Convexity Notes

John A. Gubner
Department of Electrical and Computer Engineering
University of Wisconsin–Madison

Abstract

Basic results about convex functions on finite and infinite-dimensional spaces are derived. Much of the text assumes that the reader is familiar with advanced calculus and the projection theorem for Hilbert space. However, the section on subgradients can be read without such knowledge and without reading any of the earlier sections.

If you find this writeup useful, or if you find typos or mistakes, please let me know at John.Gubner@wisc.edu

Contents

1. The Epigraph

1.1. Motivation

Let X be a real inner-product space. If $C \subset X$, then we say C is a **convex set** if for every pair of points $x, y \in C$ and every $0 \leq \lambda \leq 1$, the **convex combination**

$\lambda x + (1-\lambda)y$ also belongs to C. By writing the convex combination as $y + \lambda(x-y)$, we see that as λ goes from zero to one, the convex combination traces out a straight line starting at y and ending at x.

When C is a nonempty convex set, a real-valued function $f_0: C \to \mathbb{R}$ is said to be a **convex function** if

$$f_0\big(\lambda x + (1-\lambda)y\big) \leq \lambda f_0(x) + (1-\lambda) f_0(y) \tag{1}$$

for all $x, y \in C$ and all $0 \leq \lambda \leq 1$. The convexity of the set C is used to guarantee that whenever x and y are valid arguments for f_0, the convex combination $\lambda x + (1-\lambda)y$ is also a valid argument of f_0. The inequality (1) says when looking at the graph of f_0, the line segment joining the points $(x, f_0(x))$ and $(y, f_0(y))$ lies above the function values when the function is evaluated along the line joining x and y. This is illustrated in Figure 1 for the function $f_0(x) := -\ln x$ for $x > 0$.

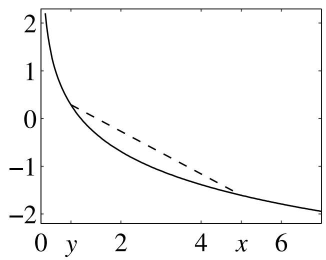

Figure 1. Illustration of (1).

Starting with a real-valued convex function f_0 on a nonempty convex set C, let us define

$$f(x) := \begin{cases} f_0(x), & x \in C, \\ \infty, & \text{otherwise.} \end{cases} \tag{2}$$

We claim that

$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda) f(y) \tag{3}$$

holds for all $x, y \in X$ and all $0 \leq \lambda \leq 1$. This is easy to see because if $x, y \in C$, then (3) reduces to (1). Otherwise, if either x or y is not in C, then (3) holds automatically because the right-hand side is infinite if $0 < \lambda < 1$. If λ is zero or one, then (3) is trivially true for all $x, y \in X$.

Since every real-valued convex function on a nonempty convex subset of X can be extended to a $(-\infty, \infty]$-valued function on X that satisfies (3), we can dispense with the set C and discuss functions $f: X \to (-\infty, \infty]$ that satisfy (3). Note, however, that this admits the new possibility $f(x) = \infty$ for all $x \in X$ that was not present before.

Definition 1.1. A function $f: X \to (-\infty, \infty]$ that is not everywhere equal to ∞, is said to be **proper**.

For a convex function $f: X \to (-\infty, \infty]$, its **conjugate function**[a] is

$$f^*(y) := \sup_{x \in X} \big[\langle x, y\rangle - f(x)\big], \quad y \in X.$$

For fixed y, consider the function of x, $\langle x, y\rangle - f(x)$, that is being maximized. If f is proper, then there is some x_0 with $f(x_0) \in \mathbb{R}$. In this case,

$$f^*(y) \geq \langle x_0, y\rangle - f(x_0) = \text{real number},$$

which implies that $f^*: X \to (-\infty, \infty]$. It also implies

$$f^*(y) = \sup_{x: f(x) < \infty} \big[\langle x, y\rangle - f(x)\big], \quad y \in X.$$

In fact, $f^*(y)$ is a convex function of y. Write

$$\begin{aligned}
f^*\big(\lambda y_1 + (1-\lambda) y_2\big) &= \sup_{x: f(x) < \infty} \big[\big\langle x, \lambda y_1 + (1-\lambda) y_2\big\rangle - f(x)\big] \\
&= \sup_{x: f(x) < \infty} \big[\lambda\{\langle x, y_1\rangle - f(x)\} + (1-\lambda)\{\langle x, y_2\rangle - f(x)\}\big] \\
&\leq \lambda \sup_{x: f(x) < \infty} \big[\langle x, y_1\rangle - f(x)\big] + (1-\lambda) \sup_{x: f(x) < \infty} \big[\langle x, y_2\rangle - f(x)\big] \\
&= \lambda f^*(y_1) + (1-\lambda) f^*(y_2).
\end{aligned}$$

Example 1.2. Consider the convex function $f(x) := -\ln x$ for $x > 0$ and $f(x) := \infty$ for $x \leq 0$. Then

$$f^*(y) = \sup_{x > 0} \big[xy + \ln x\big].$$

For $y = 0$, it is clear that $f^*(y) = \infty$. For $y > 0$, write $xy + \ln x = x\big[y + \frac{\ln x}{x}\big]$. Letting $x \to \infty$ shows $xy + \ln x \approx xy \to \infty$. For $y < 0$, calculus tells us the optimal $x = -1/y > 0$, and then $f^*(y) = -1 - \ln(-y)$. Thus,

$$f^*(y) = \begin{cases} \infty, & y \geq 0, \\ -1 - \ln(-y), & y < 0. \end{cases}$$

---

[a] In large deviations, the conjugate function is often called the **Fenchel–Legendre transform** [3].

Now suppose $f(x) = \infty$ for all x. Then $f^*(y) = -\infty$ for all y. This suggests that we start looking at functions $f\colon X \to [-\infty, \infty]$. This creates a problem if we continue trying to use (3) to define convex functions because if $f(x) = \infty$ and $f(y) = -\infty$, we may be confronted with the expression $\infty - \infty$, which is not defined! One way to avoid this problem is to introduce the **epigraph** of f.

1.2. Definitions

If $f\colon X \to [-\infty, \infty]$, then the **effective domain** of f is

$$\operatorname{dom} f := \{x \in X : f(x) < \infty\}.$$

Hence, $x \in \operatorname{dom} f$ if and only if the value of $f(x)$ is either finite or $-\infty$. It follows that $\operatorname{dom} f = \varnothing \Leftrightarrow f \equiv \infty$. The **epigraph** of f is

$$\operatorname{epi} f := \{(x,t) \in X \times \mathbb{R} : f(x) \le t\}.$$

If $(x,t) \in \operatorname{epi} f$, then $f(x) \le t < \infty$, and if $x \in \operatorname{dom} f$, then for some finite t, $(x,t) \in \operatorname{epi} f$. Thus,

$$\operatorname{epi} f = \varnothing \Leftrightarrow \operatorname{dom} f = \varnothing \Leftrightarrow f \equiv \infty. \tag{4}$$

Definition 1.3 (Convex Function). A function $f\colon X \to [-\infty, \infty]$ is **convex** if $\operatorname{epi} f$ is a convex subset of $X \times \mathbb{R}$; i.e., if for all $0 \le \lambda \le 1$ and all (x,t) and (y,s) in $\operatorname{epi} f$,

$$\lambda(x,t) + (1-\lambda)(y,s) = \big(\lambda x + (1 - \lambda y), \lambda t + (1-\lambda)y\big) \in \operatorname{epi} f,$$

or equivalently,

$$f\big(\lambda x + (1-\lambda)y\big) \le \lambda t + (1-\lambda)s. \tag{5}$$

For (x,t) and (y,s) in $\operatorname{epi} f$, we know $f(x)$ and $f(y)$ are $< \infty$. What happens if $f(x) = -\infty$? Then for every real t' no matter how negative, $f(x) \le t'$, which means $(x,t') \in \operatorname{epi} f$. Applying (5) with t' instead of t and taking $0 < \lambda < 1$, we conclude that $f\big(\lambda x + (1-\lambda)y\big) = -\infty$. Now suppose $f(x)$ and $f(y)$ are finite real numbers. Then we can specialize to $t = f(x)$ and $s = f(y)$ in (5) to get our original formula,

$$f\big(\lambda x + (1-\lambda)y\big) \le \lambda f(x) + (1-\lambda) f(y). \tag{6}$$

Hence, convexity of $\operatorname{epi} f$ implies (6) holds for all $x, y \in \operatorname{dom} f$, even when $f(x)$ or $f(y)$ is $-\infty$.

Proposition 1.4. *If $f: X \to (-\infty, \infty]$ and (6) holds for all $x, y \in X$ and $0 \le \lambda \le 1$, then* epi f *is convex.*

Proof. Suppose (x,t) and (y,s) are in epi f. Then $f(x) \le t$ and $f(y) \le s$, and using this in (6) yields (5). $\square$

1.3. Projection onto the Epigraph and Lower Semicontinuity

If X is complete; i.e., X is a Hilbert space, then the **Projection Theorem** tells us that for every nonempty, closed, convex set $C \subset X$ and every $x \in X$, there is a unique point $\widehat{x} \in C$ satisfying both

$$\|x - \widehat{x}\| \le \|x - y\|, \quad y \in C,$$

and

$$\langle x - \widehat{x}, y - \widehat{x} \rangle \le 0, \quad y \in C. \tag{7}$$

Since X is an inner-product space, we can make $X \times \mathbb{R}$ into an inner-product space by setting

$$\big\langle (x,t), (y,s) \big\rangle := \langle x, y \rangle + ts.$$

The corresponding norm squared on $X \times \mathbb{R}$ is

$$\|(x,t) - (y,s)\|^2 = \|x-y\|^2 + |t-s|^2,$$

which implies[b]

$$\|(x,t) - (y,s)\| \ge \begin{cases} \|x-y\| \\ |t-s| \end{cases} \quad \text{and} \quad \|(x,t)-(y,s)\| \le \|x-y\| + |t-s|.$$

Hence, two points (x,t) and (y,s) are close if and only if x is close to y in the norm on X and $|t-s|$ is also small. If X is a Hilbert space, then it is not hard to show that $X \times \mathbb{R}$ is also a Hilbert space. Since we will be interested in projections onto epi f, it is important to have conditions on f under which epi f will be nonempty, closed, and convex. Proposition 1.4 was a first step.

If epi f is closed, then for each real t, the **level set** $A := \{x \in X : f(x) \le t\}$ is closed. To see this, suppose $x_n \in A$ and $x_n \to x$. Then $\|(x_n,t) - (x,t)\| = \|x_n - x\|$,

[b] The right-hand inequality follows by noting that

$$\big(\|x-y\| + |t-s|\big)^2 = \|x-y\|^2 + 2\|x-y\|\,|t-s| + |t-s|^2 \ge \|x-y\|^2 + |t-s|^2 = \|(x,t)-(y,s)\|^2.$$

which implies $(x_n,t) \to (x,t)$. Since $\operatorname{epi} f$ is closed, we conclude $(x,t) \in \operatorname{epi} f$, which means that $f(x) \le t$; i.e., $x \in A$.

The level sets $\{x \in X : f(x) \le t\}$ are all closed if and only if their complements, $\{x \in X : f(x) > t\}$, are all open. A function with this latter property is said to be **lower semicontinuous**. Hence, lower semicontinuity is equivalent to all the level sets being closed.

We now show that if f is lower semicontinuous, then $\operatorname{epi} f$ is closed. Suppose $(x_n,t_n) \in \operatorname{epi} f$ and $(x_n,t_n) \to (x,t)$. We must show $(x,t) \in \operatorname{epi} f$, or $f(x) \le t$. Since $(x_n,t_n) \to (x,t)$, $x_n \to x$ and $t_n \to t$. Now suppose $f(x) > t$. Then for small $\varepsilon > 0$, $f(x) > t+\varepsilon$ as well. For large n, x_n will be close to x, and by lower semicontinuity, $f(x_n) > t+\varepsilon$. Hence, $\underline{\lim}_{n\to\infty} f(x_n) \ge t+\varepsilon > t$. But $(x_n,t_n) \in \operatorname{epi} f$ implies $f(x_n) \le t_n$, which implies the contradiction

$$\underline{\lim}_{n\to\infty} f(x_n) \le \underline{\lim}_{n\to\infty} t_n = \lim_{n\to\infty} t_n = t.$$

We have now established the following result.

Theorem 1.5. *If $f\colon X \to [-\infty,\infty]$, then*

$$\operatorname{epi} f \text{ is closed} \Leftrightarrow \text{all level sets are closed} \Leftrightarrow f \text{ is lower semicontinuous.}$$

Example 1.6. Recall the function $f(x) := -\ln x$ for $x > 0$ and $f(x) := \infty$ for $x \le 0$. It is easy to show that f is lower semicontinuous on $\mathbb{R}$; just observe that

$$\{x \in \mathbb{R} : f(x) > t\} = (-\infty, e^{-t})$$

is open.

Example 1.7. Let $f\colon X \to (-\infty,\infty]$ and $g\colon X \to (-\infty,\infty]$. If f and g are lower semicontinuous, show that $h := f+g$ is lower semicontinuous.

Solution. Suppose $h(x_0) > t$ for some real t. We must show that for all x near x_0, $h(x) > t$ as well. First suppose $g(x_0)$ is finite. Then for small $\varepsilon > 0$, $f(x_0)+g(x_0) > t$ implies $f(x_0) > t - g(x_0) + \varepsilon$. Since f and g are lower semicontinuous, for x near x_0, we can have $f(x) > t - g(x_0) + \varepsilon$ and $g(x) > g(x_0) - \varepsilon$. For such x, $h(x) = f(x)+g(x) > t$. If $f(x_0)$ and $g(x_0)$ are both ∞, then $f(x_0) > t/2$ and $g(x_0) > t/2$, and for x near x_0, $f(x) > t/2$ and $g(x) > t/2$, which implies $h(x) > t$.

---

2. Minimization of Convex Functions

2.1. The Finite-Dimensional Case

In this section, convexity of f is not needed or assumed.

Theorem 2.1. *Let $f\colon \mathbb{R}^d \to [-\infty,\infty]$ be lower semicontinuous, and let C be a nonempty, closed and bounded subset of $\mathbb{R}^d$. Then f achieves its minimum on C; i.e., there is some $x_0 \in C$ with $f(x_0) = \inf_{x\in C} f(x)$.*

Remark 2.2. If $f > -\infty$ on C, then since the minimum is achieved, we conclude that $\inf_{x\in C} f(x) = f(x_0) > -\infty$. If there is an $x_1 \in C$ with $f(x_1) < \infty$, then $\inf_{x\in C} f(x) \le f(x_1) < \infty$.

Before proving Theorem 2.1, we need the following alternative characterization of lower semicontinuity using sequences. We say that f is ***sequentially*** **lower semicontinuous** if whenever $x_n \to x$ and $f(x_n) \to L$, we have $f(x) \le L$, where L may be finite or $\pm\infty$.

Theorem 2.3. *If $f\colon X \to [-\infty,\infty]$, then f is sequentially lower semicontinuous $\Leftrightarrow$ f is lower semicontinuous.*

Proof. By Theorem 1.5, the condition "f is lower semicontinuous" is equivalent to "$\operatorname{epi} f$ is closed." First suppose f is sequentially lower semicontinuous and that $(x_n,t_n) \to (x,t)$, where $(x_n,t_n) \in \operatorname{epi} f$ and $(x,t) \in X \times \mathbb{R}$. Then $x_n \to x$ and $f(x_n) \le t_n \to t$. Put $L := \underline{\lim}_{n\to\infty} f(x_n)$. Then there is a subsequence $f(x_{n_k}) \to L$, and we have $x_{n_k} \to x$ and $f(x_{n_k}) \le t_{n_k} \to t$. By sequential lower semicontinuity,

$$f(x) \le L = \lim_{k\to\infty} f(x_{n_k}) \le \lim_{k\to\infty} t_{n_k} = t,$$

which says $(x,t) \in \operatorname{epi} f$ as required.

For the converse, suppose $\operatorname{epi} f$ is closed. If $x_n \to x$ and $L := \lim_{n\to\infty} f(x_n)$, we must show that $f(x) \le L$. If $L = \infty$, the result is trivially true. Otherwise, L is finite or $-\infty$. If L is finite, then for large n, $f(x_n)$ is finite, and we have $(x_n, f(x_n)) \in \operatorname{epi} f$ and converging to (x,L), which must belong to $\operatorname{epi} f$. Hence, $f(x) \le L$. If $L = -\infty$, let t be any real number. Then for large n we must have $f(x_n) \le t$; i.e., $(x_n,t) \in \operatorname{epi} f$ and converges to $(x,t) \in \operatorname{epi} f$. Thus, $f(x) \le t$. Now let $t \to -\infty$ to get $f(x) \le -\infty = L$. $\square$

Proof of Theorem 2.1. Put $L := \inf_{x\in C} f(x)$. Then there is a sequence $x_n \in C$ with $f(x_n) \to L$. Since C is a closed and bounded subset of IR^d, C is **sequentially compact**, which implies there is a subsequence x_{n_k} converging to a point $x_0 \in C$. Since $f(x_{n_k}) \to L$, Theorem 2.3 implies $f(x_0) \le L$, which, by its definition, is less than or equal to $f(x)$ for all $x \in C$. □

2.2. The Infinite-Dimensional Case

The proof of Theorem 2.1 used the fact that closed and bounded subsets of IR^d are sequentially compact. Recall that a set C is **sequentially compact** if every sequence in C has a converging subsequence that converges to a point in C.

In infinite-dimensional spaces, closed and bounded sets need not be sequentially compact. However, a bounded sequence in a Hilbert space always has a ***weakly* convergent subsequence** [1, p. 26, Theorem 1.8.1].

Definition 2.4. A sequence x_n in an inner-product space X **converges *weakly*** to x, denoted by $x_n \rightharpoonup x$, if $\langle x_n, z\rangle \to \langle x, z\rangle$ for all $z \in X$.

Lemma 2.5. *A closed, convex subset C of a Hilbert space is* ***weakly closed*** *in the sense that if $x_n \in C$ converges weakly to x, then $x \in C$.*

Proof. If C is empty, there is nothing to prove. Otherwise, by the Projection Theorem, we can write

$$\langle x - \widehat{x}, y - \widehat{x}\rangle \le 0, \quad y \in C.$$

Replace y by $x_n \in C$ and take limits to get

$$0 \ge \lim_{n\to\infty} \langle x - \widehat{x}, x_n - \widehat{x}\rangle = \langle x - \widehat{x}, x - \widehat{x}\rangle = \|x - \widehat{x}\|^2.$$

Hence, $x = \widehat{x} \in C$. □

We define a function f to be ***weakly* sequentially lower semicontinuous** if whenever $x_n \rightharpoonup x$ and $f(x_n) \to L$, we have $f(x) \le L$, where L may be finite or $\pm\infty$.

Theorem 2.6. *Let X be a Hilbert space, and suppose $f{:}X \to [-\infty, \infty]$ is convex. Then f is weakly sequentially lower semicontinuous $\Leftrightarrow$ f is lower semicontinuous.*

Proof. Suppose f is weakly sequentially lower semicontinuous. We will show that f is sequentially lower semicontinuous. Suppose $x_n \to x$ and $f(x_n) \to L$. We must show that $f(x) \le L$. To see this, just use the Cauchy–Schwarz inequality to show that $x_n \to x$ implies $x_n \rightharpoonup x$, and then *weak* sequential lower semicontinuity gives $f(x) \le L$.

For the converse, suppose $\operatorname{epi} f$ is closed. If $x_n \rightharpoonup x$ and $L := \lim_{n\to\infty} f(x_n)$, we must show that $f(x) \le L$. If $L = \infty$, the result is trivially true. Otherwise, L is finite or $-\infty$. If L is finite, then for large n, $f(x_n)$ is finite, and we have $(x_n, f(x_n)) \in \operatorname{epi} f$ and converging *weakly* to (x, L), which must belong to $\operatorname{epi} f$ by Lemma 2.5 (the convexity of f means $\operatorname{epi} f$ is convex, which is required to apply the lemma). Hence, $f(x) \le L$. If $L = -\infty$, let t be any real number. Then for large n we must have $f(x_n) \le t$; i.e., $(x_n, t) \in \operatorname{epi} f$ and converges *weakly* to $(x, t) \in \operatorname{epi} f$. Thus, $f(x) \le t$. Now let $t \to -\infty$ to get $f(x) \le -\infty = L$. □

Theorem 2.7. *Let X be a Hilbert space, and suppose $f: X \to [-\infty, \infty]$ is lower semicontinuous and convex. If C is a closed, bounded, and convex subset of X, then f achieves its minimum on C; i.e., there is some $x_0 \in C$ with $f(x_0) = \inf_{x\in C} f(x)$.*

Proof. Put $L := \inf_{x\in C} f(x)$. Then there is a sequence $x_n \in C$ with $f(x_n) \to L$. Since C is bounded and X is a Hilbert space, there is a subsequence x_{n_k} that converges *weakly* to some x, which must lie in C by Lemma 2.5 (this is where the convexity of C is used). Since $f(x_{n_k}) \to L$, Theorem 2.6 (which uses the convexity and lower semicontinuity of f) implies $f(x_0) \le L$, which, by its definition, is less than or equal to $f(x)$ for all $x \in C$. □

Remark 2.2 applies to Theorem 2.7.

In some cases, the set C is not bounded, and so we cannot apply Theorem 2.7. However, suppose $f(x) \to \infty$ as $\|x\| \to \infty$. Such a function is said to be **coercive**. If f is coercive and $f(x_1) < \infty$ for some $x_1 \in C$, then any minimizer of f on C must occur for x inside some closed ball of finite radius r — we only need r large enough that $\|x\| > r$ implies $f(x) > f(x_1)$. If C is closed and convex, then so is the intersection of C with a closed ball. We can then apply Theorem 2.7 to this intersection.

Proposition 2.8 (Prop. 11.11 in [2]). *Let $f: X \to [-\infty, \infty]$. If the level sets $\{x \in X : f(x) \le k\}$ are bounded for $k = 1, 2, \ldots$, then f is coercive, and if f is coercive, then for all real t, level sets $\{x \in X : f(x) \le t\}$ are bounded.*

Proof. Given any positive integer M, we must show that for all sufficiently large $\|x\|$, $f(x) > M$. Since the level set $\{x \in X : f(x) \le M\}$ is bounded, there is some L

such that $\{x \in X : f(x) \leq M\} \subset \{x \in X : \|x\| \leq L\}$; equivalently, for all $\|x\| > L$, we have $f(x) > M$.

Now suppose f is coercive, but for some real t, $\{x \in X : f(x) \leq t\}$ is *not* bounded. Then for every $n = 1, 2, \ldots$, there is an x_n with $f(x_n) \leq t$, but $\|x_n\| > n$, contradicting the coercivity of f. □

3. Applications of the Projection Theorem

3.1. Decay of Convex Functions and the Proximal Mapping

Some convex functions such as $f(x) := x^2$ and $f(x) := e^x$ are bounded below, while others such as $f(x) := -\ln x$ tend to $-\infty$ as $x \to \infty$. However, most convex functions cannot tend to $-\infty$ faster than a linear function of x. For example, $-\ln x \geq 1 - x$ is a nice illustration of this observation. Here is a precise statement of the general result.

Theorem 3.1. *Let f be a proper, convex, lower-semicontinuous function defined on a real Hilbert space X. Then there exist $\widehat{x}_0, x_1 \in X$ such that $f(\widehat{x}_0)$ is finite and*

$$f(x) \geq f(\widehat{x}_0) + \langle x - \widehat{x}_0, x_1 \rangle, \quad x \in X. \tag{8}$$

Proof. Since f is not everywhere infinite, let x_0 be such that $f(x_0)$ is finite. Since $(x_0, f(x_0)) \in \operatorname{epi} f$, $\operatorname{epi} f$ is nonempty. By Theorem 1.5, $\operatorname{epi} f$ is closed in $X \times \mathbb{R}$, and by definition of convex function, $\operatorname{epi} f$ is convex. Now choose any $-\infty < t_0 < f(x_0)$ so that $(x_0, t_0) \notin \operatorname{epi} f$. By the Projection Theorem, there exists $(\widehat{x}_0, \widehat{t}_0) \in \operatorname{epi} f$ with (cf. (7))

$$\big\langle (x_0, t_0) - (\widehat{x}_0, \widehat{t}_0), (x, t) - (\widehat{x}_0, \widehat{t}_0) \big\rangle \leq 0, \quad (x, t) \in \operatorname{epi} f,$$

which we write more explicitly as

$$\langle x_0 - \widehat{x}_0, x - \widehat{x}_0 \rangle + (t_0 - \widehat{t}_0)(t - \widehat{t}_0) \leq 0, \quad t \in \mathbb{R} \text{ and } f(x) \leq t. \tag{9}$$

Specializing to $x = x_0$ and $t = f(x_0) + \lambda$ for $\lambda \geq 0$ yields

$$\|x_0 - \widehat{x}_0\|^2 + (t_0 - \widehat{t}_0)(f(x_0) + \lambda - \widehat{t}_0) \leq 0. \tag{10}$$

Letting $\lambda \to \infty$ implies $t_0 \leq \widehat{t}_0$. We show below that

$$t_0 < \widehat{t}_0 \quad \text{and} \quad f(\widehat{x}_0) = \widehat{t}_0. \tag{11}$$

Accepting this for now, consider any x for which $f(x)$ is finite and put $t = f(x)$ in (9) and use the equality in (11) to replace the right-most appearance of $\widehat{t_0}$. This yields

$$\langle x_0 - \widehat{x_0}, x - \widehat{x_0}\rangle + (t_0 - \widehat{t_0})(f(x) - f(\widehat{x_0})) \le 0.$$

By the strict inequality in (11), we can divide through by the *negative* quantity $t_0 - \widehat{t_0}$ and rearrange to get (8) with $x_1 := (x_0 - \widehat{x_0})/(\widehat{t_0} - t_0)$. At this point we have proved (8) only for those x with $f(x)$ finite. But if $f(x) = \infty$, then (8) is trivially true.

It remains to establish (11). With regard to the inequality, as noted above (11), we already know that $t_0 \le \widehat{t_0}$. Suppose $t_0 = \widehat{t_0}$. Then (10) implies $x_0 = \widehat{x_0}$. We thus arrive at $(x_0, t_0) = (\widehat{x_0}, \widehat{t_0}) \in \operatorname{epi} f$, contradicting our choice of t_0 so that $(x_0, t_0) \notin \operatorname{epi} f$. Therefore, $t_0 < \widehat{t_0}$. To prove the equality in (11), first note that $(\widehat{x_0}, \widehat{t_0}) \in \operatorname{epi} f$ implies $f(\widehat{x_0}) \le \widehat{t_0}$. Suppose $f(\widehat{x_0}) < \widehat{t_0}$. Then taking $x = \widehat{x_0}$ and $t = f(\widehat{x_0})$ in (9) tells us that $(t_0 - \widehat{t_0})(f(\widehat{x_0}) - \widehat{t_0}) \le 0$, which is impossible for the product of two negative numbers. $\square$

We say that a convex function f is **strongly convex** if $g(x) := f(x) - \frac{\mu}{2}\|x\|^2$ is convex for some $\mu > 0$.

Corollary 3.2. *Let f be a proper, convex, lower-semicontinuous function defined on a real Hilbert space X. If f is strongly convex, then f is coercive.*

Proof. Consider the convex function $g(x) := f(x) - \frac{\mu}{2}\|x\|^2$, which is proper and lower semicontinuous. Apply the preceding theorem and then substitute the definition of g to get

$$f(x) \ge f(\widehat{x_0}) + \tfrac{\mu}{2}\left[\|x\|^2 - \|\widehat{x_0}\|^2\right] + \langle x - \widehat{x_0}, x_1\rangle.$$

By the Cauchy–Schwarz inequality, $\langle x, x_1\rangle \ge -\|x\|\,\|x_1\|$, and so

$$f(x) \ge f(\widehat{x_0}) + \tfrac{\mu}{2}\left[\|x\|^2 - \|\widehat{x_0}\|^2\right] - \|x\|\,\|x_1\| - \langle \widehat{x_0}, x_1\rangle,$$

which tends to infinity as $\|x\| \to \infty$. $\square$

Corollary 3.3. *Let f be a proper, convex, lower-semicontinuous function defined on a real Hilbert space X. Then $f(x) + \frac{\mu}{2}\|x - x_0\|^2$ is coercive.*

Proof. Observe that

$$\left[f(x) + \tfrac{\mu}{2}\|x - x_0\|^2\right] - \tfrac{\mu}{2}\|x\|^2 = f(x) - \mu\langle x, x_0\rangle + \tfrac{\mu}{2}\|x_0\|^2$$

is convex since it is the sum of the convex function f, the linear function $-\mu\langle x, x_0\rangle$, and a constant. Now apply the previous corollary. $\square$

For a proper, convex, lower-semicontinuous function f on a real Hilbert space X, the **proximal mapping** is

$$(\operatorname{prox} f)(x) := \operatorname*{argmin}_{y \in X}\left[f(y) + \tfrac{1}{2}\|y - x\|^2\right], \quad x \in X.$$

3.2. Subgradients

Let f be a convex function on an inner-product space X. Given a point $x \in X$, if there exists a point $z \in X$ such that

$$f(y) \geq f(x) + \langle y - x, z \rangle, \quad y \in X,$$

then we say that z is a **subgradient** of f at x. Theorem 3.1 gives conditions under which there is at least one point $\widehat{x}_0$ where f has a subgradient.

Example 3.4. Consider the function $f\colon \mathrm{I\!R}^d \to \mathrm{I\!R}$ defined by $f(x) := \|x\|_1 := \sum_{k=1}^d |x_k|$, where $x = [x_1, \ldots, x_d]^\mathsf{T} \in \mathrm{I\!R}^d$. We show that if $z = [z_1, \ldots, z_d]^\mathsf{T}$ with $z_k := \operatorname{sgn}(x_k)$, then z is a subgradient of f at x. We must show that

$$\sum_{k=1}^d |y_k| \geq \sum_{k=1}^d |x_k| + \sum_{k=1}^d (y_k - x_k) z_k.$$

The first step is to rewrite the right-hand side as

$$\sum_{k=1}^d \{ |x_k| - x_k z_k \} + \sum_{k=1}^d y_k z_k.$$

Since $x_k z_k = x_k \operatorname{sgn}(x_k) = |x_k|$, the preceding display reduces to

$$\sum_{k=1}^d y_k z_k \leq \left| \sum_{k=1}^d y_k z_k \right| \leq \sum_{k=1}^d |y_k z_k| \leq \sum_{k=1}^d |y_k|,$$

where the last step uses the fact that $|z_k| \leq 1$. Careful analysis of the foregoing steps reveals that if $x_k = 0$, we can allow z_k to be any number in the interval $[-1, 1]$.

Example 3.5. Let A be a $d \times p$ matrix of real numbers, and consider the function $f\colon \mathrm{I\!R}^p \to \mathrm{I\!R}$ defined by $f(x) := \|Ax\|_1$, where $x = [x_1, \ldots, x_p]^\mathsf{T} \in \mathrm{I\!R}^p$. We show that $z := A^\mathsf{T} u$, where $u_i := \operatorname{sgn}((Ax)_i)$ is a subgradient of f at x. We must show that

$$\begin{aligned} \|Ay\|_1 &\geq \|Ax\|_1 + \langle y - x, A^\mathsf{T} u \rangle \\ &= \|Ax\|_1 + \langle A(y - x), u \rangle, \end{aligned}$$

which we can rewrite as

$$\sum_{i=1}^d |(Ay)_i| \geq \sum_{i=1}^d |(Ax)_i| + \sum_{i=1}^d [(Ay)_i - (Ax)_i] u_i.$$

The first step is to rewrite the right-hand side as

$$\sum_{i=1}^{d}\big\{|(Ax)_i| - (Ax)_i u_i\big\} + \sum_{i=1}^{d}(Ay)_i u_i.$$

Since $(Ax)_i u_i = (Ax)_i \operatorname{sgn}((Ax)_i) = |(Ax)_i|$, the preceding display reduces to

$$\sum_{i=1}^{d}(Ay)_i u_i \leq \left|\sum_{i=1}^{d}(Ay)_i u_i\right| \leq \sum_{i=1}^{d}|(Ay)_i u_i| \leq \sum_{i=1}^{d}|(Ay)_i|,$$

where the last step uses the fact that $|u_i| \leq 1$. Careful analysis of the foregoing steps reveals that if $(Ax)_i = 0$, we can allow u_i to be any number in the interval $[-1,1]$.

Example 3.6. Suppose

$$f_{\max}(y) := \max_{1\leq k\leq n} f_k(y),$$

where each f_k has a subgradient z_k at a point x; i.e.,

$$f_k(y) \geq f_k(x) + \langle y-x, z_k\rangle, \quad \text{for all } y. \tag{12}$$

Put $K(x) := \{k : f_k(x) = f_{\max}(x)\}$. For each $k \in K(x)$, let λ_k be nonnegative with $\sum_{k\in K(x)} \lambda_k = 1$. We claim that

$$z_* := \sum_{k\in K(x)} \lambda_k z_k$$

is a subgradient of $f_{\max}$ at x. To prove this, first note that for any $k \in K(x)$,

$$\begin{aligned} f_{\max}(y) &\geq f_k(y), && \text{by the definition of } f_{\max}, \\ &\geq f_k(x) + \langle y-x, z_k\rangle, && \text{by (12)}, \\ &= f_{\max}(x) + \langle y-x, z_k\rangle, && \text{since } k \in K(x). \end{aligned}$$

Multiply through by λ_k and sum over $k \in K(x)$ to get

$$f_{\max}(y) \geq f_{\max}(x) + \Big\langle y-x, \underbrace{\sum_{k\in K(x)} \lambda_k z_k}_{:=z_*}\Big\rangle.$$

When each f_k is differentiable at x, we have $z_k = \nabla f_k(x)$.

References

[1] A. V. Balakrishnan, *Applied Functional Analysis*, 2nd ed. New York: Springer, 1981.
[2] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York: Springer, 2011.
[3] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. New York: Springer, 1998.
[4] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. New York: Wiley, 1999.
[5] I. Gohberg and S. Goldberg, *Basic Operator Theory*. Boston: Birkhäuser, 1980.
[6] S. R. Lay, *Convex Sets and Their Applications*. Mineola, NY: Dover, 2007.
[7] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.
[8] D. G. Luenberger and Yinyu Ye, *Linear and Nonlinear Programming*, 3rd ed. New York: Springer 2010.
[9] J. Nocedal and S. J. Wright *Numerical Optimization*, 2nd ed. New York: Springer, 2006.
[10] O. L. Mangasarian, *Nonlinear Programming*. New York: McGraw-Hill, 1969.
[11] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.
[12] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill, 1976.
[13] J. van Tiel, *Convex Analysis: An Introductory Text*. Chichester: Wiley, 1984.
[14] P. Whittle, *Optimization under Constraints: Theory and Applications of Nonlinear Programming*. London, U.K.: Wiley, 1971.

Index