CPT Lecture Notes 5: Convexity

Levent Ülkü

August 30, 2019

Convexity: Basic definitions

A nonempty set $S \subseteq \Re^m$ is convex if for every $x, x' \in S$ and every $t \in [0,1]$, $tx + (1-t)x' \in S$. A convex combination of vectors $x_1, \ldots, x_n$ is a vector of the form $\sum_{i=1}^n \alpha_i x_i$ where $\alpha_1, \ldots, \alpha_n$ are nonnegative numbers which add up to 1. Define

$$\tilde{S} := \begin{array}{c} \{\sum_{i=1}^n \alpha_i x_i : n \in \mathbb{N},\ x_i \in S \text{ for all } i, \\ \alpha_i \geq 0 \text{ for all } i, \text{ and } \sum_{i=1}^n \alpha_i = 1\}. \end{array}$$

Hence $\tilde{S}$ is the set of all convex combinations of members of S. Taking $n = 1$ (and $\alpha_1 = 1$ necessarily) we observe that $S \subseteq \tilde{S}$. We also have:

Theorem: $S \subseteq \Re^m$ is convex if and only if $S = \tilde{S}$, i.e., if and only if it contains all convex combinations of its elements.

Proof: HW.

More HW... Show that:

1. Arbitrary intersections of convex sets are convex.

2. $S + T := \{s + t : s \in S, t \in T\}$ is convex if S and T are convex.

3. For every scalar $\lambda \geq 0$, $\lambda S := \{\lambda s : s \in S\}$ is convex if S is convex.

4. The closure and interior (using the Euclidean metric) of a convex set are convex.

The convex hull of a set $S \subseteq \Re^m$, denoted coS, is the "smallest" convex superset of S, i.e,

$$coS := \cap\{G \subseteq \Re^m : S \subseteq G \text{ and } G \text{ is convex}\}.$$

Note that coS is convex and $S \subseteq coS$.

Theorem: $coS = \tilde{S}$.

Proof: HW

Consequently, we have:

$$S \text{ is convex } \Leftrightarrow coS \subseteq S$$

Analogy: S is closed if and only if it contains its closure.

The Caratheodory Theorem

The Caratheodory theorem is a much sharper version of the statement $coS \subseteq \tilde{S}$. Recall:

1. A collection of vectors $x_1, ..., x_k$ in $\Re^m$ is linearly independent if for every collection of numbers $\alpha_1, ..., \alpha_k$, we have $\sum_{i=1}^k \alpha_i x_i = 0 \Rightarrow \alpha_1 = ... = \alpha_k = 0$.

2. Any collection of k vectors in $\Re^m$ is linearly dependent if $k > m$. This means that there are numbers $\alpha_1, ..., \alpha_k$, not all zero, such that $\sum_{i=1}^k \alpha_i x_i = 0$.

Theorem: (Carathéodory)

Let $S \subseteq \Re^m$ be nonempty. If $x \in coS$, then x can be written as a convex combination of no more than $m+1$ members of S, i.e., there exist $x_1, x_2, ..., x_{m+1} \in S$ and $\alpha_1, \alpha_2, ..., \alpha_{m+1} \geq 0$ with $\sum_{i=1}^{m+1} \alpha_i = 1$ such that $x = \sum_{i=1}^{m+1} \alpha_i x_i$.

Proof:

Fix $x \in coS$.

Let $A = \{n \in \mathbb{N} : x \text{ is a convex combination of } n \text{ vectors in } S\}$.

$A \neq \emptyset$ since $coS = \tilde{S}$. Let $k = \min A$. We need to show that $k \leq m + 1$.

Suppose not: $k > m + 1$. Pick $x_1, ..., x_k \in S$ and strictly positive (why?) constants $\alpha_1, ..., \alpha_k$ adding up to 1 such that $x = \sum_{i=1}^{k} \alpha_i x_i$.

The $k - 1$ vectors $x_2 - x_1, x_3 - x_1, ..., x_k - x_1 \in \mathbb{R}^m$ are linearly dependent since $k - 1 > m$.

Pick constants $\theta_2, ..., \theta_k$ at least one strictly positive such that $\sum_{i=2}^{k} \theta_i (x_i - x_1) = 0$. (Why can we pick θ's such that at least one is strictly positive?)

Let $\sigma_1 = -\sum_{i=2}^{k} \theta_i$ and $\sigma_i = \theta_i$ for $i = 2, ..., k$.

Then (1) $\sigma_j > 0$ (since $\theta_j > 0$), (2)$\sum_{i=1}^{k} \sigma_i x_i = 0$ and (3) $\sum_{i=1}^{k} \sigma_i = 0$.

Let $\beta = \min\{\frac{\alpha_i}{\sigma_i} : \sigma_i > 0\}$.

Note that $\beta > 0$ since $\alpha_i > 0$ for all i.
Pick some l such that $\frac{\alpha_l}{\sigma_l} = \beta > 0$.

Note that (1) $\alpha_i - \beta\sigma_i \geq 0$ for each i, (2) $\alpha_l - \beta\sigma_l = 0$, (3) $\sum_{i=1}^{k} (\alpha_i - \beta\sigma_i) = 1$ and (4) $x = \sum_{i=1}^{k} (\alpha_i - \beta\sigma_i) x_i$.

Thus $k \neq \min A$, a contradiction.■

More HW... Show that:

1. $co(\sum_{i=1}^{n} S_i) = \sum_{i=1}^{n} coS_i$

2. If $A \subset \mathbb{R}^m$ is open, then coA is open.

3. The convex hull of a closed set in $\mathbb{R}^m$ need not be closed. But if $K \subset \mathbb{R}^m$ is compact, then coK is compact.

The Shapley-Folkman Theorem

Theorem: (Shapley-Folkman)

Let $S_i \subseteq \Re^m$ for every $i = 1, ..., n$, and let $x \in co \sum_{i=1}^n S_i$. Then there exist $x_1, ..., x_n$ such that

1. $x_i \in coS_i$ for every i,
2. $x = \sum_{i=1}^n x_i$, and
3. $\#\{i : x_i \notin S_i\} \leq m$.

Remark: Without 3, the theorem is trivial as $co(\sum_{i=1}^{n} S_i) \subseteq \sum_{i=1}^{n} coS_i$. (Homework.) With 3, it says quite a bit. Fix m and let n be large. Then $co \sum_{i=1}^{n} S_i$ is "almost a subset of" $\sum_{i=1}^{n} S_i$ making $\sum_{i=1}^{n} S_i$ "almost convex."

Remark: If $x \in \Re^m$ can be written as $x = \sum_{i=1}^{k} \alpha_i x_i$ where $\alpha_1, ..., \alpha_k \in \Re_+$, $x_1, ..., x_k \in \Re^m$ and if $k > m$, then there exist $\beta_1, ..., \beta_k \in \Re_+$ with $\#\{i : \beta_i > 0\} \leq m$ such that $x = \sum_{i=1}^{k} \beta_i x_i$. This has nothing to do with Caratheodory, since $\sum_{i=1}^{k} \alpha_i x_i$ and $\sum_{i=1}^{k} \beta_i x_i$ are linear combinations. (This means the scalars α_i and β_i do not need to add up to 1.)

Proof of the SF Theorem: (Lin Zhou)

Fix $x \in co(\sum_{i=1}^{n} S_i)$.

Since $co(\sum_{i=1}^{n} S_i) \subseteq \sum_{i=1}^{n} coS_i$, there exist $x_1, ..., x_n$ such that $x_i \in coS_i$ for every i and $x = \sum_i x_i$.

If $n \leq m$, then the proof is complete as $\#\{i : x_i \notin S_i\} \leq n$.

Suppose that $n > m$ and, using the Caratheodory Theorem, write

$$x = \sum_{i=1}^{n} \underbrace{\sum_{j=1}^{m+1} \alpha_{ij} x_{ij}}_{=x_i}$$

where, for every i, $x_{ij} \in S_i$, $\alpha_{ij} \geq 0$ and $\sum_{j=1}^{m+1} \alpha_{ij} = 1$.

Define the following vectors in $\Re^{m+n}$:

$$z = \begin{bmatrix} x \\ \mathbf{1}_{n\times 1} \end{bmatrix}, \; z_{ij} = \begin{bmatrix} x_{ij} \\ \mathbf{e}^{i}_{n\times 1} \end{bmatrix} \text{ for every } i = 1, \ldots, n \text{ and } j = 1, \ldots, m+1$$

Note that $\mathbf{1} = \sum_{i=1}^{n} \sum_{j=1}^{m+1} \alpha_{ij} \mathbf{e}^{i}$. Hence $z = \sum_{i=1}^{n} \sum_{j=1}^{m+1} \alpha_{ij} z_{ij}$.

Since $n(m+1) > m+n$, there must exist β_{ij}, $i = 1, \ldots, n$ and $j = 1, \ldots, m+1$ such that (recall the remark before the proof)

1. $\beta_{ij} \geq 0$ and $\beta_{ij} \neq 0$ for at most $m+n$ of the $n(m+1)$ indices ij, and

2. $z = \sum_{i=1}^{n} \sum_{j=1}^{m+1} \beta_{ij} z_{ij}$

Note that, by construction, $\sum_{j=1}^{m+1} \beta_{ij} = 1$ for every $i = 1, ..., n$. Again by construction, we have:

$$x = \sum_{i=1}^{n} \underbrace{\sum_{j=1}^{m+1} \beta_{ij} x_{ij}}_{= y_i \in coS_i}, \ \beta_{ij} \geq 0 \text{ and } \sum_{j=1}^{m+1} \beta_{ij} = 1.$$

Now a little counting is in order.

For every i there is some j such that $\beta_{ij} > 0$.

There exist at most m more indices ij with $\beta_{ij} > 0$.

Hence $\{i : \sum_{j=1}^{m+1} \beta_{ij} x_{ij} \in S_i\}$ has at least $n - m$ elements.

This finishes the proof. (Why?) ■

The Seperating Hyperplane Theorem via Minkowski

Fix $p \in \Re^m$ and $\alpha \in \Re$.

The hyperplane formed by p and α:

$$H(p;\alpha) := \{x \in \Re^m : p \cdot x = \alpha\}.$$

p is called the normal vector of $H(p;\alpha)$.

We are interested in separating convex sets with hyperplanes.

Theorem: (Minkowski)

Let $S \subseteq \Re^m$ be nonempty, convex and closed and let $\bar{x} \notin S$. There exists $p \in \Re^m \backslash \{0\}$ and $x_0 \in S$ such that $p \cdot \bar{x} > p \cdot x_0 \geq p \cdot x$ for every $x \in S$.

Proof:

Step 1: Define $g : \Re^m \longrightarrow \Re$ by $g(x) = (x - \bar{x}) \cdot (x - \bar{x})$.

Note that g is continuous.

Fix $r > 0$ such that $clB(\bar{x}, r) \cap S \neq \varnothing$.

Since $clB(\bar{x}, r) \cap S$ is closed and bounded, there exists some

$$x_0 \in \arg\min_{x \in clB(\bar{x},r) \cap S} g(x).$$

We have $g(x_0) \leq r^2 < g(y)$ for every $y \in S \backslash clB(\bar{x}, r)$.

Thus $x_0 \in \arg\min_{x \in S} g(x)$.

Step 2: Let $p = \bar{x} - x_0$.

Note that $p \neq 0$ and therefore that

$$\begin{aligned} 0 &< (\bar{x} - x_0) \cdot (\bar{x} - x_0) \\ &= p \cdot (\bar{x} - x_0). \end{aligned}$$

Step 3: Now we need to show that for every $x \in S$, $p \cdot x_0 \geq p \cdot x$.

Fix $x \in S$ and $t \in (0,1)$. Since $tx + (1-t)x_0 \in S$ we have

$$\begin{aligned}
g(x_0) &= (x_0 - \bar{x}) \cdot (x_0 - \bar{x}) \\
&\leq g(tx + (1-t)x_0) \\
&= (tx + (1-t)x_0 - \bar{x}) \cdot (tx + (1-t)x_0 - \bar{x}) \\
&= (x_0 - \bar{x} + t(x - x_0)) \cdot (x_0 - \bar{x} + t(x - x_0)) \\
&= (x_0 - \bar{x}) \cdot (x_0 - \bar{x}) + 2t(x - x_0) \cdot (x_0 - \bar{x}) \\
&\quad + t^2 (x - x_0) \cdot (x - x_0).
\end{aligned}$$

This gives us

$$0 \leq 2t(x - x_0) \cdot (x_0 - \bar{x}) + t^2 (x - x_0) \cdot (x - x_0)$$

or, since $t \neq 0$, $0 \leq 2(x - x_0) \cdot (x_0 - \bar{x}) + t(x - x_0) \cdot (x - x_0)$. Now letting $t \to 0$, we get, using continuity of the RHS in t, $0 \leq (x - x_0) \cdot (x_0 - \bar{x}) = (x - x_0) \cdot (-p)$, which is the desired result. ■

We know: If x_n is a sequence in $\Re^m$ with $\|x_n\| = 1$ for every n, then x_n contains a convergent subsequence. (Right?)

Theorem: Suppose that $S \subseteq \Re^m$ is nonempty and convex and that x_n is a sequence in $\Re^m \backslash clS$. If $x_n \to \bar{x}$, then there exists $p \in \Re^m \backslash \{0\}$ such that for every $x \in S$, $p \cdot x \leq p \cdot \bar{x}$.

Proof:

Fix $x \in clS$. Note that clS is convex since S is convex.

By Minkowski, for every n, there exists $q_n \neq 0$ such that $q_n \cdot x_n \geq q_n \cdot x$.

Normalizing, $\frac{q_n}{\|q_n\|} \cdot x_n \geq \frac{q_n}{\|q_n\|} \cdot x$.

Let $p_n = \frac{q_n}{\|q_n\|}$ so that $\|p_n\| = 1$.

Take a convergent subsequence p_{n_k} of p_n, with limit p. Note such that $\|p\| = 1$.

$p_{n_k} \cdot x_{n_k} \geq p_{n_k} \cdot x$. Take limits to get $p \cdot \bar{x} \geq p \cdot x$.■

Theorem: (Supporting Hyperplane) Suppose that $S \subseteq \Re^m$ is nonempty and convex. If $\bar{x} \in \partial S$, then there exists $p \in \Re^m \backslash \{0\}$ such that $p \cdot \bar{x} \geq p \cdot x$ for every $x \in S$.

Proof:

For every $n = 1, 2, ...,$ choose $x_n \in B_{\frac{1}{n}}(\bar{x}) \cap [\Re^m \backslash clS]$.

Note that $x_n \longrightarrow \bar{x}$ and $x_n \notin clS$.

By the previous theorem, there exists $p \neq 0$ such that $p \cdot \bar{x} \geq p \cdot x$ for every $x \in S$.$\blacksquare$

Theorem: Suppose that $S \subseteq \Re^m$ is a nonempty and convex set and let $\bar{x} \notin S$. Then there exists $p \in \Re^m \backslash \{0\}$ such that $p \cdot x \leq p \cdot \bar{x}$ for every $x \in S$.

Proof:

Case 1: $\bar{x} \notin clS$. The result follows from Minkowski.

Case 2: $\bar{x} \in clS$. Then $\bar{x} \in \partial S$ and the result follows from the previous theorem.■

Theorem: (Seperating Hyperplane) Let S and T be disjoint and convex subsets of $\Re^m$. There exists $p \in \Re^m \setminus \{0\}$ such that $p \cdot s \leq p \cdot t$ for every $(s, t) \in S \times T$.

Proof:

$S - T$ is convex.

$0 \notin S - T$.

For every $(s, t) \in S \times T$, $s - t \in S - T$.

By the previous theorem, there exists a nonzero vector p such that $p \cdot (s - t) \leq 0$. $\blacksquare$

Convex and Concave Functions

Let $S \subseteq \mathbb{R}^n$ be convex. A function $f : S \to \mathbb{R}$ is:

convex if $f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$
$\forall x, y \in S, t \in [0, 1]$.

strictly convex if $f(tx + (1-t)y) < tf(x) + (1-t)f(y)$
$\forall x, y \in S, t \in (0, 1)$.

concave if $f(tx + (1-t)y) \geq tf(x) + (1-t)f(y)$
$\forall x, y \in S, t \in [0, 1]$.

strictly concave if $f(tx + (1-t)y) > tf(x) + (1-t)f(y)$
$\forall x, y \in S, t \in (0, 1)$.

Homework: Let $S \subseteq \mathbb{R}^n$ be convex and $f : S \to \mathbb{R}$. Show that (1) f is convex if and only if $\{(x, y) \in \mathbb{R}^{n+1} : y \geq f(x)\}$ is convex; (2) f is concave if and only if $\{(x, y) \in \mathbb{R}^{n+1} : y \leq f(x)\}$ is convex.

Suppose $S \subseteq \mathbb{R}^n$ is convex and $f : S \to \mathbb{R}$.

A vector $p \in \mathbb{R}^n$ is a subgradient for f at $x \in S$ if

$$f(y) \geq f(x) + p \cdot (y - x) \text{ for all } y \in S.$$

Supergradient defined with the reverse inequality.

Theorem: Suppose that $S \subseteq \mathbb{R}^n$ is convex and $f : S \rightarrow \mathbb{R}$. If f has a subgradient at every $x \in S$, then f is convex.

Proof:

Fix $x, y \in S$ and $t \in [0, 1]$. Let $z = tx + (1 - t)y$. Note $z \in S$.

Let p be a subgradient of f at z, i.e.,

$$\begin{aligned} f(x) &\geq f(z) + p \cdot (x - z) \\ f(y) &\geq f(z) + p \cdot (y - z) \end{aligned}$$

Multiply the first inequality with t, the second with $1 - t$ and sum:

$$f(z) \leq tf(x) + (1 - t)f(y). \blacksquare$$

We will skip the proof of the following result.

Theorem: Suppose that $S \subseteq \mathbb{R}^n$ is convex, $f : S \to \mathbb{R}$ is convex, and $x \in intS$. Then f is continuous at x.

Theorem: Suppose that $S \subseteq \mathbb{R}^n$ is convex, $f : S \to \mathbb{R}$ is convex, and $x \in intS$. Then f has a subgradient at x.

Proof:

Choose $x \in intS$.

Let $A = \{(z, y) \in \mathbb{R}^{n+1} : y \geq f(z)\}$.

Note that A is convex because f is convex (right?) but A need not be closed.

Step 1: We will show that for every $\varepsilon > 0$, $(x, f(x) - \varepsilon) \notin clA$.

If not, there exists $\{(x_k, y_k)\}$ in A with limit $(x, f(x) - \varepsilon) \in A$.

Hence $y_k \geq f(x_k)$ for every k.

Since f is continuous at x (because it is convex, see the previous result), the inequality is preserved at the limit: $f(x) - \varepsilon \geq f(x)$, a contradiction to $\varepsilon > 0$.

Step 2: By Step 1, $\{(x, f(x) - \frac{1}{k})\}$ is a sequence outside clA.

By a theorem we proved in the build-up to the Separating Hyperplane Theorem, the limit $(x, f(x))$ and the set clA (and therefore the set A) can be separated.

Hence there exists $(q, r) \in \mathbb{R}^{n+1} \backslash \{0\}$ such that $q \cdot a + rb \geq q \cdot x + rf(x)$ whenever $(a, b) \in A$, i.e., whenever $b \geq f(a)$.

Step 3: We will show that $r > 0$.

Since $(x, f(x) + 1) \in A$, $q \cdot x + r(f(x) + 1) \geq q \cdot x + rf(x)$, giving $r \geq 0$.

Suppose towards a contradiction that $r = 0$.

Then $q \cdot a \geq q \cdot x$ whenever $a \in S$.

$x \in intS : \exists \varepsilon > 0$ such that $x \pm \varepsilon e_i \in S$ for every $i = 1, ..., n$.

Hence $q \cdot (x + \varepsilon e_i) \geq q \cdot x$ and $q \cdot (x - \varepsilon e_i) \geq q \cdot x$ for every i.

Hence $\varepsilon q_i \geq 0 \geq \varepsilon q_i$ for every i, where q_i is the ith coordinate of q.

Hence $q = 0 \in \mathbb{R}^n$, and consequently $(q, r) = 0 \in \mathbb{R}^{n+1}$, a contradiction.

Step 4:

Let $p = \frac{-1}{r} q \in \mathbb{R}^n$.

We will show that p is a subgradient of f at x.

Take any $y \in S$.

By Step 3, $q \cdot y + rf(y) \geq q \cdot x + rf(x)$.

Multplying both hand sides with $\frac{-1}{r}$ gives $p \cdot y - f(y) \leq p \cdot x - f(x)$.

Rearranging, $f(y) \geq f(x) + p \cdot (y - x)$. $\blacksquare$

Theorem: Suppose that $S \subseteq \mathbb{R}^n$ is convex, $f : S \rightarrow \mathbb{R}$ is convex, $x \in intS$ and f is differentiable at x. Then $\nabla f(x)$ is the unique subgradient of f at x.

Proof:

Choose $\varepsilon > 0$ such that $B(x, \varepsilon) \subset S$.

Apply the last result to conclude that there exists p such that $f(y) - p \cdot y \geq f(x) - p \cdot x$ for every $y \in B(x, \varepsilon)$.

Let $g(y) = f(y) - p \cdot y$.

It follows that x minimizes $g(y)$ on $B(x, \varepsilon)$ and $\nabla g(x) = 0$, i.e., $p = \nabla f(x)$. ■

Theorem: Suppose that $S \subseteq \mathbb{R}^n$ is convex and $f : S \to \mathbb{R}$. If p_1 is a subgradient of f at x_1 and p_2 is a subgradient of f at x_2, then $(p_1 - p_2) \cdot (x_1 - x_2) \geq 0$.

Proof: If p_i is a subgradient of f at x_i, then $f(x_2) \geq f(x_1) + p_1 \cdot (x_2 - x_1)$ and $f(x_1) \geq f(x_2) + p_2 \cdot (x_1 - x_2)$. Manipulating, $(p_1 - p_2) \cdot (x_1 - x_2) \geq 0$. ■

Theorem: Suppose that $S \subseteq \mathbb{R}^n$ is open and convex, and $f : S \to \mathbb{R}$ is differentiable and convex. Then $(\nabla f(x_1) - \nabla f(x_2)) \cdot (x_1 - x_2) \geq 0$.

Proof: Skip.

Note that if $n = 1$ and if f is differentiable and convex, $x_1 \geq x_2$ implies $f'(x_1) \geq f'(x_2)$. Hence $(x_1 - x_2)(f'(x_1) - f'(x_2)) \geq 0$. The previous theorem is the generalization of this observation to $n > 1$.