Analysis of Convex Sets and Functions

Robert M. Freund

April, 2004

1 Convex Sets - Basics

A set $S \subset \mathbb{R}^n$ is defined to be a *convex* set if for any $x^1 \in S$, $x^2 \in S$, and any scalar λ satisfying $0 \leq \lambda \leq 1$, $\lambda x^1 + (1-\lambda)x^2 \in S$. Points of the form $\lambda x^1 + (1-\lambda)x^2$ are said to be a *convex combination* of x^1 and x^2, if $0 \leq \lambda \leq 1$.

A *hyperplane* $H \in \mathbb{R}^n$ is a set of the form $\{x \in R^n \mid p^t x = \alpha\}$ for some fixed $p \in \mathbb{R}^n$, $p \neq 0$, and $\alpha \in R$.

A *closed halfspace* is a set of the form $H^+ = \{x \mid p^t x \geq \alpha\}$ or $H^- = \{x \mid p^t x \leq \alpha\}$. An open halfspace is defined analogously, with $\geq$ and $\leq$ replaced by $<$ and $>$.

- Any hyperplane or closed or open halfspace is a convex set
- The ball $\{x \in \mathbb{R}^n \mid x_1^2 + \ldots + x_n^2 \leq 5\}$ is a convex set.
- The intersection of an arbitrary number of convex sets is a convex set.
- A polyhedron $\{x \in \mathbb{R}^n \mid Ax \leq b\}$ is a convex set.
- $X = \{x \mid x = \alpha x_1 + \beta x_2$, where $x_1 \in S_1$ and $x_2 \in S_2\}$ is a convex set, if S_1 and S_2 are convex sets.

Let $S \subset \mathbb{R}^n$. The convex hull of S, denoted $H(S)$, is the collection of all convex combinations of points of S, i.e., $x \in H(S)$ if x can be represented as

$$x = \sum_{j=1}^{k} \lambda_j x^j$$

where

$$\sum_{j=1}^{k} \lambda_j = 1, \lambda_j \geq 0, j = 1, \ldots, k,$$

for some positive integer k and vectors $x^1, \ldots, x^k \in S$.

Lemma 1 *If $S \subset \mathbb{R}^n$, then $H(S)$ is the smallest convex set containing S, and $H(S)$ is the intersection of all convex sets containing S.*

The convex hull of a finite number of points $x^1, \ldots, x^{k+1} \in \mathbb{R}^n$ is called a *polytope*.

If the vectors $x^2 - x^1, x^3 - x^1, \ldots, x^{k+1} - x^1$, are linearly independent, then $H(x^1, \ldots, x^{k+1})$ is called a *k-dimensional simplex*, with vertices $x^1, \ldots, x^{k+1}$.

Carathéodory Theorem. Let $S \subset \mathbb{R}^n$, and let $x \in H(S)$ be given. Then $x \in H(x^1, \ldots, x^{n+1})$ for $n+1$ suitably chosen points $x^j \in S$, $j = 1, \ldots, n+1$. In other words, x can be represented as

$$x = \sum_{j=1}^{n+1} \lambda_j x^j$$

where

$$\sum_{j=1}^{n+1} \lambda_j = 1,\ \ \lambda_j \geq 0, j = 1, \ldots, n,\ \text{ for some } x^j \in S, j = 1, \ldots, n+1\ .$$

Proof: Since $x \in H(S)$, there exist k points $x^1, \ldots, x^k$ that satisfy

$$\sum_{j=1}^{k} \lambda_j x^j = x$$

and

$$\sum_{j=1}^{k} \lambda_j = 1, \lambda_1, \ldots, \lambda_k \geq 0\ .$$

If $k \leq n+1$, we are done. If not, let us consider the above linear system as a system of $n+1$ equations in the nonnegative variable $\lambda \geq 0$, i.e., as a system $A\lambda = b, \lambda \geq 0$, where

$$A = \begin{bmatrix} x^1 & x^2 & \ldots & x^k \\ 1 & 1 & \ldots & 1 \end{bmatrix},\ b = \begin{bmatrix} x \\ 1 \end{bmatrix}.$$

Let $e = (1, \ldots, 1)^t$.

Notice that this system has $n+1$ equations in k nonnegative variables. From the theory of linear programming, this system has a basic feasible solution, and we can presume, without loss of generality, that the basis consists of columns $1, \ldots, n+1$ of A. Thus, there exists $\lambda \geq 0$, $e^t\lambda = 1$, $\lambda_{n+2}, \ldots, \lambda_k = 0$ such that

$$\sum_{j=1}^{n+1} \lambda_j x^j = x, \quad \sum_{j+1}^{n+1} \lambda_j = 1, \ \lambda \geq 0 \ .$$

∎

2 Closure and Interior of Convex Sets

Let $x \in \mathbb{R}^n$. Define $N_\epsilon(x) = \{y \in \mathbb{R}^n \mid \|y - x\| < \epsilon\}$, where $\|\cdot\|$ is the Euclidean norm:

$$\|z\| := \sqrt{\sum_{j=1}^{n} z_j^2} = \sqrt{z^t z} \ .$$

Let S be an arbitrary set in R^n. $x \in \text{cl}S$, the *closure* of S, if $S \cap N_\epsilon(x) \neq \emptyset$ for every $\epsilon > 0$. If $S = \text{cl}S$, then S is a *closed* set. $x \in \text{int}S$, the *interior* of S, if there exists $\epsilon > 0$ for which $N_\epsilon(x) \subset S$. If $S = \text{int}S$, then S is an *open* set. $x \in \partial S$, the *boundary* of S, if $N_\epsilon(x)$ contains a point in S and a point not in S, for all $\epsilon > 0$.

Theorem 2 *Let S be a convex set in R^n. If $x^1 \in \text{cl}S$ and $x^2 \in \text{int}S$, then $\lambda x^1 + (1-\lambda)x^2 \in \text{int}S$ for every $\lambda \in (0,1)$.*

Proof: Since $x^2 \in \text{int}S$, there exists $\epsilon > 0$ such that $\|z - x^2\| < \epsilon$ implies $z \in S$. Also, since $x^1 \in \text{cl}S$, for any $\delta > 0$, there exists a z with $\|z - x^1\| < \delta$ and $z \in S$. Now let $y = \lambda x^1 + (1-\lambda)x^2$ be given, with $\lambda \in (0,1)$. Let $r = (1-\lambda)\epsilon$. I claim that for any z with $\|z - y\| < r$, then $z \in S$. This means that $y \in \text{int}S$, proving the theorem.

To prove the claim, consider such a z with $\|z - y\| < r$. Let $\delta = (1-\lambda)\epsilon - \|z - y\|$. Then $\delta > 0$. Then there exists z^1 with $\|z^1 - x^1\| < \delta$ and $z^1 \in S$. Let $z^2 = (z - \lambda z^1)/(1-\lambda)$.

$$\begin{aligned} \|z^2 - x^2\| = \|\tfrac{(z-\lambda z^1)}{(1-\lambda)} - x^2\| &= \|\tfrac{(z-\lambda z^1)-(y-\lambda x^1)}{1-\lambda}\| \\ &= \left(\tfrac{1}{1-\lambda}\right)\|z - y - \lambda(z^1 - x^1)\| \\ &\leq \tfrac{1}{1-\lambda}(\|z-y\| + \lambda\|z^1 - x^1\|) \\ &< \tfrac{1}{1-\lambda}\left(\|z-y\| + \lambda\left((1-\lambda)\epsilon - \|z-y\|\right)\right) \\ &= \|z-y\| + \lambda\epsilon < (1-\lambda)\epsilon + \lambda\epsilon = \epsilon. \end{aligned}$$

Thus $z^2 \in S$. Also $z = \lambda z^1 + (1-\lambda)z^2$, and so $z \in S$. ∎

The proof construction is illustrated in Figure 1.

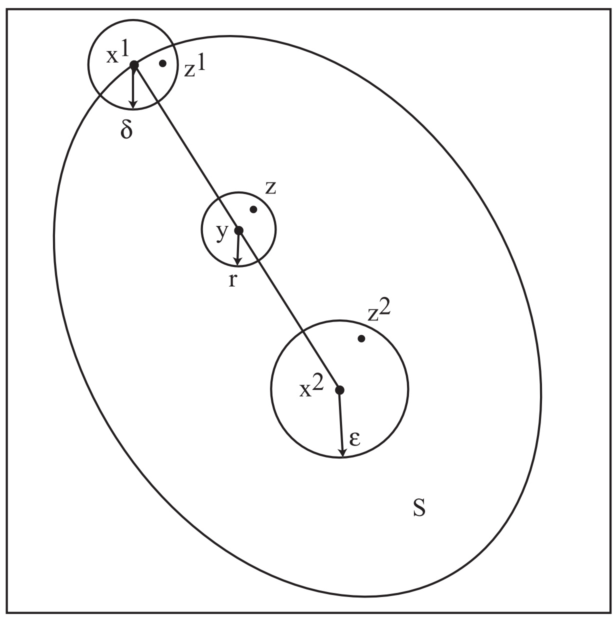

Figure 1: Construction for the proof of Theorem 2

3 Supporting Hyperplanes of Convex Sets

Let S be nonempty set in R^n, and let $\bar{x} \in \partial S$. A hyperplane $H = \{x \in \mathbb{R}^n \mid p^t x = \alpha\}$ (where $p \neq 0$) is call a *supporting hyperplane* of S at $\bar{x}$ if either $S \subset H^+$ or $S \subset H^-$, (i.e., if either $p^t x \geq \alpha$ for all $x \in S$ or $p^t x \leq \alpha$ for all $x \in S$, and $p^t \bar{x} = \alpha$.

Note that we can write H as $H = \{x \mid p^t x = p^t \bar{x}\}$, i.e., $H = \{x \mid p^t (x - \bar{x}) = 0\}$. If $S \not\subset H$, then H is called a *proper* supporting hyperplane.

Note that either $p^t \bar{x} = \inf\{p^t x \mid x \in S\}$ or $p^t \bar{x} = \sup\{p^t x \mid x \in S\}$.

Theorem 3 *If S is a nonempty convex set and $\bar{x} \in \partial S$, then there exists a supporting hyperplane to S at $\bar{x}$; that is, there exists a nonzero vector p such that $p^t x \leq p^t \bar{x}$ for all $x \in S$.*

Proof: Let $\bar{x} \in \partial S$. Then there exists a sequence of y^i, $i = 1, \ldots, \infty$, such that $y^i \to \bar{x}$, and $y^i \notin \mathrm{cl}(S)$. Now $\mathrm{cl}(S)$ is a closed convex set, and so there exists $p^i \neq 0$ and α^i such that $(p^i)^T x \leq \alpha^i$ for all $x \in \mathrm{cl}(S)$, and $(p^i)^T y^i > \alpha^i$. We can re-scale the p^i values so that $\|p^i\| = 1$. As $y^i \to \bar{x}$, choose subsequence of the p^i that converge, say, to $\bar{p}$. Then $(p^i)^T x \leq \alpha^i < (p^i)^T y^i$ for all $x \in \mathrm{cl}(S)$, and so $\bar{p}^T x \leq \bar{p}^T \bar{x}$. Thus, as $i \to \infty$, any $x \in S \subset \mathrm{cl}(S)$ satisfies $\bar{p}^T x \leq \bar{p}^T \bar{x}$. ∎

Theorem 4 *If A and B are nonempty convex sets and $A \cap B = \emptyset$, then A and B can be separated by a hyperplane. That is, there exists $p \neq 0$, α, such that $p^T x \leq \alpha$ for any $x \in A$ and $p^T x \geq \alpha$ for any $x \in B$.*

Proof: Let $S = \{x \mid x = x^1 - x^2,\ x^1 \in A,\ x^2 \in B\}$. Then S is a convex set, and $0 \notin S$. Let $T = \mathrm{cl}(S) = S \cup (\partial S)$. If $0 \in T$, then $0 \in \partial S$, and from Theorem 3, there exists $p \neq 0$, such that $p^T (x^1 - x^2) \leq 0$ for any $x^1 \in A$, $x^2 \in B$, so that $p^T x^1 \leq p^T x^2$ for any $x^1 \in A$, $x^2 \in B$. Let $\alpha = \inf\{p^T x^2 \mid x^2 \in B\}$. Then $p^T x^1 \leq \alpha \leq p^T x^2$ for any $x^1 \in A$, $x^2 \in B$.

If $0 \notin T$, then since T is a closed set, then we can apply the basic separating hyperplane theorem for convex sets: there exists $p \neq 0$ and β for which $p^T (x^1 - x^2) \leq \beta < p^T 0 = 0$ for any $x^1 \in A$, $x^2 \in B$, so that $p^T x^1 \leq p^T x^2$ for any $x^1 \in A$, $x^2 \in B$. Again let $\alpha = \inf\{p^T x^2 \mid x^2 \in B\}$. Then $p^T x^1 \leq \alpha \leq p^T x^2$ for any $x^1 \in A$, $x^2 \in B$.

∎

Corollary 5 *If S_1 and S_2 are nonempty disjoint closed convex sets, and S_1 is bounded, then S_1 and S_2 can be strongly separated.*

4 Polyhedral Convex Sets

S is a *polyhedral* set if $S = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ for some $(A,\ b)$. $\bar{x}$ is an *extreme point* of S if whenever $\bar{x}$ can be represented as $\bar{x} = \lambda x^1 + (1-\lambda)x^2$ for some $x^1, x^2 \in S$ and $\lambda \in (0,1)$, then $x^1 = x^2 = \bar{x}$.

$d \in \mathbb{R}^n$ is called a *ray* of S if whenever $x \in X$, $x + \alpha d \in S$ for all $\alpha \geq 0$. d is called an *extreme ray* of S if whenever $d = \lambda_1 d^1 + \lambda_2 d^2$ for rays d^1, d^2 of S and $\lambda_1, \lambda_2 \neq 0$, then $d_1 = \alpha d_2$ for some $\alpha \geq 0$ or $d_2 = \alpha d_1$ for some $\alpha \geq 0$.

Theorem 6 *The set $S = \{x \in \mathbb{R}^n \mid Ax = b,\ x \geq 0\}$ contains an extreme point if $S \neq \emptyset$.*

Theorem 7 *Let $S = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ be nonempty. Then S has a finite number of extreme points and extreme rays.*

Theorem 8 *Let $S = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ be nonempty. Then there exist a finite number of points $x^1, \ldots, x^K$ and a finite number of directions $d^1, \ldots, d^L$ with the property that $\bar{x} \in S$ if and only if the following inequality system has a solution in $\lambda_1, \ldots, \lambda_K$, $\mu_1, \ldots, \mu_L$:*

$$\sum_{j=1}^{K} x^j \lambda_j + \sum_{j=1}^{L} d^j \mu_j = \bar{x}$$

$$\sum_{j=1}^{K} \lambda_j = 1$$

$$\lambda_1 \geq 0, \ldots, \lambda_K \geq 0, \mu_1 \geq 0, \ldots, \mu_L \geq 0\ .$$

5 Convex Functions - Basics

Let $S \subset \mathbb{R}^n$ be a nonempty convex set. A function $f(\cdot) : S \to \mathbb{R} \cup \{-\infty, +\infty\}$ is a *convex* function on S if:

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y) \text{ for all } \lambda \in [0,1] \text{ and } x, y \in S.$$

$f(\cdot)$ is *strictly convex* if the inequality above is strict for all $\lambda \in (0,1)$ and $x \neq y$.

$f(\cdot)$ is *concave* if $-f(\cdot)$ is convex.

$f(\cdot)$ is *strictly concave* if $-f(\cdot)$ is strictly convex.

Examples:

1. $f(x) = ax + b$
2. $f(x) = |x|$
3. $f(x) = x^2 - 2x + 3$
4. $f(x) = \sqrt{x}$ for $x \ge 0$
5. $f(x_1, x_2) = 2x_1^2 + x_2^2 - 2x_1x_2$

We define $S_\alpha := \{x \in S \mid f(x) \le \alpha\}$. S_α is called the *level set* of $f(\cdot)$ at level α.

Lemma 9 *If $f(\cdot)$ is convex on S, then S_α is a convex set.*

Proof: Let $x, y \in S_\alpha$. Then $f(x) \le \alpha$ and $f(y) \le \alpha$, and so $\lambda \in [0,1]$ implies

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y) \le \lambda\alpha + (1-\lambda)\alpha = \alpha \ ,$$

and so $\lambda x + (1-\lambda)y \in S_\alpha$. Therefore S_α is a convex set. ∎

Note, the converse Lemma 9 is false.

Theorem 10 *If $S \subset \mathbb{R}^n$ is an open convex set and $f(\cdot)$ is a real-valued convex function on S, then $f(\cdot)$ is continuous on S.*

Proof: We will prove this in three steps.

Step 1. If $f(\cdot)$ is convex on S and $x^1, \ldots, x^k \in S$ and $\lambda_1, \ldots, \lambda_k \geq 0$ and $\sum_{j=1}^{n} \lambda_j = 1$, then

$$f\left(\sum_{j=1}^{k} \lambda_j x^j\right) \leq \sum_{j=1}^{k} \lambda_j f(x^j) \ .$$

Step 2. For each $x \in S$, there exists a $\beta > 0$ such that $x \pm \beta e_i \in S$, for $i = 1, \ldots, n$, where e_i is the i^{th} unit vector. Call these vectors $z_1, \ldots, z_{2n}$. Let $M_x := \max\{f(z_1), \ldots, f(z_{2n})\}$.

Since x lies in the interior of the convex hull of $z_1, \ldots, z_{2n}$, there exists $t_x > 0$ such that $B(x, t_x) := \{y \mid \|y - x\| \leq t_x\}$ also lies in the convex hull of $z_1, \ldots, z_{2n}$. Then for any y satisfying $\|y - x\| \leq t_x$ it holds that $y = \sum_{j=1}^{2n} \lambda_j z_j$ for some appropriate $\lambda_1, \ldots, \lambda_{2n} \geq 0$ and $\sum_{j=1}^{2n} \lambda_j = 1$ and so

$$f(y) = f\left(\sum_{j=1}^{2n} \lambda_j z_j\right) \leq \sum_{j=1}^{2n} \lambda_j f(z_j) \leq M_x \ .$$

Step 3. (The proof of the theorem). Without loss of generality, we assume that $f(0) = 0$ and we want to prove $f(x)$ is continuous at $x = 0$. For any $\epsilon > 0$, we must exhibit a $\delta > 0$ such that if $\|y\| < \delta$ then $-\epsilon \leq f(y) \leq \epsilon$. Let $t > 0$ and M be chosen so that $\|y\| \leq t$ implies $f(y) \leq M$ (from *Step 2*). Now let $\delta = t\epsilon/M$. Let y satisfy $\|y\| < \delta$. Also, we can assume that $M > \epsilon$ and write:

$$y = \left(1 - \frac{\epsilon}{M}\right) 0 + \left(\frac{\epsilon}{M}\right)\left(\frac{M}{\epsilon} y\right) \ .$$

Since $\| \pm \frac{M}{\epsilon} y\| = \frac{M}{\epsilon}\|y\| < t$, it follows that $f\left(\pm \frac{M}{\epsilon} y\right) \leq M$. Therefore:

$$f(y) \leq \left(1 - \frac{\epsilon}{M}\right) f(0) + \frac{\epsilon}{M} f\left(\frac{M}{\epsilon} y\right) \leq 0 + \frac{\epsilon}{M} \times M = \epsilon \ .$$

We also have:

$$0 = \frac{\frac{\epsilon}{M}}{1+\frac{\epsilon}{M}}\left(-\frac{M}{\epsilon}y\right) + \frac{1}{1+\frac{\epsilon}{M}}y \ ,$$

and so

$$0 = f(0) \le \frac{\frac{\epsilon}{M}}{1+\frac{\epsilon}{M}} f\left(-\frac{M}{\epsilon}y\right) + \frac{1}{1+\frac{\epsilon}{M}} f(y) \le \frac{\frac{\epsilon}{M}}{1+\frac{\epsilon}{M}} M + \frac{f(y)}{1+\frac{\epsilon}{M}} \ .$$

Therefore $f(y) \ge -\epsilon$, and so $-\epsilon \le f(y) \le \epsilon$. ∎

The *directional derivative* of $f(\cdot)$ at $\bar{x}$ in the direction d, denoted by $f^{'}(\bar{x}, d)$ is defined as:

$$f^{'}(\bar{x}, d) = \lim_{\lambda \to 0^+} \frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda} \ ,$$

when this limit exists.

Lemma 5.1 *Suppose that S is convex and $f(\cdot) : S \to \mathbb{R}$ is convex, $\bar{x} \in S$ and $\bar{x} + \lambda d \in S$ for all $\lambda > 0$ and sufficiently small. Then $f^{'}(\bar{x}, d)$ exists.*

Proof: Let $\lambda_2 > \lambda_1 > 0$ be sufficiently small. Then

$$\begin{aligned} f(\bar{x} + \lambda_1 d) &= f\left(\tfrac{\lambda_1}{\lambda_2}(\bar{x} + \lambda_2 d) + \left(1 - \tfrac{\lambda_1}{\lambda_2}\right)(\bar{x})\right) \\ &\le \tfrac{\lambda_1}{\lambda_2} f(\bar{x} + \lambda_2 d) + \left(1 - \tfrac{\lambda_1}{\lambda_2}\right) f(\bar{x}) \end{aligned}$$

Rearranging the above, we have:

$$\frac{f(\bar{x} + \lambda_1 d) - f(\bar{x})}{\lambda_1} \le \frac{f(\bar{x} + \lambda_2 d) - f(\bar{x})}{\lambda_2} \ .$$

Thus the term $\frac{f(\bar{x}+\lambda d) - f(\bar{x})}{\lambda}$ is nondecreasing in λ, and so the limit exists, and is

$$f^{'}(\bar{x}, d) = \inf_{\lambda > 0} \left\{ \frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda} \right\} \ ,$$

so long as we allow $f^{'}(\bar{x}, d) = -\infty$. ∎

6 Subgradients of Convex Functions

Let $S \subset \mathbb{R}^n$, and $f(\cdot) : S \to \mathbb{R}$ be given. The *epigraph* of $f(\cdot)$, denoted epi $f(\cdot)$, is a subset of $\mathbb{R}^{n+1}$ and is defined to be:

$$\text{epi } f(\cdot) := \{(x, \alpha) \in \mathbb{R}^{n+1} \mid x \in S,\ \alpha \in \mathbb{R},\ f(x) \leq \alpha\} \ .$$

The *hypograph* of $f(\cdot)$, denoted hyp $f(\cdot)$, is defined analogously as

$$\text{hyp } f\{(x, \alpha) \in \mathbb{R}^{n+1} \mid x \in S,\ \alpha \in \mathbb{R},\ f(x) \geq \alpha\} \ .$$

Theorem 11 *Let S be a nonempty convex set in $\mathbb{R}^n$, and $f(\cdot) : S \to \mathbb{R}$. Then $f(\cdot)$ is a convex function if and only if* epi $f(\cdot)$ *is a convex set.*

Proof: Assume $f(\cdot)$ is convex. Let (x_1, α_1), $(x_2, \alpha_2) \in$ epi $f(\cdot)$. Then if $\lambda \in (0, 1)$, $\lambda x_1 + (1 - \lambda)x_2 \in S$, because $x_1 \in S$, $x_2 \in S$, and S is convex. Also $\alpha_1 \geq f(x_1)$, $\alpha_2 \geq f(x_2)$, $f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \leq \lambda\alpha_1 + (1 - \lambda)\alpha_2$. Thus $(\lambda x_1 + (1 - \lambda)x_2)$, $(\lambda\alpha_1 + (1 - \lambda)\alpha_2) \in$ epi $f(\cdot)$, and so epi $f(\cdot)$ is a convex set.

Conversely, assume that epi $f(\cdot)$ is a convex set. Let $x_1,\ x_2 \in S$, and let $\alpha_1 = f(x_1),\ \alpha_2 = f(x_2)$. Then $(x_1, \alpha_1) \in$ epi $f(\cdot)$ and $(x_2, \alpha_2) \in$ epi $f(\cdot)$.

Thus $(\lambda x_1 + (1 - \lambda)x_2,\ \lambda\alpha_1 + (1 - \lambda)\alpha_2) \in$ epi $f(\cdot)$.

This means $f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\alpha_1 + (1-\lambda)\alpha_2 = \lambda f(x_1) + (1-\lambda)f(x_2)$, and so $f(\cdot)$ is a convex function. ∎

If $f(\cdot) : S \to \mathbb{R}$ is a convex function, then $\xi \in \mathbb{R}^n$ is a *subgradient* of $f(\cdot)$ at $\bar{x} \in S$ if

$$f(x) \geq f(\bar{x}) + \xi^t(x - \bar{x}) \text{ for all } x \in S \ .$$

Example: $f(x) = x^2$ and let $\bar{x} = 3$ and $\xi = 6$. Notice that from the basic inequality

$$0 \leq (x - 3)^2 = x^2 - 6x + 9$$

we obtain for any x:

$$f(x) = x^2 \geq 6x - 9 = 9 + 6(x - 3) = f(\bar{x}) + 6(x - \bar{x}) \ ,$$

which shows that $\xi = 6$ is a subgradient of $f(\cdot)$ at $x = \bar{x} = 3$.

Example: $f(x) = |x|$. Then it is straightforward to show that:

- If $\bar{x} > 0$, then $\xi = 1$ is a subgradient of $f(\cdot)$ at $\bar{x}$
- If $\bar{x} < 0$, then $\xi = -1$ is a subgradient of $f(\cdot)$ at $\bar{x}$
- If $\bar{x} = 0$, then $\xi \in [-1, 1]$ is a subgradient of $f(\cdot)$ at $\bar{x}$.

Theorem 12 *Let $S \subset \mathbb{R}^n$ be a convex set and $f(\cdot) : S \rightarrow \mathbb{R}$ be a convex function. If $\bar{x} \in \text{int}S$, then there exists a vector $\xi \in \mathbb{R}^n$ such that the hyperplane $H = \{(x, y) \in \mathbb{R}^{n+1} \mid y = f(\bar{x}) + \xi^t(x - \bar{x})\}$ supports* epi $f(\cdot)$ *at $(\bar{x}, f(\bar{x}))$. In particular, $f(x) \geq f(\bar{x}) + \xi^t(x - \bar{x})$ for all $x \in S$, and so ξ is a subgradient of $f(\cdot)$ at $\bar{x}$.*

Proof: Because epi $f(\cdot)$ is a convex set, and $(\bar{x}, f(\bar{x}))$ belongs to the boundary of epi $f(\cdot)$, there exists a supporting hyperplane to epi $f(\cdot)$ at $(\bar{x}, f(\bar{x}))$. Thus there exists a nonzero vector $(\xi, u) \in \mathbb{R}^{n+1}$ such that

$$\xi^t x + u\alpha \leq \xi^t \bar{x} + uf(\bar{x}) \text{ for all } (x, \alpha) \in \text{epi } f(\cdot) \ .$$

Let α be any scalar larger that $f(x)$. Then as we make α arbitrarily large, the inequality must still hold. Thus $u \leq 0$. If $u < 0$, we can re-scale so that $u = -1$. Then $\xi^t x - \alpha \leq \xi^t \bar{x} - f(\bar{x})$, which upon rearranging yields:

$$\alpha \geq f(\bar{x}) + \xi^t(x - \bar{x}) \text{ for all } (x, \alpha) \in \text{epi } f(\cdot) \ .$$

In particular, with $y = f(x)$, $f(x) \geq f(\bar{x}) + \xi^t(x - \bar{x})$, proving the theorem.

It remains to show that $u = 0$ is an impossibility. If $u = 0$, then $\xi^t x \leq \xi^t \bar{x}$ for all $x \in S$. But since $\bar{x} \in$ int S, $\bar{x} + \theta\xi \in S$ for $\theta > 0$ and sufficiently small. Thus $\xi^t(\bar{x} + \theta\xi) \leq \xi^t \bar{x}$, and so $\theta\xi^t\xi \leq 0$. But this is a contradiction unless $\xi = 0$, which cannot be true since $(\xi, u) \neq (0, 0)$. ∎

The collection of subgradients of $f(x)$ at $x = \bar{x}$ is denoted by $\partial f(\bar{x})$ and is called the *subdifferential* of $f(\cdot)$. We write $\xi \in \partial f(\bar{x})$ if ξ is a subgradient of $f(x)$ at $x = \bar{x}$.

Theorem 13 *Let $S \subset \mathbb{R}^n$ be a convex set, and let $f(\cdot) : S \to \mathbb{R}$ be a function defined on S. Suppose that for each $\bar{x} \in \text{int}S$ there exists a subgradient vector ξ. That is, for each $\bar{x} \in S$ there exists a vector ξ such that $f(x) \geq f(\bar{x}) + \xi^t(x - \bar{x})$ for all $x \in S$. Then $f(\cdot)$ is a convex function on* $\text{int}S$. ∎

Theorem 14 *Let $f(\cdot)$ be convex on $\mathbb{R}^n$, let S be a convex set and consider the problem:*

$$P: \quad \min_x \quad f(x)$$

$$s.t. \quad x \in S \ .$$

Then $\bar{x} \in S$ is an optimal solution of P if and only if $f(\cdot)$ has a subgradient ξ at $\bar{x}$ such that $\xi^t(x - \bar{x}) \geq 0$ for all $x \in S$.

Proof: Suppose $f(\cdot)$ has a subgradient ξ at $\bar{x}$ for which $\xi^t(x - \bar{x}) \geq 0$ for all $x \in S$. Then $f(x) \geq f(\bar{x}) + \xi^t(x - \bar{x}) \geq f(\bar{x})$ for all $x \in S$. Hence $\bar{x}$ is an optimal solution of P.

Conversely, suppose $\bar{x}$ is an optimal solution of P. Define the following sets:

$$A = \{(d, \alpha) \in \mathbb{R}^{n+1} \mid f(\bar{x} + d) < \alpha + f(\bar{x})\}$$

and

$$B = \{(d, \alpha) \in \mathbb{R}^{n+1} \mid \bar{x} + d \in S, \ \alpha \leq 0\} \ .$$

Both A and B are convex sets. Also $A \cap B = \emptyset$. (If not, there is a d and an α such that $f(\bar{x} + d) < \alpha + f(\bar{x}) \leq f(\bar{x})$, $\bar{x} + d \in S$, so $\bar{x}$ is not an optimal solution.)

Therefore A and B can be separated by a hyperplane $H = \{(d, \alpha) \in \mathbb{R}^{n+1} \mid \xi^t d + u\alpha = \beta\}$ where $(\xi, u) \neq 0$, such that:

- $f(\bar{x} + d) < \alpha + f(\bar{x}) \Rightarrow \xi^t d + u\alpha \leq \beta$
- $\bar{x} + d \in S, \ \alpha \leq 0 \Rightarrow \xi^t d + u\alpha \geq \beta$

In the first implication α can be made arbitrarily large, so this means $u \leq 0$. Also in the first implication setting $d = 0$ and $\alpha = \epsilon > 0$ implies that $\beta \geq \epsilon u$. In the second implication setting $d = 0$ and $\alpha = 0$ implies that $\beta \leq 0$. Thus $\beta = 0$ and $u \leq 0$. In the second implication setting $\alpha = 0$ we have $\xi^t d \geq 0$ whenever $\bar{x} + d \in S$, and so $\xi^t(\bar{x} + d - \bar{x}) \geq 0$ whenever $\bar{x} + d \in S$. Put another way, we have $x \in S$ implies that $\xi^t(x - \bar{x}) \geq 0$.

It only remains to show that ξ is subgradient. Note that $u < 0$, for if $u = 0$ it would follow from the first implication that $\xi^t d \leq 0$ for any d, a contradiction. Since $u < 0$, we can re-scale so that $u = -1$.

Now let d be given so that $(\bar{x} + d) \in S$ and let $\alpha = f(\bar{x} + d) - f(\bar{x}) + \epsilon$ for some $\epsilon > 0$. Then $f(\bar{x} + d) < \alpha + f(\bar{x})$, and it follows from the first implication that $\xi^t d - f(\bar{x} + d) + f(\bar{x}) - \epsilon \leq 0$ for all $\epsilon > 0$. Thus $f(\bar{x} + d) \geq f(\bar{x}) + \xi^t d$ for all $\bar{x} + d \in S$. Setting $x = \bar{x} + d$, we have that if $x \in S$,

$$f(x) \geq f(\bar{x}) + \xi^t(x - \bar{x}),$$

and so ξ is a subgradient of $f(\cdot)$ at $\bar{x}$. ∎

7 Differentiable Convex Functions

Let $f(\cdot) : S \to \mathbb{R}$ be given. $f(\cdot)$ is *differentiable* at $\bar{x} \in \text{int}S$ if there is a vector $\nabla f(\bar{x})$, the *gradient vector*, such that $f(x) = f(\bar{x}) + \nabla f(\bar{x})^t(x - \bar{x}) + \|x - \bar{x}\|\alpha(\bar{x}; x - \bar{x})$, where $\lim_{x \to \bar{x}} \alpha(\bar{x}; x - \bar{x}) = 0$. Note that

$$\nabla f(\bar{x}) = \left(\frac{\partial f(\bar{x})}{\partial x_1}, \ldots, \frac{\partial f(\bar{x})}{\partial x_n} \right) .$$

Lemma 7.1 *Let $f(\cdot) : S \to \mathbb{R}$ be convex. If $f(\cdot)$ is differentiable at $\bar{x} \in \text{int}S$, then the collection of subgradients of $f(\cdot)$ at $\bar{x}$ is the singleton set $\{\nabla f(\bar{x})\}$, i.e., $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$.*

Proof: The set of subgradients of $f(\cdot)$ at $\bar{x}$ is nonempty by Theorem 12. Let ξ be a subgradient of $f(\cdot)$ at $\bar{x}$. Then for any d and any $\lambda > 0$ we have:

$$f(\bar{x} + \lambda d) \geq f(\bar{x}) + \xi^t(\bar{x} + \lambda d - \bar{x}) = f(\bar{x}) + \lambda \xi^t d \ .$$

Also

$$f(\bar{x}+\lambda d) = f(\bar{x}) + \lambda \nabla f(\bar{x})^t d + \lambda \|d\| \alpha(\bar{x};\lambda d) \ .$$

Subtracting, we have

$$0 \geq \lambda(\xi - \nabla f(\bar{x}))^t d - \lambda \|d\| \alpha(\bar{x};\lambda d)$$

which is equivalent to:

$$0 \geq (\xi - \nabla f(\bar{x}))^t d - \|d\| \alpha(\bar{x};\lambda d) \ .$$

As $\lambda \to 0$ we have $\alpha(\bar{x};\lambda d) \to 0$. So $(\xi - \nabla f(\bar{x}))^t d \leq 0$ for any d. This can only mean that $\xi - \nabla f(\bar{x})$. ∎

Theorem 15 *If $f(\cdot) : S \to \mathbb{R}$ is differentiable, then $f(\cdot)$ is convex if and only if*

$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^t (x - \bar{x}) \text{ for all } x, \bar{x} \in \text{int } S.$$

∎

8 Exercises on Convex Sets and Functions

1. Let K^* denote the *dual cone* of the closed convex cone $K \subset \mathbb{R}^n$, defined by:

$$K^* := \{y \in \mathbb{R}^n \mid y^T x \geq 0 \text{ for all } x \in K\} \ .$$

   Prove that $(K^*)^* = K$, thus demonstrating that the dual of the dual of a closed convex cone is the original cone.

2. Let $\mathbb{R}^n_+$ denote the nonnegative orthant, namely $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x_j \geq 0, j = 1, \ldots, n\}$. Considering $\mathbb{R}^n_+$ as a cone, prove that $\left(\mathbb{R}^n_+\right)^* = \mathbb{R}^n_+$, thus showing that $\mathbb{R}^n_+$ is self-dual.

3. Let $Q^n = \left\{x \in \mathbb{R}^n \mid x_1 \geq \sqrt{\sum_{j=2}^n x_j^2}\right\}$. Q^n is called the second-order cone, the Lorentz cone, or the ice-cream cone (I am not making this up). Considering Q^n as a cone, prove that $(Q^n)^* = Q^n$, thus showing that Q^n is self-dual.

4. Prove Lemma 1 of the notes on Analysis of Convex Sets.

5. Let S_1 and S_2 be two nonempty sets in $\mathbb{R}^n$, and define $S_1 \bigoplus S_2 := \{x \mid x = x_1 + x_2 \text{ for some } x_1 \in S_1, x_2 \in S_2\}$.

   (a) Show that $S_1 \bigoplus S_2$ is a convex set if both S_1 and S_2 are convex sets.

   (b) Show by an example that $S_1 \bigoplus S_2$ is not necessarily a closed set, even if both S_1 and S_2 are closed convex sets.

   (c) Show that if either S_1 or S_2 is a bounded convex set, and both S_1 and S_2 are closed sets, then $S_1 \bigoplus S_2$ is a closed set.

6. Suppose that $f(\cdot)$ is a convex function on $\mathbb{R}^n$. Prove that ξ is a subgradient of $f(\cdot)$ at $\bar{x}$ if and only if

$$f^{'}(\bar{x}, d) \geq \xi^t d$$

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

7. Prove Theorem 13 of the notes.

8. Consider the function $f(\cdot) : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$. Such a function is called an *extended* real-valued function. The *epigraph* of $f(\cdot)$ is defined as:

$$\text{epi} f(\cdot) := \{(x, \alpha) \mid f(x) \leq \alpha\} \ .$$

   We define $f(\cdot)$ to be a convex function if epi$f(\cdot)$ is a convex set. Show that if $f(\cdot) : \mathbb{R}^n \to \mathbb{R}$, then this definition of a convex function is equivalent to the standard definition of a convex function.

9. In class, we stated the meta result that the study of convex functions reduces to the study of convex functions on the real line. This exercise formalizes this observation and shows how we might use it to obtain results about convex function from results about convex functions on the real line.

   (a) Let $f(x)$ be a real-valued function defined on an open set $X \in \mathbb{R}^n$. Show that $f(x)$ is a convex function if and only if for any two points $x, y \in X$, the function $g(\theta) := f(\theta x + (1 - \theta)y)$ of the scalar θ is convex on the open interval $\theta \in (0, 1)$.

(b) Suppose that we have proved that a twice differentiable function $g(\theta)$ of the scalar θ is convex on the open interval $\theta \in (0, 1)$ if and only if its second derivative is nonnegative for every point in this interval. Use this fact, part **(a)**, and the chain rule to show that a twice differentiable real-valued function $f(x)$ defined on an open set $X \in \mathbb{R}^n$ is convex if and only if its Hessian matrix is positive semi-definite at every point in X.

10. **(a)** Let $f(x)$ be a real-valued function defined on an open interval $I = (l, u)$ of the real line. For any two points $a < b$ in (l, u) , let $S(a, b) := \frac{f(b)-f(a)}{b-a}$ be the secant slope of $f(x)$ between the points a and b. Prove the following result:

    Three Cord Lemma: $f(x)$ is a convex function on an the interval $I = (l, u)$ of the real line if and only if for any three points $a < b < c$ in I we have $S(a, b) \leq S(a, c) \leq S(b, c)$.■

    (b) Use the Three Cord Lemma and the mean value theorem to show that a twice differentiable function $f(x)$ of the scalar x is convex on the open interval $I = (l, u)$ of the real line if and only if its second derivative is nonnegative for every point in this interval.

    (Hint: What does the Three Cord Lemma in part **(a)** say about the relationship between the derivative of $f(x)$ at the points a and b ?)

11. Let S be a nonempty bounded convex set in $\mathbb{R}^n$, and define $f(\cdot) : \mathbb{R}^n \to \mathbb{R}$ as follows:

$$f(y) := \sup_x \{y^t x \mid x \in S\} .$$

    The function $f(\cdot)$ is called the *support function* of S.

    (a) Prove that $f(\cdot)$ is a convex function.

    (b) Suppose that $f(\bar{y}) = \bar{y}^t \bar{x}$ for some $\bar{x} \in S$. Show that $\bar{x} \in \partial f(\bar{y})$.

12. Let $f(\cdot) : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Show that if $\xi \in \partial f(\bar{x})$, then the hyperplane

$$H := \{(x, \alpha) \mid \alpha = f(\bar{x}) + \xi^t (x - \bar{x})\}$$

    is a supporting hyperplane of epi$f(\cdot)$ at $(\bar{x}, f(\bar{x}))$.

13. Let $f(\cdot) : \mathbb{R}^n \to \mathbb{R}$ be a convex function, and let $\bar{x}$ be given. Show that $\partial f(\bar{x})$ is a closed convex set.

14. Let $f_1(\cdot), \ldots, f_k(\cdot) : \mathbb{R}^n \to \mathbb{R}$ be differentiable convex functions, and define:
$$f(x) := \max\{f_1(x), \ldots, f_k(x)\} \ .$$
Let $I(x) := \{i \mid f(x) = f_i(x)\}$. Show that
$$\partial f(\bar{x}) = \left\{ \xi \mid \xi = \sum_{i \in I(x)} \lambda_i \nabla f_i(\bar{x}), \sum_{i \in I(x)} \lambda_i = 1, \lambda_i \geq 0 \text{ for } i \in I(x) \right\} \ .$$

15. Consider the function $L^*(u)$ defined by the following optimization problem, where X is a compact polyhedral set:
$$\begin{array}{rl} L^*(u) := \min_x & c^t x + u^t(Ax - b) \\ \\ \text{s.t.} & x \in X \ . \end{array}$$
   (a) Show that $L^*(\cdot)$ is a concave function.
   (b) Characterize the set of subgradients of $L^*(\cdot)$ at any given $u = \bar{u}$.