s_i^{(n)} \in \mathbb{R}$, $i = 1, \ldots, k$. Hence

$$\lim_{n\to\infty} \sum_{i=1}^{k} s_i^{(n)} x_i = \lim_{n\to\infty} w_n = -x_0 = \sum_{i=1}^{k} \left(-\frac{1}{k}\right) x_i.$$

By Exercise 86 this implies that $\lim_{n\to\infty} s_i^{(n)} = -\frac{1}{k}$ for every $i = 1, \ldots, k$. Fix $n \in \mathbb{N}$ so large that $s_i^{(n)} < 0$ for every $i = 1, \ldots, k$ and set

$$s := \sum_{i=1}^{k} s_i^{(n)} < 0.$$

By the convexity of the set E we have that

$$0 = \frac{1}{1-s} w_n + \sum_{i=1}^{k} \left(-\frac{s_i^{(n)}}{1-s}\right) x_i \in E,$$

which is a contradiction.

This shows that $-x_0 \notin \overline{E}$. We are now in a position to apply Theorem 84 to the closed set $\overline{E}$ and the compact set $\{-x_0\}$ to find a vector $b \in \mathbb{R}^N \setminus \{0\}$ and two numbers $\alpha \in \mathbb{R}$ and $\varepsilon > 0$ such that

$$b \cdot x \le \alpha - \varepsilon \quad \text{for all } x \in \overline{E} \text{ and } b \cdot (-x_0) \ge \alpha + \varepsilon.$$

By the definition of E, for any $x \in \mathrm{ri_{aff}}(C)$ and $t > 0$ we have that $tx \in E$, and so from the previous inequality we get

$$b \cdot x \le \frac{\alpha - \varepsilon}{t} \quad \text{for all } t > 0.$$

Letting $t \to 0^+$ and $t \to \infty$ yield $\alpha - \varepsilon \ge 0$ and $b \cdot x \le 0$, respectively. Moreover, $b \cdot x_0 \le -(\alpha + \varepsilon) < 0$. Hence we have proved that $b \cdot x \le 0$ for all $x \in \mathrm{ri_{aff}}(C)$ with the strict inequality at $x_0 \in \mathrm{ri_{aff}}(C)$.

Step 2: To prove the converse implication assume that there exist a vector $b \in \mathbb{R}^N \setminus \{0\}$ and $\alpha \in \mathbb{R}$ such that

$$b \cdot x \leq \alpha \quad \text{for all } x \in C_1 \text{ and } b \cdot x \geq \alpha \quad \text{for all } x \in C_2 \tag{23}$$

and $C_1 \cup C_2$ is not contained in the hyperplane $\left\{x \in \mathbb{R}^N : b \cdot x = \alpha\right\}$. As in the previous step define

$$C := C_1 - C_2.$$

By Exercise 80,

$$\text{ri}_{\text{aff}}(C) = \text{ri}_{\text{aff}}(C_1) - \text{ri}_{\text{aff}}(C_2).$$

Thus it suffices to show that $0 \notin \text{ri}_{\text{aff}}(C)$. By (23),

$$b \cdot x \leq 0 \quad \text{for all } x \in C$$

with the strict inequality for at least one element $x_0 \in C$. Assume by contradiction that $0 \in \text{ri}_{\text{aff}}(C)$. Then by Exercise 87,

$$0 + \varepsilon (0 - x_0) \in C$$

for all $\varepsilon > 0$ sufficiently small. This implies that $b \cdot (-\varepsilon x_0) \leq 0$, which is a contradiction since $b \cdot x_0 < 0$. ■

As a consequence of the previous separation theorem we obtain the following result.

Corollary 89 *Let $C_1 \subset C_2 \subset \mathbb{R}^N$ be nonempty convex sets. Then there exist $b \in \mathbb{R}^N \setminus \{0\}$ and $\alpha \in \mathbb{R}$ such that*

$$b \cdot x = \alpha \quad \text{for all } x \in C_1 \text{ and } b \cdot x \geq \alpha \quad \text{for all } x \in C_2$$

if and only $C_1 \cap \text{ri}_{\text{aff}}(C_2) = \emptyset$.

Exercise 90 *Prove the previous corollary.*

Definition 91 *Let $C \subset \mathbb{R}^N$ be a convex set. A point $x_0 \in C$ is called an* extreme point *of C if x_0 is not an interior point of any segment contained in C, that is, if there do not exist x_1, $x_2 \in C$ and $\theta \in (0,1)$ such that $x_0 = \theta x_1 + (1-\theta)\,x_2$.*

Example 92 *A line has no extreme points. The extreme points of a closed ball are the its boundary, while the extreme points of a closed cube are its vertices.*

Theorem 93 *Let $C \subset \mathbb{R}^N$ be convex and compact. Then C is the convex hull of its extreme points.*

Proof. The proof is done by induction on the dimension m of the set C. If $m = 0$ or $m = 1$, then C is a point and a closed segment, respectively, and so there is nothing to prove. Assume that the result is true for any convex and compact set of dimension at most m, where $m \leq N-1$ and let $C \subset \mathbb{R}^N$ be a convex and compact set of dimension $m+1$. As in Remark 76 we may assume that $C \subset \mathbb{R}^{m+1}$. If $x_0 \in \partial C$, then by the previous corollary (with $C_1 = \{x_0\}$ and $C_2 = C^\circ$) we may find $b \in \mathbb{R}^{m+1} \setminus \{0\}$ and $\alpha \in \mathbb{R}$ such that

$$b \cdot x_0 = \alpha \quad \text{and } b \cdot x \geq \alpha \quad \text{for all } x \in C^\circ.$$

Let H be the hyperplane (in $\mathbb{R}^{m+1}$) of equation $b \cdot x = \alpha$. Then $H \cap C$ is compact and has dimension at most m. Hence by the induction hypothesis x_0 may be written as a convex combination of extreme points of $H \cap C$. To conclude the proof in this case, it remains to show that an extreme point of $H \cap C$ is also an extreme point of C. Thus let $x \in H \cap C$ be an extreme point of $H \cap C$ and assume that there exist x_1, $x_2 \in C$ and $\theta \in (0,1)$ such that $x = \theta x_1 + (1-\theta)\,x_2$. Then at least one of x_1, $x_2 \in C$, say x_1, does not belong to H, so that $b \cdot x_1 > \alpha$, while $b \cdot x_2 \geq \alpha$. Multiply the first inequality by $\theta \in (0,1)$ and the second by $(1-\theta)$ and add them to find

$$b \cdot x = b \cdot (\theta x_1 + (1-\theta)\,x_2) > \alpha,$$

which contradicts the fact that $x \in H$. This shows that if $x_0 \in \partial C$, then x_0 may be written as a convex combination of extreme points of C. On the other hand, if $x_0 \in C^\circ$, then any line through x_0 intersects ∂C in two points x_1, $x_2 \in C$. Since x_1, x_2 are convex combinations of extreme points of C and x_0 is a convex combination of x_1, x_2, it follows that x_0 is a convex combinations of extreme points of C. ■

3 Convex Functions

We now turn to the study of convex functions.

Definition 94 *Given a vector space V, a function $f : V \to [-\infty, \infty]$ is said to be* *(i)* convex *if*

$$f(\theta v_1 + (1-\theta) v_2) \le \theta f(v_1) + (1-\theta) f(v_2) \tag{24}$$

for all v_1, $v_2 \in V$ *and* $\theta \in (0,1)$ *for which the right-hand side is well-defined;*

(ii) strictly convex *if*

$$f(\theta v_1 + (1-\theta) v_2) < \theta f(v_1) + (1-\theta) f(v_2)$$

for all v_1, $v_2 \in V$, $v_1 \neq v_2$, *and* $\theta \in (0,1)$ *for which the right-hand side is well-defined;*

(iii) proper *if it is convex, does not take the value* $-\infty$, *and is not identically* ∞;

(iv) concave *(respectively* strictly concave*) if* $-f$ *is convex (respectively strictly convex).*

In (i) and (ii) the right-hand side is not defined only when $f(v_1) = \pm\infty$ and $f(v_2) = \mp\infty$.

If E is a subset of the vector space V, then a function $f : E \to [-\infty, \infty]$ is said to be *convex* if the extension

$$\bar{f}(v) := \begin{cases} f(v) & \text{if } v \in E, \\ \infty & \text{if } v \notin E, \end{cases}$$

is a convex function in V. Analogous definitions apply to the concept of strict convexity, concavity, and strict concavity.

Example 95 *Prove that:*

(i) The function $f : \mathbb{R}^N \to \mathbb{R}$ *defined by*

$$f(x) := Ax \cdot x,$$

where A *is a symmetric matrix in* $\mathbb{R}^{N\times N}$, *is convex if and only if* A *is positive semidefinite.*

(ii) If V *is a vector space, the* indicator function *of a set* $E \subset V$ *defined by*

$$f(v) = I_E(v) := \begin{cases} 0 & \text{if } v \in E, \\ \infty & \text{if } v \notin E, \end{cases}$$

is a convex function if and only if the set E *is convex.*

(iii) The function $f : \mathbb{R} \to [-\infty, \infty]$ *defined by*

$$f(x) := \begin{cases} -\infty & \text{if } |x| < 1, \\ 0 & \text{if } |x| = 1, \\ \infty & \text{if } |x| > 1, \end{cases}$$

is convex and not proper.

Definition 96 *Let E be a set and let $f : E \to [-\infty, \infty]$ be any function. The* epigraph *of f is the (possibly empty) set*

$$\operatorname{epi} f := \{(v,t) \in E \times \mathbb{R} : \, f(v) \leq t\}.$$

Proposition 97 *Let V be a vector space. A function $f : V \to [-\infty, \infty]$ is convex if and only if* $\operatorname{epi} f$ *is a convex set.*

Proof. Assume that f is convex and let (v_1, s), $(v_2, t) \in \operatorname{epi} f$ and $\theta \in (0,1)$. We claim that

$$\theta (v_1, s) + (1-\theta)(v_2, t) \in \operatorname{epi} f.$$

Indeed, since $s \geq f(v_1)$, $t \geq f(x)$, it follows that

$$\theta s + (1-\theta) t \geq \theta f(v_1) + (1-\theta) f(v_2) \geq f(\theta v_1 + (1-\theta) v_2),$$

where we have used the convexity of f. Hence the claim is proved.

Conversely, let $\operatorname{epi} f$ be a convex set, let v_1, $v_2 \in V$, and let $\theta \in (0,1)$. If $f(v_1) = \pm\infty$ and $f(v_2) = \mp\infty$, then the right-hand side of (24) is not well-defined.

If $f(v_1) = \infty$ and $f(v_2) > -\infty$ or $f(v_1) > -\infty$ and $f(v_2) = \infty$, then (24) holds. Thus assume that $f(v_1)$, $f(v_2) < \infty$ and let $s \geq f(v_1)$, $t \geq f(v_2)$. Since (v_1, s), $(v_2, t) \in \operatorname{epi} f$ it follows from the convexity of $\operatorname{epi} f$ that $\theta(v_1, s) + (1-\theta)(v_2, t) \in \operatorname{epi} f$, that is,

$$f(\theta v_1 + (1-\theta) v_2) \leq \theta s + (1-\theta) t.$$

The convexity of f follows by letting $s \searrow f(v_1)$ and $t \searrow f(v_2)$. ∎

Let V be a vector space. The *effective domain* of a function $f : V \to [-\infty, \infty]$ is the set

$$\operatorname{dom}_e f := \{v \in V : \, f(v) < \infty\}.$$

We observe that if f is convex, then the effective domain of f is a convex set, and that

$$\operatorname{epi} f := \{(v,t) \in \operatorname{dom}_e f \times \mathbb{R} : \, f(v) \leq t\}.$$

Exercise 98 *Prove that if $f : \mathbb{R}^N \to [-\infty, \infty]$ is a convex function, then*

$$\operatorname{ri}_{\text{aff}}(\operatorname{epi} f) = \{(v,t) \in \operatorname{ri}_{\text{aff}}(\operatorname{dom}_e f) \times \mathbb{R} : \, f(v) \leq t\}$$

Wednesday, February 27, 2008

Remark 99 *Note that if $f:\mathbb{R}^N\to[-\infty,\infty]$ is a convex function, finite at some point in the interior of* $\operatorname{dom}_e f$*, then f is proper. Indeed, suppose that $f(x_0)\in\mathbb{R}$ for some x_0 in the interior of* $\operatorname{dom}_e f$*, and that $f(x)=-\infty$ for some $x\in\mathbb{R}^N$. Let $B(x_0,r)\subset\operatorname{dom}_e f$. Let $\varepsilon>0$ be so small that $|\varepsilon(x_0-x)|<r$, i.e., $x_0+\varepsilon(x_0-x)\in B(x_0,r)$. Set $\theta:=\frac{\varepsilon}{1+\varepsilon}$. Note that*

$$x_0=\theta x+(1-\theta)(x_0+\varepsilon(x_0-x)),$$

and due to the convexity of f we conclude that $f(x_0)=-\infty$, which is a contradiction.

Note that if f is convex and $c>0$, then cf is still convex, the sum of two convex functions f and g is convex (we set $(f+g)(v):=+\infty$ whenever $f(v)=\pm\infty$ and $g(v)=\mp\infty$), and the pointwise supremum of an arbitrary family of convex functions is again a convex function. If f is convex and if $g:[-\infty,\infty]\to[-\infty,\infty]$ is convex and nondecreasing, then $g\circ f$ is convex.

Most of the result in Subsection 1.3 continue to hold.

Theorem 100 *Let V be a vector space, let $f:V\to[-\infty,\infty]$ and $g:V\to[-\infty,\infty]$ be convex, and let $\alpha\geq 0$. Then $f+g$ and αf are convex.*

Proof. Since f and g are convex,

$$f(\theta v+(1-\theta)w)\leq\theta f(v)+(1-\theta)f(w),$$
$$g(\theta v+(1-\theta)w)\leq\theta g(v)+(1-\theta)g(w)$$

for all v, $w\in V$ and $\theta\in(0,1)$. The result now follows by summing the two inequalities and by multiplying the second by $\alpha\geq 0$. ■

Theorem 101 *Let V be a vector space, let $f:V\to\mathbb{R}$ convex and $g:\mathbb{R}\to(-\infty,\infty]$ be convex and increasing. Then $g\circ f:V\to(-\infty,\infty]$ is convex.*

Proof. For all v, $w\in V$ and $\theta\in(0,1)$, we have

$$\begin{aligned}(g\circ f)(\theta v+(1-\theta)w)&=g(f(\theta v+(1-\theta)w))\\&\overset{g\text{ increasing}}{\leq}g(\theta f(v)+(1-\theta)f(w))\\&\overset{g\text{ convex}}{\leq}\theta g(f(v))+(1-\theta)g(f(w)).\end{aligned}$$

■

Example 102 *If $f:\mathbb{R}^N\to[0,\infty]$ is convex, then taking*

$$g(t):=\begin{cases}t^p & \textit{if } t\geq 0,\\ 0 & \textit{if } t<0,\end{cases}$$

where $p \geq 1$*, or*

$$g(t) := \begin{cases} \sqrt{t^2+1} & \textit{if } t \geq 0, \\ 1 & \textit{if } t < 0, \end{cases}$$

it follows that the functions

$$\psi_1(x) := (f(x))^p, \quad \psi_2(x) := \sqrt{f^2(x)+1}$$

are convex. In particular, if $f(x) := |x|$*, then the functions* $\psi_1(x) := |x|^p$ *and* $\psi_2(x) := \sqrt{|x|^2+1}$ *are convex.*

Theorem 103 *Let* V *be a vector space, let* $f_\alpha : V \to [-\infty, \infty]$*,* $\alpha \in \Lambda$*, be an arbitrary family of convex functions, and let*

$$f(v) := \sup_{\alpha \in \Lambda} f_a(v), \quad v \in V.$$

Then $f : V \to [-\infty, \infty]$ *is convex.*

Proof. Since f_α is convex, for all v, $w \in V$ and $\theta \in (0,1)$,

$$\begin{aligned} f(\theta v + (1-\theta) w) &= \sup_{\alpha \in \Lambda} f_\alpha(\theta v + (1-\theta) w) \leq \sup_{\alpha \in \Lambda} [\theta f_\alpha(v) + (1-\theta) f_\alpha(w)] \\ &\leq \theta \sup_{\alpha \in \Lambda} f_\alpha(v) + (1-\theta) \sup_{\alpha \in \Lambda} f_\alpha(w) = \theta f(v) + (1-\theta) f(w). \end{aligned}$$

■

Theorem 104 *Let* V *and* W *be vector spaces, let* $L : W \to V$ *be a linear transformation and let* $f : V \to [-\infty, \infty]$ *be convex. Then the function* $g : W \to [-\infty, \infty]$*, defined by*

$$g(w) := f(L(w)), \quad w \in W,$$

is convex.

Proof. Since f is convex, for all w_1, $w_2 \in W$ and $\theta \in (0,1)$,

$$\begin{aligned} g(\theta w_1 + (1-\theta) w_2) &= f(L(\theta w_1 + (1-\theta) w_2)) = f(\theta L(w_1) + (1-\theta) L(w_2)) \\ &\leq \theta f(L(w_1)) + (1-\theta) f(L(w_2)) = \theta g(w_1) + (1-\theta) g(w_2). \end{aligned}$$

■

Theorem 105 *Let* V *and* W *be vector spaces, let* $L : V \to W$ *be a linear onto transformation and let* $f : V \to [-\infty, \infty]$ *be convex. Then the function* $\psi : W \to [-\infty, \infty]$*, defined by*

$$\psi(w) := \inf \{ f(v) : L(v) = w \}, \quad w \in W,$$

is convex.

Proof. Fix $w_1,\ w_2 \in W$ and $\theta \in (0,1)$. If $\psi(w_1) = \infty$ or $\psi(w_2) = \infty$, then there is nothing to prove. Thus, assume that $\psi(w_1) < \infty$ or $\psi(w_2) < \infty$. Let $v_1,\ v_2 \in V$ be such that $L(v_1) = w_1$ and $L(v_2) = w_2$ and $f(v_1) < \infty$, $f(v_2) < \infty$. Since L is linear,

$$L(\theta v_1 + (1-\theta) v_2) = \theta w_1 + (1-\theta) w_2,$$

and so

$$\psi(\theta w_1 + (1-\theta) w_2) \le f(\theta v_1 + (1-\theta) v_2) \le \theta f(v_1) + (1-\theta) f(v_2).$$

Taking the infimum over all such $v_1,\ v_2 \in V$ yields

$$\psi(\theta w_1 + (1-\theta) w_2) \le \theta \psi(w_1) + (1-\theta) \psi(w_2).$$

■

3.1 Regularity of Convex Functions

In this section we address continuity and differentiability properties of convex functions.

In the first result we prove that real-valued convex functions on finite-dimensional spaces are locally Lipschitz. This is hinged on the following characterization of convex functions on the real line.

If E is a subset of $\mathbb{R}^N$ and $f : E \to \mathbb{R}$, then the *oscillation* of f on E is defined by

$$\operatorname{osc}(f; E) := \sup\{|f(x_1) - f(x_2)| : x_1, x_2 \in E\}.$$

Theorem 106 *If $f : B(x_0, 2r) \subset \mathbb{R}^N \to \mathbb{R}$ is convex, then*

$$\operatorname{Lip}(f; B(x_0, r)) \le \frac{\operatorname{osc}(f; B(x_0, 2r))}{r}.$$

In particular, any convex function $f : \mathbb{R}^N \to (-\infty, \infty]$ is locally Lipschitz in the interior of its effective domain $\operatorname{dom}_e f$.

Proof. Step 1: Without loss of generality we may assume that $x_0 = 0$.[2] To see this, assume that $\operatorname{osc}(f; B(x_0, 2r))$ is finite and let w, $x \in B(x_0, r)$. Fix $0 < \varepsilon < r$. Suppose that $f(x) \geq f(w)$ and choose ξ to be a point of intersection of $\partial B(x_0, 2r - \varepsilon)$ with the ray from w through x such that $|w - \xi| \geq r - \varepsilon$. Define $g(t) := f(w + t\nu)$, where $\nu = (x - w) / |x - w|$. Since f is convex, then g is convex, and so, using the fact that $|\xi - w| \geq |x - w|$ we have

$$\frac{g(|x-w|) - g(0)}{|x-w|} \leq \frac{g(|\xi - w|) - g(0)}{|\xi - w|},$$

or equivalently

$$\frac{f(x) - f(w)}{|x-w|} \leq \frac{f(\xi) - f(w)}{|\xi - w|} \leq \frac{\operatorname{osc}(f; B(0, 2r))}{r - \varepsilon},$$

where we have used the fact that $|w - \xi| \geq r - \varepsilon$. Letting $\varepsilon \to 0^+$ yields

$$\frac{f(x) - f(w)}{|x-w|} \leq \frac{\operatorname{osc}(f; B(0, 2r))}{r} = L. \tag{25}$$

Step 2: Here we prove that if x_0 belongs to the interior of $\operatorname{dom}_e f$, then there exists a neighborhood of x_0 on which f is bounded and thus its oscillation is finite.

Without loss of generality we may assume that $x_0 = 0$, and consider in $\mathbb{R}^N$ the equivalent norm

$$\|x\|_\infty := \max\{|x_i| : i = 1, \ldots, N\}.$$

Let $\varepsilon > 0$ be such that $B_\infty(0, 2\varepsilon) \subset \operatorname{dom}_e f$ and set

$$a := \max\{f(x) : x_i \in \{-\varepsilon, 0, \varepsilon\}, i = 1, \ldots, N\}.$$

We claim that

$$f(x) \leq a \quad \text{for all } x \in \overline{B_\infty(0, \varepsilon)}. \tag{26}$$

Indeed, let x, $w \in \overline{B_\infty(0, \varepsilon)}$ with $w_i \in \{-\varepsilon, 0, \varepsilon\}$, $i = 1, \ldots, N$. If $x_N \neq 0$ write

$$x_N = \frac{|x_N|}{\varepsilon}(\operatorname{sgn} x_N)\varepsilon + \left(1 - \frac{|x_N|}{\varepsilon}\right)0.$$

By the convexity of f we have

$$\begin{aligned} f(w_1, \ldots, w_{N-1}, x_N) \leq & \frac{|x_N|}{\varepsilon} f(w_1, \ldots, w_{N-1}, (\operatorname{sgn} x_N)\varepsilon) \\ & + \left(1 - \frac{|x_N|}{\varepsilon}\right) f(w_1, \ldots, w_{N-1}, 0) \leq a. \end{aligned}$$

[2] In the proof of this theorem, $x^{(i)}$ denotes a vector of $\mathbb{R}^N$, while x_i is the ith component of a vector $x \in \mathbb{R}^N$.

The same inequality holds if $x_N = 0$. Similarly, if $x_{N-1} \neq 0$, then we have

$$\begin{aligned} f(w_1, \dots, &w_{N-2}, x_{N-1}, x_N) \\ &\leq \frac{|x_{N-1}|}{\varepsilon} f\left(w_1, \dots, w_{N-2}, (\operatorname{sgn} x_{N-1})\, \varepsilon, x_N\right) \\ &\quad + \left(1 - \frac{|x_{N-1}|}{\varepsilon}\right) f\left(w_1, \dots, w_{N-2}, 0, x_N\right) \leq a, \end{aligned}$$

where we have used the previous inequality. Recursively we obtain (26).

Next we show that

$$f(x) \geq 2f(0) - a \quad \text{for all } x \in \overline{B_\infty(0,\varepsilon)}. \tag{27}$$

Writing

$$0 = \frac{1}{2}x + \frac{1}{2}(-x),$$

we have

$$f(0) = f\left(\frac{1}{2}x + \frac{1}{2}(-x)\right) \leq \frac{1}{2}f(x) + \frac{1}{2}f(-x) \leq \frac{1}{2}f(x) + \frac{1}{2}a,$$

and so

$$f(x) \geq 2f(0) - a,$$

where we have used (26). ■

Remark 107 *It follows from the previous proof (see (27)) that if $f : B_\infty(x_0, 2r) \subset \mathbb{R}^N \to \mathbb{R}$ is convex, then*

$$\inf_{B_\infty(x_0,r)} f \geq 2f(x_0) - \sup_{B_\infty(x_0,r)} f.$$

Corollary 108 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be a proper convex function. Then the restriction of f to $\operatorname{ri}_{\mathrm{aff}}(\operatorname{dom}_e f)$ is locally Lipschitz. In particular, if $\operatorname{dom}_e f$ is affine, then the restriction of f to $\operatorname{dom}_e f$ is locally Lipschitz.*

Proof. If $\operatorname{dom}_e f$ consists of a point, then there is nothing to prove. If $\operatorname{dom}_e f$ has more than one point, by Proposition 59,

$$\operatorname{aff}(\operatorname{dom}_e f) = x_0 + W,$$

where $x_0 \in \operatorname{dom}_e f$ and W is a subspace of $\mathbb{R}^N$ of dimension $1 \leq \ell \leq N$. Construct an affine transformation $T : \mathbb{R}^\ell \to \mathbb{R}^N$ such that $T\left(\mathbb{R}^\ell\right) = x_0 + W$ and $T : \mathbb{R}^\ell \to x_0 + W$ is bijective. Define

$$g(w) := f(T(w)), \quad w \in \mathbb{R}^\ell.$$

Then $\operatorname{ri}_{\mathrm{aff}}(\operatorname{dom}_e g)$ reduces simply to the interior of $\operatorname{dom}_e g$. If $K \subset \operatorname{ri}_{\mathrm{aff}}(\operatorname{dom}_e f)$ is any compact set, then $T^{-1}(K)$ is compact (since $T : \mathbb{R}^\ell \to x_0 + W$ is bijective) and so by Theorem 106 applied to g, there exists a constant $L_K > 0$ such that

$$|g(w_1) - g(w_2)| \leq L_K |w_1 - w_2|$$

for all w_1, $w_2 \in T^{-1}(K)$, or equivalently,

$$|f(T(w_1)) - f(T(w_2))| \leq L_K |w_1 - w_2|$$

for all w_1, $w_2 \in T^{-1}(K)$. Hence,

$$|f(x_1) - f(x_2)| \leq L_K \left|T^{-1}(x_1) - T^{-1}(x_2)\right| \leq L_K C |x_1 - x_2|$$

for all x_1, $x_2 \in K$. ■

Exercise 109 *The previous corollary cannot be improved in general. Indeed, let $N = 2$ and consider the function*

$$f(x) = f(x_1, x_2) = \begin{cases} \dfrac{(x_2)^2}{2x_1} & \text{if } x_1 > 0, \\ 0 & \text{if } x_1 = x_2 = 0, \\ \infty & \text{otherwise.} \end{cases}$$

Prove that f is convex and lower semicontinuous in $\mathbb{R}^2$ and continuous everywhere except at the origin.

Corollary 110 *Let $A \subset \mathbb{R}^N$ be a relatively open convex set and let $f_\alpha : A \to \mathbb{R}$, $\alpha \in I$, be an arbitrary family of convex functions such that for every $x \in A$,*

$$\sup_{\alpha \in I} |f_\alpha(x)| < \infty.$$

Then for every compact set $K \subset A$ there exist two constants L_K, $M_K > 0$ such that

$$|f_\alpha(x)| \leq M_k$$

for all $x \in K$ and all $\alpha \in I$, and

$$|f_\alpha(x) - f_\alpha(y)| \leq L_k |x - y|$$

for all x, $y \in K$ and all $\alpha \in I$.

Proof. As in the previous corollary, without loss of generality we may assume that A is open. Fix any ball $\overline{B_\infty(x_0, 2r)} \subset A$. By Theorem 103 the function

$$g(x) := \sup_{\alpha \in I} f_\alpha(x)$$

is convex, and since by hypothesis it is finite on A, it is continuous by the previous theorem. Hence there exists $M > 0$ such that

$$g(x) = \sup_{\alpha \in I} f_\alpha(x) \leq M$$

for all $x \in \overline{B_\infty(x_0, 2r)}$. Since

$$\inf_{\alpha \in I} f_\alpha(0) > -\infty,$$

it follows by (27) that

$$f_\alpha(x) \geq \inf_{\alpha\in I} f_\alpha(0) - M =: m \quad \text{for all } x \in \overline{B_\infty(0, 2\varepsilon)}$$

and all $\alpha \in I$. In turn, by (25),

$$|f_\alpha(x) - f_\alpha(y)| \leq (M - m)\,|x - y|$$

for all x, $y \in \overline{B_\infty(x_0, r)}$ and all $\alpha \in I$. A simple compactness argument now gives the result for an arbitrary compact set $K \subset A$. ■

Exercise 111 *Prove that the previous corollary continues to hold if we only assume that*

$$\sup_{\alpha\in I} f_\alpha(x) < \infty$$

for all x in a dense subset of A and

$$\inf_{\alpha\in I} f_\alpha(x_0) > -\infty$$

for some $x_0 \in A$.

Theorem 112 *Let $A \subset \mathbb{R}^N$ be a relatively open convex set and let $f_n : A \to \mathbb{R}$, $n \in \mathbb{N}$, be a sequence of convex functions. Assume that there exists a set $E \subset A$ such that $\overline{E} \supset A$ and for every $x \in E$ there exists in $\mathbb{R}$ the limit*

$$f(x) := \lim_{n\to\infty} f_n(x)\,.$$

Then the limit $f(x) := \lim_{n\to\infty} f_n(x)$ exists in in $\mathbb{R}$ for every $x \in A$, the function $f : A \to \mathbb{R}$ is convex, and $\{f_n\}$ converges uniformly to f on any compact subset of A.

Proof. Exercise. ■

Definition 113 *Given a function $f : \mathbb{R}^N \to [-\infty, \infty]$ and a point $x_0 \in \mathbb{R}^N$ such that $f(x_0) \in \mathbb{R}$, the* one-sided directional derivative *of f at x_0 in the direction $v \in \mathbb{R}^N$ is defined by*

$$\frac{\partial^+ f}{\partial v}(x_0) := \lim_{t\to 0^+} \frac{f(x_0 + tv) - f(x_0)}{t}$$

whenever the limit exists.

Note that we allow the possibility that $\frac{\partial^+ f}{\partial v}(x_0)$ takes the values ∞ or $-\infty$.

Remark 114 *Note that*

$$\begin{aligned}\lim_{t\to 0^-} \frac{f(x_0 + tv) - f(x_0)}{t} &= -\lim_{t\to 0^-} \frac{f(x_0 - t(-v)) - f(x_0)}{-t} \\ &= -\lim_{s\to 0^+} \frac{f(x_0 + s(-v)) - f(x_0)}{s} = -\frac{\partial^+ f}{\partial(-v)}(x_0)\,,\end{aligned}$$

provided the limit exists. Hence if the directional derivative

$$\frac{\partial f}{\partial v}(x_0) := \lim_{t\to 0}\frac{f(x_0+tv)-f(x_0)}{t}$$

exists, then necessarily,

$$\begin{aligned}\frac{\partial^+ f}{\partial v}(x_0) &= \lim_{t\to 0^+}\frac{f(x_0+tv)-f(x_0)}{t} = \lim_{t\to 0^-}\frac{f(x_0+tv)-f(x_0)}{t}\\ &= -\frac{\partial^+ f}{\partial(-v)}(x_0).\end{aligned}$$

Monday, March 3, 2008

Theorem 115 *Let $f : \mathbb{R}^N \to [-\infty, \infty]$ be convex and let $x_0 \in \mathbb{R}^N$ be such that $f(x_0) \in \mathbb{R}$. Then $\frac{\partial^+ f}{\partial v}(x_0)$ exists for all $v \in \mathbb{R}^N$. Moreover the function $p : \mathbb{R}^N \to [-\infty, \infty]$ defined by*

$$p(v) := \frac{\partial^+ f}{\partial v}(x_0), \quad v \in \mathbb{R}^N,$$

is convex, positively homogeneous of degree one, and

$$-p(-v) \le p(v) \tag{28}$$

for all $v \in \mathbb{R}^N$.

Proof. To prove the existence of $\frac{\partial^+ f}{\partial v}(x_0)$, it suffices to consider the case $v \neq 0$, since $\frac{\partial^+ f}{\partial 0}(x_0) = 0$. Consider the function $g : \mathbb{R} \to [-\infty, \infty]$ defined by

$$g(t) := f(x_0 + tv), \quad t \in \mathbb{R}.$$

Then g is convex and

$$\frac{g(t) - g(0)}{t} = \frac{f(x_0 + tv) - f(x_0)}{t}$$

for all $t \in \mathbb{R}$. Thus to prove that $\frac{\partial^+ f}{\partial v}(x_0)$ exists, it remains to show that $g'_+(0)$ exists. If there exists $t_1 > 0$ such that $g(t_1) = -\infty$, then by convexity $g(t) = -\infty$ for all $t \in (0, t_1]$ so that

$$\frac{g(t) - g(0)}{t} = -\infty \to -\infty$$

as $t \to 0^+$. If there exists $t_1 > 0$ such that $g(t_1) = \infty$, then by convexity $g(t) = \infty$ for all $t > t_1$. Hence, if there is a decreasing sequence $t_n \to 0^+$ such that $g(t_n) = \infty$, then $g(t) = \infty$ for all $t \in (0, t_1]$, and so

$$\frac{g(t) - g(0)}{t} = \infty \to \infty.$$

The only case left is the case $g(t) \in \mathbb{R}$ for $t > 0$ near 0, which we already treated in Theorem .

To prove that p is positively homogeneous of degree one, fix $\lambda > 0$. Then

$$\begin{aligned} p(\lambda v) &= \frac{\partial^+ f}{\partial (\lambda v)}(x_0) = \lim_{t \to 0^+} \frac{f(x_0 + \lambda t v) - f(x_0)}{t} = \\ &= \lambda \lim_{t \to 0^+} \frac{f(x_0 + \lambda t v) - f(x_0)}{\lambda t} = \lambda \lim_{s \to 0^+} \frac{f(x_0 + s v) - f(x_0)}{s} \\ &= \lambda \frac{\partial^+ f}{\partial v}(x_0) = \lambda p(v). \end{aligned}$$

Next we prove that p is convex, let v_1, $v_2 \in \mathbb{R}^N$, $\theta \in (0,1)$, and $t > 0$. Assume that the sum

$$\theta p(v_1) + (1-\theta) p(v_2)$$

is well-defined. Then for all $t > 0$ sufficiently small, so is the sum

$$\theta \frac{f(x_0 + tv_1) - f(x_0)}{t} + (1-\theta) \frac{f(x_0 + tv_2) - f(x_0)}{t}.$$

Hence, from the convexity of f,

$$\begin{aligned}
&\frac{f(x_0 + t(\theta v_1 + (1-\theta) v_2)) - f(x_0)}{t} \\
&= \frac{f(\theta(x_0 + tv_1) + (1-\theta)(x_0 + tv_2)) - f(x_0)}{t} \\
&\leq \theta \frac{f(x_0 + tv_1) - f(x_0)}{t} + (1-\theta) \frac{f(x_0 + tv_2) - f(x_0)}{t}.
\end{aligned}$$

Letting $t \to 0^+$ yields

$$p(\theta v_1 + (1-\theta) v_2) \leq \theta p(v_1) + (1-\theta) p(v_2).$$

Finally, to prove that $-p(-v) \leq p(v)$ it suffices to assume that $p(v) < \infty$. Then by convexity,

$$0 = p(0) = p\left(\frac{1}{2} v + \frac{1}{2}(-v)\right) \leq \frac{1}{2} p(v) + \frac{1}{2} p(-v),$$

which gives the desired result. ■

Remark 116 *Under the hypotheses of the previous theorem, we have that the function p is subadditive in the sense that*

$$p(v_1 + v_2) \leq p(v_1) + p(v_2) \tag{29}$$

for all v_1, $v_2 \in \mathbb{R}^N$ for which the right-hand side makes sense. Indeed, by convexity and positive homogeneity of p we have

$$p(v_1 + v_2) = p\left(\frac{1}{2}(2v_1) + \frac{1}{2}(2v_2)\right) \leq \frac{1}{2} p(2v_1) + \frac{1}{2} p(2v_2) = p(v_1) + p(v_2).$$

Note that if f is differentiable at x_0, then p should be real-valued and linear. In particular, equality should hold in (28) and (29).

Theorem 117 *Let $f : \mathbb{R}^N \to [-\infty, \infty]$ be convex and let $x_0 \in \mathbb{R}^N$ be such that $f(x_0) \in \mathbb{R}$. If all the the one-sided directional derivatives $\frac{\partial^+ f}{\partial e_i}(x_0)$ and $\frac{\partial^+ f}{\partial(-e_i)}(x_0)$, $i = 1, \ldots, N$, are finite, then f is real-valued in a neighborhood of x_0. Moreover, if all the partial derivatives $\frac{\partial f}{\partial x_i}(x_0)$, $i = 1, \ldots, N$, exist (finite), then f is differentiable at x_0. In particular, any convex function $f : \mathbb{R}^N \to (-\infty, \infty]$ is differentiable $\mathcal{L}^N$ a.e. in the interior of its effective domain* $\operatorname{dom}_e f$.

Proof. Step 1: Since

$$\lim_{t\to 0^+} \frac{f(x_0+te_i)-f(x_0)}{t} = \frac{\partial^+ f}{\partial e_i}(x_0) \in \mathbb{R}$$
$$\lim_{t\to 0^+} \frac{f(x_0-te_i)-f(x_0)}{t} = \frac{\partial^+ f}{\partial(-e_i)}(x_0) \in \mathbb{R}$$

for all $i = 1, \ldots, N$, there exists $\delta > 0$ such that

$$\left|\frac{f(x_0 \pm te_i)-f(x_0)}{t} - \frac{\partial^+ f}{\partial(\pm e_i)}(x_0)\right| \le 1$$

for all $0 < t \le \delta$ and all $i = 1, \ldots, N$. In particular,

$$|f(x_0 \pm te_i)| \le C$$

for all $0 < t \le \delta$ and all $i = 1, \ldots, N$.

Define

$$\psi(h) := f(x_0+h), \quad h \in B\left(0, \frac{\delta}{N}\right).$$

Then the function ψ is convex. Hence, if we write

$$h = (h_1, \ldots, h_N) = \frac{1}{N}(Nh_1e_1 + \ldots + Nh_Ne_N),$$

for all $h \in B\left(0, \frac{\delta}{N}\right)$ we have

$$\begin{aligned}\psi(h) &= \psi\left(\frac{1}{N}(Nh_1e_1 + \ldots + Nh_Ne_N)\right)\\ &\le \frac{1}{N}(\psi(Nh_1e_1) + \ldots + \psi(Nh_Ne_N)) \le C.\end{aligned}$$

From the definition of ψ and its convexity,

$$f(x_0) = \psi(0) = \psi\left(\frac{h+(-h)}{2}\right) \le \frac{1}{2}[\psi(h)+\psi(-h)] \le \frac{1}{2}[\psi(h)+C],$$

for all $h \in B\left(0, \frac{\delta}{N}\right)$, we have that

$$\psi(h) \ge 2f(x_0) - C$$

for all $h \in B\left(0, \frac{\delta}{N}\right)$.

Step 2: By the previous step f is finite in some $B(x_0, r)$ for some $r > 0$. Define

$$g(h) := f(x_0+h) - f(x_0) - \nabla f(x_0) \cdot h, \quad h \in B(0, r).$$

Then the function g is convex, since it is given by the sum of a convex function and a linear function. Hence, if we write

$$h = (h_1, \ldots, h_N) = \frac{1}{N}(Nh_1e_1 + \ldots + Nh_Ne_N),$$

for all $h \in B\left(0, \frac{r}{N}\right)$ we have

$$g(h) = g\left(\frac{1}{N}(Nh_1e_1 + \ldots + Nh_Ne_N)\right) \leq \frac{1}{N}(g(Nh_1e_1) + \ldots + g(Nh_Ne_N)). \tag{30}$$

Since

$$g(Nh_ie_i) = f(x_0 + Nh_ie_i) - f(x_0) - \frac{\partial f}{\partial x_i}(x_0)\,Nh_i,$$

it follows from the definition of partial derivative that

$$\lim_{h_i \to 0} \frac{g(Nh_ie_i)}{Nh_i} = 0. \tag{31}$$

For any $x, y \in \mathbb{R}^N$ by the Cauchy-Schwarz inequality

$$x_1y_1 + \ldots + x_Ny_N \leq |x|\,|y| \leq |x|\,(|y_1| + \ldots + |y_N|).$$

Hence from (30) and using the fact that $0 = g(0) = 0$, we get

$$g(h) \leq \sum_{i:\, h_i \neq 0} h_i \frac{g(Nh_ie_i)}{h_iN} \leq |h| \sum_{i:\, h_i \neq 0} \left|\frac{g(Nh_ie_i)}{h_iN}\right|.$$

Similarly,

$$g(-h) \leq |h| \sum_{i:\, h_i \neq 0} \left|\frac{g(-Nh_ie_i)}{h_iN}\right|.$$

From the definition of g and its convexity,

$$0 = g(0) = g\left(\frac{h + (-h)}{2}\right) \leq \frac{1}{2}[g(h) + g(-h)],$$

or equivalently $g(h) \geq -g(-h)$. Thus,

$$-|h| \sum_{i:\, h_i \neq 0} \left|\frac{g(-Nh_ie_i)}{h_iN}\right| \leq -g(-h) \leq g(h) \leq |h| \sum_{i:\, h_i \neq 0} \left|\frac{g(Nh_ie_i)}{h_iN}\right|.$$

Diving by $|h|$ and using (31) gives

$$0 = \lim_{h \to 0} \frac{g(h)}{|h|} = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0) - \nabla f(x_0) \cdot h}{|h|}.$$

Step 3: Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be convex and for every fixed $i = 1, \ldots, N$, let

$$E_i := \left\{x \in (\operatorname{dom}_e f)^\circ : \frac{\partial f}{\partial x_i} \text{ exists at } x \text{ for all }\right\}.$$

We claim that $\mathcal{L}^N\left((\operatorname{dom}_e f)^\circ \setminus E_i\right) = 0$. For simplicity in the notation we assume $i = N$. Write

$$x = (x', t) \in \mathbb{R}^{N-1} \times \mathbb{R}.$$

Fix any $x' \in \mathbb{R}^{N-1}$ and consider the function

$$g^{x'}(t) := f(x', t), \quad t \in \mathbb{R}.$$

If the line $\{x'\} \times \mathbb{R}$ intersects $(\operatorname{dom}_e f)^\circ$, fix any $t_0 \in \mathbb{R}$ such that $(x', t_0) \in (\operatorname{dom}_e f)^\circ$ and let $I \subset \mathbb{R}$ be the largest segment such that $\{x'\} \times I \subset (\operatorname{dom}_e f)^\circ$. Since the function $g^{x'} : I \to \mathbb{R}$ is convex, it follows from Corollary 10 that $g^{x'}$ is differentiable in I except on a countable number of points. Note that if $g^{x'}$ is differentiable at $t_0 \in I$, then

$$\left(g^{x'}\right)'(t_0) = \lim_{t \to t_0} \frac{g^{x'}(t) - g^{x'}(t_0)}{t - t_0} = \lim_{t \to t_0} \frac{f(x', t) - f(x', t_0)}{t - t_0} = \frac{\partial f}{\partial x_N}(x', t_0).$$

Since the section

$$\left((\operatorname{dom}_e f)^\circ\right)_{x'} := \left\{t \in \mathbb{R} : (x', t) \in (\operatorname{dom}_e f)^\circ\right\}$$

is an open set in $\mathbb{R}$ we may write as a disjoint union of open intervals. Thus, we have shown that $\frac{\partial f}{\partial x_N}(x', \cdot)$ exists in $\left((\operatorname{dom}_e f)^\circ\right)_{x'}$ except on a countable number of points, that is, that the set $\left((\operatorname{dom}_e f)^\circ \setminus E_N\right)_{x'}$ is countable. In particular, $\mathcal{L}^1\left(\left((\operatorname{dom}_e f)^\circ \setminus E_i\right)_{x'}\right) = 0$. By Tonelli's theorem

$$\mathcal{L}^N\left((\operatorname{dom}_e f)^\circ \setminus E_i\right) = \int_{\mathbb{R}^{N-1}} \mathcal{L}^1\left(\left((\operatorname{dom}_e f)^\circ \setminus E_i\right)_{x'}\right) dx' = 0.$$

This concludes the proof. ■

Remark 118 *(i) If a function $f : B(x_0, r) \subset \mathbb{R}^N \to \mathbb{R}$ is differentiable at x_0, then*

(a) all directional derivatives $\frac{\partial f}{\partial v}(x_0)$ exist;

(b) $v \in \mathbb{R}^N \mapsto \frac{\partial f}{\partial v}(x_0)$ is linear, that is,

$$\frac{\partial f}{\partial v}(x_0) = \nabla f(x_0) \cdot v, \quad v \in \mathbb{R}^N.$$

It is easy to construct (nonconvex) functions for which (a) and (b) hold, but the function f is not differentiable there. For example, the function

$$f(x, y) := \begin{cases} x & \text{if } y = x^2, \\ 0 & \text{otherwise} \end{cases}$$

satisfies (a) and (b), but it is not differentiable at $(0, 0)$.

(ii) The second part of the proof follows also from Theorem 106 and Rademacher's theorem.

(iii) It can actually be proved that

$$\left\{x \in (\mathrm{dom}_e\, f)^\circ : \, f \text{ is not differentiable at } x\right\} \subset \bigcup_{n=1}^{\infty} K_n,$$

where K_n is compact and $\mathcal{H}^{N-1}(K_n) < \infty$. (Anderson and Klee, Theorem 3.1 page 353).

(iv) It follows by the previous theorem that if $x_0 \in \mathrm{dom}_e\, f \setminus (\mathrm{dom}_e\, f)^\circ$, for any *orthonormal basis $\{e_1, \ldots, e_N\}$, one of the one-sided directional derivatives $\frac{\partial^+ f}{\partial e_i}(x_0)$ and $\frac{\partial^+ f}{\partial(-e_i)}(x_0)$ must be $\pm\infty$ for some $i = 1, \ldots, N$.*

Wednesday, March 5, 2008

Next we prove that differentiable convex functions are of class C^1.

Theorem 119 *Let $B \subset \mathbb{R}^N$ be an open ball. If $f : B \to \mathbb{R}$ is convex and E is the set of points in B at which f is differentiable, then $\nabla f : E \to \mathbb{R}^N$ is continuous.*

Proof. Let $x_0 \in E$ and define

$$g(x) := f(x) - f(x_0) - \nabla f(x_0) \cdot (x - x_0), \quad x \in B.$$

Then the function g is convex and differentiable in E. By Theorem 106,

$$|\nabla f(x) - \nabla f(x_0)| = |\nabla g(x)| \leq \operatorname{Lip}(g; B(x_0, r)) \leq \frac{\operatorname{osc}(g; B(x_0, 2r))}{r}$$

for any $x \in E$ with $|x - x_0| < r$ and with $B(x_0, 2r) \subset B$. Since f is differentiable at x_0 we have that

$$\lim_{r \to 0^+} \sup_{x \in E,\, |x - x_0| < r} |\nabla f(x) - \nabla f(x_0)| \leq \lim_{r \to 0^+} \frac{\operatorname{osc}(g; B(x_0, 2r))}{r} = 0,$$

and thus $\nabla f(x) \to \nabla f(x_0)$ as $x \to x_0$, $x \in E$. ■

Remark 120 *The (nonconvex) function*

$$f(x) = \begin{cases} x^2 \sin^2 \frac{1}{x} - 1 & \text{if } x \neq 0, \\ -1 & \text{if } x = 0 \end{cases}$$

is differentiable in $\mathbb{R}$, but f' is not continuous in $\mathbb{R}$.

We will actually prove more, namely, that ∇f is a function of bounded variation.

Definition 121 *Let $\Omega \subset \mathbb{R}^N$ be an open set. We say that a function $g \in L^1(\Omega)$ has* bounded variation *and we write $g \in BV(\Omega)$, if there exist finite signed Radon measures $\mu_1, \dots, \mu_N \in \mathcal{M}(\Omega; \mathbb{R})$ such that*

$$\int_\Omega \frac{\partial \psi}{\partial x_i} g \, dx = -\int_\Omega \psi \, d\mu_i$$

for all $i = 1, \dots, N$. We say that a function $g \in L^1_{\mathrm{loc}}(\Omega)$ has locally bounded variation *if $g \in BV(\Omega')$ for every open set $\Omega' \subset\subset \Omega$.*

Theorem 122 *If $f : B(x_0, r) \subset \mathbb{R}^N \to \mathbb{R}$ is convex, then $\frac{\partial f}{\partial x_i} \in BV_{\mathrm{loc}}(B(x_0, r))$, $i = 1, \dots, N$.*

Proof. Without loss of generality assume that $x_0 = 0$. Let $\psi \in C_c^\infty(B(0,r))$ and let $0 < \varepsilon < \operatorname{dist}(\operatorname{supp}\psi, \partial B(0,r))$. Let φ_ε be a standard mollifier, and define

$$f_\varepsilon(x) := \int_{B(0,r)} \varphi_\varepsilon(x-y)\, f(y)\, dy, \quad x \in B(0, r-\varepsilon).$$

We claim that f_ε is convex. To see this, let $x_1, x_2 \in B(0, r-\varepsilon)$ and $\theta \in (0,1)$. By a simple change of variables

$$f_\varepsilon(x) := \int_{B(0,r)} \varphi_\varepsilon(z)\, f(x-z)\, dz, \quad x \in B(0, r-\varepsilon),$$

and so by the convexity of f and the fact that $\varphi_\varepsilon \geq 0$,

$$\begin{aligned}
f_\varepsilon(\theta x_1 + (1-\theta)x_2) &= \int_{B(0,r)} \varphi_\varepsilon(z)\, f(\theta x_1 + (1-\theta)x_2 - z)\, dz \\
&= \int_{B(0,r)} \varphi_\varepsilon(z)\, f(\theta(x_1 - z) + (1-\theta)(x_2 - z))\, dz \\
&\leq \theta \int_{B(0,r)} \varphi_\varepsilon(z)\, f(x_1 - z)\, dz + (1-\theta)\int_{B(0,r)} \varphi_\varepsilon(z)\, f(x_2 - z)\, dz \\
&= \theta f_\varepsilon(x_1) + (1-\theta) f_\varepsilon(x_2).
\end{aligned}$$

Since $f_\varepsilon \in C^\infty(B(0, r-\varepsilon))$, we have that its Hessian matrix $\left(\frac{\partial^2 f_\varepsilon}{\partial x_i \partial x_j}\right)_{i,j=1}^N$ is semipositive definite. Hence for any vector $\nu \in \mathbb{R}^N$

$$\sum_{i,j=1}^N \frac{\partial^2 f_\varepsilon}{\partial x_i \partial x_j}(x)\, \nu_i \nu_j \geq 0, \quad x \in B(0, r-\varepsilon).$$

Moreover, integrating by parts twice yields

$$\int_{B(0,r)} \psi(x) \sum_{i,j=1}^N \frac{\partial^2 f_\varepsilon}{\partial x_i \partial x_j}(x)\, \nu_i \nu_j\, dx = \int_{B(0,r)} f_\varepsilon(x) \sum_{i,j=1}^N \frac{\partial^2 \psi}{\partial x_i \partial x_j}(x)\, \nu_i \nu_j\, dx.$$

Note that if $\psi \geq 0$, then so is the integral. Since f is continuous, we have that f_ε converges uniformly to f in $\operatorname{supp}\psi$, and so letting $\varepsilon \to 0^+$ we obtain that

$$L_\nu(\psi) := \int_{B(0,r)} f(x) \sum_{i,j=1}^N \frac{\partial^2 \psi}{\partial x_i \partial x_j}(x)\, \nu_i \nu_j\, dx$$

for all $\psi \in C_c^\infty(B(0,r))$, with $L_\nu(\psi) \geq 0$ whenever $\psi \geq 0$. Since $L_\nu : \psi \in C_c^\infty(B(0,r)) \to \mathbb{R}$ is linear and nonnegative, it follows by the Riesz representation theorem that there exists a measure $\mu_\nu : \mathcal{B}(B(0,r)) \to [0,\infty]$ finite on compact sets, such that

$$L_\nu(\psi) = \int_{B(0,r)} \psi\, d\mu_\nu$$

for all $\psi \in C_c^\infty(B(0,r))$.

Fix i, $j = 1,\ldots,N$. If $i = j$ define $\mu_{ii} := \mu_{e_i}$. If $i \neq j$ consider the vector $y = \frac{e_i+e_j}{2}$. In this case,

$$\sum_{i,j=1}^{N} \frac{\partial^2 \psi}{\partial x_i \partial x_j}(x)\,\nu_i \nu_j = \frac{1}{2}\left[\frac{\partial^2 \psi}{\partial x_i \partial x_i}(x) + 2\frac{\partial^2 \psi}{\partial x_i \partial x_j}(x) + \frac{\partial^2 \psi}{\partial x_j \partial x_j}(x)\right],$$

and so

$$\begin{aligned}
\int_{B(0,r)} f\frac{\partial^2 \psi}{\partial x_i \partial x_j}\,dx &= \int_{B(0,r)} f\sum_{i,j=1}^{N}\frac{\partial^2 \psi}{\partial x_i \partial x_j}\nu_i\nu_j\,dx \\
&\quad - \frac{1}{2}\left[\int_{B(0,r)} f\frac{\partial^2 \psi}{\partial x_i \partial x_i}\,dx + \int_{B(0,r)} f\frac{\partial^2 \psi}{\partial x_j \partial x_j}\,dx\right] \\
&= \int_{B(0,r)} \psi\,d\mu_{\frac{e_i+e_j}{2}} - \frac{1}{2}\left[\int_{B(0,r)} \psi\,d\mu_{e_i} + \int_{B(0,r)} \psi\,d\mu_{e_j}\right] \\
&= \int_{B(0,r)} \psi\,d\mu_{ij},
\end{aligned}$$

where $\mu_{ij} := \mu_{\frac{e_i+e_j}{2}} - \frac{1}{2}\mu_{e_i} - \frac{1}{2}\mu_{ej}$. Then we have proved that for all i, $j = 1,\ldots,N$ and for all $\psi \in C_c^\infty(B(0,r))$,

$$\int_{B(0,r)} f\frac{\partial^2 \psi}{\partial x_i \partial x_j}\,dx = \int_{B(0,r)} \psi\,d\mu_{ij}.$$

Integrating by parts, yields

$$-\int_{B(0,r)} \frac{\partial f}{\partial x_i}\frac{\partial \psi}{\partial x_j}\,dx = \int_{B(0,r)} \psi\,d\mu_{ij},$$

which implies that for every $i = 1,\ldots,N$, the weak partial derivatives of $\frac{\partial f}{\partial x_i}$ are the signed measures μ_{ij}, $j = 1,\ldots,N$, that is, that $\frac{\partial f}{\partial x_i} \in BV_{\text{loc}}(B(0,r))$. ■

Monday, March 17, 2008

3.2 Lower Semicontinuous Functions

Definition 123 *A function $f : \mathbb{R}^N \to [-\infty, \infty]$ is said to be*

(i) lower semicontinuous *if the set $\left\{x \in \mathbb{R}^N : f(x) \le t\right\}$ is closed for every $t \in \mathbb{R}$.*

(ii) upper semicontinuous *if $-f$ is lower semicontinuous (respectively sequentially lower semicontinuous).*

Exercise 124 *Let $f : \mathbb{R}^N \to [-\infty, \infty]$. Prove that the following conditions are equivalent:*

(i) f is lower semicontinuous;

(ii) epi f *is closed;*

(iii) for every $x_0 \in \mathbb{R}^N$

$$f(x_0) \le \liminf_{x \to x_0} f(x).$$

Exercise 125 *Let $\{f_\alpha\}$ be a (possibly uncountable) family of lower semicontinuous functions, $f_\alpha : \mathbb{R}^N \to [-\infty, \infty]$.*

(i) Prove that the function

$$f_+ := \sup_\alpha f_\alpha$$

is still lower semicontinuous.

(ii) Prove that if the family $\{f_\alpha\}$ is finite, then the function

$$f_- := \min_\alpha f_\alpha$$

is still lower semicontinuous.

Definition 126 *Given a function $f : \mathbb{R}^N \to [-\infty, \infty]$, the* lower semicontinuous envelope $\operatorname{lsc} f : \mathbb{R}^N \to [-\infty, \infty]$ *of f is defined by*

$$\operatorname{lsc} f(x) := \sup \left\{g(x) : g : \mathbb{R}^N \to [-\infty, \infty] \text{ is lower semicontinuous, } g \le f\right\}, \quad x \in \mathbb{R}^N.$$

We now relate the various types of convex envelopes.

Definition 127 *Let $f : \mathbb{R}^N \to [-\infty, \infty]$. The* lower semicontinuous envelope $\operatorname{lsc} f : \mathbb{R}^N \to [-\infty, \infty]$ *of f is defined by*

$$\operatorname{lsc} f(x) := \sup \left\{g(x) : g : \mathbb{R}^N \to [-\infty, \infty] \text{ is lower semicontinuous, } g \le f\right\}.$$

Exercise 128 *Prove that for any* $f : \mathbb{R}^N \to [-\infty, \infty]$,

$$\begin{aligned}\operatorname{lsc} f(x) &= \inf_{\{x_n\}} \left\{ \liminf_{n\to\infty} f(x_n) : \{x_n\} \subset \mathbb{R}^N,\ x_n \to x \right\} \\ &= \min \left\{ f(x), \liminf_{y\to x} f(y) \right\}.\end{aligned}$$

and that $\operatorname{epi}(\operatorname{lsc} f) = \overline{\operatorname{epi} f}$.

Proposition 129 *Let* $f : \mathbb{R}^N \to [-\infty, \infty]$ *be convex. Then* $\operatorname{lsc} f$ *is convex, and*

$$\operatorname{epi}(\operatorname{lsc} f) = \overline{\operatorname{epi} f}.$$

Proof. Since f is convex, then $\operatorname{epi} f$ is convex by Proposition 97, and by the previous exercise we have

$$\operatorname{epi}(\operatorname{lsc} f) = \overline{\operatorname{epi} f},$$

hence $\operatorname{epi}(\operatorname{lsc} f)$ is convex because it is the closure of a convex set, i.e., $\operatorname{lsc} f$ is convex. ■

Corollary 108 implies in particular that the lower semicontinuous envelope of a proper convex function coincides with the function except at most on the relative boundary of its effective domain. Indeed, we have the following result.

Theorem 130 *Let* $f : \mathbb{R}^N \to (-\infty, \infty]$ *be a proper convex function. Then* $\operatorname{lsc} f$ *agrees with* f *everywhere except possibly on* $\operatorname{rb}_{\text{aff}}(\operatorname{dom}_e f)$. *Moreover, for any fixed* $x_0 \in \operatorname{ri}_{\text{aff}}(\operatorname{dom}_e f)$ *and for any* $x \in \mathbb{R}^N$,[3]

$$f(x) \geq \operatorname{lsc} f(x) = \lim_{\theta \to 1^-} f((1-\theta)x_0 + \theta x). \tag{32}$$

Proof. Step 1: Fix any

$$x \notin \operatorname{rb}_{\text{aff}}(\operatorname{dom}_e f) = \overline{\operatorname{dom}_e f} \setminus \operatorname{ri}_{\text{aff}}(\operatorname{dom}_e f).$$

By the previous exercise, for every $x \in \mathbb{R}^N$,

$$\operatorname{lsc} f(x) = \inf_{\{x_n\}} \left\{ \liminf_{n\to\infty} f(x_n) : \{x_n\} \subset \mathbb{R}^N,\ x_n \to x \right\},$$

and so we may find $\{x_n\} \subset \mathbb{R}^N$ such that $x_n \to x$ and

$$\operatorname{lsc} f(x) = \lim_{n\to\infty} f(x_n).$$

We now distinguish two cases.

If $x \in \operatorname{ri}_{\text{aff}}(\operatorname{dom}_e f)$, then since

$$\infty > f(x) \geq \operatorname{lsc} f(x) = \lim_{n\to\infty} f(x_n),$$

[3] esercizio?

it follows that $x_n \in \mathrm{ri}_{\mathrm{aff}}(\mathrm{dom}_e f)$ for all n sufficiently large, and thus, using the continuity of f in $\mathrm{ri}_{\mathrm{aff}}(\mathrm{dom}_e f)$ (see Corollary 108), we obtain that

$$\mathrm{lsc}\, f(x) = \lim_{n\to\infty} f(x_n) = f(x).$$

If $x \notin \overline{\mathrm{dom}_e f}$, then $x_n \notin \mathrm{dom}_e f$ for all n sufficiently large, and so

$$f(x) \geq \mathrm{lsc}\, f(x) = \lim_{n\to\infty} f(x_n) = \infty,$$

which concludes the first part of the theorem.

Step 2: To prove (32) fix any $x_0 \in \mathrm{ri}_{\mathrm{aff}}(\mathrm{dom}_e f)$ (note that since f is proper, by Proposition 77 we may always find at least one). We again distinguish two cases.

If $x \in \overline{\mathrm{dom}_e f}$, then in view of the convexity of $\mathrm{dom}_e f$, by Proposition 77 we have

$$\theta x + (1-\theta)x_0 \in \mathrm{ri}_{\mathrm{aff}}(\mathrm{dom}_e f)$$

for all $0 \leq \theta < 1$. Hence by Step 1,

$$f(\theta x + (1-\theta)x_0) = \mathrm{lsc}\, f(\theta x + (1-\theta)x_0)$$

for all $0 \leq \theta < 1$. The lower semicontinuous function

$$g(\theta) := \mathrm{lsc}\, f(\theta x + (1-\theta)x_0), \quad \theta \in [0,1],$$

is convex and real-valued (except possibly at $\theta = 1$), and so continuous in $[0,1]$. Therefore

$$\begin{aligned} f(x) \geq \mathrm{lsc}\, f(x) &= g(1) = \lim_{\theta\to 1^-} g(\theta) \\ &= \lim_{\theta\to 1^-} \mathrm{lsc}\, f(\theta x + (1-\theta)x_0) = \lim_{\theta\to 1^-} f(\theta x + (1-\theta)x_0). \end{aligned}$$

Finally, if $x \notin \overline{\mathrm{dom}_e f}$, then $\theta x + (1-\theta)x_0 \notin \mathrm{dom}_e f$ for all θ sufficiently close to one, and so again by Step 1,

$$\infty = f(\theta x + (1-\theta)x_0) = \mathrm{lsc}\, f(\theta x + (1-\theta)x_0)$$

for all θ sufficiently close to one, which yields (32). ■

Example 131 *The previous theorem can be used to show that a function is convex. As an example, consider the function*

$$g(x) := \begin{cases} -\sqrt{1-|x|^2} & \text{for } |x| \leq 1 \\ \infty & \text{otherwise.} \end{cases}$$

Then $\mathrm{dom}_e g = \overline{B(0,1)}$. *Using the second partial derivative condition, we can check that g is convex in $B(0,1)$. If we now define*

$$f(x) := \begin{cases} -\sqrt{1-|x|^2} & \text{for } |x| < 1, \\ \infty & \text{otherwise,} \end{cases}$$

we have that f is convex. In view of the previous theorem, it follows that $g = \mathrm{lsc}\, f$, *and so g is convex.*

Corollary 132 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be a proper lower semicontinuous convex function. Then for any fixed $x_0 \in \operatorname{dom}_e f$ and for any $x \in \mathbb{R}^N$,*

$$f(x) = \lim_{\theta \to 1^-} f((1-\theta)x_0 + \theta x).$$

Proof. The function

$$g(t) := f(tx + (1-t)x_0), \quad t \in \mathbb{R},$$

is proper, convex, lower semicontinuous, with $g(0) = f(x_0) \in \mathbb{R}$ and $g(1) = f(x)$. The effective domain of g is an interval I. If $I^\circ \cap [0,1] \neq \emptyset$, then, taking any $t_1 \in I^\circ \cap [0,1]$, for $t > t_1$ we have that $t = (1-\theta)t_1 + \theta 1$, and so by the previous theorem

$$g(1) = \operatorname{lsc} g(1) = \lim_{\theta \to 1^-} g((1-\theta)t_1 + \theta 1) = \lim_{t \to 1^-} g(t).$$

If $I^\circ \cap [0,1] = \emptyset$, then $g(t) \equiv \infty$ in $[0,1]$, since g is lower semicontinuous. ■

Next we prove that if f is lower semicontinuous, then f can be written as the pointwise supremum of a countable number of affine functions below it.

In the case that f is real-valued it is possible to give an explicit characterization of the coefficients a_i and b_i in the previous proposition. Let $f : \mathbb{R}^N \to \mathbb{R}$ be a convex function and let $\varphi \in C_c^1\left(\mathbb{R}^N\right)$ be any function with $\varphi \geq 0$ and $\int_{\mathbb{R}^N} \varphi(x)\, dx = 1$. Define

$$\begin{aligned} a_\varphi &:= \int_{\mathbb{R}^N} f(x)\left((N+1)\varphi(x) + \nabla\varphi(x)\cdot x\right) dx, \\ b_\varphi &:= -\int_{\mathbb{R}^N} f(x)\nabla\varphi(x)\, dx. \end{aligned}$$

When necessary we will also write $a_\varphi(f)$ and $b_\varphi(f)$ to highlight the dependence on f.

Theorem 133 (De Giorgi) *Let f and φ be as above. Then*

(i) $f(x) \geq a_\varphi + b_\varphi \cdot x$ for all $x \in \mathbb{R}^N$;

(ii) $f(x) = \sup_{k\in\mathbb{N},\, q\in\mathbb{Q}^N}\left\{a_{\varphi_{k,q}} + b_{\varphi_{k,q}}\cdot x\right\}$ for all $x \in \mathbb{R}^N$, where

$$\varphi_{k,q}(x) := k^N \varphi(k(q-x)), \quad x \in \mathbb{R}^N. \tag{33}$$

Proof. (i) Assume first that $f \in C^1\left(\mathbb{R}^N\right)$. Since f is differentiable, for any for any $x, y \in \mathbb{R}^N$ the function $g : \mathbb{R} \to \mathbb{R}$, defined by

$$g(t) := f(tx + (1-t)y), \quad t \in \mathbb{R},$$

is differentiable and

$$g'(t) = (\nabla f(tx + (1-t)y))\cdot(x-y).$$

By Theorem 16,

$$g(1) \geq g(0) + g'(0)(1-0),$$

that is,

$$f(x) \geq f(y) + \nabla f(y)\cdot(x-y).$$

Multiply the previous inequality by $\varphi(y)$ and integrate in y over $\mathbb{R}^N$ to obtain

$$f(x) \geq \int_{\mathbb{R}^N}(f(y) - \nabla f(y)\cdot y)\varphi(y)\, dy + x\cdot\int_{\mathbb{R}^N}\nabla f(y)\varphi(y)\, dy.$$

Integrating by parts now yields

$$f(x) \geq \int_{\mathbb{R}^N} f(y)\left((N+1)\varphi(y) + \nabla\varphi(y)\cdot y\right) dy - x\cdot\int_{\mathbb{R}^N} f(y)\nabla\varphi(y)\, dy.$$

This proves (i) when $f \in C^1\left(\mathbb{R}^N\right)$. In the general case, let ψ_ε be a standard mollifier. Applying the previous inequality to the smooth convex function $f_\varepsilon :=$ O Yosida

$\psi_\varepsilon * f$ gives

$$f_\varepsilon(x) \geq \int_{\mathbb{R}^N} f_\varepsilon(y)\left((N+1)\varphi(y) + \nabla\varphi(y)\cdot y\right)\, dy - x\cdot\int_{\mathbb{R}^N} f_\varepsilon(y)\nabla\varphi(y)\, dy.$$

Since φ has compact support, by Theorem 112 we may now let $\varepsilon \to 0^+$.

(ii) Let $k \in \mathbb{N}$, $q \in \mathbb{Q}^N$. By replacing the function φ with $\varphi_{k,q}$ in (i) we obtain

$$f(x) \geq a_{\varphi_{k,q}} + b_{\varphi_{k,q}} \cdot x \text{ for all } x \in \mathbb{R}^N$$

and hence $f \geq g$, where

$$g(x) := \sup_{k\in\mathbb{N},\, q\in\mathbb{Q}^N} \left\{a_{\varphi_{k,q}} + b_{\varphi_{k,q}} \cdot x\right\}.$$

Since g is everywhere finite and convex, it is continuous, and so is f (see Corollary 108). Hence, to prove (ii) it suffices to show that

$$f(q) = g(q) \text{ for all } q \in \mathbb{Q}^N,$$

since $\mathbb{Q}^N$ is dense in $\mathbb{R}^N$.

For $k \in \mathbb{N}$, $q \in \mathbb{Q}^N$, we have

$$\begin{aligned} & a_{\varphi_{k,q}} + b_{\varphi_{k,q}} \cdot x \\ &= \int_{\mathbb{R}^N} f(y)(N+1)k^N\varphi(k(q-y)) - k^{N+1}\nabla\varphi(k(q-y))\cdot(y-x)\, dy \\ &= \int_{\mathbb{R}^N} f\left(q - \frac{w}{k}\right)\left((N+1)\varphi(w) - \nabla\varphi(w)\cdot(k(q-x)-w)\right)\, dw, \end{aligned}$$

where we have made the change of variables $w = k(q-y)$. Taking $x = q$ we obtain

$$a_{\varphi_{k,q}} + b_{\varphi_{k,q}} \cdot q = \int_{\mathbb{R}^N} f\left(q - \frac{w}{k}\right)\left((N+1)\varphi(w) + \nabla\varphi(w)\cdot w\right)\, dw.$$

Since the integrand is continuous and with compact support, we may let $k \to \infty$ to get

$$\lim_{k\to\infty} a_{\varphi_{k,q}} + b_{\varphi_{k,q}} \cdot q = f(q)\int_{\mathbb{R}^N}\left((N+1)\varphi(w) + \nabla\varphi(w)\cdot w\right)\, dw = f(q),$$

where we have used the facts that

$$\int_{\mathbb{R}^N} \varphi(w)\, dw = 1, \quad \int_{\mathbb{R}^N} \nabla\varphi(w)\cdot w\, dw = -N\int_{\mathbb{R}^N}\varphi(w)\, dw = -N.$$

Since

$$f(q) = \lim_{k\to\infty} a_{\varphi_{k,q}} + b_{\varphi_{k,q}} \cdot q \leq g(q),$$

(ii) follows. ■

Friday, March 21, 2008

To extend the previous theorem to the case in which f can take the value ∞, we need the following approximation result that we will prove in the next section.

Theorem 134 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be a proper convex lower semicontinuous function. For every $\varepsilon > 0$, let*

$$f_\varepsilon(x) := \inf_{y \in \mathbb{R}^N} \left\{ f(y) + \frac{1}{2\varepsilon} |x - y|^2 \right\}, \quad x \in \mathbb{R}^N.$$

Then $f_\varepsilon : \mathbb{R}^N \to \mathbb{R}$ is convex, differentiable, and $f_\varepsilon \nearrow f$ as $\varepsilon \to 0^+$.

The function f_ε is called the Moreau-Yosida approximation of f.

Proposition 135 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be convex and lower semicontinuous. Then*

$$f(x) = \sup_{i \in \mathbb{N}} \{a_i + b_i \cdot x\} \tag{34}$$

for all $x \in \mathbb{R}^N$ and for some $a_i \in \mathbb{R}$, $b_i \in \mathbb{R}^N$.

Proof. Consider any sequence $\varepsilon_n \to 0^+$. By the previous theorem the Moreau-Yosida approximation f_{ε_n} is real-valued, and so

$$f_{\varepsilon_n}(x) = \sup_{k \in \mathbb{N}} \{a_{k,n} + b_{k,n} \cdot x\}$$

for all $x \in \mathbb{R}^N$. Hence

$$f(x) = \sup_{n \in \mathbb{N}} f_{\varepsilon_n}(x) = \sup_{n,\, k \in \mathbb{N}} \{a_{k,n} + b_{k,n} \cdot x\}$$

for all $x \in \mathbb{R}^N$. ■

Remark 136 *If f takes the value $-\infty$, then the previous result fails, since there cannot be any affine function below f. Note that there exist functions $f : \mathbb{R}^N \to \{-\infty, \infty\}$ that are convex and lower semicontinuous with $f \not\equiv -\infty$. As an example let $N = 1$ and define*

$$f(x) := \begin{cases} \infty & \text{if } x > 0, \\ -\infty & \text{if } x \le 0. \end{cases}$$

3.3 Subdifferentiability

Next we study differentiability properties of convex functions. We begin by introducing the notion of subdifferential.

Definition 137 *Let* $f : \mathbb{R}^N \to [-\infty, \infty]$*, and let* $x_0 \in \mathbb{R}^N$ *be such that* $f(x_0) \in \mathbb{R}$*. The function* f *is said to be* subdifferentiable *at* x_0 *if there exists* $y_0 \in \mathbb{R}^N$ *such that*

$$f(x) \geq f(x_0) + y_0 \cdot (x - x_0) \quad \textit{for all } x \in \mathbb{R}^N.$$

The element y_0 *is called a* subgradient *of* f *at* x_0*, and the set of all subgradients at* x_0 *is called the* subdifferential *of* f *at* x_0 *and is denoted by* $\partial f(x_0)$*. Precisely,*

$$\partial f(x_0) = \left\{ y_0 \in \mathbb{R}^N : f(x) \geq f(x_0) + y_0 \cdot (x - x_0) \quad \textit{for all } x \in \mathbb{R}^N \right\}.$$

If f *is not subdifferentiable at* x_0*, then* $\partial f(x_0) := \emptyset$*.*

Proposition 138 *If* $f : \mathbb{R}^N \to [-\infty, \infty]$ *is subdifferentiable at some* $x_0 \in \mathbb{R}^N$*, then* $\partial f(x_0)$ *is a closed and convex set.*

Proof. If $y_1, y_2 \in \partial f(x_0)$, then

$$f(x) \geq f(x_0) + y_i \cdot (x - x_0) \quad \text{for all } x \in \mathbb{R}^N \text{ and all } i = 1, 2.$$

If $\theta \in (0, 1)$, multiplying the first inequality by θ and the second by $1 - \theta$ and adding them shows that

$$f(x) \geq f(x_0) + (\theta y_1 + (1 - \theta) y_2) \cdot (x - x_0) \quad \text{for all } x \in \mathbb{R}^N.$$

Thus $\theta y_1 + (1 - \theta) y_2 \in \partial f(x_0)$, which shows that $\partial f(x_0)$ is convex. Moreover, if $y_n \in \partial f(x_0)$ and $y_n \to y_0$, then

$$f(x) \geq f(x_0) + y_n \cdot (x - x_0) \quad \text{for all } x \in \mathbb{R}^N \text{ and all } n \in \mathbb{N},$$

and so letting $n \to \infty$ gives

$$f(x) \geq f(x_0) + y_0 \cdot (x - x_0) \quad \text{for all } x \in \mathbb{R}^N \text{ and all } n \in \mathbb{N},$$

which shows that $y_0 \in \partial f(x_0)$. Thus $\partial f(x_0)$ is closed. ■

Note that

$$f(x_0) = \min_{x \in \mathbb{R}^N} f(x) \quad \text{if and only if } 0 \in \partial f(x_0). \tag{35}$$

Exercise 139 *Find the subdifferential of the following functions.*

(i) $f(x) = |x|$, $x \in \mathbb{R}^N$.

(ii) $f(x) = \max\{|x_1|, \ldots, |x_N|\}$, $x \in \mathbb{R}^N$.

Exercise 140 *Show that the convex function* $f : \mathbb{R}^N \to (-\infty, \infty]$ *defined by*

$$f(x) := \begin{cases} -\left(1 - |x|^2\right)^{\frac{1}{2}} & \textit{if } |x| \leq 1, \\ \infty & \textit{otherwise,} \end{cases}$$

is differentiable, and so subdifferentiable (see Theorem 147 below) in the open unit ball $\left\{x \in \mathbb{R}^N : |x| < 1\right\}$ *but it is not subdifferentiable at points* x *with* $|x| = 1$*.*

We now study the existence of the subdifferential. This relies on the following theorem. We now study the existence of the subdifferential. This relies on the following theorem.

Theorem 141 (Hahn-Banach) *Let V be a finite dimensional vector space and let $g : V \to (-\infty, \infty]$ be a proper convex function finite in a neighborhood of 0. Let $V_1 \subset V$ be a subspace of V and let $L : V_1 \to \mathbb{R}$ be a linear function such that*

$$g(v) \geq L(v)$$

for all $v \in V_1$. Then there exists a linear extension $L_1 : V \to \mathbb{R}$ of L such that

$$g(v) \geq L_1(v) \quad \text{for all } v \in V.$$

Proof. If $V_1 = V$, then there is nothing to prove. Thus let $w_0 \in V \setminus V_1$, with $w_0 \neq 0$ and consider the subspace W the linear span of $V_1 \cup \{w\}$. If $w \in W$, then $w = v + tw_0$ where $v \in V_1$ and $t \in \mathbb{R}$ (this decomposition is unique). Define

$$\hat{L}(v + tw_0) := L(v) + tc,$$

where $c \in \mathbb{R}$ has to be chosen appropriately. We would like

$$g(v + tw) \geq L(v) + tc$$

for all $v \in V_1$ and $t \in \mathbb{R}$. If $t > 0$, then the previous inequality is equivalent to

$$\frac{g(v + tw) - L(v)}{t} \geq c,$$

that is,

$$\inf_{v \in V_1,\, t > 0} \frac{g(v + tw) - L(v)}{t} \geq c.$$

On the other hand, if $t < 0$, then writing $s := -t > 0$, we have that

$$g(v - sw) \geq L(v) - sc$$

for all $v \in V_1$ and $s > 0$, that is

$$\sup_{v \in V_1,\, s > 0} \frac{L(v) - g(v - sw)}{s} \leq c.$$

Thus, to prove the existence of c it is necessary that

$$\inf_{v \in V_1,\, t > 0} \frac{g(v + tw) - L(v)}{t} \geq \sup_{v \in V_1,\, s > 0} \frac{L(v) - g(v - sw)}{s}, \tag{36}$$

that is,

$$\frac{g(v_1 + tw) - L(v_1)}{t} \geq \frac{L(v_2) - g(v_2 - sw)}{s}$$

for all $v_1,\, v_2 \in V_1$ and for all $s,\, t > 0$. This inequality may be rewritten as

$$sg\left(v_1 + tw\right) + tg\left(v_2 - sw\right) \geq L\left(tv_2 + sv_1\right).$$

Since g is convex and L linear,

$$\begin{aligned}
&sg\left(v_1 + tw\right) + tg\left(v_2 - sw\right) \\
&\quad = (s+t)\left[\frac{s}{s+t} g\left(v_1 + tw\right) + \frac{t}{s+t} g\left(v_2 - sw\right)\right] \\
&\quad \geq (s+t)\, g\left(\frac{s}{s+t}\left(v_1 + tw\right) + \frac{t}{s+t}\left(v_2 - sw\right)\right) \\
&\quad = (s+t)\, g\left(\frac{s}{s+t} v_1 + \frac{t}{s+t} v_2\right) \geq (s+t)\, L\left(\frac{s}{s+t} v_1 + \frac{t}{s+t} v_2\right) \\
&\quad = L\left(tv_2 + sv_1\right).
\end{aligned}$$

Hence (36) holds. Since g is finite in a neighborhood of 0, taking $t_1,\, s_1 > 0$ sufficiently small, we have that $g\left(t_1 w\right),\, g\left(-s_1 w\right) \in \mathbb{R}$. Hence by (36),

$$\begin{aligned}
\infty > \frac{g\left(0 + t_1 w\right) - 0}{t_1} &\geq \inf_{v \in V_1,\, t > 0} \frac{g\left(v + tw\right) - L\left(v\right)}{t} \\
&\geq \sup_{v \in V_1,\, s > 0} \frac{L\left(v\right) - g\left(v - sw\right)}{s} \geq \frac{0 - g\left(0 - s_1 w\right)}{s_1} > -\infty,
\end{aligned}$$

which shows that the numbers in (36) are real and thus we can choose any real number c between them.

Hence, we have extended L to W. If the dimension of W is not the dimension of V, then we repeat the process until we extend L to all of V. ■

Theorem 142 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be a proper convex function. If $x_0 \in$ $\mathrm{ri}_{\mathrm{aff}}\left(\mathrm{dom}_e\, f\right)$, then $\partial f\left(x_0\right) \neq \emptyset$. In particular, if f is real-valued, then $\partial f\left(x\right) \neq \emptyset$ for every $x \in \mathbb{R}^N$.*

Proof. Fix a point $x_0 \in \mathrm{ri}_{\mathrm{aff}}\left(\mathrm{dom}_e\, f\right)$. If $\mathrm{dom}_e\, f$ consists only of x_0, then

$$f\left(x_0\right) = \min_{x \in \mathbb{R}^N} f\left(x\right),$$

and so $0 \in \partial f\left(x_0\right)$ by (35).

If $\mathrm{dom}_e\, f$ consists of at least two points, then by Proposition 59 we may write $\mathrm{aff}\left(\mathrm{dom}_e\, f\right) = x_1 + V$, where V is a subspace of $\mathbb{R}^N$. Since $x_0 \in \mathrm{ri}_{\mathrm{aff}}\left(\mathrm{dom}_e\, f\right)$ we have that $x_0 = x_1 + v_0$ for some vector $v_0 \in V$. Let ν be any vector in $V \setminus \{0\}$. Since $x_0 \in \mathrm{ri}_{\mathrm{aff}}\left(\mathrm{dom}_e\, f\right)$, the line

$$\{x_0 + t\nu :\, t \in \mathbb{R}\}$$

intersects the convex set $\mathrm{dom}_e\, f$ into an interval. Let $I := \{t \in \mathbb{R} :\, x_0 + t\nu \in \mathrm{dom}_e\, f\}$. Then $0 \in I^{\cdot}$. The function $h : I \to \mathbb{R}$ defined by

$$h\left(t\right) := f\left(x_0 + t\nu\right), \quad t \in I, \tag{37}$$

is convex, and so by Corollary 18 it is subdifferentiable at $t = 0$. Thus for any $c \in \partial h(0)$,

$$h(t) \geq h(0) + ct \quad \text{for all } t \in I,$$

or equivalently,

$$f(x_0 + t\nu) - f(x_0) \geq ct \quad \text{for all } t \in I.$$

Since $f(x_0 + t\nu) = \infty$ for all $t \notin I$, the previous inequality holds for all $t \in \mathbb{R}$. Thus we may apply the Hahn–Banach theorem to the convex function $g(v) := f(x_0 + v) - f(x_0)$, $v \in V$ and with $L(t\nu) := ct$, $t \in \mathbb{R}$, to find a linear extension $L_1 : V \to \mathbb{R}$ of L such that

$$g(v) \geq L_1(v) \quad \text{for all } v \in V,$$

that is

$$f(x_0 + v) - f(x_0) \geq L_1(v) \quad \text{for all } v \in V.$$

Since L_1 is linear, there exists $w_0 \in V$ such that $L_1(v) = w_0 \cdot v$ for all $v \in V$, with $L_1(\nu) = L(\nu) = c$. Hence

$$f(x_0 + v) - f(x_0) \geq w_0 \cdot v$$

for all $v \in V$ and

$$w_0 \cdot \nu = c. \tag{38}$$

Since $x_0 + V = x_1 + v_0 + V = x_1 + V = \operatorname{aff}(\operatorname{dom}_e f)$, we have proved that

$$f(x) \geq f(x_0) + w_0 \cdot (x - x_0)$$

for all $x \in \operatorname{aff}(\operatorname{dom}_e f)$. Since $f = \infty$ outside $\operatorname{aff}(\operatorname{dom}_e f)$, we have proved that f is subdifferentiable at x_0.

In particular, taking $c = h'_+(0)$, from (38) we get that

$$w_0 \cdot \nu = \frac{\partial^+ f}{\partial \nu}(x_0). \tag{39}$$

■

Remark 143 *Actually one can prove that if f is not subdifferentiable at x_0, then*

$$\frac{\partial^+ f}{\partial v}(x_0) = -\frac{\partial^+ f}{\partial(-v)}(x_0) = -\infty$$

for all $v \in \mathbb{R}^N$ such that $x_0 + v \in \mathrm{ri}_{\mathrm{aff}}(\operatorname{dom}_e f)$.

Corollary 144 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be a proper convex function. If f is subdifferentiable at $x_0 \in \operatorname{dom}_e f$, then for every $v \in \mathbb{R}^N$,*

$$\frac{\partial^+ f}{\partial v}(x_0) \geq \sup_{y \in \partial f(x_0)} y \cdot v \geq \inf_{y \in \partial f(x_0)} y \cdot v \geq -\frac{\partial^+ f}{\partial (-v)}(x_0).$$

Moreover, if $x_0 \in \operatorname{ri}_{\mathrm{aff}}(\operatorname{dom}_e f)$, then

$$\frac{\partial^+ f}{\partial v}(x_0) = \max_{y \in \partial f(x_0)} y \cdot v \geq \min_{y \in \partial f(x_0)} y \cdot v = -\frac{\partial^+ f}{\partial (-v)}(x_0)$$

for all $v \in \mathbb{R}^N$ such that $x_0 + v \in \operatorname{aff}(\operatorname{dom}_e f)$.

Proof. Fix any $y_0 \in \partial f(x_0)$. Then

$$f(x) \geq f(x_0) + y_0 \cdot (x - x_0) \quad \text{for all } x \in \mathbb{R}^N.$$

Hence for any $v \in \mathbb{R}^N$ and $t > 0$ we have that

$$\frac{f(x_0 + tv) - f(x_0)}{t} \geq y_0 \cdot v.$$

Letting $t \to 0^+$ yields

$$\frac{\partial^+ f}{\partial v}(x_0) \geq y_0 \cdot v.$$

Replacing v with $-v$ gives

$$\frac{\partial^+ f}{\partial (-v)}(x_0) \geq -y_0 \cdot v,$$

and so

$$\frac{\partial^+ f}{\partial v}(x_0) \geq \sup_{y \in \partial f(x_0)} y \cdot v \geq \inf_{y \in \partial f(x_0)} y \cdot v \geq -\frac{\partial^+ f}{\partial (-v)}(x_0).$$

Moreover, from the proof of the previous theorem, see (38), if $x_0 \in \operatorname{ri}_{\mathrm{aff}}(\operatorname{dom}_e f)$ and $x_0 + v \in \operatorname{aff}(\operatorname{dom}_e f)$, then there exists $y_0 \in \partial f(x_0)$ such that $\frac{\partial^+ f}{\partial v}(x_0) = y_0 \cdot v$. Similarly, there exists $y_1 \in \partial f(x_0)$ such that $y_1 \cdot v = -\frac{\partial^+ f}{\partial(-v)}(x_0)$. This proves the last part of the statement. ■

Exercise 145 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be a proper convex function and let $x_0 \in \mathbb{R}^N$. Prove that $\partial f(x_0)$ is a nonempty bounded set if and only if $x_0 \in (\operatorname{dom}_e f)^\circ$.*

Exercise 146 *The set of points at which a convex function is subdifferentiable may be larger than $\operatorname{ri}_{\mathrm{aff}}(\operatorname{dom}_e f)$. Indeed, let $N = 2$ and consider the function*

$$f(x) = f(x_1, x_2) = \begin{cases} \max\left\{1 - x_1^{\frac{1}{2}}, |x_2|\right\} & \text{if } x_1 \geq 0, \\ \infty & \text{otherwise.} \end{cases}$$

Prove that f is subdifferentiable everywhere in the half-plane $\{x_1 \geq 0\}$ except in the relative interior of the segment joining $(0, 1)$ and $(0, -1)$. Note that the set on which f is subdifferentiable is not convex.

Now we study the relation between differentiability and subdifferentiability.

Theorem 147 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be convex and let $x_0 \in \mathbb{R}^N$ be such that $f(x_0) \in \mathbb{R}$. If f is differentiable at x_0, then it is subdifferentiable at x_0 and $\partial f(x_0) = \{\nabla f(x_0)\}$. Conversely, if f is subdifferentiable at x_0 and the subdifferential of f at x_0 is a singleton, then f is differentiable at x_0.*

Proof. Assume that f is differentiable at x_0. Then $x_0 \in (\operatorname{dom}_e f)^\circ$ and so by Theorem 142, $\partial f(x_0)$ is nonempty. Fix any $y_0 \in \partial f(x_0)$. Then by Corollary 144

$$\frac{\partial^+ f}{\partial v}(x_0) \geq y_0 \cdot v \geq -\frac{\partial^+ f}{\partial(-v)}(x_0).$$

But since f is differentiable at x_0 we have that

$$\frac{\partial^+ f}{\partial v}(x_0) = -\frac{\partial^+ f}{\partial(-v)}(x_0) = \nabla f(x_0) \cdot v,$$

and so

$$y_0 \cdot v = \nabla f(x_0) \cdot v$$

for all $v \in \mathbb{R}^N$, which implies that $y_0 = \nabla f(x_0)$.

Conversely, assume that f is subdifferentiable at x_0 and the subdifferential of f at x_0 is a singleton, $\partial f(x_0) = \{y_0\}$. By Exercise 145, $x_0 \in (\operatorname{dom}_e f)^\circ$, and so by Corollary 144, for any fixed $v \in \mathbb{R}^N$

$$\frac{\partial^+ f}{\partial v}(x_0) = \max_{y \in \partial f(x_0)} y \cdot v = y_0 \cdot v = \min_{y \in \partial f(x_0)} y \cdot v = -\frac{\partial^+ f}{\partial(-v)}(x_0).$$

Hence

$$\frac{\partial^+ f}{\partial v}(x_0) = -\frac{\partial^+ f}{\partial(-v)}(x_0) \in \mathbb{R}$$

for all $v \in \mathbb{R}^N$, which shows that f is differentiable at x_0 by Theorem 117. ■

Next we study some continuity properties of the subdifferential.

Theorem 148 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be a proper convex function. Then for any compact set $K \subset (\operatorname{dom}_e f)^\circ$, the set*

$$\partial f(K) = \bigcup_{x \in K} \partial f(x)$$

is compact.

Proof. We prove first that $\partial f(K)$ is closed. Let $\{y_n\} \subset \partial f(K)$ be such that $y_n \to y_0$ as $n \to \infty$. Find $\{x_n\} \subset K$ such that and $y_n \in \partial f(x_n)$. By the compactness of K, there exists a subsequence $\{x_{n_k}\}$ converging to some $x_0 \in K$. By the subdifferentiability of f at x_{n_k} we get that

$$f(x) \geq f(x_{n_k}) + (x - x_{n_k}) \cdot y_{n_k}$$

for all $k \in \mathbb{N}$ and for all $x \in \mathbb{R}^N$. Since f is continuous at $x_0 \in (\operatorname{dom}_e f)^\circ$, letting $k \to \infty$ in the previous inequality yields

$$f(x) \geq f(x_0) + (x - x_0) \cdot y_0$$

for all $x \in \mathbb{R}^N$, which shows that $y_0 \in \partial f(x_0) \subset \partial f(K)$. Thus $\partial f(K)$ is closed.

To prove that $\partial f(K)$ is bounded, for every $x_0 \in K$ and any $y_0 \in \partial f(x_0) \backslash \{0\}$ we have

$$f(x) - f(x_0) \geq (x - x_0) \cdot y_0$$

for all $x \in \mathbb{R}^N$. Taking $x = x_0 + r\frac{y_0}{|y_0|}$, where $r > 0$ is taken so small that $\overline{B(x_0, r)} \subset (\operatorname{dom}_e f)^\circ$, gives

$$f\left(x_0 + r\frac{y_0}{|y_0|}\right) - f(x_0) \geq r\frac{y_0}{|y_0|} \cdot y_0 = r|y_0|.$$

On the other hand, since f is Lipschitz in $K \cup \overline{B(x_0, r)}$, there exists $L_K > 0$ such that

$$|f(x_1) - f(x_2)| \leq L_K |x_1 - x_2|$$

for all x_1, $x_2 \in K \cup \overline{B(x_0, r)}$, and so from the previous inequality we get

$$rL_K = L_K \left|x_0 + r\frac{y_0}{|y_0|} - x_0\right| \geq f\left(x_0 + r\frac{y_0}{|y_0|}\right) - f(x_0) \geq r|y_0|.$$

This shows that $|y| \leq L_K$ for all $y \in \partial f(K)$. ■

As a corollary of the previous theorem we obtain that ∂f is upper semicontinuous.

Definition 149 *A multifunction* $\Gamma : \mathbb{R}^N \to \mathcal{P}\left(\mathbb{R}^N\right)$ *is called* upper semicontinuous at $x_0 \in \mathbb{R}^N$ *if for any open set* $V \subset \mathbb{R}^N$ *containing* $\Gamma(x_0)$*, there exists a neighborhood* U *of* x_0 *such that*

$$\Gamma(U) := \bigcup_{x \in U} \Gamma(x) \subset V.$$

Theorem 150 *Let* $f : \mathbb{R}^N \to (-\infty, \infty]$ *be a proper convex function. Then* ∂f *is upper semicontinuous in* $(\operatorname{dom}_e f)^\circ$.

Proof. Fix $x_0 \in (\operatorname{dom}_e f)^\circ$ and any open set $V \subset \mathbb{R}^N$ containing $\partial f(x_0)$. We claim that there exists $r > 0$ such that $B(x_0, r) \subset \operatorname{dom}_e f$ and

$$\partial f(B(x_0, r)) = \bigcup_{x \in B(x_0, r)} \partial f(x) \subset V.$$

Indeed, if not, then there exist $\{x_n\} \subset (\operatorname{dom}_e f)^\circ$ and $y_n \in \partial f(x_n) \setminus V$ such that $x_n \to x_0$ as $n \to \infty$. Since the set $\partial f\left(\overline{B(x_0, r)}\right)$ is compact for $r > 0$ small by the previous theorem, and $y_n \in \partial f\left(\overline{B(x_0, r)}\right)$ for all n large, there exists a subsequence $\{y_{n_k}\}$ converging to some $y_0 \in \partial f\left(\overline{B(x_0, r)}\right)$. By the subdifferentiability of f at x_{n_k} we get that

$$f(x) \geq f(x_{n_k}) + (x - x_{n_k}) \cdot y_{n_k}$$

for all $k \in \mathbb{N}$ and for all $x \in \mathbb{R}^N$. Since f is continuous at $x_0 \in (\operatorname{dom}_e f)^\circ$, letting $k \to \infty$ in the previous inequality yields

$$f(x) \geq f(x_0) + (x - x_0) \cdot y_0$$

for all $x \in \mathbb{R}^N$, which shows that $y_0 \in \partial f(x_0) \subset V$, which is a contradiction, since $y_{n_k} \in \partial f(x_{n_k}) \setminus V$. ■

As a corollary of the previous result we obtain uniform convergence of gradients of differentiable convex functions.

Theorem 151 *Let* $A \subset \mathbb{R}^N$ *be an open convex set and let* $f_n : A \to \mathbb{R}$*,* $n \in \mathbb{N}$*, be a sequence of convex functions converging pointwise in* A *to a convex function* $f : A \to \mathbb{R}$*. Then for any* $x_0 \in A$*, for any sequence* $\{x_n\} \subset A$ *converging to* x_0*, and for any open set* $V \subset \mathbb{R}^N$ *containing* $\partial f(x_0)$*,*

$$\partial f_n(x_n) \subset V \tag{40}$$

for all n *sufficiently large.*

Proof. By Theorem 112, $\{f_n\}$ converges uniformly on compact sets of A. Let $r > 0$ be so small that $\overline{B(x_0, 2r)} \subset A$. Then for all n sufficiently large, $x_n \in \overline{B(x_0, 2r)} \subset A$. For every such n and for every $y \in \partial f_n(x_n) \setminus \{0\}$ we have

$$f_n(x) - f_n(x_n) \geq (x - x_n) \cdot y$$

for all $x \in \mathbb{R}^N$. Taking $x = x_n + r\frac{y}{|y|}$ gives

$$f_n\left(x_n + r\frac{y}{|y|}\right) - f_n(x_n) \geq r\frac{y}{|y|} \cdot y = r|y|.$$

By uniform convergence in $\overline{B(x_0, 2r)}$ we have that

$$f\left(x_n + r\frac{y}{|y|}\right) - f(x_n) + 1 \geq r\frac{y}{|y|} \cdot y = r|y|$$

for all n sufficiently large. On the other hand, since f is Lipschitz in $\overline{B(x_0, 2r)}$, there exists $L > 0$ such that

$$|f(x_1) - f(x_2)| \leq L|x_1 - x_2|$$

for all x_1, $x_2 \in K$, and so from the previous inequality we get

$$\begin{aligned} rL + 1 &= L\left|x_n + r\frac{y}{|y|} - x_n\right| + 1 \\ &\geq f\left(x_n + r\frac{y}{|y|}\right) - f(x_n) + 1 \geq r|y|. \end{aligned}$$

This shows that $|y| \leq L + \frac{1}{r}$ for all n sufficiently large and all $y \in \partial f_n(x_n)$.

Now suppose by contradiction that (40) does not hold. Then there exist infinitely many n and $y_n \in \partial f_n(x_n)$ such that $y_n \notin V$. By the previous part of the proof, there exists a subsequence $\{y_{n_k}\}$ converging to some $y_0 \in \mathbb{R}^N$. Since

$$f_{n_k}(x) \geq f_{n_k}(x_{n_k}) + (x - x_{n_k}) \cdot y_{n_k}$$

for all $x \in \mathbb{R}^N$. Using once more uniform convergence and letting $k \to \infty$, we get

$$f(x) \geq f(x) + (x - x_0) \cdot y_0$$

for all $x \in \mathbb{R}^N$, which shows that $y_0 \in \partial f(x_0) \subset V$. This is a contradiction. ■

Corollary 152 *Let $A \subset \mathbb{R}^N$ be an open convex set and let $f_n : A \to \mathbb{R}$, $n \in \mathbb{N}$, be a sequence of differentiable convex functions converging pointwise in A to a differentiable convex function $f : A \to \mathbb{R}$. Then $\{\nabla f_n\}$ converges uniformly to ∇f on compact sets.*

Proof. Let $K \subset A$ be a compact set and assume by contradiction that there exist $\varepsilon > 0$ and a sequence $\{x_{n_k}\} \subset K$ such that

$$|\nabla f_{n_k}(x_{n_k}) - \nabla f(x_{n_k})| > \varepsilon$$

for all $k \in \mathbb{N}$. Extracting a subsequence (not relabelled), we may assume that $x_{n_k} \to x_0 \in K$. Since ∇f is continuous, we that

$$\left|\nabla f_{n_k}\left(x_{n_k}\right) - \nabla f\left(x_0\right)\right| > \frac{\varepsilon}{2}$$

for all k sufficiently large.

By the previous theorem, applied to the open set $V := B\left(\nabla f\left(x_0\right), \frac{\varepsilon}{2}\right)$ (that contains $\partial f\left(x_0\right) = \left\{\nabla f\left(x_0\right)\right\}$) ,

$$\partial f_{n_k}\left(x_{n_k}\right) \subset B\left(\nabla f\left(x_0\right), \frac{\varepsilon}{2}\right)$$

for all k sufficiently large. This contradiction completes the proof. ■

Next we study the subdifferentiability of the sum of two convex functions.

Proposition 153 *Let f_1, $f_2 : \mathbb{R}^N \to (-\infty, \infty]$ be two proper convex functions. Then*

$$\partial\left(f_1 + f_2\right) \supset \partial f_1 + \partial f_2.$$

Moreover, if

$$\operatorname{ri}_{\operatorname{aff}}\left(\operatorname{dom}_e f_1\right) \cap \operatorname{ri}_{\operatorname{aff}}\left(\operatorname{dom}_e f_2\right) \neq \emptyset, \tag{41}$$

then

$$\partial\left(f_1 + f_2\right) = \partial f_1 + \partial f_2.$$

Proof. Step 1: Let $x_0 \in \mathbb{R}^N$. If $y_1 \in \partial f_1\left(x_0\right)$ and $y_2 \in \partial f_2\left(x_0\right)$, then

$$\begin{aligned} f_1(x) &\geq f_1\left(x_0\right) + y_1 \cdot \left(x - x_0\right) \quad \text{for all } x \in \mathbb{R}^N, \\ f_2(x) &\geq f_2\left(x_0\right) + y_2 \cdot \left(x - x_0\right) \quad \text{for all } x \in \mathbb{R}^N, \end{aligned}$$

and so, adding the two inequalities, we conclude that $y_1 + y_2 \in \partial\left(f_1 + f_2\right)\left(x_0\right)$. Hence if $\partial f_1\left(x_0\right)$ and $\partial f_2\left(x_0\right)$ are nonempty, then $\partial\left(f_1 + f_2\right)\left(x_0\right)$ is nonempty and

$$\partial\left(f_1 + f_2\right)\left(x_0\right) \supset \partial f_1\left(x_0\right) + \partial f_2\left(x_0\right).$$

Step 2: Conversely, assume that (41) holds. If $\partial\left(f_1 + f_2\right)\left(x_0\right)$ is nonempty, let $y \in \partial\left(f_1 + f_2\right)\left(x_0\right)$. We claim that $y \in \partial f_1\left(x_0\right) + \partial f_2\left(x_0\right)$. Replacing f_1 and f_2 with the convex functions

$$\begin{aligned} g_1(x) &:= f_1\left(x + x_0\right) - f_1\left(x_0\right) - y \cdot x_0, \quad x \in \mathbb{R}^N, \\ g_2(x) &:= f_2\left(x + x_0\right) - f_2\left(x_0\right), \quad x \in \mathbb{R}^N, \end{aligned}$$

we can assume, without loss of generality, that

$$x_0 = y = 0, \quad f_1(0) = f_2(0) = 0.$$

Since $0 \in \partial\left(f_1 + f_2\right)(0)$, it follows by (35) that

$$0 = \left(f_1 + f_2\right)(0) = \min_{x \in \mathbb{R}^N}\left(f_1 + f_2\right)(x).$$

Consider the convex sets

$$C_1 := \left\{(x,t) \in \mathbb{R}^N \times \mathbb{R} : \, t \geq f_1(x)\right\},$$
$$C_2 := \left\{(x,t) \in \mathbb{R}^N \times \mathbb{R} : \, t \leq -f_2(x)\right\}.$$

Then (exercise)

$$\mathrm{ri}_{\mathrm{aff}}\, C_1 := \left\{(x,t) \in \mathbb{R}^N \times \mathbb{R} : \, x \in \mathrm{ri}_{\mathrm{aff}}(\mathrm{dom}_e\, f_1),\, t > f_1(x)\right\},$$
$$\mathrm{ri}_{\mathrm{aff}}\, C_2 := \left\{(x,t) \in \mathbb{R}^N \times \mathbb{R} : \, x \in \mathrm{ri}_{\mathrm{aff}}(\mathrm{dom}_e\, f_2),\, t < -f_2(x)\right\}.$$

Note that if $(x,t) \in \mathrm{ri}_{\mathrm{aff}}\, C_1 \cap \mathrm{ri}_{\mathrm{aff}}\, C_2$, then $-f_2(x) > t > f_1(x)$, which implies that $(f_1 + f_2)(x) < 0$. This contradicts the fact that the minimum of $f_1 + f_2$ is zero. Hence $\mathrm{ri}_{\mathrm{aff}}\, C_1 \cap \mathrm{ri}_{\mathrm{aff}}\, C_2 = \emptyset$. By Theorem 88Then there exist a vector $(b,c) \in \mathbb{R}^N \times \mathbb{R} \setminus \{(0,0)\}$ and $\alpha \in \mathbb{R}$ such that

$$b \cdot x + ct \leq \alpha \quad \text{for all } (x,t) \in C_1 \text{ and } b \cdot x + ct \geq \alpha \quad \text{for all } (x,t) \in C_2,$$

and $C_1 \cup C_2$ is not contained in the hyperplane $\left\{(x,t) \in \mathbb{R}^N \times \mathbb{R} : \, b \cdot x + ct = \alpha\right\}$. If $c = 0$, then the hyperplane $\left\{x \in \mathbb{R}^N : \, b \cdot x = \alpha\right\}$ would separate C_1 and C_2, which is impossible since $\mathrm{ri}_{\mathrm{aff}}(\mathrm{dom}_e\, f_1) \cap \mathrm{ri}_{\mathrm{aff}}(\mathrm{dom}_e\, f_2) \neq \emptyset$. Thus $c \neq 0$, and so letting $t \to \infty$ in the first inequality, it follows that $c < 0$, while taking $x = 0$ and $t \to 0$, we get that $a = 0$. Thus

$$-\frac{b}{c} \cdot x \leq t \quad \text{for all } (x,t) \in C_1,$$
$$\frac{b}{c} \cdot x \leq -t \quad \text{for all } (x,t) \in C_2,$$

or equivalently,

$$-\frac{b}{c} \cdot x + f_1(0) = \frac{b}{-c} \cdot x \leq f_1(x) \quad \text{for all } x \in \mathbb{R}^N,$$
$$\frac{b}{c} \cdot x + f_1(0) = \frac{b}{c} \cdot x \leq f_2(x) \quad \text{for all } x \in \mathbb{R}^N,$$

which implies that $-\frac{b}{c} \in \partial f_1(0)$, while $\frac{b}{c} \in \partial f_2(0)$, and so $0 = -\frac{b}{c} + \frac{b}{c} \in \partial f_1(0) + \partial f_2(0)$.

Step 3: To conclude the proof we observe that if either $\partial f_1(x_0)$ or $\partial f_2(x_0)$ is empty, then by Step 2 so must be $\partial(f_1 + f_2)(x_0)$. If both $\partial f_1(x_0)$ and $\partial f_2(x_0)$ are nonempty, then by Steps 1 and 2, $\partial(f_1 + f_2)(x_0)$ is also nonempty and

$$\partial(f_1 + f_2)(x_0) = \partial f_1(x_0) + \partial f_2(x_0).$$

This completes the proof. ■

We recall that a multifunction $\Gamma : \mathbb{R}^N \to \mathcal{P}\left(\mathbb{R}^N\right)$ is called *monotone* if

$$(x_2 - x_1) \cdot (y_2 - y_1) \geq 0$$

for all (x_1, y_1), $(x_2, y_2) \in \operatorname{graph}\Gamma$. A monotone $\Gamma : \mathbb{R}^N \to \mathcal{P}\left(\mathbb{R}^N\right)$ is called *maximal* if its graph is not a proper subset of the graph of a monotone multifunction.

Theorem 154 *Let $\Gamma : \mathbb{R}^N \to \mathcal{P}\left(\mathbb{R}^N\right)$ be monotone. Then*

(i) for every $\varepsilon > 0$, $\Gamma + \varepsilon I$, and $(\Gamma + \varepsilon I)^{-1}$ are monotone multifunctions;

(ii) for every $\varepsilon > 0$, $(\Gamma + \varepsilon I)^{-1}$ is univalued and Lipschitz with Lipschitz constant at most $\frac{1}{\varepsilon}$;

(iii) if the domain of $(\Gamma + \varepsilon I)^{-1}$ is $\mathbb{R}^N$ for some $\varepsilon > 0$, then Γ is maximal monotone.

Proof. It is enough to consider the case $\varepsilon = 1$. Let (x_1, z_1), $(x_2, z_2) \in \operatorname{graph}(\Gamma + \varepsilon I)$. Then $z_i = y_i + \varepsilon x_i$, where $y_i \in \Gamma(x_i)$, $i = 1, 2$. By the monotonicity of Γ,

$$(y_2 - y_1) \cdot (x_2 - x_1) \geq 0,$$

and so

$$\begin{aligned}(z_2 - z_1) \cdot (x_2 - x_1) &= \left[(y_2 - y_1) + \varepsilon (x_2 - x_1)\right] \cdot (x_2 - x_1) \\ &= (y_2 - y_1) \cdot (x_2 - x_1) + \varepsilon |x_2 - x_1|^2 \geq \varepsilon |x_2 - x_1|^2 \geq 0,\end{aligned} \tag{42}$$

which shows that $\Gamma + \varepsilon I$, and, in turn, $(\Gamma + \varepsilon I)^{-1}$ are monotone multifunctions. Note that the previous inequality implies, in particular, that

$$\begin{aligned}\varepsilon |x_2 - x_1|^2 &\leq \left[(y_2 - y_1) + \varepsilon (x_2 - x_1)\right] \cdot (x_2 - x_1) \\ &\leq |(y_2 + \varepsilon x_2) - (y_1 + \varepsilon x_1)| \, |x_2 - x_1|,\end{aligned}$$

and so

$$\varepsilon |x_2 - x_1| \leq |(y_2 + \varepsilon x_2) - (y_1 + \varepsilon x_1)| = |z_2 - z_1|. \tag{43}$$

To prove (ii), recall that by (16), for all $z \in \mathbb{R}^N$,

$$(\Gamma + \varepsilon I)^{-1}(z) := \left\{x \in \mathbb{R}^N : z \in (\Gamma + \varepsilon I)(x)\right\}.$$

To prove that $(\Gamma + \varepsilon I)^{-1}$ is univalued, fix $z \in \mathbb{R}^N$ and assume that $(\Gamma + \varepsilon I)^{-1}(z)$ is nonempty. If x_1, $x_2 \in (\Gamma + \varepsilon I)^{-1}(z)$, then $z \in (\Gamma + \varepsilon I)(x_i)$, $i = 1, 2$, and so we may write

$$z = y_1 + \varepsilon x_1 = y_1 + \varepsilon x_1,$$

where $y_i \in \Gamma(x_i)$, $i = 1, 2$, and from the inequality (43) we get that $x_1 = x_2$. Thus, $(\Gamma + \varepsilon I)^{-1}$ is univalued. Hence, from now on the set $\operatorname{dom}(\Gamma + \varepsilon I)^{-1}$ we may identify the multifunction $(\Gamma + \varepsilon I)^{-1}$ with the function

$$z \in \operatorname{dom}(\Gamma + \varepsilon I)^{-1} \mapsto x,$$

where $x \in \mathbb{R}^N$ is the unique element such that $z \in (\Gamma + \varepsilon I)(x)$. Inequality (43) implies that this function is Lipschitz continuous with Lipschitz constant at most $\frac{1}{\varepsilon}$.

Finally, to prove (iii), assume that the domain of $(\Gamma + \varepsilon I)^{-1}$ is $\mathbb{R}^N$. Let $(x_1, y_1) \in \mathbb{R}^N \times \mathbb{R}^N$ be such that

$$(y_2 - y_1) \cdot (x_2 - x_1) \geq 0$$

for all $(x_2, y_2) \in \operatorname{graph}\Gamma$. We claim that (x_1, y_1) belongs to $\operatorname{graph}\Gamma$. Since the domain of $(\Gamma + \varepsilon I)^{-1}$ is $\mathbb{R}^N$ there exists a unique $x \in \mathbb{R}^N$ such that $y_1 + \varepsilon x_1 \in (\Gamma + \varepsilon I)(x)$. Hence $y_1 + \varepsilon x_1 = w_1 + \varepsilon x$, where $w_1 \in \Gamma(x)$. Since $(x, w_1) \in \operatorname{graph}\Gamma$, we have that

$$0 \leq (w_1 - y_1) \cdot (x - x_1) = -(x - x_1) \cdot (x - x_1) = -|x - x_1|^2,$$

which implies that $x = x_1$ and, in turn, that $y_1 = w_1$. Thus $(x_1, y_1) = (x, w_1) \in \operatorname{graph}\Gamma$ and the proof is complete. ■

We now study some monotonicity properties of the subgradient.

Theorem 155 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be a proper convex function. Then*

(i) ∂f is a monotone multifunction;

(ii) if f is lower semicontinuous and not identically ∞, then the domain of ∂f is maximal monotone.

Proof. Let (x_1, y_1), $(x_2, y_2) \in \operatorname{graph}\partial f$. Then $y_i \in \partial f(x_i)$, $i = 1, 2$, and so

$$f(x) \geq f(x_1) + y_1 \cdot (x - x_1),$$
$$f(x) \geq f(x_2) + y_2 \cdot (x - x_2)$$

for all $x \in \mathbb{R}^N$. Taking $x = x_2$ in the first inequality, $x = x_1$ in the second and adding the resulting inequalities yields

$$(y_2 - y_1) \cdot (x_2 - x_1) \geq 0,$$

which gives (i). To prove (ii), assume that f is convex and lower semicontinuous, with $f \not\equiv \infty$. In view of the previous theorem, it is enough to show that the domain of $(\partial f + I)^{-1}$ is $\mathbb{R}^N$. Thus fix $y \in \mathbb{R}^N$ and consider the function

$$g(x) := f(x) + \frac{1}{2}|x|^2 - y \cdot x, \quad x \in \mathbb{R}^N.$$

Then g is convex and lower semicontinuous. We claim that g is bounded from below. Indeed, since $f \not\equiv \infty$, we have that $\mathrm{ri}_{\mathrm{aff}}(\mathrm{dom}_e f)$ is nonempty, and so by Corollary **??** if $x_0 \in \mathrm{ri}_{\mathrm{aff}}(\mathrm{dom}_e f)$, then $\partial f(x_0) \neq \emptyset$ and so

$$f(x) \geq f(x_0) + y_0 \cdot (x - x_0)$$

for all $x \in \mathbb{R}^N$ and for any $y_0 \in \partial f(x_0)$. Hence

$$g(x) \geq f(x_0) + y_0 \cdot (x - x_0) \frac{1}{2} |x|^2 - y \cdot x, \quad x \in \mathbb{R}^N,$$

which proves the claim. Thus g admits a minimum at some point $x \in \mathbb{R}^N$. By (35), $0 \in \partial g(x)$, that is

$$0 \in \partial g(x) = \partial f(x) + x - y,$$

where we have used Proposition 153. This shows that $y \in (\partial f + I)(x)$, and so the domain of $(\partial f + I)^{-1}$ is $\mathbb{R}^N$. ■

As a corollary we have some characterizations of convex functions.

Theorem 156 *Let $E \subset \mathbb{R}^N$ be a convex set and let $f : \mathbb{R}^N \to [-\infty, \infty]$ be differentiable in E. Then the following three conditions are equivalent:*

(i) $f : E \to \mathbb{R}$ *is convex;*

(ii) for all $x, y \in E$,
$$f(x) \geq f(y) + \nabla f(y) \cdot (x - y);$$

(iii) for all $x, y \in E$,
$$(\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq 0.$$

Proof. Assume that (i) holds. Since f is differentiable in E, by the previous theorem f is subdifferentiable at every $y \in E$ and $\partial f(y) = \{\nabla f(y)\}$. Hence (ii) holds.

Assume next that (ii) holds. Then (iii) follows from the previous theorem.

Finally, assume that (iii) holds and fix $x, y \in E$. Since f is differentiable in E, the function $g : [0, 1] \to \mathbb{R}$, defined by

$$g(t) := f(tx + (1 - t) y), \quad t \in [0, 1],$$

is differentiable and

$$g'(t) = (\nabla f(tx + (1 - t) y)) \cdot (x - y).$$

If $s > t$, then

$$\begin{aligned} g'(s) - g'(t) &= (\nabla f(sx + (1 - s) y) - \nabla f(tx + (1 - t) y)) \cdot (x - y) \\ &= \frac{1}{s - t} (\nabla f(sx + (1 - s) y) - \nabla f(tx + (1 - t) y)) \cdot \\ &\quad ((sx + (1 - s) y) - (tx + (1 - t) y)) \geq 0. \end{aligned}$$

Hence g' is nondecreasing, and so g is convex. In particular,

$$f(tx + (1-t)y) = g(t) \leq (1-t)g(0) + tg(1) = (1-t)f(y) + tf(x),$$

which implies the convexity of f. ■

Remark 157 *A similar result holds for strictly convex functions provided we require the inequalities (i) and (ii) to be strict when $x \neq y$.*

Monday, March 31, 2008

Next we prove that the subdifferential uniquely defines a lower semicontinuous function up to an additive constant. We begin with a regularization result that is of interest in itself.

Theorem 158 *Let $f:\mathbb{R}^N\to(-\infty,\infty]$ be a proper convex lower semicontinuous function. For every $\varepsilon>0$, let*

$$f_\varepsilon(x):=\inf_{y\in\mathbb{R}^N}\left\{f(y)+\frac{1}{2\varepsilon}|x-y|^2\right\},\quad x\in\mathbb{R}^N.$$

Then $f_\varepsilon:\mathbb{R}^N\to\mathbb{R}$ is convex, differentiable,

$$\partial f_\varepsilon=\left(\varepsilon I+(\partial f)^{-1}\right)^{-1},$$

and $f_\varepsilon\nearrow f$ as $\varepsilon\to 0^+$.

Proof. Step 1: We begin by showing that the infimum is attained. Fix $x\in\mathbb{R}^N$ and consider the function

$$g_{\varepsilon,x}(y):=f(y)+\frac{1}{2\varepsilon}|x-y|^2,\quad y\in\mathbb{R}^N.$$

Then g is convex and lower semicontinuous. We claim that g is bounded from below. Indeed, since $f\not\equiv\infty$, we have that $\operatorname{ri}_{\text{aff}}(\operatorname{dom}_e f)$ is nonempty, and so by Corollary **??** if $x_0\in\operatorname{ri}_{\text{aff}}(\operatorname{dom}_e f)$, then $\partial f(x_0)\neq\emptyset$ and so

$$f(y)\geq f(x_0)+z_0\cdot(y-x_0)$$

for all $y\in\mathbb{R}^N$ and for any $z_0\in\partial f(x_0)$. Hence

$$g_{\varepsilon,x}(y)\geq f(x_0)+z_0\cdot(y-x_0)+\frac{1}{2\varepsilon}|x-y|^2,\quad y\in\mathbb{R}^N, \tag{44}$$

which proves the claim. Thus $g_{\varepsilon,x}$ admits a minimum at some point $y_{\varepsilon,x}\in\mathbb{R}^N$. By (**??**), $0\in\partial g_\varepsilon(y_{\varepsilon,x})$, that is

$$0\in\partial g_{\varepsilon,x}(y_{\varepsilon,x})=\partial f(y_{\varepsilon,x})+\left\{\frac{1}{\varepsilon}(y_{\varepsilon,x}-x)\right\},$$

where we have used Proposition 153. In particular, $z_{\varepsilon,x}:=\frac{1}{\varepsilon}(x-y_{\varepsilon,x})\in\partial f(y_{\varepsilon,x})$, and

$$f_\varepsilon(x)=f(y_{\varepsilon,x})+\frac{\varepsilon}{2}|z_{\varepsilon,x}|^2. \tag{45}$$

Step 2: We claim that the mapping $x\in\mathbb{R}^N\mapsto z_{\varepsilon,x}$ is Lipschitz, with Lipschitz constant less than or equal $\frac{1}{\varepsilon}$. Since ∂f is maximal, $(\partial f)^{-1}$ is maximal. In turn

$\left(\varepsilon I + (\partial f)^{-1}\right)^{-1}$ is univalued and Lipschitz with Lipschitz constant at most $\frac{1}{\varepsilon}$. To conclude, it suffices to prove that $z_{\varepsilon,x} \in \left(\varepsilon I + (\partial f)^{-1}\right)^{-1}(x)$. Indeed,

$$\begin{aligned} z_{\varepsilon,x} \in \left(\varepsilon I + (\partial f)^{-1}\right)^{-1}(x) &\Leftrightarrow x \in \left(\varepsilon I + (\partial f)^{-1}\right)(z_{\varepsilon,x}) = \varepsilon z_{\varepsilon,x} + (\partial f)^{-1}(z_{\varepsilon,x}) \\ &\Leftrightarrow x - \varepsilon z_{\varepsilon,x} \in (\partial f)^{-1}(z_{\varepsilon,x}) \Leftrightarrow z_{\varepsilon,x} \in \partial f(x - \varepsilon z_{\varepsilon,x}) = \partial f(y_{\varepsilon,x}). \end{aligned}$$

Step 3: The convexity of f_ε follows as in your midterm. If x, $x_1 \in \mathbb{R}^N$, then $z_{\varepsilon,x_1} \in \partial f(y_{\varepsilon,x_1})$, and so

$$f(y_{\varepsilon,x}) - f(y_{\varepsilon,x_1}) \geq z_{\varepsilon,x_1} \cdot (y_{\varepsilon,x} - y_{\varepsilon,x_1}).$$

In turn

$$\begin{aligned} f_\varepsilon(x) - f_\varepsilon(x_1) &= f(y_{\varepsilon,x}) + \frac{\varepsilon}{2}|z_{\varepsilon,x}|^2 - f(y_{\varepsilon,x_1}) - \frac{\varepsilon}{2}|z_{\varepsilon,x_1}|^2 \\ &\geq \frac{\varepsilon}{2}|z_{\varepsilon,x}|^2 - \frac{\varepsilon}{2}|z_{\varepsilon,x_1}|^2 + z_{\varepsilon,x_1} \cdot (y_{\varepsilon,x} - y_{\varepsilon,x_1}) \\ &= \frac{\varepsilon}{2}\left[|z_{\varepsilon,x}|^2 - |z_{\varepsilon,x_1}|^2 + 2z_{\varepsilon,x_1} \cdot \left(\frac{y_{\varepsilon,x} \mp x \pm x_1 - y_{\varepsilon,x_1}}{\varepsilon}\right)\right] \\ &= \frac{\varepsilon}{2}\left[|z_{\varepsilon,x}|^2 - |z_{\varepsilon,x_1}|^2 + 2z_{\varepsilon,x_1} \cdot (z_{\varepsilon,x} - z_{\varepsilon,x_1})\right] + z_{\varepsilon,x_1} \cdot (x - x_1), \end{aligned}$$

and so

$$f_\varepsilon(x) - f_\varepsilon(x_1) - z_{\varepsilon,x_1} \cdot (x - x_1) \geq \frac{\varepsilon}{2}|z_{\varepsilon,x} - z_{\varepsilon,x_1}|^2 \geq 0.$$

By interchanging x an x_1 we get

$$f_\varepsilon(x_1) - f_\varepsilon(x) - z_{\varepsilon,x} \cdot (x_1 - x) \geq 0,$$

or equivalently

$$f_\varepsilon(x) - f_\varepsilon(x_1) - z_{\varepsilon,x_1} \cdot (x - x_1) - (z_{\varepsilon,x} - z_{\varepsilon,x_1}) \cdot (x - x_1) \leq 0.$$

Hence

$$0 \leq f_\varepsilon(x) - f_\varepsilon(x_1) - z_{\varepsilon,x_1} \cdot (x - x_1) \leq (z_{\varepsilon,x} - z_{\varepsilon,x_1}) \cdot (x - x_1),$$

and so

$$\frac{|f_\varepsilon(x) - f_\varepsilon(x_1) - z_{\varepsilon,x_1} \cdot (x - x_1)|}{|x - x_1|} \leq |z_{\varepsilon,x} - z_{\varepsilon,x_1}| \leq \frac{1}{\varepsilon}|x - x_1| \to 0$$

as $x \to x_1$. Thus f_ε is differentiable and $\nabla f_\varepsilon(x_1) = z_{\varepsilon,x_1} = \frac{1}{\varepsilon}(y_{\varepsilon,x_1} - x_1) \in \partial f(y_{\varepsilon,x_1})$.

Step 4: It remains to show that $f_\varepsilon \nearrow f$ as $\varepsilon \to 0^+$. If $0 < \varepsilon_1 < \varepsilon_2$, we have that $\frac{1}{2\varepsilon_1} > \frac{1}{2\varepsilon_2}$, and so for every $x \in \mathbb{R}^N$,

$$\begin{aligned} f(x) \geq f_{\varepsilon_1}(x) &= \inf_{y \in \mathbb{R}^N}\left\{f(y) + \frac{1}{2\varepsilon_1}|x - y|^2\right\} \\ &\geq \inf_{y \in \mathbb{R}^N}\left\{f(y) + \frac{1}{2\varepsilon_2}|x - y|^2\right\} = f_{\varepsilon_2}(x). \end{aligned}$$

Thus

$$f(x) \geq \lim_{\varepsilon \to 0^+} f_\varepsilon(x).$$

To prove the opposite inequality, it suffices to assume that

$$\ell := \lim_{\varepsilon \to 0^+} f_\varepsilon(x) < \infty.$$

Then by (45) and (44),

$$\begin{aligned}\ell \geq f_\varepsilon(x) &\geq f(y_{\varepsilon,x}) + \frac{\varepsilon}{2}|z_{\varepsilon,x}|^2 = f(y_{\varepsilon,x}) + \frac{1}{2\varepsilon}|x - y_{\varepsilon,x}|^2 \\ &\geq f(x_0) + z_0 \cdot (y_{\varepsilon,x} - x_0) + \frac{1}{2\varepsilon}|x - y_{\varepsilon,x}|^2,\end{aligned}$$

which implies that $y_{\varepsilon,x} \to x$ as $\varepsilon \to 0^+$. In turn, by the lower semicontinuity of f, and the fact that $f_\varepsilon(x) \geq f(y_{\varepsilon,x})$, we have that

$$\lim_{\varepsilon \to 0^+} f_\varepsilon(x) \geq \liminf_{\varepsilon \to 0^+} f(y_{\varepsilon,x}) \geq f(x).$$

■

Corollary 159 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ and $g : \mathbb{R}^N \to (-\infty, \infty]$ be two proper convex lower semicontinuous functions such that*

$$\partial f = \partial g.$$

Then $g = f + const$.

Proof. By the previous theorem $\nabla f_\varepsilon = \nabla g_\varepsilon$. Hence there exists $c_\varepsilon \in \mathbb{R}$ such that $g_\varepsilon(x) = f_\varepsilon(x) + c_\varepsilon$ for all $x \in \mathbb{R}^N$. Fix $x_0 \in \operatorname{dom} \partial f = \operatorname{dom} \partial g$. Then $c_\varepsilon = g_\varepsilon(x_0) - f_\varepsilon(x_0) \to g(x_0) - f(x_0) =: c$, and so for all $x \in \mathbb{R}^N$,

$$g(x) = \lim_{\varepsilon \to 0^+} g_\varepsilon(x) = \lim_{\varepsilon \to 0^+} (f_\varepsilon(x) + c_\varepsilon) = f(x) + c.$$

■

Definition 160 *A multifunction* $\Gamma : \mathbb{R}^N \to \mathcal{P}\left(\mathbb{R}^N\right)$ *is called* cyclically monotone *if*

$$(x_1 - x_0) \cdot y_0 + (x_2 - x_1) \cdot y_1 + \cdots + (x_0 - x_m) \cdot y_m \leq 0$$

for all $m \in \mathbb{N}$*, and all* (x_0, y_0)*, …,* $(x_m, y_m) \in \operatorname{graph} \Gamma$*. A cyclically monotone* $\Gamma : \mathbb{R}^N \to \mathcal{P}\left(\mathbb{R}^N\right)$ *is called* maximal *if its graph is not a proper subset of the graph of a cyclically monotone multifunction.*

Proposition 161 *Let* $f : \mathbb{R}^N \to (-\infty, \infty]$ *be a proper convex function. Then* $\partial f : \mathbb{R}^N \to \mathcal{P}\left(\mathbb{R}^N\right)$ *is cyclically monotone.*

Proof. Let $m \in \mathbb{N}$, and all (x_0, y_0), …, $(x_m, y_m) \in \operatorname{graph} \partial f$, i.e., $y_i \in \partial f(x_i)$ for all $i = 0, \ldots, m$. Define $x_{m+1} := x_0$ and $y_{m+1} := y_0$. By the subdifferentiability of f at x_i we get that

$$f(x) - f(x_i) \geq (x - x_i) \cdot y_i$$

for all $i = 0, \ldots, m$ and for all $x \in \mathbb{R}^N$. Taking $x = x_{i+1}$ and summing all the inequalities gives

$$0 = f(x_{m+1}) - f(x_0) = \sum_{i=0}^{m} \left[f(x_{i+1}) - f(x_i)\right] \geq \sum_{i=0}^{m} (x_{i+1} - x_i) \cdot y_i.$$

■

Theorem 162 (Rockafellar) *Let* $\Gamma : \mathbb{R}^N \to \mathcal{P}\left(\mathbb{R}^N\right)$ *be a multifunction. Then there exists a proper convex function* $f : \mathbb{R}^N \to (-\infty, \infty]$ *such that* $\Gamma(x) \subset \partial f(x)$ *for all* $x \in \mathbb{R}^N$ *if and only if* Γ *is cyclically monotone.*

Proof. Let $\Gamma : \mathbb{R}^N \to \mathcal{P}\left(\mathbb{R}^N\right)$ be a multifunction and assume that there exists a proper convex function $f : \mathbb{R}^N \to (-\infty, \infty]$ such that $\Gamma(x) \subset \partial f(x)$ for all $x \in \mathbb{R}^N$. Since ∂f by the previous proposition, then the same must be true for Γ.

Conversely, assume that $\Gamma : \mathbb{R}^N \to \mathcal{P}\left(\mathbb{R}^N\right)$ is cyclically monotone and fix $(x_0, y_0) \in \operatorname{graph} \Gamma$. Define the function

$$\begin{aligned} f(x) := \sup \{&(x_1 - x_0) \cdot y_0 + \cdots + (x_m - x_{m-1}) \cdot y_{m-1} + (x - x_m) \cdot y_m \\ &m \in \mathbb{N},\ (x_1, y_1), \ldots, (x_m, y_m) \in \operatorname{graph} \Gamma\}, \quad x \in \mathbb{R}^N. \end{aligned}$$

Then f is convex and lower semicontinuous, since it is the supremum of a family of affine functions. Since $\Gamma : \mathbb{R}^N \to \mathcal{P}\left(\mathbb{R}^N\right)$ is cyclically monotone, we have that $f(x_0) = 0$, which implies that f is not identically ∞. It follows from the definition that f never takes the value $-\infty$, and so f is proper. It remains to show that if $(\overline{x}, \overline{y}) \in \operatorname{graph} \Gamma$, then $\overline{y} \in \partial f(\overline{x})$. Fix any $t < f(\overline{x})$ and by the definition of f find $m \in \mathbb{N}$, and (x_0, y_0), …, $(x_m, y_m) \in \operatorname{graph} \Gamma$ such that

$$t < (x_1 - x_0) \cdot y_0 + \cdots + (x_m - x_{m-1}) \cdot y_{m-1} + (\overline{x} - x_m) \cdot y_m.$$

Define $x_{m+1} := \overline{x}$ and $y_{m+1} := \overline{y}$. Again by the definition of f for any $x \in \mathbb{R}^N$,

$$\begin{aligned} f(x) &\geq (x_1 - x_0) \cdot y_0 + \cdots + (x_m - x_{m-1}) \cdot y_{m-1} + (\overline{x} - x_m) \cdot y_m + (x - \overline{x}) \cdot \overline{y} \\ &> t + (x - \overline{x}) \cdot \overline{y}. \end{aligned}$$

Hence $f(x) \geq t + (x - \overline{x}) \cdot \overline{y}$ for all $x \in \mathbb{R}^N$ and all $t < f(\overline{x})$. Letting $t \nearrow f(\overline{x})$ we conclude that

$$f(x) \geq f(\overline{x}) + (x - \overline{x}) \cdot \overline{y}$$

for all $x \in \mathbb{R}^N$, which implies that f is subdifferentiable at $\overline{x}$ with $\overline{y} \in \partial f(\overline{x})$. ■

Note that the convex function f constructed in the previous theorem is lower semicontinuous. Next we prove that maximal cyclically monotone multifunctions are exactly the class of subgradients of proper convex lower semicontinuous functions.

Theorem 163 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be a proper convex lower semicontinuous function. Then $\partial f : \mathbb{R}^N \to \mathcal{P}(\mathbb{R}^N)$ is a maximal cyclically monotone multifunction. Conversely, given a maximal cyclically monotone multifunction $\Gamma : \mathbb{R}^N \to \mathcal{P}(\mathbb{R}^N)$, up to an additive constant, there exists a unique proper convex lower semicontinuous function $f : \mathbb{R}^N \to (-\infty, \infty]$ such that $\Gamma(x) = \partial f(x)$ for all $x \in \mathbb{R}^N$.*

Proof. Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be a proper convex lower semicontinuous function. We claim that $\partial f : \mathbb{R}^N \to \mathcal{P}(\mathbb{R}^N)$ is a maximal cyclically monotone multifunction. In view of the previous theorem, it remains to show that ∂f is maximal. Let $\Gamma : \mathbb{R}^N \to \mathcal{P}(\mathbb{R}^N)$ be a cyclically monotone multifunction such that $\Gamma(x) \supset \partial f(x)$ for all $x \in \mathbb{R}^N$. Since Γ is cyclically monotone, in particular, Γ is monotone. On the other hand, ∂f is a maximal monotone multifunction by Theorem 155, and thus $\partial f = \Gamma$. ■

Exercise 164 (Monotone and cyclically monotone multifunction) *(i) Prove that a multifunction $\Gamma : \mathbb{R} \to \mathcal{P}(\mathbb{R})$ is monotone if and only if it is cyclically monotone.*

(ii) Let A be an $N \times N$ matrix, $N \geq 2$, and consider the multifunction

$$\begin{aligned} &\Gamma : \mathbb{R}^N \to \mathcal{P}(\mathbb{R}^N), \\ &x \mapsto \Gamma(x) := \{Ax\}. \end{aligned}$$

Prove that Γ is cyclically monotone if and only if A is symmetric and positive-semidefinite.

(iii) Prove that if $A + A^T$ is positive-semidefinite, then Γ is monotone.

(iv) Construct a multifunction Γ that is monotone but not cyclically monotone.

We conclude this subsection with a proof of Alexandrov's theorem.

Theorem 165 (Alexandrov) *If $f : B(x_0, r) \subset \mathbb{R}^N \to \mathbb{R}$ is convex, and*

$$E := \{x \in B(x_0, r) : f \text{ is differentiable at } x\},$$

then $\nabla f : E \to \mathbb{R}^N$ is differentiable $\mathcal{L}^N$ a.e. in E.

Remark 166 *Note that Theorem 122 does not imply Alexandrov's theorem, unless $N = 1$. Indeed, by a Theorem of Serrin, a function in BV has a representative for which the partial derivatives exist $\mathcal{L}^N$ a.e., but this is not enough to conclude differentiability.*

Friday, April 4, 2008

Exercise 167 *Let $C \subset \mathbb{R}^N$ be a convex set and let $f : C \to \mathbb{R}$ be convex and Lipschitz. Consider the function*

$$g(x) := \inf \{f(y) + (\operatorname{Lip} f) |x - y| : y \in C\}, \quad x \in \mathbb{R}^N.$$

(i) Prove that g is Lipschitz with $\operatorname{Lip} g = \operatorname{Lip} f$ and that $g = f$ in C.

(ii) Prove that g is convex.

(iii) Give an example of a convex function $f : [-1, 1] \to \mathbb{R}$ that can be extended to a convex function from $\mathbb{R}$ into $\mathbb{R}$ in two different ways.

Proof of Alexandrov's theorem. Since the result is local, by restricting f to a smaller ball and then applying the previous exercise, we may assume that f is defined in the whole $\mathbb{R}^N$ (as a real-valued convex function). Let

$$E_1 = \left\{x \in \mathbb{R}^N : f \text{ is differentiable at } x\right\}.$$

By Theorem 117, the set E_1 has full measure, that is, $\mathcal{L}^N\left(\mathbb{R}^N \setminus E_1\right) = 0$. Moreover, by the previous theorem the function $G := (\partial f + I)^{-1} : \mathbb{R}^N \to \mathbb{R}^N$ is Lipschitz continuous with Lipschitz constant at most one and onto, since $\partial f(x)$ is nonempty for every $x \in \mathbb{R}^N$. It follows by Rademacher's theorem that G is differentiable $\mathcal{L}^N$ a.e.. Moreover,G maps null sets to null sets, and the set

$$G\left(\left\{y \in \mathbb{R}^N : G \text{ is differentiable at } y \text{ and } \det G(y) = 0\right\}\right)$$

has Lebesgue measure zero (see Rudin). Hence the set

$$E_2 = \{G(y) : G \text{ is differentiable at } y \text{ and } \det G(y) \neq 0\}$$

has full measure, that is, $\mathcal{L}^N\left(\mathbb{R}^N \setminus E_2\right) = 0$.

We claim that ∇f is differentiable in $E_1 \cap E_2$. For any $x \in E_1 \cap E_2$. Then $x = G(y)$ for some $y \in \mathbb{R}^N$, G is differentiable at y and $\det G(y) \neq 0$. By the definition of G, we have

$$\begin{aligned} \nabla f(x) &= \nabla f(G(y)) = \nabla f(G(y)) \pm G(y) \qquad (46) \\ &= (\nabla f + I)(\partial f + I)^{-1}(y) - G(y) = y - G(y). \end{aligned}$$

Fix $x_0 \in E_1 \cap E_2$ and let $x_0 = G(y_0)$ for some $y_0 \in \mathbb{R}^N$. Take Moreover, if $(G(y_0) + h) \in E_1$, then $\nabla f(G(y_0) + h)$ exists. Since G is Lipschitz and $\det G(y_0) \neq 0$, if h is sufficiently small, there exists $z_h \in \mathbb{R}^N$ such that (see Rudin, proof of Theorem notes)

$$G(y_0 + z_h) = G(y_0) + h$$

and we may choose $z_{x,h}$ to satisfy $|z_{x,h}| \leq K|h|$ for some constant $K > 0$. On the other, hand

$$|h| = |G(y_0 + z_h) - G(y_0)| \leq |z_h|.$$

Using (46) twice, for all h sufficiently small we have

$$\begin{aligned}\nabla f(x_0+h) &= \nabla f(G(y_0)+h) = \nabla f(G(y_0+z_h)) \\ &= y_0 + z_h - G(y_0+z_h) \\ &= \nabla f(G(y_0)) + G(y_0) - G(y_0+z_h) + z_h \\ &= \nabla f(x_0) - \nabla G(y_0) z_h + o(h) + z_h,\end{aligned}$$

where in the last equality we have used the fact that G is differentiable at y_0 and that $o(z_h) = o(h)$. Since

$$G(y_0) + \nabla G(y_0) z_h + o(z_h) = G(y_0+z_h) = G(y_0) + h,$$

it follows that

$$z_h = (\nabla G(y_0))^{-1}(h + o(z_h)) = (\nabla G(y_0))^{-1} h + o(h),$$

and so we obtain

$$\begin{aligned}\nabla f(x_0+h) &= \nabla f(x_0) + z_h - \nabla G(y_0) z_h + o(h) \\ &= \nabla f(x_0) + \left((\nabla G(y_0))^{-1} - I\right) h + o(h),\end{aligned}$$

which shows that

$$\lim_{h\to 0} \frac{\nabla f(x_0+h) - \nabla f(x_0) - \left((\nabla G(y_0))^{-1} - I\right) h}{|h|} = 0,$$

namely that ∇f is differentiable at x_0, with

$$\nabla^2 f(x_0) = (\nabla G(y_0))^{-1} - I.$$

We also show that a second order Taylor's formula holds, precisely, that

$$f(x_0+h) = f(x_0) + \nabla f(x_0) h + \frac{1}{2}\left\langle \nabla^2 f(x_0) h, h\right\rangle + o\left(|h|^2\right).$$

Define

$$\begin{aligned}&\psi(h) := f(x_0+h), \\ &\widetilde{\psi}(h) := f(x_0) + \nabla f(x_0) h + \frac{1}{2}\left\langle \left((\nabla G(y_0))^{-1} - I\right) h, h\right\rangle.\end{aligned}$$

Then $\psi(0) = \widetilde{\psi}(0)$ and for $\mathcal{L}^N$ a.e. small h,

$$\begin{aligned}\nabla\psi(h) &= \nabla f(x_0+h) = \nabla f(x_0) - \left((\nabla G(y_0))^{-1} - I\right) h + o(h) \\ &= \nabla\widetilde{\psi}(h) + o(h),\end{aligned}$$

and so the function $\Psi(h) := \psi(h) - \widetilde{\psi}(h)$ is locally Lipschitz continuous and satisfies $\Psi(0) = 0$, $\nabla\Psi(h) = o(h)$ for $\mathcal{L}^N$ a.e. small h. Hence $\Psi(h) = o\left(|h|^2\right)$, which completes the proof. ■

Exercise 168 *Consider the function* $f:\mathbb{R}^2\to\mathbb{R}$*, defined by*

$$f(x,y) := \sum_{j=0}^{4} a_j \frac{x^j y^{4-j}}{x^2+y^2} \quad \text{if } (x,y)\neq(0,0),$$
$$f(0,0) := 0,$$

where a_0*,* a_1*,* a_2*,* a_3*,* $a_4\in\mathbb{R}$*.*

(i) Calculate the Hessian matrix

$$Hf(x,y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(x,y) & \frac{\partial^2 f}{\partial x \partial y}(x,y) \\ \frac{\partial^2 f}{\partial y \partial x}(x,y) & \frac{\partial^2 f}{\partial y^2}(x,y) \end{pmatrix}$$

for all $(x,y)\in\mathbb{R}^2$ *and find a necessary and sufficient condition on* a_0*,* a_1*,* a_2*,* a_3*,* a_4 *for* Hf *to be symmetric.*

(ii) Find a necessary and sufficient condition on a_0*,* a_1*,* a_2*,* a_3*,* a_4 *for* ∇f *to be everywhere differentiable.*

(iii) Prove that if $n\in\mathbb{N}$ *is sufficiently large, then the function*

$$g(x,y) := f(x,y) + n\left(x^2+y^2\right)$$

is convex, but for appropriate values of a_0*,* a_1*,* a_2*,* a_3*,* a_4*,* $Hg(0,0)$ *is not symmetric or* ∇g *is not everywhere differentiable.*

3.4 Conjugate Functions

Definition 169 *Given a function* $f:\mathbb{R}^N\to[-\infty,\infty]$*, the* conjugate function $f^*:\mathbb{R}^N\to[-\infty,\infty]$ *of* f *is defined by*

$$f^*(y) := \sup_{x\in\mathbb{R}^N}\{y\cdot x - f(x)\}, \quad y\in\mathbb{R}^N,$$

and the biconjugate function $f^{**}:\mathbb{R}^N\to[-\infty,\infty]$ *of* f *is defined by* $f^{**} := (f^*)^*$*.*

Remark 170 *It follows from the definition of* f^**, that*

$$f^*(y) := \sup_{x\in\operatorname{dom}_e f}\{y\cdot x - f(x)\}, \quad y\in\mathbb{R}^N,$$

and so f^* *never takes the value* $-\infty$*, unless* $f\equiv\infty$*.*

Since f^* is the supremum of a family of continuous and convex functions, f^* is convex and lower semicontinuous. The same holds true for f^{**}, and $f^{**}\le f$. Even when f is convex and lower semicontinuous it may happen that $f^{**}\neq f$. Indeed, if f takes the value $-\infty$ at some point and $f\not\equiv-\infty$, then $f^*\equiv\infty$, and in turn, $f^{**}\equiv-\infty$.

Exercise 171 *Prove that:*

(i) If $1 \leq p < \infty$, then the polar function of $f(x) = \frac{|x|^p}{p}$ is

$$f^*(y) = \frac{|y|^{p'}}{p'} \quad y \in \mathbb{R}^N,$$

for $p > 1$ and

$$f^*(y) = \begin{cases} \infty & \text{if } |y| > 1, \\ 0 & \text{if } |y| \leq 1, \end{cases}$$

if $p = 1$.

(ii) The polar function of $f(x) := \sqrt{|x|^2 + 1}$ is

$$f^*(y) = \begin{cases} \infty & \text{if } |y| > 1, \\ -\sqrt{1 - |y|^2} & \text{if } |y| \leq 1. \end{cases}$$

From the definition of f^* we have that if $f \not\equiv \infty$,

$$f^*(y) \geq y \cdot x - f(x)$$

for all $y \in \mathbb{R}^N$ and $x \in \mathbb{R}^N$, or equivalently

$$f^*(y) + f(x) \geq y \cdot x,$$

provided we exclude the case $f^*(y) = \infty$ and $f(x) = -\infty$. The next result characterizes pairs $(x, y) \in \mathbb{R}^N \times \mathbb{R}^N$ for which equality holds.

Theorem 172 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be a proper convex function, and let $(x, y) \in \mathbb{R}^N \times \mathbb{R}^N$. Then $y \in \partial f(x)$ if and only if*

$$f(x) + f^*(y) = y \cdot x. \tag{47}$$

Proof. Fix $(x_0, y_0) \in \mathbb{R}^N \times \mathbb{R}^N$. If $y_0 \in \partial f(x_0)$, then

$$f(x) \geq f(x_0) + y_0 \cdot (x - x_0) \quad \text{for all } x \in \mathbb{R}^N,$$

or equivalently,

$$y_0 \cdot x_0 - f(x_0) \geq f^*(y_0) = \sup_{x \in \mathbb{R}^N} \{y_0 \cdot x - f(x)\}.$$

Since the opposite inequality holds by definition of f^*, equality (47) follows.

Conversely, assume that (47) holds at (x_0, y_0). In particular, $f(x_0) \in \mathbb{R}$. By definition of $f^*(y_0)$ we have that for all $x \in \mathbb{R}^N$,

$$f(x) - y_0 \cdot x \geq -f^*(y_0) = f(x_0) - y_0 \cdot x_0,$$

that is,

$$f(x) \geq f(x_0) + y_0 \cdot (x - x_0) \quad \text{for all } x \in \mathbb{R}^N,$$

which is equivalent to $y_0 \in \partial f(x_0)$. ■

Theorem 173 *Let $f : \mathbb{R}^N \to \mathbb{R}$ be a proper convex function. Then*

$$\operatorname{graph}(\partial f)^{-1} \subset \operatorname{graph}\partial f^*.$$

Moreover, if f is also lower semicontinuous, then

$$(\partial f)^{-1} = \partial f^*, \tag{48}$$

and $f^{**} = f$.

Proof. The proof is very similar to the one of Theorem 54, with the only difference that we use Corollary 159 to conclude from $(\partial f)^{-1} = \partial f^*$ that $f^{**} = f + \text{const}$. We omit the details. ■

Theorem 174 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be a proper, lower semicontinuous, convex function. Then $f^* : \mathbb{R}^N \to (-\infty, \infty]$ is differentiable in $(\operatorname{dom}_e f^*)^\circ$ if and only if f is strictly convex in all convex sets contained in*

$$\bigcup_{y \in (\operatorname{dom}_e f^*)^\circ} \partial f^*(y).$$

Proof. Assume that f is strictly convex in all convex sets contained in

$$E := \bigcup_{y \in (\operatorname{dom}_e f^*)^\circ} \partial(f^*)(y),$$

let $y_0 \in (\operatorname{dom}_e f^*)^\circ$ and assume by contradiction that f^* is not differentiable at y_0. Then by Theorems 147 and 142, $\partial f^*(y_0)$ contains at least two distinct elements x_1, x_2. By the previous theorem it follows that $y_0 \in \partial f(x_1) \cap \partial f(x_1)$. By the equality case in the Young inequality, we have

$$x_1 \cdot y_0 = f(x_1) + f^*(y_0), \quad x_2 \cdot y_0 = f(x_2) + f^*(y_0).$$

Hence, for any $\theta \in [0,1]$ we have

$$\begin{aligned}\theta f(x_1) + (1-\theta) f(x_2) + f^*(y_0) &= \theta(f(x_1) + f^*(y_0)) + (1-\theta)(f(x_2) + f^*(y_0)) \\ &= \theta(x_1 \cdot y_0) + (1-\theta)(x_2 \cdot y_0) \\ &= (\theta x_1 + (1-\theta) x_2) \cdot y_0 \\ &\le f(\theta x_1 + (1-\theta) x_2) + f^*(y_0),\end{aligned}$$

where in the last inequality we have used the Young inequality. It follows that

$$\theta f(x_1) + (1-\theta) f(x_2) = f(\theta x_1 + (1-\theta) x_2),$$

which contradicts the strict convexity of f along the segment $\overline{x_1 x_2} \subset E$.

Conversely, assume that $f^* : (\operatorname{dom}_e f^*)^\circ \to \mathbb{R}$ is differentiable in $(\operatorname{dom}_e f^*)^\circ$. By the previous theorem, $f = f^{**}$. Let C be any convex set contained in E and assume by contradiction that f is not strictly convex in C. Then there exist

$x_1 \neq x_2$ in C such that f is affine along the segment $\overline{x_1 x_2}$. Let $x := \frac{x_1+x_2}{2} \in C \subset E$. By the definition of E, there exists $y \in (\mathrm{dom}_e\, f^*)^\circ$ such that $x \in \partial\,(f^*)\,(y)$, and since f^* is differentiable in $(\mathrm{dom}_e\, f^*)^\circ$, it follows that $\nabla f^*\,(y) = x$. Using the facts that f is affine in $(\mathrm{dom}_e\, f^*)^\circ$ and that $\nabla f^*\,(y) = x$ it follows from the equality in the Young inequality that (recall that $f = f^{**}$) we have

$$0 = f\,(x) + f^*\,(y) - x \cdot y = \frac{1}{2}\,(f\,(x_1) + f^*\,(y) - x_1 \cdot y) + \frac{1}{2}\,(f\,(x_2) + f^*\,(y) - x_2 \cdot y)\,.$$

By the Young inequality, necessarily $f\,(x_1) + f^*\,(y) - x_1 y = 0$ and $f\,(x_2) + f^*\,(y) - x_2 y = 0$, which implies that $x_1,\, x_2 \in \partial f^*\,(y) = \{\nabla f^*\,(y)\}$. This is a contradiction. ∎

Next we extend Theorem **??** to $\mathbb{R}^N$.

Theorem 175 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be a proper, lower semicontinuous, convex function. Assume that ∇f exists and is differentiable in a neighborhood of some point $x_0 \in \mathbb{R}^N$, that that $\nabla^2 f$ is continuous at x_0, and that the $N \times N$ matrix $\nabla^2 f(x_0)$ is nonsingular. Let $y_0 := \nabla f(x_0)$. Then f^* is differentiable in a neighborhood of y_0, ∇f^* is differentiable at y_0, and*

$$\nabla^2 f^*(y_0) = \left(\nabla^2 f(x_0)\right)^{-1}.$$

Moreover, if f is of class C^2 in a neighborhood of x_0, then f^* is of class C^2 in a neighborhood of y_0.

Proof. Step 1: We claim that $\partial f^*(y_0) = \{x_0\}$. Indeed, since $y_0 \in \partial f(x_0)$, by Theorem 173 we have that

$$(\partial f)^{-1}(y_0) = \{x \in \operatorname{dom} \partial f : y_0 \in \partial f(x)\} \subset \partial f^*(y_0),$$

and so have that $x_0 \in \partial f^*(y_0)$. If the convex set $\partial f^*(y_0)$ contains another element, say $x_0 + v$ for some $v \in \mathbb{R}^N \setminus \{0\}$, then it contains the segment $x_0 + v[0,1]$. Again by Theorem 173 we have that $y_0 \in \partial f(x_0 + tv)$ for all all $t \in [0,1]$. Since f is differentiable in a neighborhood of x_0, there exists $t_0 \in (0,1]$ such that

$$\nabla f(x_0 + tv) = y_0$$

for all $t \in [0, t_0]$. By differentiating with respect to t we obtain that

$$\nabla^2 f(x_0)\, v = 0,$$

which contradicts the fact that $\nabla^2 f(x_0)$ is nonsingular. Thus the claim holds and $\partial f^*(y_0) = \{x_0\}$. By Theorem 147, we have that f^* is differentiable at y_0, and, in turn, that $y_0 \in (\operatorname{dom}_e f^*)^\circ$. Note that in this part we have not used the fact that $\nabla^2 f$ is continuous at x_0.

Step 2: We prove that f^* is differentiable in a neighborhood of y_0. Since $\nabla^2 f$ is continuous at x_0 and $\nabla^2 f(x_0)$ is nonsingular, there exists $r > 0$ such that ∇f is differentiable and $\nabla^2 f$ is nonsingular in $B(x_0, r)$. By Theorem 150 and the fact that $y_0 \in (\operatorname{dom}_e f^*)^\circ$, we have that ∂f^* is upper semicontinuous, so there exists $\delta > 0$ such that $B(y_0, \delta) \subset \operatorname{dom}_e f^*$ and

$$\partial f^*(B(y_0, \delta)) = \bigcup_{y \in B(y_0, \delta)} \partial f^*(y) \subset B(x_0, r).$$

We claim that f^* is differentiable in $B(y_0, \delta)$. Since $B(y_0, \delta) \subset \operatorname{dom}_e f^*$, by Theorem 147 we have that $\partial f^*(y)$ is nonempty for each $y \in B(y_0, \delta)$, and so it remains to show that $\partial f^*(y)$ is a singleton. If $x \in \partial f^*(y)$, then $x \in B(x_0, r)$, and thus by the previous step applied to x instead of x_0 we get the desired result.

Step 3: Define $W := \partial f^* (B (y_0, \delta)) \subset B (x_0, r)$. We claim that W is a neighborhood of x_0. Since ∇f is continuous at x_0, there exists $r_1 \in (0, r)$ such that if $x \in B (x_0, r_1)$, then $\nabla f (x) \in B (y_0, \delta)$ (recall that $y_0 = \nabla f (x_0)$). By Theorem 173, $(\partial f)^{-1} = \partial f^*$, and so for all $x \in B (x_0, r_1)$,

$$x \in \partial f^* (\nabla f (x)) \subset \partial f^* (B (y_0, \delta)) = W.$$

Next we prove that ∇f is bijective from W onto $B (y_0, \delta)$. If $x \in W$, then there exists at least one $y \in B (y_0, \delta)$ such that $x = \nabla f^* (y)$. Hence $y \in \partial f (x)$, but since $x \in B (x_0, r)$, then necessarily, $y = \nabla f (x)$. This shows that ∇f is bijective from W onto $B (y_0, \delta)$ and its inverse is ∇f^*. We are now in a position to apply the inverse function theorem to conclude that ∇f^* is differentiable at y_0, and

$$\nabla^2 f^* (y_0) = \left(\nabla^2 f (x_0)\right)^{-1}.$$

■

Wednesday, April 9, 2008

3.5 Convex Envelopes

As we will see in relaxation problems, in the case of nonconvex integrands f one is interested in "convexifying" f. This brings us to various notions of convex envelopes.

Definition 176 *Let V be a vector space and let $f : V \to [-\infty, \infty]$. The* convex envelope $\operatorname{co} f : V \to [-\infty, \infty]$ *of f is defined by*

$$(\operatorname{co} f)(v) := \sup \{g(v) : g : V \to [-\infty, \infty] \text{ convex, } g \leq f\}.$$

Remark 177 *To obtain $\operatorname{co} f$, we should first convexify the epigraph of f, that is consider the set $\operatorname{co}(\operatorname{epi} f)$, and then consider a function whose epigraph is $\operatorname{co}(\operatorname{epi} f)$. We will see that to make this precise we need to somewhat "close" the bottom of $\operatorname{co}(\operatorname{epi} f)$. See the next proposition and Remark 180 below.*

Let V be a vector space $f : V \to [-\infty, \infty]$. Define

$$\begin{aligned}\left(\operatorname{co}^1 f\right)(v) := \inf \{&\theta f(v_1) + (1-\theta) f(v_2) : \theta \in [0,1], v_1, v_2 \in \operatorname{dom}_e f, \\ &v = \theta v_1 + (1-\theta) v_2\},\end{aligned}$$

where if $\theta = 0$ we set $\theta f(v_1) := 0$ even if $f(v_1) = -\infty$ (and similarly if $(1-\theta) = 0$). Note that $\operatorname{co} f \leq \operatorname{co}^1 f \leq f$, but in general $\operatorname{co}^1 f$ is not convex. Moreover, if f is convex, then $\operatorname{co} f = \operatorname{co}^1 f = f$.

For every $n \in \mathbb{N}$, define recursively

$$\operatorname{co}^{n+1} f := \operatorname{co}^1 f(\operatorname{co}^n f).$$

Exercise 178 *Let $x_1, x_2, x_3 \in \mathbb{R}^2$ be three points that are not alligned, let $E := \{x_1, x_2, x_3\}$ and let $f = I_E$. Find $\operatorname{co}^1 f$ and $\operatorname{co} f$.*

Proposition 179 (Dal Maso) *Let V be a vector space $f : V \to [-\infty, \infty]$. Then $\operatorname{co} f \leq \operatorname{co}^{n+1} f \leq \operatorname{co}^n f \leq f$ for every $n \in \mathbb{N}$ and*

$$\begin{aligned}(\operatorname{co} f)(v) = \lim_{n\to\infty} (\operatorname{co}^n f)(v) = \inf \Bigg\{&\sum_{i=1}^m \theta_i f(v_i) : m \in \mathbb{N}, \theta_i \in [0,1], v_i \in \operatorname{dom}_e f, \\ &i = 1, \ldots, m, \sum_{i=1}^m \theta_i = 1, \sum_{i=1}^m \theta_i v_i = v\Bigg\},\end{aligned}$$

for all $v \in V$.

Proof. The inequalities $\operatorname{co} f \leq \operatorname{co}^{n+1} f \leq \operatorname{co}^n f \leq f$ follows from the definitions. Define $f_\infty(v) := \lim_{n\to\infty} (\operatorname{co}^n f)(v)$. Then $\operatorname{co} f \leq f_\infty \leq f$. Note that

$$f_\infty(v) = \inf_{n\in\mathbb{N}} (\mathrm{co}^n f)(v) = \inf \left\{ \sum_{i=1}^m \theta_i f(v_i) : \, m \in \mathbb{N}, \, \theta_i \in [0,1], \, v_i \in \mathrm{dom}_e f, \right.$$
$$\left. i = 1, \dots, m, \, \sum_{i=1}^m \theta_i = 1, \, \sum_{i=1}^m \theta_i v_i = v \right\}.$$

To prove that $f_\infty \le \mathrm{co}\, f$, it remains to show that f_∞ is convex. To see this, let $v_1, \; v_2 \in V$ and $\theta \in (0,1)$. If $f_\infty(v_1)$ or $f_\infty(v_2)$ are infinite, there is nothing to prove, so assume that $f_\infty(v_1) < \infty$ and $f_\infty(v_2) < \infty$. In turn, $(\mathrm{co}^n f)(v_1) < \infty$ and $(\mathrm{co}^n f)(v_2) < \infty$ for all n sufficiently large. Fix any such n. By the definition of $(\mathrm{co}^n f)(v)$, , we have

$$\left(\mathrm{co}^{n+1} f\right)(\theta v_1 + (1-\theta) v_2) \le \theta (\mathrm{co}^n f)(v_1) + (1-\theta)(\mathrm{co}^n f)(v_2),$$

whenever the right hand-side is well-defined. Letting $n \to \infty$, yields

$$f_\infty(\theta v_1 + (1-\theta) v_2) \le \theta f_\infty(v_1) + (1-\theta) f_\infty(v_2),$$

which shows that f_∞ is convex. Hence $f_\infty \le \mathrm{co}\, f$. ■

Remark 180 *Note that Proposition 179 implies that*

$$\mathrm{dom}_e\, \mathrm{co}\, f = \mathrm{co}(\mathrm{dom}_e f)$$

and that

$$\mathrm{co}(\mathrm{epi}\, f) \subset \mathrm{epi}(\mathrm{co}\, f).$$

In general, the strict inclusion possible. The problem is the bottom of the set $\mathrm{co}(\mathrm{epi}\, f)$*.*

Exercise 181 *Prove that if* $f : \mathbb{R}^N \to [-\infty, \infty]$*, then*

$$\mathrm{co}(\mathrm{epi}\, f) \subset \mathrm{epi}(\mathrm{co}\, f) \subset \overline{\mathrm{co}(\mathrm{epi}\, f)}.$$

Exercise 182 *Let* $f : \mathbb{R} \to \mathbb{R}$*, defined by*

$$f(x) := \begin{cases} |x| & \text{if } x \ne 0, \\ 1 & \text{if } x = 0. \end{cases}$$

Find $\mathrm{co}(\mathrm{epi}\, f)$ *and* $\mathrm{epi}(\mathrm{co}\, f)$*.*

If we restrict our attention to the space $V = \mathbb{R}^N$, then by Carathéodory's theorem one may restrict the number of convex combinations in $\mathrm{co}\, f$.

Corollary 183 *Let* $f : \mathbb{R}^N \to [-\infty, \infty]$*. Then for all* $x \in \mathbb{R}^N$*,*

$$(\mathrm{co}\, f)(x) = \inf \left\{ \sum_{i=1}^{N+2} \theta_i f(x_i) : \, \theta_i \in [0,1], \, x_i \in \mathrm{dom}_e f, \, i = 1, \dots, N+2, \right.$$
$$\left. \sum_{i=1}^{N+2} \theta_i = 1, \, \sum_{i=1}^{N+2} \theta_i x_i = x \right\}.$$

Proof. Let $g(x)$ denote the the right-hand side of the previous identity. By the previous theorem we have that $\operatorname{co} f \leq g$. To prove the opposite inequality, let x be such that

$$x = \sum_{i=1}^{n} \theta_i x_i$$

for some $n \in \mathbb{N}$, $\sum_{i=1}^{n} \theta_i = 1$, $\theta_i > 0$, $x_i \in \operatorname{dom}_e f$. Then for any $t_i \geq f(x_i)$, with $t_i \in \mathbb{R}$, (note that $f(x_i)$ could be $-\infty$) the point

$$\left(\sum_{i=1}^{n} \theta_i x_i, \sum_{i=1}^{n} \theta_i t_i \right)$$

belongs to $\operatorname{co}(\operatorname{epi} f)$, and so, by Carathéodory's theorem applied to $\operatorname{co}(\operatorname{epi} f)$, it can be written as a convex combination of $N+2$ elements, say

$$\left(\sum_{i=1}^{n} \theta_i x_i, \sum_{i=1}^{n} \theta_i t_i \right) = \left(\sum_{j=1}^{N+2} \lambda_j y_j, \sum_{i=i}^{N+2} \lambda_j s_j \right),$$

where $s_j \geq f(y_j)$. Hence

$$g(x) \leq \sum_{j=i}^{N+2} \lambda_j f(y_j) \leq \sum_{j=i}^{N+2} \lambda_j s_j = \sum_{i=1}^{n} \theta_i t_i.$$

Letting $t_i \searrow f(x_i)$, gives

$$g(x) \leq \sum_{i=1}^{n} \theta_i f(x_i),$$

and taking the infimum on the right-hand side yields $g(x) \leq \operatorname{co} f(x)$. ■

Actually, it is possible to prove that in the previous corollary one can replace $N+2$ with $N+1$. This follows from the following auxiliary result.

Proposition 184 *Let $E \subset \mathbb{R}^N$ be a nonempty set and let $x_0 \in \operatorname{co}E \cap \partial(\operatorname{co}E)$. Then x can be represented as a convex combination of N elements of E.*

Proof. If $\operatorname{co}E$ has dimension less than N, then there is nothing to prove. Thus assume that $\operatorname{co}E$ has dimension less than N, so that By Theorem 89 with $C_1 = \{x_0\}$ and $C_2 = \operatorname{co} E$ there exist $b \in \mathbb{R}^N \setminus \{0\}$ and $\alpha \in \mathbb{R}$ such that

$$b \cdot x_0 - \alpha = 0 \text{ and } b \cdot x - \alpha \geq 0 \quad \text{for all } x \in \operatorname{co} E. \tag{49}$$

On the other hand, by Carathéodory's theorem there exist $x_i \in E$, $i = 1, \ldots, N+1$, and $\theta_i \geq 0$, $i = 1, \ldots, N+1$, with $\sum_{i=1}^{N+1} \theta_i = 1$ and

$$x_0 = \sum_{i=1}^{N+1} \theta_i x_i.$$

Without loss of generality we may assume that $\theta_i > 0$ for all $i = 1, \ldots, N+1$, since otherwise there is nothing to prove. Taking $x = x_i$ in (49) gives $b \cdot x_i - \alpha \geq 0$, $i = 1, \ldots, N+1$, and summing we obtain

$$0 = b \cdot x_0 - \alpha = \sum_{i=1}^{N+1} \theta_i \, (b \cdot x_i - \alpha) \geq 0.$$

Hence $b \cdot x_i - \alpha = 0$ for all $i = 1, \ldots, N+1$. Hence all the x_i belong the hyperplane $b \cdot x = \alpha$, and since this has dimension $N-1$ we can write x_0 as as a convex combination of N elements of $E \cap \{b \cdot x = \alpha\}$. ∎

Theorem 185 *Let $f : \mathbb{R}^N \to [-\infty, \infty]$. Then for all $x \in \mathbb{R}^N$,*

$$(\operatorname{co} f)(x) = \inf \left\{ \sum_{i=1}^{N+1} \theta_i f(x_i) : \theta_i \in [0,1], x_i \in \operatorname{dom}_e f, i = 1, \dots, N+1, \right.$$
$$\left. \sum_{i=1}^{N+1} \theta_i = 1, \sum_{i=1}^{N+1} \theta_i x_i = x \right\}.$$

Exercise 186 *Prove the previous theorem.*

The next proposition gives another characterization of $\operatorname{co} f$.

Corollary 187 *Let V be a vector space and let $f : V \to [-\infty, \infty]$. Then*

$$\operatorname{co} f(v) = \inf \{ t \in \mathbb{R} : (v,t) \in \operatorname{co}(\operatorname{epi} f) \}$$

for all $v \in V$.

Proof. Define

$$h(v) := \inf \{ t \in \mathbb{R} : (v,t) \in \operatorname{co}(\operatorname{epi} f) \}, \quad v \in V.$$

If $(v,t) \in \operatorname{co}(\operatorname{epi} f)$, then by Proposition 72, (v,t) can be written as a convex combination of elements of $\operatorname{epi} f$, that is,

$$v = \sum_{i=1}^{n} \theta_i v_i, \quad t = \sum_{i=1}^{n} \theta_i t_i$$

for some $n \in \mathbb{N}$, $\sum_{i=1}^{n} \theta_i = 1$, $\theta_i \geq 0$, $v_i \in \operatorname{dom}_e f$, $t_i \geq f(v_i)$. Hence,

$$t \geq \sum_{i=1}^{n} \theta_i f(v_i) \geq \operatorname{co} f(v).$$

Taking the infimum over all t such that $(v,t) \in \operatorname{co}(\operatorname{epi} f)$, gives

$$h(v) \geq \operatorname{co} f(v).$$

Conversely, if $v = \sum_{i=1}^{n} \theta_i v_i$, for some $n \in \mathbb{N}$, $\sum_{i=1}^{n} \theta_i = 1$, $\theta_i \geq 0$, $v_i \in \operatorname{dom}_e f$, then for any $t_i > f(v_i)$, define

$$t := \sum_{i=1}^{n} \theta_i t_i.$$

Then $(v,t) \in \operatorname{co}(\operatorname{epi} f)$, and so

$$\sum_{i=1}^{n} \theta_i t_i = t \geq h(v).$$

Letting $t_i \searrow f(v_i)$ yields

$$\sum_{i=1}^{n} \theta_i f(v_i) \geq h(v),$$

and taking the infimum on the left-hand side gives $\operatorname{co} f(v) \geq h(v)$. ■

We conclude this section with some regularity results for the convex envelope $\operatorname{co} f$ of a smooth function f.

Theorem 188 *Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be a continuous function. Assume that f is differentiable in $\operatorname{dom}_e f$. Then its convex envelope $\operatorname{co} f$ is C^1 in a neighborhood of each point $x_0 \in \mathbb{R}^N$ satisfying*

$$(\operatorname{co} f)(x_0) < \liminf_{|x| \to \infty} f(x).$$

Moreover, if ∇f is locally Hölder continuous with exponent $0 < \alpha < 1$ or locally Lipschitz in $\operatorname{dom}_e f$, then $\nabla(\operatorname{co} f)$ has the same (local) regularity in the open set

$$\left\{ w \in \mathbb{R}^N : (\operatorname{co} f)(w) < \liminf_{|x| \to \infty} f(x) \right\}.$$

Lemma 189 *Let $B \subset \mathbb{R}^N$ be any open ball. If $g : B \to \mathbb{R}$ is convex, and $f : B \to \mathbb{R}$ is differentiable at $x_0 \in B$, $g \leq f$, $f(x_0) = g(x_0)$, then g is differentiable at x_0 and $\nabla g(x_0) = \nabla f(x_0)$.*

Proof. Proof 1: Since $g \leq f$ and f is differentiable at x_0, we have

$$\begin{aligned} &\limsup_{x \to x_0} \frac{g(x) - g(x_0) - \nabla f(x_0) \cdot (x - x_0)}{|x - x_0|} \\ &\qquad \leq \lim_{x \to x_0} \frac{f(x) - f(x_0) - \nabla f(x_0) \cdot (x - x_0)}{|x - x_0|} = 0. \end{aligned} \tag{50}$$

Conversely, using Remark 107(i), for $\varepsilon > 0$ sufficiently small we have

$$\begin{aligned} &\inf_{y \in B_\infty(x_0, \varepsilon)} \{ g(y) - g(x_0) - \nabla f(x_0) \cdot (y - x_0) \} \\ &\geq -\left(2^N - 1\right) \sup_{y \in B_\infty(x_0, \varepsilon)} \{ g(y) - g(x_0) - \nabla f(x_0) \cdot (y - x_0) \} \\ &\geq -\left(2^N - 1\right) \sup_{y \in B_\infty(x_0, \varepsilon)} \{ f(y) - f(x_0) - \nabla f(x_0) \cdot (y - x_0) \} \\ &\geq -\left(2^N - 1\right) \sup_{y \in B_\infty(x_0, \varepsilon)} \left\{ \frac{|f(y) - f(x_0) - \nabla f(x_0) \cdot (y - x_0)|}{\|y - x_0\|_\infty} \|y - x_0\|_\infty \right\} \\ &\geq -\left(2^N - 1\right) \varepsilon \sup_{y \in B_\infty(x_0, \varepsilon)} \left\{ \frac{|f(y) - f(x_0) - \nabla f(x_0) \cdot (y - x_0)|}{\|y - x_0\|_\infty} \right\} \\ &= -\left(2^N - 1\right) \varepsilon o(1), \end{aligned}$$

where in the last identity, we have used the fact that f is differentiable at x_0. Hence if $\|x - x_0\|_\infty = \varepsilon$, then

$$\begin{aligned}
&\frac{g(x) - g(x_0) - \nabla f(x_0) \cdot (x - x_0)}{\|x - x_0\|_\infty} \\
&\geq \frac{\inf_{y \in B_\infty(x_0, \varepsilon)} \{g(y) - g(x_0) - \nabla f(x_0) \cdot (y - x_0)\}}{\varepsilon} \\
&\geq -\left(2^N - 1\right) o(1),
\end{aligned}$$

and so, since $\|\cdot\|_\infty$ is equivalent to $|\cdot|$,

$$\liminf_{x \to x_0} \frac{g(x) - g(x_0) - \nabla f(x_0) \cdot (x - x_0)}{|x - x_0|} \geq 0.$$

This, together with (50), implies that g is differentiable at x_0 and $\nabla g(x_0) = \nabla f(x_0)$.

Proof 2 (suggested by Pietro): Since g is subdifferentiable at x_0 for any $y_0 \in \partial g(x_0)$,

$$g(x) \geq g(x_0) + y_0 \cdot (x - x_0)$$

for all $x \in B$. On the other hand, since f is differentiable at x_0,

$$f(x) = f(x_0) + \nabla f(x_0) \cdot (x - x_0) + o(|x - x_0|),$$

and so

$$\begin{aligned}
&f(x_0) + \nabla f(x_0) \cdot (x - x_0) + o(|x - x_0|) \\
&= f(x) \geq g(x) \geq g(x_0) + y_0 \cdot (x - x_0),
\end{aligned}$$

which implies that

$$o(|x - x_0|) \geq (y_0 - \nabla f(x_0)) \cdot (x - x_0)$$

for all $x \in \mathbb{R}^N$, with $|x - x_0|$ sufficiently small. This implies that $y_0 = \nabla f(x_0)$ (why?). Thus $\partial g(x_0) = \{\nabla f(x_0)\}$. ■

Example 190 *The next two examples show the sharpness of the previous theorem.*

(i) The function $f(x) = f(x_1, x_2) = \sqrt{\exp(-x_1^2) + x_2^2}$ *shows that the condition* $(\operatorname{co} f)(x) < \liminf_{|x| \to \infty} f(x)$ *cannot be eliminated.*

(ii) Note that in general one cannot go beyond the regularity stated in the previous theorem. Indeed, any smooth function $f : \mathbb{R} \to [0, \infty)$ *such that* $\operatorname{co} f = f$ *outside* $[-1, 1]$, $f^{-1}(\{0\}) = \{-1, 1\}$, *and* $f''(\pm 1) > 0$ *shows that* $\operatorname{co} f$ *may not be of class* C^2 *even if* f *is.*

Proof of Theorem 188. Step 1: Let

$$A := \left\{ x \in \mathbb{R}^N : (\operatorname{co} f)(x) < \liminf_{|w| \to \infty} f(w) \right\}$$

If A is nonempty, then

$$\liminf_{|x| \to \infty} f(x) > -\infty,$$

and since f is continuous, we deduce that f must be bounded from below by some constant c. By replacing f with $f - c$, without loss of generality, we may assume that $f \geq 0$.

Step 2: Next we claim that $\operatorname{co} f$ is differentiable in A. Indeed, fix $x_0 \in A$. By Theorem 185, let $\left\{ \left(\theta_i^{(n)}, x_i^{(n)} \right) \right\} \subset [0,1] \times \mathbb{R}^N$, $i = 1, \ldots, N+1$, $n \in \mathbb{N}$, be a minimizing sequence such that

$$\sum_{i=1}^{N+1} \theta_i^{(n)} = 1, \quad \sum_{i=1}^{N+1} \theta_i^{(n)} x_i^{(n)} = x_0,$$

and

$$\sum_{i=1}^{N+1} \theta_i^{(n)} f\left(x_i^{(n)} \right) \to \operatorname{co} f(x_0) \tag{51}$$

as $n \to \infty$.

Upon extracting a subsequence if necessary, for each $i = 1, \ldots, N+1$ we may assume that $\theta_i^{(n)} \to \theta_i$, and that either $\left| x_i^{(n)} \right| \to \infty$ or $x_i^{(n)} \to x_i$ as $n \to \infty$.

Fix

$$(\operatorname{co} f)(x_0) < s < t < \liminf_{|w| \to \infty} f(w),$$

let $0 < \varepsilon_0 < 1$ be so small that

$$t(1 - \varepsilon_0) > s, \tag{52}$$

and find $L > 0$ such that

$$f(x) \geq t \quad \text{for all } |x| \geq L. \tag{53}$$

Define

$$I := \left\{ i = 1, \ldots, N+1 : \left| x_i^{(n)} \right| \geq L \text{ for all } n \text{ large} \right\},$$
$$J := \{ i = 1, \ldots, N+1 \} \setminus I.$$

We claim that there exists $i \in J$ such that $\theta_i \geq \frac{\varepsilon_0}{N+1}$. Indeed, if this is not the case, then $\theta_i < \frac{\varepsilon_0}{N+1}$ for all $i \in J$, and so

$$\sum_{i \in I} \theta_i = 1 - \sum_{i \in J} \theta_i \geq 1 - \varepsilon_0 > 0. \tag{54}$$

By (53), for any $i \in I$ we have that $f\left(x_i^{(n)}\right) \geq t$ for all n sufficiently large, and so, using the fact that $f \geq 0$,

$$\sum_{i=1}^{N+1} \theta_i^{(n)} f\left(x_i^{(n)}\right) \geq \sum_{i \in I} \theta_i^{(n)} f\left(x_i^{(n)}\right) \geq t \sum_{i \in I} \theta_i^{(n)}.$$

Letting $n \to \infty$, by (51) and (54) we get

$$s > \operatorname{co} f\left(x_0\right) \geq t \sum_{i \in I} \theta_i \geq t\left(1-\varepsilon_0\right),$$

which contradicts (52) and proves the claim. Hence, without loss of generality, we may assume that $\theta_1^{(n)} \to \theta_1 \geq \frac{\varepsilon_0}{N+1}$, $x_1^{(n)} \to x_1 \in \overline{B(0, L)}$ as $n \to \infty$. Since $f \geq 0$ we have that

$$\theta_1^{(n)} f\left(x_1^{(n)}\right) \leq \sum_{i=1}^{N+1} \theta_i^{(n)} f\left(x_i^{(n)}\right),$$

and so letting $n \to \infty$ by (51) and using the continuity of f we get

$$\frac{\varepsilon_0}{N+1} f\left(x_1\right) \leq \theta_1 f\left(x_1\right)=\lim_{n \to \infty} \theta_1^{(n)} f\left(x_1^{(n)}\right) \leq \operatorname{co} f\left(x_0\right)<s. \tag{55}$$

This shows that $x_1 \in \operatorname{dom}_e f$, and thus f is differentiable at x_1 by assumption. Note also that $\left\{x_1^{(n)}\right\} \subset \operatorname{dom}_e f$ for all n sufficiently large.

By the convexity of $\operatorname{co} f$ and since for any $h \in \mathbb{R}^N$,

$$x_0+h=\theta_1^{(n)}\left(x_1^{(n)}+\frac{h}{\theta_1^{(n)}}\right)+\sum_{i=2}^{N+1} \theta_i^{(n)} x_i^{(n)},$$

for all n sufficiently large we obtain

$$\begin{aligned}
(\operatorname{co} f)\left(x_0+h\right)-(\operatorname{co} f)\left(x_0\right) \leq & \theta_1^{(n)}(\operatorname{co} f)\left(x_1^{(n)}+\frac{h}{\theta_1^{(n)}}\right) \\
& +\sum_{i=2}^{N+1} \theta_i^{(n)}(\operatorname{co} f)\left(x_i^{(n)}\right)-(\operatorname{co} f)\left(x_0\right) \\
\leq & \theta_1^{(n)}\left[f\left(x_1^{(n)}+\frac{h}{\theta_1^{(n)}}\right)-f\left(x_1^{(n)}\right)\right] \\
& +\left[\sum_{i=1}^{N+1} \theta_i^{(n)} f\left(x_i^{(n)}\right)-(\operatorname{co} f)\left(x_0\right)\right],
\end{aligned}$$

where we have used the fact that $\left\{x_1^{(n)}\right\} \subset \operatorname{dom}_e f$ for all n sufficiently large. Letting $n \to \infty$ in the previous inequality yields

$$(\operatorname{co} f)\left(x_0+h\right)-(\operatorname{co} f)\left(x_0\right) \leq \theta_1\left[f\left(x_1+\frac{h}{\theta_1}\right)-f\left(x_1\right)\right] \tag{56}$$

for all $h \in \mathbb{R}^N$. Since by assumption f is differentiable at x_1, it follows in particular that the right-hand side is finite for all h sufficiently small, say $|h| < r$. In turn, the nonnegative convex function $(\operatorname{co} f)(x_0 + \cdot)$ is finite for the same values of h. Since the left-hand side is a convex function in the variable h, the previous lemma implies that $(\operatorname{co} f)(x_0 + \cdot)$ is differentiable at 0 and $\nabla(\operatorname{co} f)(x_0) = \nabla f(x_1)$.

Thus we have shown that $\operatorname{co} f$ is differentiable in A, and by Theorem 119(i) it follows that $\nabla(\operatorname{co} f)$ is continuous on A.

Step 3: Finally, assume that ∇f is locally Hölder continuous with exponent $0 < \alpha < 1$ or locally Lipschitz in $\operatorname{dom}_e f$ and let U be an open set compactly contained in A. Find $U \subset\subset D \subset\subset A$. By the continuity of $\operatorname{co} f$ and the definition of the set A we may find s,

$$0 < s < \liminf_{|x|\to\infty} f(x),$$

such that $(\operatorname{co} f)(x) < s$ for all $x \in \overline{D}$. Fix

$$s < t < \liminf_{|x|\to\infty} f(x),$$

and let $\varepsilon_0 > 0$ and $L > 0$ be as in (52) and (53). By the previous step, for any $x \in \overline{D}$ we may find $x_1^{(x)} \in \operatorname{dom}_e f \cap \overline{B(0,L)}$ and $\frac{\varepsilon_0}{N+1} \le \theta_1 \le 1$ such that $\nabla(\operatorname{co} f)(x) = \nabla f\left(x_1^{(x)}\right)$ and (55) and (56) hold. We claim that there exists an open set U_1 compactly contained in $\operatorname{dom}_e f$ such that $x_1^{(x)} \in U_1$ for all $x \in \overline{D}$. Indeed, if not, then we may find a sequence $\{x_k\} \subset \overline{D}$ converging to some $x \in \overline{D}$ such that $x_1^{(x_k)} \to x_1 \in \overline{B(0,L)} \setminus \operatorname{dom}_e f$. But by (55),

$$\frac{\varepsilon_0}{N+1} f\left(x_1^{(x_k)}\right) \le (\operatorname{co} f)(x_k) < s.$$

Letting $k \to \infty$ and using the continuity of f we obtain a contradiction since $x_1 \notin \operatorname{dom}_e f$. Hence the claim holds.

Let $U_1 \subset\subset U_2 \subset\subset \operatorname{dom}_e f$ and by hypothesis let $C = C(U_2) > 0$ be such that

$$|\nabla f(x) - \nabla f(w)| \le C|x - w|^\alpha \tag{57}$$

for all $x, w \in U_2$. Let $r > 0$ be so small that $w + \frac{(N+1)h}{\varepsilon_0} \in U_2$ for all $|h| < r$ and all $w \in U_1$.

If $x \in \overline{U}$, then by what we just proved, $x_1^{(x)} \in U_1$, and so $x_1^{(x)} + \frac{h}{\theta_1^{(x)}} \in U_2$ for all $|h| < r$. By the mean value theorem and the fact that $\nabla(\operatorname{co} f)(x) =$

$\nabla f\left(x_1^{(x)}\right)$ we obtain

$$
\begin{aligned}
&(\operatorname{co} f)(x+h)-(\operatorname{co} f)(x)-\nabla(\operatorname{co} f)(x)\cdot h\\
&\qquad\qquad\le \theta_1^{(x)}\left[f\left(x_1^{(x)}+\frac{h}{\theta_1^{(x)}}\right)-f\left(x_1^{(x)}\right)-\nabla f\left(x_1^{(x)}\right)\cdot\frac{h}{\theta_1^{(x)}}\right]\\
&\qquad\qquad= \theta_1^{(x)}\left[\nabla f\left(w_1^{(x,h)}\right)\cdot\frac{h}{\theta_1^{(x)}}-\nabla f\left(x_1^{(x)}\right)\cdot\frac{h}{\theta_1^{(x)}}\right]\\
&\qquad\qquad\le \theta_1^{(x)} C\left|\frac{h}{\theta_1^{(x)}}\right|^{1+\alpha}\le\frac{(N+1)^\alpha}{\varepsilon_0^\alpha}C|h|^{1+\alpha}
\end{aligned}
$$

for some $w_1^{(x,h)}$ on the segment of endpoints $x_1^{(x)}$ and $x_1^{(x)}+\frac{h}{\theta_1^{(x)}}$, and where we have used (56), (57), and the fact that $\theta_1^{(x)}\in\left[\frac{\varepsilon_0}{N+1},1\right]$. Hence also by Theorem 156,

$$
0\le(\operatorname{co} f)(x+h)-(\operatorname{co} f)(x)-\nabla(\operatorname{co} f)(x)\cdot h\le\frac{(N+1)^\alpha}{\varepsilon_0^\alpha}C|h|^{1+\alpha}
$$

for all $|h|<r$. By Remark 107(iii) applied to the convex function

$$
g(h):=(\operatorname{co} f)(x+h)-(\operatorname{co} f)(x)-\nabla(\operatorname{co} f)(x)\cdot h
$$

we obtain that

$$
\begin{aligned}
&|\nabla(\operatorname{co} f)(x+h)-\nabla(\operatorname{co} f)(x)| \\
&\qquad\qquad=|\nabla g(h)|\le\operatorname{Lip}(g;B(0,2|h|))\le\frac{\operatorname{osc}(g;B(0,4|h|))}{2|h|}\\
&\qquad\qquad\le 4^{1+\alpha}\frac{(N+1)^\alpha}{2\varepsilon_0^\alpha}C|h|^\alpha
\end{aligned}\tag{58}
$$

for all $|h|<\frac{1}{4}r$.

Fix $\bar{x}\in\overline{U}$ and let $r^{(\bar{x})}>0$ be so small that $B\left(\bar{x},r^{(\bar{x})}\right)\subset D$ and $r^{(\bar{x})}<\frac{r}{8}$. We claim that $\nabla(\operatorname{co} f)$ is Hölder continuous with exponent $0<\alpha<1$ or Lipschitz in $B\left(\bar{x},r^{(\bar{x})}\right)$. To see this, let x, $w\in B\left(\bar{x},r^{(\bar{x})}\right)$ and write

$$
w=x+h,
$$

where $h:=w-x$ is such that

$$
|h|=|w-x|\le|w-\bar{x}|+|x-\bar{x}|<2\bar{r}<\frac{r}{4}.
$$

By (58),

$$
|\nabla(\operatorname{co} f)(w)-\nabla(\operatorname{co} f)(x)|\,4^{1+\alpha}\frac{(m+1)^\alpha}{2\varepsilon_0^\alpha}C|w-x|^\alpha,
$$

which proves the claim.

Since the family of balls $\left\{B\left(\bar{x}, r^{(\bar{x})}\right)\right\}_{\bar{x}\in\overline{U}}$ is an open cover for the compact set $\overline{U}$, we can find a finite number of balls that still cover $\overline{U}$.

Hence $\nabla \operatorname{co} f$ is locally Hölder continuous with exponent $0 < \alpha < 1$ or locally Lipschitz. ■

3.6 An Application

Theorem 191 *Let E be a Borel subset of $\mathbb{R}^N$ with finite measure, let $1 \leq p \leq \infty$, and let $f : \mathbb{R}^m \to \mathbb{R}$ be a Borel function bounded from below by an affine function. If $z_0 \in \mathbb{R}^m$, then*

$$\inf \left\{ \int_E f(u(x))\, dx : \, u \in L^p(E; \mathbb{R}^m), \frac{1}{|E|} \int_E u(x)\, dx = z_0 \right\} = (\operatorname{co} f)(z_0) |E|,$$

and the infimum is attained if and only if

$$z_0 \in \operatorname{co} M_{z_0},$$

where

$$M_{z_0} := \{ z \in \mathbb{R}^m : \, f(z) = (\operatorname{co} f)(z_0) + \beta \cdot (z - z_0) \text{ for all } \beta \in \partial (\operatorname{co} f)(z_0) \}.$$

Lemma 192 *Let E be a Borel subset of $\mathbb{R}^N$ with positive finite measure, and let $v \in L^1(E; \mathbb{R}^m)$. Then*

$$\frac{1}{\mathcal{L}^N(E)} \int_E v\, dx \in \operatorname{co} \{ v(x) : \, x \in E, \, x \text{ is a Lebesgue point of } v \}.$$

Proof. The proof is by induction on m. Let

$$z_0 := \frac{1}{\mathcal{L}^N(E)} \int_E v\, dx, \quad G := \{ v(x) : \, x \in E, \, x \text{ a Lebesgue point of } v \}.$$

For $m = 1$ it is not difficult to show that $\operatorname{co} G$ is the (possibly infinite) interval of endpoints $\operatorname{essinf}_E v$ and $\operatorname{esssup}_E v$. If $z_0 \notin \operatorname{co} G$, then either $z_0 \geq \operatorname{esssup}_E v$ or $z_0 \leq \operatorname{essinf}_E v$. Assume that $z_0 \geq \operatorname{esssup}_E v$. Then $z_0 - v(x) \geq 0$ for $\mathcal{L}^N$ a.e. $x \in E$, and so since

$$\frac{1}{\mathcal{L}^N(E)} \int_E (z_0 - v(x))\, dx = 0,$$

we deduce that $v(x) = z_0$ for $\mathcal{L}^N$ a.e. $x \in E$. In turn, $G = \{z_0\}$, which contradicts the fact that $z_0 \notin \operatorname{co} G$. The case $z_0 \leq \operatorname{essinf}_E v$ is treated in an analogous way.

Assume that the result is true for functions with values in $\mathbb{R}^{m-1}$ and let $v \in L^1(E; \mathbb{R}^m)$. If $z_0 \notin \operatorname{co} G$, then by Theorem 88 (with $C_1 = \{z_0\}$ and $C_2 = \operatorname{co} G$) we may find a half-space

$$H = \{ z \in \mathbb{R}^m : \, b \cdot (z - z_0) \geq 0 \}$$

through z_0 containing $\operatorname{co} G$, where $b \in \mathbb{R}^m$, $b \neq 0$. Then from the definition of z_0 and G and since $G \subset H$,

$$\begin{aligned} 0 = \int_E b \cdot (v(x) - z_0)\, dx &= \int_{\{y \in E : \, b \cdot (v(y) - z_0) \geq 0\}} b \cdot (v(x) - z_0)\, dx \\ &+ \int_{\{y \in E : \, b \cdot (v(y) - z_0) < 0\}} b \cdot (v(x) - z_0)\, dx. \end{aligned}$$

Hence

$$\int_E |b\cdot(v(x)-z_0)|\,dx = 2\int_{\{y\in E:\, b\cdot(v(y)-z_0)<0\}} |b\cdot(v(x)-z_0)|\,dx = 0,$$

since $v(x)\in H$ for $\mathcal{L}^N$ a.e. $x\in E$. This implies that

$$v(x)\in\{z\in\mathbb{R}^m:\, b\cdot(z-z_0)=0\}$$

for $\mathcal{L}^N$ a.e. $x\in E$, and thus the function v takes values on an $(m-1)$-dimensional hyperplane. By the induction hypothesis we have that $z_0\in\operatorname{co}G$, which is a contradiction. ■

Proof of Theorem 191. We begin by observing that by Theorem 185,

$$\begin{aligned}
&\inf\left\{\int_E f(s)\,dx:\, s\in L^p(E;\mathbb{R}^m),\ s\text{ simple},\ \frac{1}{|E|}\int_E s\,dx = z_0\right\}\\
&\quad=\inf\left\{\sum_{i=1}^n \theta_i f(z_i):\, n\in\mathbb{N},\ \theta_i\in[0,1],\ z_i\in\mathbb{R}^m,\ i=1,\dots,n,\right.\\
&\qquad\left.\sum_{i=1}^n\theta_i=1,\ \sum_{i=1}^n\theta_i z_i = z_0\right\} = (\operatorname{co}f)(z_0)\,|E|,
\end{aligned}$$

and so

$$\inf\left\{\int_E f(u)\,dx:\, u\in L^p(E;\mathbb{R}^m),\ \frac{1}{|E|}\int_E u\,dx = z_0\right\}\le(\operatorname{co}f)(z_0)\,|E|.$$

To prove the opposite inequality, observe that since f is real-valued and bounded from below by an affine function, we have that $\operatorname{co}f:\mathbb{R}^m\to\mathbb{R}$, and thus $\operatorname{co}f$ is subdifferentiable at z_0. Hence for all $\xi\in\partial(\operatorname{co}f)(z_0)$

$$f(z)\ge(\operatorname{co}f)(z)\ge(\operatorname{co}f)(z_0)+\xi\cdot(z-z_0).$$

Taking $z=u(x)$ and integrating over E yields

$$\begin{aligned}
\int_E f(u(x))\,dx &\ge \int_E(\operatorname{co}f)(u(x))\,dx \ge (\operatorname{co}f)(z_0)\,|E| + \xi\cdot\int_E(u(x)-z_0)\,dx \qquad (59)\\
&= (\operatorname{co}f)(z_0)\,|E|,
\end{aligned}$$

where we have used the fact that $\frac{1}{|E|}\int_E u\,dx = z_0$. Hence

$$\inf\left\{\int_E f(v)\,dx:\, v\in L^p(E;\mathbb{R}^m),\ \frac{1}{|E|}\int_E v\,dx = z_0\right\}\ge(\operatorname{co}f)(z_0)\,|E|.$$

Suppose now that the infimum is attained at some function $v\in L^p(E;\mathbb{R}^m)$ with

$$\frac{1}{|E|}\int_E v(x)\,dx = z_0. \qquad (60)$$

By (59) we have

$$\int_E f(v)\,dx = \int_E (\operatorname{co} f)(v)\,dx = (\operatorname{co} f)(z_0)\,|E|\,, \tag{61}$$

and so $f(v(x)) = (\operatorname{co} f)(v(x))$ for $\mathcal{L}^N$ a.e. $x \in E$.

Therefore, given any $\beta \in \partial(\operatorname{co} f)(z_0)$ (recall Theorem **??**), it follows that

$$f(v(x)) = (\operatorname{co} f)(v(x)) \geq (\operatorname{co} f)(z_0) + \beta \cdot (v(x) - z_0)\,,$$

and the inequality must be an equality for $\mathcal{L}^N$ a.e. $x \in E$ or else, in view of (60), (61) would be violated. We deduce that

$$v(x) \in \{z \in \mathbb{R}^m : f(z) = (\operatorname{co} f)(z) = (\operatorname{co} f) + \beta \cdot (z - z_0)\}$$

for $\mathcal{L}^N$ a.e. $x \in E$, which, together with (60) and Lemma 192, yields

$$z_0 \in \operatorname{co}\{z \in \mathbb{R}^m : f(z) = (\operatorname{co} f)(z) = (\operatorname{co} f)(z_0) + \beta \cdot (z - z_0)\}.$$

Conversely, assume that $z_0 \in \operatorname{co} M_{z_0}$ and write

$$z_0 = \sum_{i=1}^{m+1} \theta_i z_i, \tag{62}$$

where $\theta_i \in [0,1]$, $z_i \in M_{z_0}$, $i = 1, \ldots, m+1$, and $\sum_{i=1}^{m+1} \theta_i = 1$. Since the Lebesgue measure is nonatomic, we may find a partition of E into measurable subsets E_i such that

$$|E_i| = \theta_i\,|E|\,,$$

$i = 1, \ldots, m+1$, and define

$$v := \sum_{i=1}^{m+1} \chi_{E_i} z_i.$$

By (62), v is admissible, and for a fixed $\beta \in (\operatorname{co} f)(z_0)$ we have

$$\begin{aligned}\int_E f(v)\,dx &= \sum_{i=1}^{m+1} |E_i|\, f(z_i) = \sum_{i=1}^{m+1} |E_i|\,((\operatorname{co} f)(z_0) + \beta \cdot (z_i - z_0)) \\ &= |E|\,(\operatorname{co} f)(z_0)\,,\end{aligned}$$

where we have used (62) and the fact that $z_i \in M_{z_0}$. ■

3.7 Biconjugate Functions

Proposition 193 *Let $f : \mathbb{R}^N \to [-\infty, \infty]$ be convex and lower semicontinuous. If f takes the value $-\infty$, then $f : \mathbb{R}^N \to \{-\infty, \infty\}$*

Proof. Assume by contradiction that there exists $x_0 \in \mathbb{R}^N$ such that $f(x_0) \in \mathbb{R}$ and let $f(x_1) = -\infty$. Consider the function $g : \mathbb{R} \to [-\infty, \infty]$ defined by

$$g(t) := f(tx_0 + (1-t)x_1), \quad t \in \mathbb{R}.$$

Then g is convex and $g(0) = -\infty$ and $g(1) \in \mathbb{R}$, so by convexity $g(t) = -\infty$ for all $t \in [0,1)$. But since f is lower semicontinuous, then

$$-\infty = \liminf_{t \to 1^-} g(t) \geq g(1),$$

which is a contradiction. ■

Theorem 194 *Let* $f : \mathbb{R}^N \to [-\infty, \infty]$, $f \not\equiv \infty$. *Then*

(i) $f^{**}(x) = \sup\{g(x) : g$ *affine,* $g \leq f\}$ *for all* $x \in \mathbb{R}^N$. *In particular, if* f^{**} *takes the value* $-\infty$, *then* $f^{**} \equiv -\infty$;

(ii) $f^{**} \leq \operatorname{lsc}(\operatorname{co} f) \leq \operatorname{co}(\operatorname{lsc} f) \leq \operatorname{co} f \leq f$;

(iii) *if, in addition, there exists an affine function below* f, *then* $f^{**} = \operatorname{lsc}(\operatorname{co} f)$.

Proof. Step 1: Set

$$\tilde{f}(x) := \sup\{g(x) : g \text{ affine}, g \le f\}, \quad x \in \mathbb{R}^N.$$

We first prove that $\tilde{f} \equiv -\infty$, i.e., the family of admissible functions g in the definition of $\tilde{f}$ is empty, if and only if $f^* \equiv \infty$. Indeed, if there exist $y \in \mathbb{R}^N$ and $\alpha \in \mathbb{R}$ such that

$$y \cdot x + \alpha \le f(x)$$

for every $x \in \mathbb{R}^N$, then, equivalently,

$$y \cdot x - f(x) \le -\alpha$$

for every $x \in \mathbb{R}^N$. Therefore

$$f^*(y) \le -\alpha. \tag{63}$$

Conversely, if there exists $y \in \mathbb{R}^N$ such that $f^*(y) < \infty$, then $f^*(y) \in \mathbb{R}$ since $f \not\equiv \infty$. In view of the definition of $f^*(y)$, it follows that

$$y \cdot x - f^*(y) \le f(x)$$

for every $x \in \mathbb{R}^N$ and thus

$$g(x) := y \cdot x - f^*(y) \le \tilde{f}(x). \tag{64}$$

Step 2: We prove (i). If $f^* \equiv \infty$, then $f^{**} \equiv -\infty$, and by Step 1, property (i) holds. Suppose now that there exists $y \in \mathbb{R}^N$ such that $f^*(y) < \infty$. Taking the supremum in (64) over all such y yields $f^{**}(x) \le \tilde{f}(x)$.

Conversely, by Step 1 there is at least one admissible function $g(x) = y \cdot x + \alpha$ in the definition of $\tilde{f}$. As in (63) we obtain $f^*(y) \le -\alpha$, and we deduce that

$$f^{**}(x) \ge y \cdot x - f^*(y) \ge y \cdot x + \alpha.$$

Taking the supremum over all such pairs (y, α), we conclude that $f^{**}(x) \ge \tilde{f}(x)$.
(ii) The last two inequalities are immediate. Since $\operatorname{co} f \le f$, then $\operatorname{lsc}(\operatorname{co} f) \le \operatorname{lsc} f$, and using the fact that $\operatorname{lsc}(\operatorname{co} f)$ is convex by Proposition 129, we obtain that $\operatorname{lsc}(\operatorname{co} f) \le \operatorname{co}(\operatorname{lsc} f)$.

Since f^{**} is lower semicontinuous, convex, and below f, we have that $f^{**} \le \operatorname{lsc}(\operatorname{co} f)$.
(iii) Since $\operatorname{lsc}(\operatorname{co} f)$ is convex, lower semicontinuous, and above an affine function, then by Proposition **??**, invoking (ii), it follows that

$$\begin{aligned} \operatorname{lsc}(\operatorname{co} f)(x) &= \sup\{g(x) : g \text{ affine}, g \le \operatorname{lsc}(\operatorname{co} f)(x)\} \\ &= (\operatorname{lsc}(\operatorname{co} f))^{**}(x), \end{aligned} \tag{65}$$

where in the last equality we used part (i). Since $\operatorname{lsc}(\operatorname{co} f) \le f$ by (65) we conclude that $\operatorname{lsc}(\operatorname{co} f) \le f^{**}$. ∎

Remark 195 *(i) Note that if there is no affine function below f, then Theorem 194(iii) does not hold in general. Observe that Remark 136 exhibits an example in which $f^{**} \equiv -\infty \lneqq \operatorname{lsc}(\operatorname{co} f) = f$.*

(ii) From Theorem 194 it follows that if there exists an affine function below f and if $\operatorname{co} f$ is lower semicontinuous, then $f^{**} = \operatorname{lsc}(\operatorname{co} f) = \operatorname{co}(\operatorname{lsc} f) = \operatorname{co} f$. In particular, if $f : \mathbb{R}^N \to \mathbb{R}$ admits an affine function g such that $f \geq g$, then

$$-\infty < g \leq f^{**} \leq \operatorname{co}(\operatorname{lsc} f) \leq \operatorname{co} f \leq f < \infty,$$

*and so $\operatorname{co} f : V \to \mathbb{R}$. By Theorem **??** it follows that $\operatorname{co} f$ is continuous, and so $f^{**} = \operatorname{lsc}(\operatorname{co} f) = \operatorname{co}(\operatorname{lsc} f) = \operatorname{co} f$.*

(iii) Note that when $f : \mathbb{R}^N \to \mathbb{R}$ is bounded from below by an affine function g, then $\operatorname{co} f$ is continuous by Corollary 108, and so by (ii) we have that $f^{**} = \operatorname{lsc}(\operatorname{co} f) = \operatorname{co}(\operatorname{lsc} f) = \operatorname{co} f$. However, when f takes the value ∞, then by Theorem 130, $f^{**} = \operatorname{lsc}(\operatorname{co} f)$ agrees with $\operatorname{co} f$ except possibly on $\operatorname{rb}_{\text{aff}}(\operatorname{dom}_e \operatorname{co} f)$, and in particular, it may happen that $f^{**} \lneqq \operatorname{co}(\operatorname{lsc} f)$ on $\operatorname{rb}_{\text{aff}}(\operatorname{dom}_e \operatorname{co} f)$ as shown by the following example.

Exercise 196 *Let $N = 2$ and consider the function*

$$f(x) = f(x_1, x_2) = \begin{cases} x_2 - x_1 e^{x_2} & \text{if } x_2 \geq 0 \text{ and } 0 < x_1 \leq x_2 e^{-x_2}, \\ 0 & \text{if } x_2 \geq 0 \text{ and } x_2 e^{-x_2} < x_1, \\ \infty & \text{otherwise.} \end{cases}$$

Prove that

$$f^{**}(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 \geq 0 \text{ and } x_2 \geq 0, \\ \infty & \text{otherwise,} \end{cases}$$

while

$$\operatorname{co}(\operatorname{lsc} f)(x_1, x_2) = \begin{cases} x_2 & \text{if } x_1 = 0 \text{ and } x_2 \geq 0, \\ 0 & \text{if } x_1 > 0 \text{ and } x_2 \geq 0, \\ \infty & \text{otherwise.} \end{cases}$$

Note that $\operatorname{co}(\operatorname{lsc} f)$ is convex but not lower semicontinuous.

Corollary 197 *Let $f : \mathbb{R}^N \to [-\infty, \infty]$, $f \not\equiv \infty$.*

(i) If f is subdifferentiable at some $x_0 \in \mathbb{R}^N$, then $f(x_0) = f^{**}(x_0)$.

(ii) If $f(x_0) \in \mathbb{R}$ and $f^{**}(x_0) \in \mathbb{R}$ for some $x_0 \in \mathbb{R}^N$, then $\partial f(x_0) \subset \partial f^{**}(x_0)$. Moreover, if $f(x_0) = f^{**}(x_0)$, then $\partial f(x_0) = \partial f^{**}(x_0)$.

Proof. (i) If f is subdifferentiable at x_0, then for every $y_0 \in \partial f(x_0)$,

$$f(x) \geq f(x_0) + y_0 \cdot (x - x_0)$$

for all $x \in \mathbb{R}^N$, and so by (i) and (ii),

$$f(x) \geq f^{**}(x) \geq f(x_0) + y_0 \cdot (x - x_0)$$

for all $x \in \mathbb{R}^N$. In particular, taking $x = x_0$ gives $f(x_0) = f^{**}(x_0)$.

(ii) If $f(x_0) \in \mathbb{R}$ and $f^{**}(x_0) \in \mathbb{R}$ for some $x_0 \in \mathbb{R}^N$, then for any $y_0 \in \partial f(x_0)$, we have If $f(x_0) = f^{**}(x_0)$, then the previous inequality becomes

$$\begin{aligned} f(x) &\geq f^{**}(x) \geq f(x_0) + y_0 \cdot (x - x_0) \\ &\geq f^{**}(x_0) + y_0 \cdot (x - x_0), \end{aligned}$$

which implies that $y_0 \in \partial f^{**}(x_0)$. Hence $\partial f(x_0) \subset \partial f^{**}(x_0)$. On the other hand, if $f(x_0) = f^{**}(x_0)$ and $y_0 \in \partial f^{**}(x_0)$, then

$$\begin{aligned} f(x) &\geq f^{**}(x) \geq f^{**}(x_0) + y_0 \cdot (x - x_0) \\ &= f(x_0) + y_0 \cdot (x - x_0), \end{aligned}$$

and so $y_0 \in \partial f(x_0)$. Hence $\partial f(x_0) = \partial f^{**}(x_0)$. ■

Wednesday, April 23, 2008

3.8 An Application

Theorem 198 *Let $E \subset \mathbb{R}^N$ be a Lebesgue measurable set, let $1 \leq p < \infty$, and let $f : E \times \mathbb{R}^m \to (-\infty, \infty]$ be a Borel function. Assume that there exist a nonnegative function $\gamma \in L^1(E)$ and a constant $C > 0$ such that*

$$f(x, z) \geq -C|z|^p - \gamma(x) \text{ for } \mathcal{L}^N \text{ a.e. } x \in E \text{ and for all } z \in \mathbb{R}^m.$$

Then the functional

$$v \in L^p(E; \mathbb{R}^m) \mapsto \int_E f(x, v(x))\, dx$$

is sequentially lower semicontinuous with respect to weak convergence in $L^p(E; \mathbb{R}^m)$ if and only if

(i) $f(x, \cdot)$ is convex in $\mathbb{R}^m$ for $\mathcal{L}^N$ a.e. $x \in E$;

(ii) there exist two functions $a \in L^1(E)$ and $b \in L^{p'}(E; \mathbb{R}^m)$ such that

$$f(x, z) \geq a(x) + b(x) \cdot z$$

for $\mathcal{L}^N$ a.e. $x \in E$ and all $z \in \mathbb{R}^m$;

(iii) $f(x, \cdot)$ is lower semicontinuous in $\mathbb{R}^m$ for $\mathcal{L}^N$ a.e. $x \in E$.

Proof. We prove only the sufficency part. Thus, assume that (i)-(iii) hold.

Step 1: Suppose first that f is nonnegative. Let $\{v_n\} \subset L^p(E; \mathbb{R}^m)$ be a sequence weakly converging to some $v \in L^p(E; \mathbb{R}^m)$. Without loss of generality we may assume that

$$\liminf_{n\to\infty} \int_E f(x, v_n(x))\, dx = \lim_{n\to\infty} \int_E f(x, v_n(x))\, dx < \infty$$

and that

$$\sup_n \int_E f(x, v_n(x))\, dx < \infty.$$

Define the measures

$$\mu_n(B) := \int_{B\cap E} f(x, v_n(x))\, dx, \quad B \in \mathcal{B}\left(\mathbb{R}^N\right).$$

Then

$$\sup_n \mu_n\left(\mathbb{R}^N\right) < \infty,$$

and so, passing to a subsequence if necessary, there exists a (positive) Radon measure μ such that

$$\mu_n \overset{*}{\rightharpoonup} \mu \quad \text{in } \mathcal{M}\left(\mathbb{R}^N; \mathbb{R}\right) \tag{66}$$

as $n \to \infty$. We claim that

$$\frac{d\mu}{d\mathcal{L}^N}(x_0) \geq \chi_E(x_0) f(x_0, v(x_0)) \qquad \text{for } \mathcal{L}^N \text{ a.e. } x_0 \in \mathbb{R}^N. \tag{67}$$

If (67) holds, then the conclusion of the theorem follows. Indeed, since by the Radon–Nikodym and Lebesgue decomposition theorems

$$\mu = \frac{d\mu}{d\mathcal{L}^N} \mathcal{L}^N + \mu_s,$$

where $\mu_s \geq 0$, by the lower semicontinuity of the norms, we have

$$\lim_{n\to\infty} \int_E f(x, v_n)\, dx = \lim_{n\to\infty} \mu_n \left(\mathbb{R}^N\right) \geq \mu \left(\mathbb{R}^N\right) \geq \int_{\mathbb{R}^N} \frac{d\mu}{d\mathcal{L}^N}\, dx \geq \int_E f(x, v)\, dx.$$

Thus, to conclude the proof of the theorem, it suffices to prove (67) for $\mathcal{L}^N$ a.e. $x_0 \in E$.

By Theorem **??** below there exist two sequences of bounded measurable functions

$$a_i : E \to \mathbb{R}, \quad b_i : E \to \mathbb{R}^m,$$

such that

$$f(x, z) = \sup_{i\in\mathbb{N}} \{a_i(x) + b_i(x) \cdot z\}$$

for $\mathcal{L}^N$ a.e. $x \in E$ and all $z \in \mathbb{R}^m$.

Fix a point $x_0 \in E$ of density one for E that satisfies

$$\frac{d\mu}{d\mathcal{L}^N}(x_0) = \lim_{\varepsilon\to 0^+} \frac{\mu(Q(x_0, \varepsilon) \cap E)}{\varepsilon^N} < \infty$$

and is a Lebesgue point of all the L^1_{loc} functions $a_i(\cdot)\,\chi_E(\cdot)$ and $(b_i(\cdot) \cdot v(\cdot))\,\chi_E(\cdot)$. Choose $\varepsilon_k \searrow 0$ such that $\mu(\partial Q(x_0, \varepsilon_k)) = 0$. Then

$$\begin{aligned}
\frac{d\mu}{d\mathcal{L}^N}(x_0) &= \lim_{k\to\infty} \frac{\mu(Q(x_0, \varepsilon_k))}{\varepsilon_k^N} = \lim_{k\to\infty} \lim_{n\to\infty} \frac{\mu_n(Q(x_0, \varepsilon_k))}{\varepsilon_k^N} \\
&= \lim_{k\to\infty} \lim_{n\to\infty} \frac{1}{\varepsilon_k^N} \int_{Q(x_0,\varepsilon_k)\cap E} f(x, v_n(x))\, dx \\
&\geq \liminf_{k\to\infty} \liminf_{n\to\infty} \frac{1}{\varepsilon_k^N} \int_{Q(x_0,\varepsilon_k)\cap E} (a_i(x) + b_i(x) \cdot v_n(x))\, dx \\
&= \liminf_{k\to\infty} \frac{1}{\varepsilon_k^N} \int_{Q(x_0,\varepsilon_k)\cap E} (a_i(x) + b_i(x) \cdot v(x))\, dx \\
&= a_i(x_0) + b_i(x_0) \cdot v(x_0),
\end{aligned}$$

where we have used the fact that $v_n \rightharpoonup v$ in $L^p(E; \mathbb{R}^m)$. By taking the supremum over all i we conclude that

$$\frac{d\mu}{d\mathcal{L}^N}(x_0) \geq f(x_0, v(x_0))$$

as desired. ■

Proof. Step 2: Since $f(x,z) \geq a(x) + b(x) \cdot z$ for all $\mathcal{L}^N$ a.e. $x_0 \in E$ and for all $z \in \mathbb{R}^m$, we have that the function $f(x,z) - (a(x) + b(x) \cdot z) \geq 0$ has all the properties of the previous step. By Step 1,

$$\begin{aligned}
&\liminf_{n\to\infty} \left(\int_E f(x, v_n)\, dx - \int_E (a + b \cdot v)\, dx \right) \\
&\quad = \liminf_{n\to\infty} \int_E (f(x, v_n) - (a + b \cdot v_n))\, dx \\
&\quad \geq \int_E f(x, v)\, dx - \int_E (a + b \cdot v)\, dx.
\end{aligned}$$

Since $v \in L^p(E; \mathbb{R}^m)$, the result follows. ■

Now let's discuss measurability and boundedness of the functions a_i and b_i. We begin with the case in which f is real-valued.

Theorem 199 *Let $E \subset \mathbb{R}^N$ be a Lebesgue measurable set and let $f : E \times \mathbb{R}^m \to \mathbb{R}$ be a Borel function such that $f(x, \cdot)$ is convex in $\mathbb{R}^m$ for $\mathcal{L}^N$ a.e. $x \in E$. Then there exist measurable functions $a_i : E \to \mathbb{R}$ and $b_i : E \to \mathbb{R}^m$ such that*

$$f(x,z) = \sup_{i\in\mathbb{N}} \{a_i(x) + b_i(x) \cdot z\}$$

for $\mathcal{L}^N$ a.e. $x \in E$ and for all $z \in \mathbb{R}^m$.

Moreover, if f is nonnegative, then the functions a_i and b_i may be taken to be bounded.

Proof. By De Giorgi's theorem, for $\mathcal{L}^N$ a.e. $x \in E$ and for all $z \in \mathbb{R}^m$ we may write

$$f(x,z) = \sup_{i\in\mathbb{N}} \{a_i(x) + b_i(x) \cdot z\},$$

where

$$\begin{aligned}
a_i(x) &:= \int_{\mathbb{R}^m} f(x,z)\left((m+1)\varphi_i(z) + \nabla\varphi_i(z) \cdot z\right) dz, &(68)\\
b_i(x) &:= -\int_{\mathbb{R}^m} f(x,z)\, \nabla\varphi_i(z)\, dz,
\end{aligned}$$

and the functions φ_i are of the form

$$\varphi_i(z) := k_i^m \varphi(k_i(q_i - z)), \quad z \in \mathbb{R}^m,$$

for $k_i \in \mathbb{N}$, $q_i \in \mathbb{Q}^m$, and some $\varphi \in C_c^1(\mathbb{R}^m)$ (see (33)).

To prove the measurability of a_i, note that the nonnegative functions

$$\begin{aligned}
g^+(x,z) &:= (f(x,z)\left((m+1)\varphi_i(z) + \nabla\varphi_i(z) \cdot z\right))^+, \\
g^-(x,z) &:= (f(x,z)\left((m+1)\varphi_i(z) + \nabla\varphi_i(z) \cdot z\right))^-,
\end{aligned}$$

are Borel functions. Hence, by Tonelli's theorem the functions

$$x \in E \mapsto \int_{\mathbb{R}^m} g^+(x,z)\, dz, \quad x \in E \mapsto \int_{\mathbb{R}^m} g^-(x,z)\, dz$$

are measurable. In turn, so is the function

$$x \in E \mapsto a_i(x) := \int_{\mathbb{R}^m} \left[g^+(x,z) - g^-(x,z) \right] dz.$$

Similarly, we can prove that b_i is measurable.

Finally, to prove the last part of the theorem, note that, since f is nonnegative, we may write

$$f(x,z) = \sup_{i \in \mathbb{N}} \{a_i(x) + b_i(x) \cdot z\}^+$$

for $\mathcal{L}^N$ a.e. $x \in E$ and for all $z \in \mathbb{R}^m$. For $k \in \mathbb{N}_0$ define $\sigma_0 :\equiv 0$ and

$$\sigma_k(s) := \begin{cases} 1 & s \le k-1, \\ -s+k & k-1 < s < k, \\ 0 & s > k, \end{cases}$$

and let

$$\phi_{i,k}(x) := \sigma_k(|a_i(x)| + |b_i(x)|).$$

Since $0 \le \phi_{i,k} \le 1$, it follows that

$$(a_i(x) + b_i(x) \cdot z)^+ = \sup_{k \in \mathbb{N}_0} \{\phi_{i,k}(x)\, a_i(x) + \phi_{i,k}(x)\, b_i(x) \cdot z\},$$

for $\mathcal{L}^N$ a.e. $x \in E$ and for all $z \in \mathbb{R}^m$. Note that $\phi_{i,k} a_i$ and $\phi_{i,k} b_i$ are measurable and bounded. ■

When f takes the value ∞, to find a_i and b_i we used the the Moreau-Yosida approximation of f. For every $\varepsilon > 0$, let

$$f_\varepsilon(x,z) := \inf_{y \in \mathbb{R}^m} \left\{ f(x,y) + \frac{1}{2\varepsilon} |z-y|^2 \right\},$$

where $x \in E$ and $z \in \mathbb{R}^m$. In this case to prove the measurability of f_ε we need to use the Aumann measurable selection theorem. We skip the details.

Monday, April 28, 2008

3.9 Duality

We are interested in minimizing problems of the type

$$\inf_{x\in\mathbb{R}^N}\left[f\left(x\right)+g\left(Ax\right)\right],$$

where A is an $N\times M$ matrix, $f:\mathbb{R}^N\to(-\infty,\infty]$, $g:\mathbb{R}^M\to(-\infty,\infty]$. We will see that the dual problem becomes

$$\sup_{y'\in\mathbb{R}^M}\left[-f^*\left(A^Ty'\right)-g^*\left(y'\right)\right],$$

where f^* and g^* are the conjugate functions of f and g and A^T is the transpose matrix of A.

Taking a functional $F:\mathbb{R}^N\to[-\infty,\infty]$, we consider the minimization problem

$$\inf_{x\in\mathbb{R}^N}F\left(x\right). \tag{$\mathcal{P}$}$$

We will write $\inf\mathcal{P}$ for problem $\mathcal{P}$.

We consider a perturbation problem. Consider a function $\Phi:\mathbb{R}^N\times\mathbb{R}^M\to[-\infty,\infty]$ such that

$$\Phi\left(x,0\right)=F\left(x\right) \tag{69}$$

for all $x\in\mathbb{R}^N$. For every fixed $y\in\mathbb{R}^M$ we consider the minimization problem

$$\inf_{x\in\mathbb{R}^N}\Phi\left(x,y\right). \tag{$\mathcal{P}_y$}$$

Note that $\left(\mathcal{P}\right)_0$ is our original problem $(\mathcal{P})$. We now consider the conjugate function $\Phi^*:\Phi$ of Φ, that is,

$$\Phi^*\left(x',y'\right)=\sup_{x\in\mathbb{R}^N,\,y\in\mathbb{R}^M}\left\{x'\cdot x+y'\cdot y-\Phi\left(x,y\right)\right\}.$$

The problem

$$\sup_{y'\in\mathbb{R}^M}\left\{-\Phi^*\left(0,y'\right)\right\} \tag{$\mathcal{P}^*$}$$

is called the *dual problem* of $(\mathcal{P})$. We will write $\sup\mathcal{P}^*$ for $(\mathcal{P}^*)$.

Remark 200 *Note that if we call $G\left(y'\right):=\Phi^*\left(0,y'\right)$, then problem $(\mathcal{P}^*)$ is equivalent to find*

$$\inf_{y'\in\mathbb{R}^M}G\left(y'\right),$$

which is again of the type we started with. In this case, the natural perturbation is $\Phi^*:\mathbb{R}^N\times\mathbb{R}^M\to[-\infty,\infty]$, and so, the perturbed problem of $(\mathcal{P}^*)$ becomes

$$\sup_{y'\in\mathbb{R}^M}\left\{-\Phi^*\left(x',y'\right)\right\},$$

for every fixed $x' \in \mathbb{R}^N$. Hence, we may define a dual problem for $(\mathcal{P}^*)$, the bidual problem *of $(\mathcal{P})$, namely*

$$\inf_{x\in\mathbb{R}^N} \Phi^{**}(x,0).$$

In particular, if $\Phi : \mathbb{R}^N \times \mathbb{R}^M \to (-\infty,\infty]$ is proper, convex, and lower semicontinuous, then $\Phi^{**} = \Phi$, and so the bidual problem of $(\mathcal{P})$ becomes $(\mathcal{P})$. This fact allows us to dualize to $(\mathcal{P}^*)$ any result proved for $(\mathcal{P})$.

Proposition 201 *Let F and Φ be as above. Then*

$$\inf \mathcal{P} \geq \sup \mathcal{P}^*. \tag{70}$$

Proof. Fix $y' \in \mathbb{R}^M$. Since

$$\Phi^*(0,y') = \sup_{x\in\mathbb{R}^N,\, y\in\mathbb{R}^M} \{y'\cdot y - \Phi(x,y)\},$$

for every $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^M$ we have

$$\Phi^*(0,y') \geq y'\cdot y - \Phi(x,y).$$

In particular, taking $y = 0$, we get

$$\Phi^*(0,y') \geq -\Phi(x,0) = -F(x)$$

for all $x \in \mathbb{R}^N$, or, equivalently,

$$F(x) \geq -\Phi^*(0,y')$$

for all $x \in \mathbb{R}^N$. Hence

$$\inf_{x\in\mathbb{R}^N} F(x) \geq -\Phi^*(0,y').$$

Taking the supremum over all $y' \in \mathbb{R}^M$, we get the desired result. ■

In general we can have strict inequality. See Exercise 214 below.

Next we define the auxiliary function

$$H(y) := \inf_{x\in\mathbb{R}^N} \Phi(x,y), \quad y \in \mathbb{R}^M.$$

Note that by (69),

$$H(0) = \inf_{x\in\mathbb{R}^N} \Phi(x,0) = \inf_{x\in\mathbb{R}^N} F(x). \tag{71}$$

Proposition 202 *Let F and Φ be as above. Then*

$$H^*(y') = \Phi^*(0,y'), \quad y' \in \mathbb{R}^M, \tag{72}$$

and

$$H^{**}(0) = \sup \mathcal{P}^*.$$

Proof. For each $y' \in \mathbb{R}^M$ we have

$$\begin{aligned} H^*(y') &= \sup_{y\in\mathbb{R}^M} \{y' \cdot y - H(y)\} \\ &= \sup_{y\in\mathbb{R}^M} \left\{ y' \cdot y - \inf_{x\in\mathbb{R}^N} \Phi(x,y) \right\} \\ &= \sup_{y\in\mathbb{R}^M} \sup_{x\in\mathbb{R}^N} \{y' \cdot y - \Phi(x,y)\} = \Phi^*(0,y'). \end{aligned}$$

Hence

$$\begin{aligned} \sup_{y'\in\mathbb{R}^M} \{-\Phi^*(0,y')\} &= \sup_{y'\in\mathbb{R}^M} \{-H^*(y')\} \\ &= \sup_{y'\in\mathbb{R}^M} \{0 \cdot y' - H^*(y')\} = H^{**}(0). \end{aligned}$$

■

Note that, by (71) and the previous proposition, the inequality (70) reads as

$$H(0) \geq H^{**}(0).$$

Thus we need to find conditions under which $H(0) = H^{**}(0)$.

Proposition 203 *Let F and Φ be as above. Assume that $\Phi : \mathbb{R}^N \times \mathbb{R}^M \to [-\infty,\infty]$ convex. Then $H : \mathbb{R}^M \to [-\infty,\infty]$ is convex.*

Proof. Let $y_1,\ y_2 \in \mathbb{R}^M$ and let $\theta \in (0,1)$. If $H(y_1)$ or $H(y_2)$ is infinite, there is nothing to prove, thus assume that $H(y_1) < \infty$ and $H(y_2) < \infty$ (but they could be $-\infty$). Fix two real numbers $a > H(y_1)$ and $b > H(y_2)$ and, using the definition of H, find $x_1,\ x_2 \in \mathbb{R}^N$ such that

$$H(y_1) \leq \Phi(x_1,y_1) \leq a, \quad H(y_2) \leq \Phi(x_2,y_2) \leq b.$$

Then

$$\begin{aligned} H(\theta y_1 + (1-\theta) y_2) &= \inf_{x\in\mathbb{R}^N} \Phi(x, \theta y_1 + (1-\theta) y_2) \\ &\leq \Phi(\theta x_1 + (1-\theta) x_2, \theta y_1 + (1-\theta) y_2) \\ &\leq \theta \Phi(x_1,y_1) + (1-\theta)\,\theta \Phi(x_2,y_2) \\ &\leq \theta a + (1-\theta) b. \end{aligned}$$

Letting $a \searrow H(y_1)$ and $b \searrow H(y_2)$ we obtain the desired result. ■

Unfortunately, even if Φ is lower semicontinuous, H may not be.

Definition 204 *Problem $(\mathcal{P})$ is said to be*

(i) normal *if $H(0)$ is finite and H is lower semicontinuous at 0;*

(ii) stable *if $H(0)$ is finite and H is subdifferentiable at 0.*

Roughly speaking, these conditions express the fact that the infimum changes only gradually when $(\mathcal{P})$ is perturbed. Similar notions can be defined for problem $(\mathcal{P}^*)$.

Proposition 205 *Let F and Φ be as above. Then $(\mathcal{P})$ is stable if and only if $(\mathcal{P})$ is normal,* $\inf \mathcal{P} = \sup \mathcal{P}^*$*, and $(\mathcal{P}^*)$ admits a solution.*

Proof. Step 1: We claim that the set of solutions of $(\mathcal{P}^*)$ coincides with $\partial H^{**}(0)$. Indeed, $y'' \in \mathbb{R}^M$ is a solution of $(\mathcal{P}^*)$, if and only if

$$-\Phi^*(0, y'') \geq -\Phi^*(0, y')$$

for all $y' \in \mathbb{R}^M$, that is, by (72),

$$-H^*(y'') \geq -H^*(y') = 0 \cdot y' - H^*(y')$$

for all $y' \in \mathbb{R}^M$, which can be written as

$$-H^*(y'') = \sup_{y' \in \mathbb{R}^M} \{0 \cdot y' - H^*(y')\} = H^{**}(0).$$

Since

$$-H^*(y'') + 0 \cdot y'' = H^{**}(0),$$

we have equality in Young's inequality, and this is equivalent to $y'' \in \partial H^{**}(0)$.
Step 2: If $(\mathcal{P})$ is stable, then $H(0)$ is finite and H is subdifferentiable at 0. By Corollary 197, this implies that $H(0) = H^{**}(0) \in \mathbb{R}$ and that $\partial H^{**}(0) = \partial H(0) \neq \emptyset$. Hence by the previous step, $(\mathcal{P}^*)$ has a solution.

Conversely, if $(\mathcal{P})$ is normal, $H(0) = H^{**}(0) \in \mathbb{R}$, and if $(\mathcal{P}^*)$ admits a solution, then by the previous step, $\partial H^{**}(0) \neq \emptyset$, which implies that $\partial H^{**}(0) = \partial H(0)$ by Corollary 197. Hence $(\mathcal{P})$ is stable. ■

Remark 206 *If Φ is convex, then the condition* $\inf \mathcal{P} = \sup \mathcal{P}^*$ *is superfluous in the previous proposition. Indeed, if $(\mathcal{P})$ is normal and $(\mathcal{P}^*)$ admits a solution, then by Proposition 203, the function H is convex (so that* $\operatorname{co} H = H$*), and since it is lower lower semicontinuous at* 0*, we have that* $\operatorname{lsc}(\operatorname{co} H)(0) = \operatorname{lsc} H(0) = H(0)$*. On the other hand, since $\partial H^{**}(0) \neq \emptyset$, by Theorem 194(iv), we have that* $H^{**} = \operatorname{lsc}(\operatorname{co} H)$*, and so $H^{**}(0) = H(0)$.*

Corollary 207 *Let F and Φ be as above. Then $(\mathcal{P}^*)$ is stable if and only if $(\mathcal{P}^*)$ is normal and $(\mathcal{P}^{**})$ admits a solution. In particular, if $\Phi : \mathbb{R}^N \times \mathbb{R}^M \to (-\infty, \infty]$ is proper, convex, and lower semicontinuous, then $(\mathcal{P}^*)$ is stable if and only if $(\mathcal{P}^*)$ is normal and $(\mathcal{P})$ admits a solution.*

The next result gives a sufficient condition for the stability of $(\mathcal{P})$.

Proposition 208 *Let F and Φ be as above. Assume that $\Phi : \mathbb{R}^N \times \mathbb{R}^M \to [-\infty, \infty]$ is convex, that*

$$\inf \mathcal{P} \in \mathbb{R},$$

and that there exists $x_0 \in \mathbb{R}^N$ such that $y \in \mathbb{R}^M \mapsto \Phi(x_0, y)$ is finite and continuous at $y = 0$. Then $(\mathcal{P})$ is stable. In particular, $(\mathcal{P}^*)$ admits a solution.

Proof. By Proposition 203, H is convex, and by hypothesis and (71),

$$H(0) = \inf_{x \in \mathbb{R}^N} F(x) \in \mathbb{R}.$$

Since $y \in \mathbb{R}^M \mapsto \Phi(x_0, y)$ is finite and continuous at $y = 0$ there exists $M > 0$ such that

$$\Phi(x_0, y) \le M$$

for all y in a neighborhood U of 0. Hence

$$H(y) = \inf_{x \in \mathbb{R}^N} \Phi(x, y) \le \Phi(x_0, y) \le M$$

for all $y \in U$, which implies that 0 belongs to the interior of the effective domain of H. By Remark 99, H does not take the value $-\infty$, and so by Theorem 142, H is subdifferentiable at $y = 0$. Thus $(\mathcal{P})$ is stable. ■

Next we study normality.

Proposition 209 *Let F and Φ be as above. Assume that $\Phi : \mathbb{R}^N \times \mathbb{R}^M \to (-\infty, \infty]$ is proper, convex, and lower semicontinuous. Then $(\mathcal{P})$ is normal if and only if*

$$\inf \mathcal{P} = \sup \mathcal{P}^* \in \mathbb{R}. \tag{73}$$

Proof. Assume that $(\mathcal{P})$ is normal and consider the lower semicontinuous envelope $\operatorname{lsc} H$ of H. Then

$$H^{**} \le \operatorname{lsc} H \le H. \tag{74}$$

Since H is lower semicontinuous at 0, we have that $(\operatorname{lsc} H)(0) = H(0) \in \mathbb{R}$. Since the function $\operatorname{lsc} H$ is convex, lower semicontinuous and finite at some point, it never takes the value $-\infty$. Hence by Theorem 194, $H^{**} = \operatorname{lsc} H$. In particular,

$$H^{**}(0) = (\operatorname{lsc} H)(0) = H(0) \in \mathbb{R},$$

which is exactly (73).

Conversely, if (73), then $H^{**}(0) = H(0) \in \mathbb{R}$. It follows by (74), that $H^{**}(0) = (\operatorname{lsc} H)(0) = H(0) \in \mathbb{R}$, which implies that H is lower semicontinuous at 0. Thus $(\mathcal{P})$ is normal. ■

Corollary 210 *Let F and Φ be as above. Assume that $\Phi : \mathbb{R}^N \times \mathbb{R}^M \to (-\infty, \infty]$ is proper, convex, and lower semicontinuous. Then the following three conditions are equivalent to each other:*

(i) $(\mathcal{P})$ *and* $(\mathcal{P}^*)$ *are normal and have some solutions.*

(ii) $(\mathcal{P})$ *and* $(\mathcal{P}^*)$ *are stable.*

(iii) $(\mathcal{P})$ *is stable and has some solutions.*

We now give some extremality relations.

Theorem 211 *Let F and Φ be as above. If $(\mathcal{P})$ and $(\mathcal{P}^*)$ have solutions and*

$$\inf \mathcal{P} = \sup \mathcal{P}^* \in \mathbb{R},$$

then for any solution $x_0 \in \mathbb{R}^N$ of $(\mathcal{P})$ and any solution $y_0' \in \mathbb{R}^M$ of $(\mathcal{P}^*)$ we have the extremality relation

$$\Phi(x_0, 0) + \Phi^*(0, y_0') = 0, \tag{75}$$

or, equivalently,

$$(0, y_0') \in \partial\Phi(x_0, 0).$$

Conversely, if $x_0 \in \mathbb{R}^N$ and $y_0' \in \mathbb{R}^M$ satisfy (75), then x_0 is a solution of $(\mathcal{P})$ and y_0' is a solution of $(\mathcal{P}^*)$ and $\inf \mathcal{P} = \sup \mathcal{P}^* \in \mathbb{R}$.

Proof. Assume that $x_0 \in \mathbb{R}^N$ is a solution of $(\mathcal{P})$ and $y_0' \in \mathbb{R}^M$ is a solution of $(\mathcal{P}^*)$ and (73) holds. Then

$$\begin{aligned}\Phi(x_0, 0) = F(x_0) &= \inf_{x \in \mathbb{R}^N} F(x) \\ &= \sup_{y' \in \mathbb{R}^M} \{-\Phi^*(0, y')\} = -\Phi^*(0, y_0') \in \mathbb{R},\end{aligned}$$

which can be written as

$$\Phi(x_0, 0) + \Phi^*(0, y_0') = 0 = 0 \cdot x_0 + y_0' \cdot 0$$

It now suffices to remark that equality in Young's inequality is equivalent to $(0, y_0') \in \partial\Phi(x_0, 0)$.

Conversely, if $x_0 \in \mathbb{R}^N$ and $y_0' \in \mathbb{R}^M$ satisfy (75), then, since

$$\Phi(x, 0) \geq -\Phi^*(0, y')$$

for all $x \in \mathbb{R}^N$ and all $y' \in \mathbb{R}^M$ by (70), taking $y' = y_0'$ and using (75) gives

$$\Phi(x, 0) \geq -\Phi^*(0, y_0') = \Phi(x_0, 0)$$

for all $x \in \mathbb{R}^N$, that is,

$$F(x_0) = \Phi(x_0, 0) = \min_{x \in \mathbb{R}^N} \Phi(x, 0) = \min_{x \in \mathbb{R}^N} F(x),$$

which implies that x_0 is a solution of $(\mathcal{P})$.

Similarly, taking $x = x_0$ gives

$$\Phi(x_0, 0) = -\Phi^*(0, y_0') \geq -\Phi^*(0, y')$$

for all $x \in \mathbb{R}^N$ and all $y' \in \mathbb{R}^M$, which implies that

$$\max_{y' \in \mathbb{R}^M} \{-\Phi^*(0, y')\} = -\Phi^*(0, y_0'),$$

that is, y_0' is a solution of $(\mathcal{P}^*)$.

Finally, using (70) once more,

$$\Phi(x_0, 0) = \min_{x \in \mathbb{R}^N} \Phi(x, 0) \geq \max_{y' \in \mathbb{R}^M} \{-\Phi^*(0, y')\} = -\Phi^*(0, y_0'),$$

and so all inequalities are actually inequality. ■

We now discuss some important special cases. Let A be an $N \times M$ matrix and denote by A^T its transpose. Assume that

$$F(x) = J(x, Ax), \tag{76}$$

where $J : \mathbb{R}^N \times \mathbb{R}^M \to [-\infty, \infty]$. In this case the minimization problem problem $(\mathcal{P})$ becomes

$$\inf_{x \in \mathbb{R}^N} J(x, Ax).$$

A natural choice for the function $\Phi : \mathbb{R}^N \times \mathbb{R}^M \to [-\infty, \infty]$ is

$$\Phi(x, y) := J(x, Ax - y).$$

Note that if J is convex, then so is Φ, and if J is lower semicontinuous, then so is Φ.

To calculate the dual problem, let J^* be the conjugate of J. Then for all $y' \in \mathbb{R}^M$. Since

$$\begin{aligned}
\Phi^*(0, y') &= \sup_{x \in \mathbb{R}^N,\, y \in \mathbb{R}^M} \{y' \cdot y - J(x, Ax - y)\} \\
&= \sup_{x \in \mathbb{R}^N} \sup_{y \in \mathbb{R}^M} \{y' \cdot y - J(x, Ax - y)\}.
\end{aligned}$$

For any fixed $x \in \mathbb{R}^N$, set $w := Ax - y$. Then

$$\begin{aligned}
\Phi^*(0, y') &= \sup_{x \in \mathbb{R}^N} \sup_{w \in \mathbb{R}^M} \{y' \cdot Ax - y' \cdot w - J(x, w)\} \\
&= \sup_{x \in \mathbb{R}^N,\, w \in \mathbb{R}^M} \left\{\left(A^T y'\right) \cdot x - y' \cdot w - J(x, w)\right\} \\
&= J^*\left(A^T y', -y'\right).
\end{aligned}$$

Thus the problem $(\mathcal{P}^*)$ becomes

$$\sup_{y' \in \mathbb{R}^M} -J^*\left(A^T y', -y'\right). \tag{77}$$

Proposition 208 now becomes

Proposition 212 *Let L and J be as above. Assume that $J : \mathbb{R}^N \times \mathbb{R}^M \to [-\infty, \infty]$ is convex, that*

$$\inf_{x \in \mathbb{R}^N} J(x, Ax) \in \mathbb{R},$$

and that there exists $x_0 \in \mathbb{R}^N$ such that $y \in \mathbb{R}^M \mapsto J(x_0, y)$ is finite and continuous at $L(x_0)$. Then

$$\inf \mathcal{P} = \sup \mathcal{P}^*$$

and problem $(\mathcal{P}^*)$ has a solution.

Proof. This follows from Proposition 205. ■

Theorem 211 reduces to:

Theorem 213 *Let L and J be as above. If $(\mathcal{P})$ and $(\mathcal{P}^*)$ have solutions and*

$$\inf \mathcal{P} = \sup \mathcal{P}^* \in \mathbb{R},$$

then for any solution $x_0 \in \mathbb{R}^N$ of $(\mathcal{P})$ and any solution $y_0' \in \mathbb{R}^M$ of $(\mathcal{P}^*)$ we have the extremality relation

$$J(x_0, Ax_0) + J^*\left(A^T y_0', -y_0'\right) = 0, \tag{78}$$

or, equivalently,

$$\left(A^T y_0', -y_0'\right) \in \partial J(x_0, Ax_0).$$

Conversely, if $x_0 \in \mathbb{R}^N$ and $y_0' \in \mathbb{R}^M$ satisfy (78), then x_0 is a solution of $(\mathcal{P})$ and y_0' is a solution of $(\mathcal{P}^*)$ and $\inf \mathcal{P} = \sup \mathcal{P}^* \in \mathbb{R}$.

A particular case of (76) is given when J is a sum of two functions, that is

$$F(x) = f(x) + g(Ax),$$

where as before A is an $N \times M$ matrix and

$$f : \mathbb{R}^N \to [-\infty, \infty], \quad g : \mathbb{R}^M \to [-\infty, \infty].$$

In this case the minimization problem $(\mathcal{P})$ becomes

$$\inf_{x \in \mathbb{R}^N} [f(x) + g(Ax)].$$

Since $J(x, y) = f(x) + g(y)$, if f and g are convex, then J is convex, and if f and g are proper, convex, and lower semicontinuous, then so is J. Moreover,

$$\begin{aligned} J^*(x', y') &= \sup_{x \in \mathbb{R}^N,\, y \in \mathbb{R}^M} \{x' \cdot x + y' \cdot y - f(x) - g(y)\} \\ &= f^*(x') + g^*(y'), \end{aligned}$$

so that the dual problem $(\mathcal{P}^*)$ becomes

$$\sup_{y' \in \mathbb{R}^M} \left[-f^*\left(A^T y'\right) - g^*(-y')\right],$$

where f^* and g^* are the conjugate functions of f and g respectively.

Exercise 214 *Find* $\inf \mathcal{P}$ *and* $\sup \mathcal{P}^*$ *for the following:*

(i) Let A be the identity matrix and let $f : \mathbb{R}^2 \to (-\infty, \infty]$ and $g : \mathbb{R}^2 \to (-\infty, \infty]$.be defined by

$$f(x, y) := \begin{cases} 0 & \text{if } x = 0, \\ \infty & \text{otherwise,} \end{cases} \qquad g(x, y) := \begin{cases} -\min\left\{1, \sqrt{xy}\right\} & \text{if } x \geq 0 \text{ and } y \geq 0, \\ \infty & \text{otherwise.} \end{cases}$$

(ii) Let A be the identity matrix and let $f : \mathbb{R}^2 \to (-\infty, \infty]$ and $g : \mathbb{R}^2 \to (-\infty, \infty]$.be defined by

$$f(x, y) := \begin{cases} 0 & \text{if } x = 0, \\ \infty & \text{otherwise,} \end{cases} \qquad g(x, y) := \begin{cases} -\sqrt{xy} & \text{if } x \geq 0 \text{ and } y \geq 0, \\ \infty & \text{otherwise.} \end{cases}$$

In this case Proposition 208 now becomes

Theorem 215 *Let A be an $N \times M$ matrix. Consider two convex functionals*

$$f : \mathbb{R}^N \to [-\infty, \infty], \quad g : \mathbb{R}^M \to [-\infty, \infty].$$

Assume that

$$\inf_{x \in \mathbb{R}^N} [f(x) + g(Ax)] \in \mathbb{R},$$

and that there exists $x_0 \in \mathbb{R}^N$ such that $f(x_0) \in \mathbb{R}$, $g(L(x_0)) \in \mathbb{R}$ and g is continuous at $L(x_0)$. Then

$$\inf \mathcal{P} = \sup \mathcal{P}^*,$$

and $(\mathcal{P}^*)$ admits at least a solution.

Concerning the extremality relation (78), we have

$$\begin{aligned}0 &= J\left(x_0, Ax_0\right) + J^*\left(A^T y_0', -y_0'\right)\\ &= f\left(x_0\right) + g\left(Ax_0\right) + f^*\left(A^T y_0'\right) + g^*\left(-y_0'\right)\\ &= \left[f\left(x_0\right) + f^*\left(A^T y_0'\right) - A^T y_0' \cdot x_0\right]\\ &+ \left[g\left(Ax_0\right) + g^*\left(-y_0'\right) - \left(-y_0' \cdot Ax_0\right)\right]\end{aligned}$$

and since the expressions in square brackets are nonnegative, we obtain the two extremality relations

$$f\left(x_0\right) + f^*\left(A^T y_0'\right) - A^T y_0' \cdot x_0 = 0, \tag{79}$$

$$g\left(Ax_0\right) + g^*\left(-y_0'\right) - \left(-y_0' \cdot Ax_0\right) = 0, \tag{80}$$

which are equivalent to

$$A^T y_0' \in \partial f\left(x_0\right), \quad -y_0' \in \partial g\left(Ax_0\right).$$

3.10 An example

All the results done in the previous section can be extended to the infinite-dimensional setting. It suffices to replace $\mathbb{R}^N$ and $\mathbb{R}^M$ with two topological vector spaces V and Y, and the inner products $x' \cdot x$ and $y' \cdot y$ with the duality pairings $\langle v', v\rangle_{V',V}$ and $\langle y', y\rangle_{Y',Y}$. We present here an application.

Let $\Omega \subset \mathbb{R}^N$ be an open bounded set with Lipschitz boundary. Let $g : \mathbb{R}^N \to [0, \infty)$ be a strictly convex function such that

$$0 \leq g(z) \leq C\left(1 + |z|^p\right)$$

for $z \in \mathbb{R}^N$ and for some $C > 0$ and $p > 1$. Given function $f \in L^q(\Omega)$, where $\frac{1}{p} + \frac{1}{q} = 1$, we are interested in the following minimization problem

$$\inf\left\{\int_\Omega g\left(\nabla v(x)\right)\, dx - \int_\Omega v(x)\, f(x)\, dx : \, v \in W_0^{1,p}(\Omega)\right\}.$$

Note that if $u \in W^{1,p}(\Omega)$ is a solution of this problem, then for any $\varphi \in C_c^\infty(\Omega)$ the function

$$\psi(t) := \int_\Omega g\left(\nabla u(x) + t\nabla\varphi(x)\right)\, dx - \int_\Omega \left(u(x) + t\varphi(x)\right) f(x)\, dx$$

has a minimum in $t = 0$, and so, if g is of class $C^1\left(\mathbb{R}^N\right)$ by differentiating under the integral sign (why can we do it?), we get

$$0 = \psi'(0) = \int_\Omega \nabla g\left(\nabla u(x)\right) \cdot \nabla\varphi(x)\, dx - \int_\Omega \varphi(x)\, f(x)\, dx.$$

If u and g are more regular, then we can integrate by parts to get

$$\begin{aligned}
0 &= -\int_{\Omega} \operatorname{div}\left(\nabla g\left(\nabla u\left(x\right)\right)\right)\varphi\left(x\right)\,dx \\
&\quad + \int_{\partial\Omega} \nabla g\left(\nabla u\left(x\right)\right)\cdot\nu\left(x\right)\varphi\left(x\right)\,d\mathcal{H}^{N-1} - \int_{\Omega}\varphi\left(x\right)f\left(x\right)\,dx \\
&= -\int_{\Omega}\left[\operatorname{div}\left(\nabla g\left(\nabla u\left(x\right)\right)\right) + f\left(x\right)\right]\varphi\left(x\right)\,dx,
\end{aligned}$$

where we have used the fact that $\varphi = 0$ on $\partial\Omega$. Since this is true for all $\varphi \in C_c^{\infty}\left(\Omega\right)$, we get that u solves the Dirichlet problem

$$\begin{cases} -\operatorname{div}\left(\nabla g\left(\nabla u\right)\right) = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases}$$

If the function g is not of class C^1, say,

$$g\left(z\right) := \left|z\right|^p + c\left|z\right|,$$

then we are in trouble.

To find the dual problem, we take $V = W_0^{1,p}\left(\Omega\right)$, $Y = L^p\left(\Omega;\mathbb{R}^N\right)$, and consider the linear functional

$$\begin{aligned} L : W^{1,p}\left(\Omega\right) &\to L^p\left(\Omega;\mathbb{R}^N\right) \\ v &\mapsto \nabla v \end{aligned}$$

and the convex functionals

$$F : W^{1,p}\left(\Omega\right) \to \mathbb{R}, \quad G : L^p\left(\Omega;\mathbb{R}^N\right) \to [0,\infty)$$

defined by

$$F\left(v\right) := -\int_{\Omega} v\left(x\right)f\left(x\right)\,dx \quad G\left(z\right) := \int_{\Omega} g\left(z\right)\,dx.$$

Hence, the minimization problem $(\mathcal{P})$ becomes

$$\inf_{v\in W_0^{1,p}(\Omega)}\left[\int_{\Omega} g\left(\nabla v\right)\,dx - \int_{\Omega} vf\,dx\right] = \inf_{v\in W_0^{1,p}(\Omega)}\left[F\left(v\right) + G\left(L\left(v\right)\right)\right].$$

The transpose of L is

$$L^T : L^q\left(\Omega;\mathbb{R}^N\right) \to W^{-1,q}\left(\Omega\right)$$

The dual problem $(\mathcal{P}^*)$ is given by

$$\sup_{z^*\in L^q(\Omega;\mathbb{R}^2)}\left[-F^*\left(L^T\left(z^*\right)\right) - G^*\left(-z^*\right)\right].$$

It is well-known that

$$G^*(-z^*) = \int_\Omega g^*(-z^*)\,dx,$$

while

$$\begin{aligned}
F^*(L^*(z^*)) &= \sup_{v\in W_0^{1,p}(\Omega)} \left\{ \left\langle L^T(z^*), v\right\rangle_{W^{-1,q}(\Omega),W^{1,p}(\Omega)} - F(v) \right\} \\
&= \sup_{v\in W_0^{1,p}(\Omega)} \left\{ \langle z^*, L(v)\rangle_{W^{-1,q}(\Omega),W^{1,p}(\Omega)} - F(v) \right\} \\
&= \sup_{v\in W_0^{1,p}(\Omega)} \left\{ \int_\Omega z^*\cdot\nabla v\,dx - \int_\Omega vf\,dx \right\}.
\end{aligned}$$

If z^* is sufficiently regular, we can use the divergence theorem to get

$$\begin{aligned}
\int_\Omega z^*\cdot\nabla v\,dx &= -\int_\Omega v\,\mathrm{div}\,z^*\,dx + \int_{\partial\Omega} vz^*\cdot\nu\,d\mathcal{H}^{N-1} \\
&= -\int_\Omega v\,\mathrm{div}\,z^*\,dx,
\end{aligned}$$

where we have used the fact that $v=0$ on $\partial\Omega$. Hence,

$$\begin{aligned}
F^*(L^*(z^*)) &= \sup_{v\in W_0^{1,p}(\Omega)} \left\{ \int_\Omega v\,(\mathrm{div}\,z^* + f)\,dx \right\} \\
&= \begin{cases} 0 & \text{if } \mathrm{div}\,z^* = -f \text{ in } \Omega, \\ \infty & \text{otherwise.} \end{cases}
\end{aligned}$$

Therefore problem $(\mathcal{P}^*)$ reduces to

$$\sup\left\{ -\int_\Omega g^*(-z^*)\,dx : z^* \in L^q(\mathrm{div};\Omega), \right. \tag{81}$$
$$\left. \mathrm{div}\,z^* = -f \text{ in } \Omega \right\},$$

where

$$L^q(\mathrm{div};\Omega) := \left\{ z^* \in L^q\left(\Omega;\mathbb{R}^N\right) : \mathrm{div}\,z^* \in L^q(\Omega) \right\}.$$

Taking $v_0 = 0$ we have that such that $F(0) = 0 \in \mathbb{R}$, $G(L(0)) = g(0)\,\mathcal{L}^N(\Omega) \in \mathbb{R}$ and G is continuous at 0. Hence, if $\inf\mathcal{P} > -\infty$, then

$$\inf\mathcal{P} = \sup\mathcal{P}^*,$$

and $(\mathcal{P}^*)$ admits at least a solution z_0^*.

To see what equation is solved by z_0^*, note that for any $\varphi \in C_c^\infty\left(\Omega;\mathbb{R}^N\right)$ with $\mathrm{div}\,\varphi = 0$, the function

$$\psi(t) := \int_\Omega g^*(-z_0^*(x) - t\varphi(x))\,dx$$

has a minimum in $t = 0$, and so, by differentiating under the integral sign (why can we do it?), we get

$$0 = \psi'(0) = \int_\Omega \nabla g^*(-z_0^*(x)) \cdot \varphi(x)\, dx.$$

Let $\phi \in C_c^\infty\left(\Omega; \mathbb{R}^N\right)$ and consider

$$\varphi(x) = \left(0, \ldots, \underset{i\text{-th}}{-\frac{\partial \phi}{\partial x_j}}, 0, \ldots, \underset{j\text{-th}}{-\frac{\partial \phi}{\partial x_i}}, 0, \ldots, 0\right).$$

Then $\operatorname{div} \varphi = -\frac{\partial^2 \phi}{\partial x_i \partial x_j} + \frac{\partial^2 \phi}{\partial x_i \partial x_i} = 0$, and so

$$0 = \int_\Omega \left[-\frac{\partial g^*}{\partial z_i}(-z_0^*(x)) \frac{\partial \phi}{\partial x_j}(x) + \frac{\partial g^*}{\partial z_j}(-z_0^*(x)) \frac{\partial \phi}{\partial x_i}(x)\right] dx.$$

If z_0^* and g^* are more regular, then we can integrate by parts to get

$$\begin{aligned} 0 &= \int_\Omega \left[-\frac{\partial g^*}{\partial z_i}(-z_0^*(x)) \frac{\partial \phi}{\partial x_j}(x) + \frac{\partial g^*}{\partial z_j}(-z_0^*(x)) \frac{\partial \phi}{\partial x_i}(x)\right] \\ &= \int_\Omega \left[\frac{\partial}{\partial x_j}\left(\frac{\partial g^*}{\partial z_i}(-z_0^*(x))\right) - \frac{\partial}{\partial x_i}\left(\frac{\partial g^*}{\partial z_j}(-z_0^*(x))\right)\right] \phi(x)\, dx, \end{aligned}$$

where we have used the fact that $\phi = 0$ on $\partial\Omega$. Since this is true for all $\phi \in C_c^\infty(\Omega)$, we get

$$\frac{\partial}{\partial x_j}\left(\frac{\partial g^*}{\partial z_i}(-z_0^*(x))\right) = \frac{\partial}{\partial x_i}\left(\frac{\partial g^*}{\partial z_j}(-z_0^*(x))\right),$$

which says that

$$\operatorname{curl}(\nabla g^*(-z_0^*(x))) = 0.$$

If Ω is simply connected, this implies that

$$\nabla g^*(-z_0^*(x)) = \nabla u_1(x)$$

for some function $u_1 \in W^{1,p}(\Omega)$.