Useful Facts, Theorems and Inequalities

Bharath B. N.

Dept. of ECE, Indian Institute of Science, Bangalore, India

Email: bharath@ece.iisc.ernet.in, cmurthy@ece.iisc.ernet.in

Abstract

My name is Bharath B. N. I am a PhD. student at Indian Institute of Science (IISC) Bangalore, India. The main intention of this blog is to put down some of the interesting facts and theorems that seems to be useful in communication engineering research.

I. Reverse Jensen's Inequality

Jensen's inequality is a tool typically used in probability theory to lower bound the expected value of a convex function. I came across an interesting inequality that upper bounds the expected value of a convex function (see blog of Dr. Radhakrishna Ganti or the reference there in) which is stated as a theorem below.

Theorem *1:* Let $f : \mathcal{R} \rightarrow \mathcal{R}$ be a convex function. Then

$$\mathbb{E}\{f(X)\} \leq f(\mathbb{E}X) + Var(X). \tag{1}$$

The above theorem when applied to $f(X) := \exp(-X)$ results in

$$\mathbb{E}\exp(-X) \leq \exp\{-\mathbb{E}X\} + Var(X).$$

Suppose if $X \geq 0$ a.s., then the following bound can be obtained whose proof is simple.

Theorem *2:* Let $f(X) := \exp(-X)$. If $X \geq 0$ a.s., then the following inequality is true

$$\mathbb{E}\{\exp(-X)\} \leq \exp(-\mathbb{E}X) + \frac{Var(X)}{2}. \tag{2}$$

Proof: The proof involves moment method. For $X \geq 0$, using Taylor series expansion of $\exp(-X)$ around $\mathbb{E}X$, we get

$$\exp(-X) = \exp(-\mathbb{E}X) - (X - \mathbb{E}X)\exp(-\mathbb{E}X) + \frac{(X-\mathbb{E}X)^2}{2}\exp(-X^*),\ \ X^* \in [0, \infty].$$

Taking the expectation of the above, we get

$$\mathbb{E}\exp(-X) = \exp(-\mathbb{E}X) + \frac{Var(X)}{2}\exp(-X^*) \tag{3}$$

$$\leq \exp(-\mathbb{E}X) + \frac{Var(X)}{2}, \tag{4}$$

where the last inequality above follows from the fact that $\exp(-X^*) \leq 1$ for all $X^* \in [0, \infty]$. $\square$

Note that for the special case of $X \geq 0$ a.s, the above inequality is tight compared to the inequality in Theorem 2.

II. RADON'S LEMMA

We begin with some definitions.

Definition: A set $C \subseteq \mathcal{R}^n$ is said to be convex if for all $x_1, x_2 \in C$, the convex combination $\lambda x_1 + (1-\lambda)x_2 \in C$ for all $\lambda \in [0,1]$.

Definition: Consider the set of m points denoted $S := \{s_1, \dots, s_m\} \subseteq \mathcal{R}^n$. The convex hull of S denoted $Conv(S)$ is defined as

$$Conv(S) := \{x \in \mathcal{R}^n : x = \sum_{i=1}^{m} \lambda_i s_i, \sum_{i=1}^{m} \lambda_i = 1,\ \lambda_i \geq 0\}.$$

Consider five points $S := (s_1, s_2, s_3, s_4, s_5)$ which are non-collinear (points that do not lie on a line) in $\mathcal{R}^2$. Now, imagine the convex hull of these points denoted

$$C := Conv(s_1, s_2, s_3, s_4, s_5).$$

Consider *any* two sets $S_1, S_2 \in S$ such that $|S_1| = |S_2| = 3$. I encourage you to draw picture to help visualize this. Use these sets to construct $Conv(S_1)$ and $Conv(S_2)$. Now, it is easy to see geometrically these two convex sets have at least one point in common, i.e., $Conv(S_1) \bigcap Conv(S_2) \neq \emptyset$. In fact, the intersection is a triangle. This geometrical intuition nicely generalizes to an $n-$dimensional space; this is is the essence of the following lemma due to Radon.

Theorem *3:* **(Radon's Lemma)** Suppose that $S \subseteq \mathcal{R}^n$ has only $n+2$ points in it. Then, there exists $S_1, S_2 \in S$ such that

$$Conv(S_1) \bigcap Conv(S_2) \neq \emptyset.$$

Proof: Let the $n+2$ points in S be $s_1, \dots, s_{n+2}$. We embed this into $\mathcal{R}^{n+1}$ by appending 1 to each point in S. Thus, we have

$$S' := \left\{ \begin{pmatrix} s_1 \\ 1 \end{pmatrix}, \dots, \begin{pmatrix} s_{n+2} \\ 1 \end{pmatrix} \right\} \subseteq \mathcal{R}^{n+1}. \tag{5}$$

Since $S' \subseteq \mathcal{R}^{n+1}$, there exist λ_i's, not all zeros, such that

$$\sum_{i=1}^{n+2} \lambda_i s_i = 0. \tag{6}$$

Now, we collect the positive and negative values of λ_i's in the following sets:

$$P := \{i : \lambda_i \geq 0\}, \text{ and } Q := \{i : \lambda_i < 0\}.$$

Using this, and (6), we have

$$\sum_{i \in P} \lambda_i = -\sum_{i \in Q} \lambda_i := q > 0.$$

We claim that

$$Conv(s_i : i \in P) \bigcap Conv(s_i : i \in Q) \neq \emptyset,$$

and the required sets are the one above. Now, it follows that $u := \sum_{i \in P} \frac{\lambda_i}{q} s_i = -\sum_{i \in Q} \frac{\lambda_i}{q} s_i$ is the desired common vector. □

III. Caratheodry's Theorem

In this section, I state and prove Caratheodry's theorem. Before stating the result for a general convex hull, I give results to a simple structure like convex cones.

Definition (Cone): A set $K \subseteq \mathcal{R}^n$ is said to be a cone if for all $x \in K$, $tx \in K$ for all $t \geq 0$.

Note that zero is contained in all cones. Similar to convex hull operation of given points/set in say $\mathcal{R}^n$, one can also convert these points/set to a cone. Let us call this operation as *confication* denoted $Cone(K)$, $K \subseteq \mathcal{R}^n$. Let us denote the set of all cones in $\mathcal{R}^n$ by C_K. Now, we mathematically define the conification of K.

Definition (Conification): A conification of set $K \subseteq \mathcal{R}^n$ is defined to be the smallest cone containing K. More specifically, $Cone(K) := \bigcap_{C \subseteq C_K : K \subseteq C} C$.

Note that in general $Cone(X)$ need not be a convex set (can you think of an example?). Due to this in order to make this set convex, we need to take the convex hull of $Cone(.)$, and with a slight abuse of notation, we use $Cone(.)$ to denote it. Also, we do not deal with non-convex cones in this blog.

Consider a convex-cone $conv(K) \subseteq \mathcal{R}^n$. Pick any $x \in K$. Now, the Caratheodry's theorem says that there exists $x_1, \ldots, x_m \subseteq K$ such that $x = \sum_{i=1}^m \lambda_i x_i$ for some $\lambda_i > 0$ and $\mathbf{m} \leq \mathbf{n}$. I will state and prove this in the following theorem.

Theorem 1: **(Caratheodry's theorem)** For every $x \neq 0$ in $Cone(X) \subseteq \mathcal{R}^n$ can be represented as a linear combination of *linearly independent* vectors $x_1, \ldots, x_m \in X$, and therefore $m \leq n$.

Proof: Let m be the minimum number of vectors in X such that $x = \sum_{i=1}^{m} \lambda_i x_i$, $\lambda_i \geq 0$. We need to prove that $m \leq n$. Suppose that $m > n$, then $\{x_1, \ldots, x_m\}$ is a set of linearly dependent vectors. Therefore, there exists α_i's not all of them equal to zero such that

$$\sum_{i=1}^{m} \alpha_i x_i = 0.$$

Using this, we can write $x = \sum_{i=1}^{n} (\lambda_i - \gamma\alpha_i) x_i$ for all $\gamma > 0$. The proof is complete if we show that all the coefficients above are non-negative, and one of them being equal to zero. The coefficients are non-negative if $\gamma < \lambda_i/\alpha_i$ for all $i = 1, \ldots, m$. Therefore, we can pick $\gamma_i := \inf_i \frac{\lambda_i}{\alpha}$. This results in one of the coefficients being zero retaining positivity of all the other coefficients. This leads to a contradiction as there are fewer than m (in this case at least $m-1$) vectors that represents x. $\square$

Remark: This theorem extended to a general convex hull of any set $X \subseteq \mathcal{R}^n$ is used to obtain cardinality bounds on the auxiliary random variables in information theory (see Salehi's work on this topic!). This facilitates the optimization of mutual information terms, and therefore results in a capacity characterization that is computable.

In the following, I provide an application of the Caretheodry's theorem (See Bersekas). For this part, I have followed the lecture slides of Bertsekas.

Theorem 2: Convex hull of a compact set is a compact set.

Proof: Let $X \subseteq \mathcal{R}^n$ be a compact set. Now, we need to prove that the set $Conv(X)$ is compact. It means to say that every sequence in $Conv(X)$ has a convergent subsequence and the limit belongs to $Conv(X)$. Towards this consider a sequence $x_k \in Conv(X)$. Since $x_k \in Conv(X)$, we can write $x_k := \sum_{i=1}^{n+1} \alpha_i^k x_k^i$ such that $\sum_{i=1}^{n+1} \alpha_i^k = 1$, and $x_i^k \in X$. **This is the place where we have used the Caretheodry's theorem. Had this theorem not been there then the sum would have been convex combination of infinitely many points.** Consider the following two sequences in $\mathcal{R}^{n+1}$:

$$\{\alpha_1^k, \ldots, \alpha_{n+1}^k\}$$

and

$$\{x_k^1, \ldots, x_k^{n+1}\}.$$

From the fact that $\sum_{i=1}^{n+1} \alpha_i^k = 1$, the sequence $\{\alpha_1^k, \ldots, \alpha_{n+1}^k\}$ is bounded. Further, the sequence $\{x_k^1, \ldots, x_k^{n+1}\}$ is contained in the compact set X. Using these facts, in the following, we construct a new sequence that lies in the compact set $Y \times \mathcal{R}^n$, where $Y := \{x \in \mathcal{R}^{n+1} : x^T\mathbf{1} = 1\}$:

$$\{\alpha_1^k, \ldots, \alpha_{n+1}^k; x_k^1, \ldots, x_k^{n+1}\}.$$

From the above argument this sequence is contained in a compact set and therefore has the following limit point:

$$\{\alpha_1, \ldots, \alpha_{n+1}; x^1, \ldots, x^{n+1}\}$$

such that $\sum_{i=1}^{n+1} \alpha_i = 1$. The vector $\sum_{i=1}^{n+1} \alpha_i x^i \in Conv(X)$, which shows that $Conv(X)$ is compact. $\Box$

Next, I prove some of the useful properties of a convex function. I start with the definition of a convex function where the function is assumed to be a mapping from a convex set $C \subseteq \mathcal{R}^n$ to $\mathcal{R}$.

Definition (Convex function): A function $f : C \to \mathcal{R}$ is said to be convex if and only if for all $x_1, x_2 \in C$ and $\lambda \in [0,1]$, $f(\lambda x_1 + (1-\lambda)x_2) \geq \lambda f(x_1) + (1-\lambda) f(x_2)$.

Theorem 3: A function $f : C \to \mathcal{R}$ which is twice differentiable for all $x \in C$ is convex if and only if $\nabla^2 f(x) \succcurlyeq 0$.

Proof: The proof follows from the remainder form of the Taylor series expansion of the convex function.

A. *Lower/Upper Semi-Continuous function*

In order to motivate the definition of lower/upper semi-continuous function (l/usc)

IV. Hilbert Space

Hilbert space is essentially a generalization of a finite dimensional complete inner product space to an infinite dimensional space. For example consider $\mathcal{R}^n$, and define the vector addition and multiplication of scalar with the vectors appropriately. Then this space becomes a vector space over $\mathcal{R}$. Now, let us bring in the notion of dot product by using our intuition in $\mathcal{R}^2$.

V.