Some Background Math Notes on Limsups, Sets, and Convexity

I. LIMITS

Let $f(t)$ be a real valued function of time. Suppose $f(t)$ converges to a limiting value f^* as $t \to \infty$ (where f^* is either a finite value or equal to $+\infty$ or $-\infty$). It follows that $f(t)$ also converges to the same limiting value when it is sampled over any sequence of times $\{t_1, t_2, \ldots, t_k, \ldots\}$, provided that this sequence of times grows to infinity. Specifically, the following simple fact holds:

Fact 1: Suppose $\lim_{t\to\infty} f(t) = f^*$ for some value f^*. Then for any sequence of times $\{t_k\}_{k=1}^{\infty}$ that satsifies $\lim_{k\to\infty} t_k = \infty$, we have:

$$\lim_{k\to\infty} f(t_k) = f^*. \quad \Box$$

An example is the function $f(t) = e^{-t}\cos(t)$. It is easy to see that: $\lim_{t\to\infty} e^{-t}\cos(t) = 0$. Hence, any samples of this function must also converge to zero, provided the sample times grow to infinity. However, not all functions $f(t)$ have limiting values, and in particular some functions may have different limits when sampled over different sequences of times. For example, the following functions do not have well defined limits:

- $f_1(t) = \cos(t)$
- $f_2(t) = t\cos(t)$

Specifically, if the function $f_1(t) = \cos(t)$ is sampled at times $\{0, 2\pi, 4\pi, 6\pi, \ldots\}$, then it is always equal to 1 over these samples. If it is sampled at times $\{\pi/2, 3\pi/2, 5\pi/2, 7\pi/2, \ldots\}$, then it is always equal to 0. If the function $f_2(t) = t\cos(t)$ is sampled over the times $\{0, 2\pi, 4\pi, 6\pi, \ldots\}$, then the samples converge to infinity. If it is sampled over the times $\{\pi/2, 3\pi/2, 5\pi/2, 7\pi/2, \ldots\}$, then the samples are always zero (and hence "converge" to zero). If it is sampled at other times, the limit may not converge to anything at all due to the oscillations of the cosine function.

Here, we define notions of a $\limsup$ and a $\liminf$ that are *always defined for any function* $f(t)$. This makes it easier to talk about limits of general functions. Intuitively, the $\limsup$ represents the largest limiting value of $f(t)$ over any infinitely growing sequence of times $\{t_k\}_{k=1}^{\infty}$ for which the function values $f(t_k)$ converge. Similarly, the $\liminf$ represents the smallest such limiting value. In particular, for the above functions, it holds that:

- $\limsup_{t\to\infty} \cos(t) = 1 \quad , \quad \liminf_{t\to\infty} \cos(t) = -1$
- $\limsup_{t\to\infty} t\cos(t) = \infty \quad , \quad \limsup_{t\to\infty} t\cos(t) = -\infty$

Below we formally define the $\limsup$ and $\liminf$, and describe their relation to regular limits. *It is important to note in advance that whenever the regular limit exists, both the* $\limsup$ *and* $\liminf$ *are equal to this limit. Hence, the* $\limsup$ *and* $\liminf$ *can be viewed as regular limits, and have all of the same properties of regular limits, whenever the regular limit exists. In particular, computing the* $\limsup$ *(or* $\liminf$*) is equivalent to computing the regular limit, provided that the regular limit exists.*

A. *The* sup *and* inf *definitions*

We first define the sup operator. Let $f(t)$ be any real valued function of time, and let $\mathcal{T}$ be any set of times (possibly an infinite set).

Definition 1: The *supremum* of $f(t)$ over $t \in \mathcal{T}$, denoted $\sup_{t\in\mathcal{T}} f(t)$, is defined as the smallest value x such that $f(t) \leq x$ for all $t \in \mathcal{T}$. We say that $\sup_{t\in\mathcal{T}} f(t) = \infty$ if $f(t)$ has arbitrarily large values over $t \in \mathcal{T}$.

It follows that if $\sup_{t\in\mathcal{T}} f(t)$ is equal to some value f^*, then $f(t) \leq f^*$ for all $t \in \mathcal{T}$. Further, it must be possible to choose times $\tau \in \mathcal{T}$ for which $f(\tau)$ is *arbitrarily close* to f^*. Some examples are below:

- $\sup_{t\in[0,\infty)} \cos(t) = 1$
- $\sup_{t\in[5,\infty)} \cos(t) = 1$

- $\sup_{t\in[7,\infty)}[1 - 1/(t+1)] = 1$
- $\sup_{t\in[7,\infty)} t\cos(t) = \infty$
- $\sup_{t\in[7,8]} e^{-t} = e^{-7}$
- $\sup_{t\in[7.5,8]} e^{-t} = e^{-7.5}$

The *infimum* is defined similarly:

Definition 2: The *infimum* of $f(t)$ over $t \in \mathcal{T}$, denoted $\inf_{t\in\mathcal{T}} f(t)$, is defined as the largest value x such that $f(t) \geq x$ for all $t \in \mathcal{T}$. We say that $\inf_{t\in\mathcal{T}} f(t) = -\infty$ if $f(t)$ has arbitrarily small (negative) values over $t \in \mathcal{T}$. Thus, if $\inf_{t\in\mathcal{T}} f(t)$ is equal to some value f^*, then $f(t) \geq f^*$ for all $t \in \mathcal{T}$, and $f(\tau)$ must come arbitrarily close to f^* for some values $\tau \in \mathcal{T}$.

We note that $\sup_{t\in\mathcal{T}} f(t)$ and $\inf_{t\in\mathcal{T}} f(t)$ are always defined (for any function $f(t)$ and any set of times $\mathcal{T}$). Further, we have the following simple fact, the proof of which is one line.

Fact 2: $\inf_{t\in\mathcal{T}} f(t) \leq \sup_{t\in\mathcal{T}} f(t)$. $\square$

Proof: We know that for any $\tau \in \mathcal{T}$, we have $f(\tau) \leq \sup_{t\in\mathcal{T}}$ and $f(\tau) \geq \inf_{t\in\mathcal{T}}$. Thus:

$$\inf_{t\in\mathcal{T}} f(t) \leq f(\tau) \leq \sup_{t\in\mathcal{T}} f(t)$$

proving the result. $\square$

B. *The* $\limsup$ *and* $\liminf$ *definitions*

Consider any real valued function $f(t)$.

Definition 3: The $\limsup$ of $f(t)$ as $t \to \infty$ is defined:

$$\limsup_{t\to\infty} f(t) \triangleq \lim_{t\to\infty}\left[\sup_{\tau\geq t} f(\tau)\right]$$

Note that the expression $\sup_{\tau\geq t} f(\tau)$ can be viewed as a *function of* t. Specifically, we can define $g(t) \triangleq \sup_{\tau\geq t} f(\tau)$, where $g(t)$ represents the supremum of the $f(\cdot)$ function over all times that are larger than or equal to t. This function $g(t)$ must be a *non-increasing function*, because the supremum is taken over smaller and smaller intervals as t increases. That is, the supremum over the interval $[7,\infty)$ must be greater than or equal to the supremum over the interval $[7.5,\infty)$, because the interval $[7.5,\infty)$ is a subset of the interval $[7,\infty)$. Therefore, $\sup_{\tau\in[7,\infty)} f(\tau)$ considers the supremum over all times in the larger interval $[7,\infty)$, which of course includes times $t \in [7.5,\infty)$.

Because $\sup_{\tau\geq t} f(\tau)$ can be viewed as a non-increasing function of t, it must have a limit as $t \to \infty$. This is because all non-increasing functions have well defined limits (either being finite or equal to $-\infty$). Therefore, the limit in the above $\limsup$ definition is always defined. Similarly, we have:

Definition 4: The $\liminf$ of $f(t)$ as $t \to \infty$ is defined:

$$\liminf_{t\to\infty} f(t) \triangleq \lim_{t\to\infty}\left[\inf_{\tau\geq t} f(\tau)\right]$$

The function $\inf_{\tau\geq t} f(\tau)$ can be viewed as a *non-decreasing* function of t, and hence this $\liminf$ is also always defined.

The following lemma follows immediately from Fact 2:

Lemma 1: Let $f(t)$ be any real valued function. Then:

$$\liminf_{t\to\infty} f(t) \leq \limsup_{t\to\infty} f(t). \quad \square$$

C. *Relation to the regular limit*

Lemma 2: Consider any function $f(t)$, and suppose $\limsup_{t\to\infty} f(t) = f^*$ (where f^* is possibly infinite). Then: (i) There must exist a sequence of times $\{t_k\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty} t_k = \infty$ and:

$$\lim_{k\to\infty} f(t_k) = f^*$$

(ii) For any sequence of times $\{\tau_k\}_{k=1}^{\infty}$ such that $\lim_{k\rightarrow\infty}\tau_k=\infty$ and such that the regular limit $\lim_{k\rightarrow\infty}f(\tau_k)$ exists, then:

$$\lim_{k\rightarrow\infty} f(\tau_k) \leq f^*. \quad \square$$

This lemma is the reason that the $\limsup$ of a function $f(t)$ can be viewed as the largest limiting value of the function. Similarly, the $\liminf$ of a function $f(t)$ can be viewed as its smallest limiting value.

The next lemma demonstrates that the $\limsup$, $\liminf$, and the regular limit are all equivalent whenever the regular limit exists.

Lemma 3: Consider any function $f(t)$. Then $\lim_{t\rightarrow\infty} f(t) = f^*$ (where f^* is possibly infinite) if and only if:

$$\limsup_{t\rightarrow\infty} f(t) = \liminf_{t\rightarrow\infty} f(t) = f^*$$

That is, if the regular limit exists and is equal to a value f^*, then both the $\limsup$ and $\liminf$ of the function are equal to f^*. Conversely, if $\liminf_{t\rightarrow\infty} f(t) = \limsup_{t\rightarrow\infty} f(t)$, then the regular limit also exists and is equal to the same value as the $\liminf$ and the $\limsup$. $\square$

The above lemma is very important for practical understanding of the $\liminf$ *and* $\limsup$. *It says that these limits are identical to the regular limit whenever the regular limit exists. Thus, whenever a* $\limsup$ *or* $\liminf$ *appears in an equation, the reader can view it exactly as a regular limit under the additional assumption that the regular limit exists. However, using* $\limsup$ *and* $\liminf$ *notation often makes things much easier, because there is no need to prove these limits exist (since they always exist).*

D. *Further Properties of the* $\liminf$ *and* $\limsup$

Lemma 4: Consider any two functions $f(t)$ and $g(t)$, and suppose that:

$$f(t) \leq g(t) \quad \text{for all } t$$

Then:

$$\begin{aligned} \limsup_{t\rightarrow\infty} f(t) &\leq \limsup_{t\rightarrow\infty} g(t) \\ \liminf_{t\rightarrow\infty} f(t) &\leq \liminf_{t\rightarrow\infty} g(t). \quad \square \end{aligned}$$

The $\limsup$ has the following properties (where $f(t)$ and $g(t)$ are any functions and C is any constant).

- $\limsup_{t\rightarrow\infty}[C + f(t)] = C + \limsup_{t\rightarrow\infty} f(t)$
- $\limsup_{t\rightarrow\infty} Cf(t) = C\limsup_{t\rightarrow\infty} f(t)$ (assuming $C > 0$)
- $\limsup_{t\rightarrow\infty} f(t) = -\liminf_{t\rightarrow\infty}[-f(t)]$
- $\limsup_{t\rightarrow\infty}[f(t) + g(t)] \leq \limsup_{t\rightarrow\infty} f(t) + \limsup_{t\rightarrow\infty} g(t)$ (whenever the right hand side does not yield "$\infty + -\infty$" or "$-\infty + \infty$").

The final property above is the only one that is different from regular limits. A simple example of this is as follows; Define $f(t) \triangleq \cos(t)$ and $g(t) \triangleq -\cos(t)$. Then:

$$\limsup_{t\rightarrow\infty}[f(t) + g(t)] = 0$$

$$\limsup_{t\rightarrow\infty} f(t) + \limsup_{t\rightarrow\infty} g(t) = 2$$

Two other useful properties are given below:

Lemma 5: Let $f(t)$ and $g(t)$ be any functions. Then:

$$\limsup_{t\rightarrow\infty}[f(t) + g(t)] \geq \limsup_{t\rightarrow\infty} f(t) + \liminf_{t\rightarrow\infty} g(t) \geq \liminf_{t\rightarrow\infty}[f(t) + g(t)]$$

whenever "$\limsup_{t\rightarrow\infty} f(t) + \liminf_{t\rightarrow\infty} g(t)$" does not yield "$\infty + -\infty$" or "$-\infty + \infty$." $\square$

Lemma 6: Let $f(t)$ be any function, and let $h(t)$ be a function with a well defined and limit, so that $\lim_{t\rightarrow\infty} h(t) = h^*$ for some (possibly infinite) value h^*. Then:

$$\limsup_{t\rightarrow\infty}[h(t) + f(t)] = h^* + \limsup_{t\rightarrow\infty} f(t)$$

whenever the right hand side does not yield "$\infty + -\infty$" or "$-\infty + \infty$." $\square$

E. Ungraded Exercises

Use only the basic properties described in Section I-D to prove the results in the following lemmas. (Hints: For Lemma 7, use the fact that $\limsup_{t\to\infty} f(t) = -\liminf_{t\to\infty}[-f(t)]$. For Lemma 8 use the fact that $g(t) = (g(t) - f(t)) + f(t)$):

Lemma 7: For any functions $f(t), g(t)$ and any constant $C > 0$, we have:

- $\liminf_{t\to\infty} Cf(t) = C \liminf_{t\to\infty} f(t)$
- $\liminf_{t\to\infty} f(t) = -\limsup_{t\to\infty}[-f(t)]$
- $\liminf_{t\to\infty}[f(t) + g(t)] \geq \liminf_{t\to\infty} f(t) + \liminf_{t\to\infty} g(t)$ (whenever the right hand side does not yield "$\infty + -\infty$" or "$-\infty + \infty$"). $\square$

Lemma 8: Consider any two functions $f(t), g(t)$, and suppose that:

$$\liminf_{t\to\infty}[g(t) - f(t)] \geq 0$$

Then the following two properties hold:

(i) $\limsup_{t\to\infty} f(t) \leq \limsup_{t\to\infty} g(t)$

(ii) $\liminf_{t\to\infty} f(t) \leq \liminf_{t\to\infty} g(t)$. $\square$

II. Points and Sets

Here we state the basic definitions of closed sets, limit points, and bounded sets. We also present the multi-dimensional Bolzano-Weirstrass Theorem, which ensures that every infinite sequence of points contained in a closed and bounded set has a convergent subsequence that converges to a point in the set. Further discussion of points and sets can be found in [1] [2].

A. Closed Sets

Let $\mathcal{A}$ represent a subset of N-dimensional Euclidean space (that is, $\mathcal{A} \subset \mathbb{R}^N$). Thus, $\mathcal{A}$ contains N-dimensional vectors (possibly infinitely many), where each vector has the form $\boldsymbol{x} = (x_1, \ldots, x_N)$. Such vectors are also called *points*, as they represent a single point in Euclidean space.

Definition 5: A *limit point* of a set $\mathcal{A}$ is a point $\boldsymbol{x}^* \in \mathbb{R}^N$ such that there exists an infinite sequence of points $\{\boldsymbol{x}(1), \boldsymbol{x}(2), \boldsymbol{x}(3), \ldots\}$ where $\boldsymbol{x}(k) \in \mathcal{A}$ for all $k \in \{1, 2, \ldots\}$ (and where the $\boldsymbol{x}(k)$ values can possibly repeat the same points in $\mathcal{A}$), and such that:

$$\lim_{k\to\infty} \boldsymbol{x}(k) = \boldsymbol{x}^*$$

The limit in the above equation represents a component-wise limit, so that each of the N components of the $\{\boldsymbol{x}(k)\}$ sequence converges to the corresponding component of $\boldsymbol{x}^*$. The fact that we allow points of $\mathcal{A}$ to be repeated when constructing the infinite sequence ensures that every point $\boldsymbol{x}$ that is already contained in $\mathcal{A}$ is also a limit point of $\mathcal{A}$. This can be seen by forming the trivial sequence $\{\boldsymbol{x}(k)\}_{k=1}^{\infty} = \{\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{x}, \ldots\}$.

Definition 6: A set $\mathcal{A}$ is *closed* if it contains all its limit points.

Any set $\mathcal{A}$ that is not closed can be transformed into a closed set simply by adding all of its limits points:

Definition 7: For any set $\mathcal{A}$, define the *closure of A* (denoted $Cl\{\mathcal{A}\}$), to be the set of all limit points of $\mathcal{A}$ (which also includes all original points of $\mathcal{A}$).

It is easy to see that for any set $\mathcal{A}$, the set $Cl\{\mathcal{A}\}$ is closed. For simple examples in one dimension, define:

- $\mathcal{A}_1 \triangleq (0, 1] = \{x \in \mathbb{R} \mid 0 < x \leq 1\}$
- $\mathcal{A}_2 \triangleq [0, 1] = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$
- $\mathcal{A}_3 \triangleq [0, 5.5) \cup (6.7, 9]$
- $\mathcal{A}_4 \triangleq \{1, 9.7, 10\}$

- $\mathcal{A}_5 \triangleq \{1, 1/2, 1/3, 1/4, \ldots\} = \{1/n \mid n \in \{1, 2, \ldots\}\}$

Then the only limit point of $\mathcal{A}_1$ that is not itself contained in $\mathcal{A}_1$ is the point 0. It follows that $Cl\{\mathcal{A}_1\} = [0,1] = \mathcal{A}_2$. Thus, $\mathcal{A}_2$ is closed. The only limit points of $\mathcal{A}_3$ that are not contained in $\mathcal{A}_3$ are 5.5 and 6.7. Thus, the set $\mathcal{A}_3$ is not closed, but its closure is equal to $[0, 5.5] \cup [6.7, 9]$. The set $\mathcal{A}_4$ is a finite set and hence is always closed (because the only possible limit points are members of the finite set). The set $\mathcal{A}_5$ is an infinite set of discrete points. Note that the sequence $\{1, 1/2, 1/3, \ldots\}$ converges to the point 0, but $0 \notin \mathcal{A}_5$, and hence $\mathcal{A}_5$ is not closed. This same sequence $\{1, 1/2, 1/3, \ldots\}$ is contained in $\mathcal{A}_1$, which formally shows that 0 is a limit point of $\mathcal{A}_1$.

B. Bounded Sets

Consider a subset $\mathcal{A}$ of $\mathbb{R}^N$. Recall that each point of $\mathcal{A}$ is a vector with the form $\boldsymbol{x} = (x_1, \ldots, x_N)$. Let $|\boldsymbol{x} - \boldsymbol{y}|$ denote the traditional Euclidean distance between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. Specifically:

$$|\boldsymbol{x} - \boldsymbol{y}| \triangleq \sqrt{\sum_{i=1}^N (x_i - y_i)^2}$$

Definition 8: A subset $\mathcal{A}$ is *bounded* if there exists a finite constant M such that $|\boldsymbol{x} - \boldsymbol{y}| \leq M$ for all vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ contained in $\mathcal{A}$.

Definition 9: An *N-dimensional closed hypercube centered at the origin* is a subset of $\mathbb{R}^N$ given by $[-Z/2, Z/2]^N$ for some positive constant Z, where:

$$[-Z/2, Z/2]^N \triangleq \{(x_1, \ldots, x_N) \mid -Z/2 \leq x_i \leq Z/2 \text{ for all } i \in \{1, \ldots, N\}\}$$

Such a hypercube has *edge size* Z and *volume* Z^N. It is not difficult to show from the above definitions that if a set $\mathcal{A}$ is bounded, then it can be contained in some hypercube.

Consider a bounded set $\mathcal{A} \subset \mathbb{R}^N$ and suppose we have an infinite sequence of points $\{\boldsymbol{x}(k)\}_{k=1}^\infty$, where $\boldsymbol{x}(k) \in \mathcal{A}$ for all k. This sequence possibly repeats some points of $\mathcal{A}$ many times, and might bounce around the set $\mathcal{A}$, so that $\lim_{k\rightarrow\infty} \boldsymbol{x}(k)$ may not exist. However, The next theorem proves that $\boldsymbol{x}(k)$ must have a *convergent subsequence*.

Definition 10: Let $\{\boldsymbol{x}(k)\}_{k=1}^\infty$ be an infinite sequence of points. A *subsequence* of $\{\boldsymbol{x}(k)\}_{k=1}^\infty$ is an infinite sequence $\{y_n\}_{n=1}^\infty$, where this sequence selects points from the original sequence (possibly skipping some points of the original sequence, but preserving the same order). Formally, the subsequence $\{y_n\}_{n=1}^\infty$ is defined by:

$$y_n \triangleq x(k_n)$$

where k_n is a strictly increasing function that maps the positive numbers $\{1, 2, 3, \ldots\}$ into the positive numbers $\{1, 2, 3, \ldots\}$ (so that $k_n < k_{n+1}$ for all $n \in \{1, 2, \ldots\}$).

As an example of a subsequence, let the original sequence consist of all positive integers $\{1, 2, 3, 4, \ldots\}$, so that $x(k) = k$ for $k \in \{1, 2, 3, \ldots\}$. An example subsequence of this is the sequence of all even positive integers $\{2, 4, 6, 8, \ldots\}$, so that $k_n = 2n$ and $y_n = x(k_n) = 2n$ for $n \in \{1, 2, 3, \ldots\}$.

Theorem 1: (Multi-Dimensional Bolzano-Weirstrass) Let $\mathcal{A}$ be a bounded subset of $\mathbb{R}^N$, and let $\{\boldsymbol{x}(k)\}_{k=1}^\infty$ represent an infinite sequence of points in $\mathcal{A}$ (so that $\boldsymbol{x}(k) \in \mathcal{A}$ for all k). Then $\{\boldsymbol{x}(k)\}$ has a *convergent subsequence*, i.e., a subsequence $\{\boldsymbol{x}(k_n)\}_{n=1}^\infty$ such that:

$$\lim_{n\rightarrow\infty} \boldsymbol{x}(k_n) = \boldsymbol{x}^*$$

for some fixed vector $\boldsymbol{x}^* \in \mathbb{R}^N$. $\square$

Thus, any infinite sequence of points contained in a bounded set has a convergent subsequence that converges to some point $\boldsymbol{x}^*$. Note that the vector $\boldsymbol{x}^*$ is thus a limit point of $\mathcal{A}$. Note that the point $\boldsymbol{x}^*$ itself may or may not be contained in the set $\mathcal{A}$. However, if $\mathcal{A}$ is closed, then the vector $\boldsymbol{x}^*$ would necessarily be contained in $\mathcal{A}$. In N-dimensional Euclidean space, a subset that is both closed and bounded is called a *compact set*.

For intuition about why the multi-dimensional Bolzano-Weirstrass Theorem is true, consider the simple 1-dimensional case: Let the set $\mathcal{A}$ consist of a closed interval $[a, b]$. Suppose we have a sequence of real numbers $\{x(k)\}_{k=1}^\infty$ contained in this interval. That is, $a \leq x(k) \leq b$ for all $x(k)$ values in the sequence. Consider now the intervals $[a, (a+b)/2]$ and $[(a+b)/2, b]$ (i.e., the two intervals formed by chopping $[a, b]$ into two equal length sub-intervals). Because there are an infinite number of points $x(k)$ in $[a, b]$, there must be an infinite number of these points in at least one of the two sub-intervals. Label the sub-interval with the infinite number of points $[\alpha_1, \beta_1]$ (if both sub-intervals contain an infinite number of points, arbitrarily choose one of them and rename it

$[\alpha_1, \beta_1]$). Note that this new interval $[\alpha_1, \beta_1]$ has size $(b-a)/2$. Again divide this new interval into two sub-intervals $[\alpha_1, (\alpha_1+\beta_1)/2]$ and $[(\alpha_1+\beta_1)/2, \beta_1]$. Because there are an infinite number of points $x(k)$ in $[\alpha_1, \beta_1]$, again we see there must be an infinite number of points in one of the two new sub-intervals. Label this sub-interval $[\alpha_2, \beta_2]$. Proceeding this way, we see that we can find successive points of the $\{x(k)\}$ sequence in smaller and smaller sub-intervals, where the sub-interval size is halving every step. This is a nested set of closed intervals that shrink to size 0 (where each interval has an infinite number of the points in the $\{x(k)\}$ sequence) and hence these intervals must converge about a single limit point x^*.

The multi-dimensional case can be proven by iteratively applying the single dimensional result to each dimension, or by repeating the above argument but using "hypercubes" and "sub-hypercubes" in replacement of "intervals" and "sub-intervals." Note that when a N-dimensional hypercube has each of its dimensions divided into two halves, we are left with 2^N hypercubes, each with edge size that is halved. For example, dividing a 2-dimensional square by cutting each edge in half creates $2^2 = 4$ sub-squares.

III. CONVEXITY

Definition 11: A set $\mathcal{A} \subset \mathbb{R}^N$ is said to be *convex* if for any two points $\boldsymbol{x}$ and $\boldsymbol{y}$ contained in $\mathcal{A}$, we have:

$$p\boldsymbol{x} + (1-p)\boldsymbol{y} \in \mathcal{A}$$

for all values p such that $0 \leq p \leq 1$.

Note that if $p = 0$, then $p\boldsymbol{x} + (1-p)\boldsymbol{y} = \boldsymbol{y}$, while if $p = 1$ then $p\boldsymbol{x} + (1-p)\boldsymbol{y} = \boldsymbol{x}$. Choosing intermediate values of p yields points that are on the line segment with endpoints $\boldsymbol{x}$ and $\boldsymbol{y}$. Thus, a set is convex if it contains all line segments formed by any two points of the set. An example of a convex set is a sphere (with all the inside points of the sphere included), because the line segment formed by any two points of the sphere is also inside the sphere. An example of a set that is *not* convex is a donut (more mathematically called a *torus*), where the donut hole is not included.

Exercise 1: (ungraded) Show that the set of all points $\boldsymbol{x} = (x_1, \ldots, x_N) \in \mathbb{R}^N$ that satisfy the following collection of K linear inequalities is a convex set:

$$\begin{array}{c} \sum_{i=1}^N \alpha_i^{(1)} x_i \leq b_1 \\ \sum_{i=1}^N \alpha_i^{(2)} x_i \leq b_2 \\ \cdots \\ \sum_{i=1}^N \alpha_i^{(K)} x_i \leq b_K \end{array}$$

where $\{b_1, \ldots, b_K\}$ and $\{\alpha_i^{(j)}\}$ are arbitary constants in $\mathbb{R}$.

The linear inequalities of the above exercise define a set that is called a *polyhedral convex set*.

A. *Convex Combinations and Convex Hulls*

Here we state the basic definitions of convex sets, and state two important results. A more detailed treatment of convex sets can be found in [1] [3].

Let $\mathcal{A}$ be any set in $\mathbb{R}^N$, and let $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$ be any finite set of k points in $\mathcal{A}$.

Definition 12: A *convex combination* of points in $\mathcal{A}$ is any point $\boldsymbol{x}$ of the form:

$$\boldsymbol{x} = p_1\boldsymbol{x}_1 + p_2\boldsymbol{x}_2 + \ldots + p_k\boldsymbol{x}_k \tag{1}$$

where $\{\boldsymbol{x}_i\}_{i=1}^k$ is a finite collection of points in $\mathcal{A}$, and where p_i are values such that $p_i \geq 0$ for all i and $\sum_{i=1}^k p_i = 1$.

The p_i values in the above definition can be viewed as *probabilities*, and so the right hand side of (1) can be treated as an expectation $\mathbb{E}\{\boldsymbol{X}\}$, where $\boldsymbol{X}$ is a random vector that is equal to $\boldsymbol{x}_i$ with probability p_i.

Fact 3: A set $\mathcal{A}$ is convex if and only if it contains all of its possible convex combinations. □

The above fact suggests a simple way to form a convex set out of any (potentially non-convex) set.

Definition 13: The *convex hull* of a set $\mathcal{A}$, denoted $Conv\{\mathcal{A}\}$, is defined as the set of all convex combinations of $\mathcal{A}$.

Note that the convex hull of a set $\mathcal{A}$ includes all of the original points of the set, because one can always form the trivial convex combination that weights a given point with weight $p = 1$. Thus, $\mathcal{A} \subset Conv\{\mathcal{A}\}$. Further, the convex hull itself is always convex. The convex hull of a set that is already convex is the same as the original set.

Fact 4: The convex hull of a closed set is closed. $\square$

An example of the last fact is as follows: Let $\mathcal{A} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$ be a finite set in 2-dimensional space containing three elements, where $\boldsymbol{x}_1 = (0, 0)$, $\boldsymbol{x}_2 = (2, 0)$, $\boldsymbol{x}_3 = (1, 1)$. Then $\mathcal{A}$ is closed (because it is finite), and so $Conv\{\mathcal{A}\}$ is closed. The set $Conv\{\mathcal{A}\}$ is given by the triangle (including the interior) with vertices at points $\boldsymbol{x}_1, \boldsymbol{x}_2$, and $\boldsymbol{x}_3$. Any point on the triangle can be written as a convex combination of the 3 points $\boldsymbol{x}_1, \boldsymbol{x}_2$, and $\boldsymbol{x}_3$.

Theorem 2: (Caratheodory's Theorem) Let $\mathcal{A}$ be any subset of $\mathbb{R}^N$. Let $\boldsymbol{x}$ be any convex combination of points in $\mathcal{A}$, so that $\boldsymbol{x}$ can be written as a convex combination using some finite number k of points in $\mathcal{A}$ (where k can be arbitrarily large):

$$\boldsymbol{x} = p_1\boldsymbol{x}_1 + p_2\boldsymbol{x}_2 + \ldots + p_k\boldsymbol{x}_k$$

where $\boldsymbol{x}_k \in \mathcal{A}$ for all $i \in \{1, \ldots, k\}$. Then $\boldsymbol{x}$ can also be written as a convex combination that uses only $N + 1$ points of $\mathcal{A}$:

$$\boldsymbol{x} = \tilde{p}_1\tilde{\boldsymbol{x}}_1 + \tilde{p}_2\tilde{\boldsymbol{x}}_2 + \ldots + \tilde{p}_{N+1}\tilde{\boldsymbol{x}}_{N+1}$$

for some probabilities $\{\tilde{p}_1, \ldots, \tilde{p}_{N+1}\}$ and some elements $\{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_{N+1}\}$ of $\mathcal{A}$. $\square$

Thus, if we have a set $\mathcal{A} \in \mathbb{R}^2$ that contains 10 elements: $\mathcal{A} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{10}\}$, then any element of the convex hull of $\mathcal{A}$ can be written as a convex combination that uses at most 3 of these elements.

The following is a useful result that relates to random vectors $\boldsymbol{X}$.

Fact 5: (from [4]) Let $\boldsymbol{X}$ be a random vector that takes values in some set $\mathcal{A} \subset \mathbb{R}^N$. Suppose that $\mathbb{E}\{\boldsymbol{X}\}$ is defined. Then:

$$\mathbb{E}\{\boldsymbol{X}\} \in Conv\{\mathcal{A}\}$$

Further, if the set $\mathcal{A}$ is itself convex, then $Conv\{\mathcal{A}\} = \mathcal{A}$, and so $\mathbb{E}\{\boldsymbol{X}\} \in \mathcal{A}$. $\square$

If $\mathcal{A}$ is a finite set $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$, then the above fact is easy to establish, as the expectation can then be written:

$$\mathbb{E}\{\boldsymbol{X}\} = p_1\boldsymbol{x}_1 + p_2\boldsymbol{x}_2 + \ldots + p_k\boldsymbol{x}_k$$

where $p_i = Pr[\boldsymbol{X} = \boldsymbol{x}_i]$. As this is a convex combination, it follows that $\boldsymbol{X}$ is contained in $Conv\{\mathcal{A}\}$. It is tempting to think that the general result of Fact 5 for arbitrary infinite sets $\mathcal{A}$ can be derived by writing the integral associated with the expectation as a limit of a finite sum. However, this approach does not work because the set $\mathcal{A}$ is not necessarily closed (and so the set $Conv\{\mathcal{A}\}$ is not necessarily closed). Hence, the resulting limit is not a-priori guaranteed to be contained in $Conv\{\mathcal{A}\}$. However, a simple proof of Fact 5 can be obtained using the *hyperplane separation theorem* (see, for example, [1] for a description of hyperplane separation, and the notes "Multi-Dimensional Integration Theorem" available on the link: http://www-rcf.usc.edu/∼mjneely/pdf_papers/convex_integration.pdf for a proof of a statement similar to Fact 5).

B. Convex Functions and Jensen's Inequality

Let $\boldsymbol{x}$ be a vector in $\mathbb{R}^N$. Let $f(\boldsymbol{x})$ be a function that maps the vector $\boldsymbol{x}$ to a real number.

Definition 14: A function $f(\boldsymbol{x})$ that maps $\boldsymbol{x} \in \mathbb{R}^N$ to $\mathbb{R}$ is *convex* if for all pairs of vectors $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^N$ and for all values θ_1, θ_2 such that $\theta_1 \geq 0$, $\theta_2 \geq 0$, and $\theta_1 + \theta_2 = 1$, we have:

$$f(\theta_1\boldsymbol{x}_1 + \theta_2\boldsymbol{x}_2) \leq \theta_1 f(\boldsymbol{x}_1) + \theta_2 f(\boldsymbol{x}_2)$$

Note that the value $\theta_1\boldsymbol{x}_1 + \theta_2\boldsymbol{x}_2$ can be viewed as an average of the two vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ (using probabilities θ_1 and θ_2), while the value $\theta_1 f(\boldsymbol{x}_1) + \theta_2 f(\boldsymbol{x}_2)$ can be viewed as the average value of the function applied at those two respective vectors. Thus, a function is convex if and only if the function of the average is less than or equal to the average of the function (where averages are taken over any two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$). Jensen's inequality below shows that this 2-point averaging property implies a more general (possibly infinite point) averaging property.

A function $f(\boldsymbol{x})$ that is defined only over a convex subset $\mathcal{X} \subset \mathbb{R}^N$ is said to be *convex over* $\mathcal{X}$ if the same convexity property holds, but only for all $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ in the restricted set $\mathcal{X}$ (rather than for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^N$). Note that it is important for the set $\mathcal{X}$ to be convex in this case, as otherwise the vector $\theta_1\boldsymbol{x}_1 + \theta_2\boldsymbol{x}_2$ may not be in the set $\mathcal{X}$ for all required θ_1, θ_2 values, and so the value $f(\theta_1\boldsymbol{x}_1 + \theta_2\boldsymbol{x}_2)$ may not be defined.

Definition 15: A function $f(\boldsymbol{x})$ that maps $\boldsymbol{x} \in \mathbb{R}^N$ to $\mathbb{R}$ is *concave* if $-f(\boldsymbol{x})$ is convex. A function $f(\boldsymbol{x})$ is *concave over* $\mathcal{X}$ (where $\mathcal{X}$ is a convex subset of $\mathbb{R}^N$) if $-f(\boldsymbol{x})$ is convex over $\mathcal{X}$.

Theorem 3: (Jensen's Inequality) Let $\boldsymbol{X}$ be a random vector that takes values in some convex set $\mathcal{X} \subset \mathbb{R}^N$.

(a) If $f(\boldsymbol{x})$ is a convex function over $\boldsymbol{x} \in \mathcal{X}$, then:

$$f(\mathbb{E}\{\boldsymbol{X}\}) \leq \mathbb{E}\{f(\boldsymbol{X})\}$$

(b) If $f(\boldsymbol{x})$ is a concave function over $\boldsymbol{x} \in \mathcal{X}$, then:

$$f(\mathbb{E}\{\boldsymbol{X}\}) \geq \mathbb{E}\{f(\boldsymbol{X})\}$$

Note from Fact 5 that $\mathbb{E}\{\boldsymbol{X}\} \in Conv(\mathcal{X}) = \mathcal{X}$, and so the value $f(\mathbb{E}\{\boldsymbol{X}\})$ is well defined. Note also that part (b) follows immediately from part (a) using Definition 15. Part (a) is proven as a simple consequence of Fact 5 in the next subsection.

Corollary 1: Let $\boldsymbol{X}(t)$ be a random vector process indexed by time $t \in \{0, 1, 2, \ldots\}$ such that $\boldsymbol{X}(t) \in \mathcal{X}$ for all t (where $\mathcal{X}$ is a convex set). Let $f(\boldsymbol{x})$ be a convex function over $\mathcal{X}$. Then for all t we have:

$$f\left(\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\{\boldsymbol{X}(\tau)\}\right) \leq \frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\{f(\boldsymbol{X}(\tau))\}$$

Proof: Fix a time t, and let T be an integer random variable that is uniform over $\{0, 1, 2, \ldots, t-1\}$ and that is independent of the process $\boldsymbol{X}(\tau)$ (for $\tau \in \{0, \ldots, t-1\}$). Define the random variable $\boldsymbol{Y} = \boldsymbol{X}(T)$. Apply Jensen's inequality to $f(\boldsymbol{Y})$. □

We note that a *linear function* satisfies $\theta_1 f(\boldsymbol{x}_1) + \theta_2 f(\boldsymbol{x}_2) = f(\theta_1\boldsymbol{x}_1 + \theta_2\boldsymbol{x}_2)$ for all $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, θ_1, θ_2. Thus, a linear function is both convex *and* concave. It follows by Jensen's inequality that if $f(\boldsymbol{x})$ is linear, then $f(\mathbb{E}\{\boldsymbol{X}\}) = \mathbb{E}\{f(\boldsymbol{X})\}$, and so the expectation passes through the linear function.

C. *General Proof of Jensen's Inequality*

Here we provide a simple and general proof of Jensen's inequality that uses only Fact 5. Suppose that $\mathcal{X}$ is a convex subset of $\mathbb{R}^N$, and that $f(\boldsymbol{x})$ is a convex function over $\mathcal{X}$. We want to show that if $\boldsymbol{X}$ is a vector random variable contained in $\mathcal{X}$, then $f(\mathbb{E}\{\boldsymbol{X}\}) \leq \mathbb{E}\{f(\boldsymbol{X})\}$.

To show this, define the $N+1$ dimensional set $\mathcal{Z}$ as follows:

$$\mathcal{Z} \triangleq \{(\boldsymbol{x}, y) \mid \boldsymbol{x} \in \mathcal{X}, y \geq f(\boldsymbol{x})\}$$

The set $\mathcal{Z}$ is called the *epigraph* of the function $f(\boldsymbol{x})$. Because $\mathcal{X}$ is a convex set and $f(\boldsymbol{x})$ is a convex function, it is not difficult to show that $\mathcal{Z}$ itself is a convex set. Now define the $N+1$ dimensional vector random variable $\boldsymbol{Z}$ as follows:

$$\boldsymbol{Z} \triangleq (\boldsymbol{X}, f(\boldsymbol{X}))$$

It follows that $\boldsymbol{Z} \in \mathcal{Z}$, and hence (by Fact 5) we have that $\mathbb{E}\{\boldsymbol{Z}\} \in \mathcal{Z}$. It follows by definition of the set $\mathcal{Z}$ that:

$$\mathbb{E}\{\boldsymbol{Z}\} = (\boldsymbol{x}^*, y^*)$$

for some vector $\boldsymbol{x}^* \in \mathcal{X}$ and some value y^* that satisfies:

$$y^* \geq f(\boldsymbol{x}^*) \tag{2}$$

However, by definition of the random variable $\boldsymbol{Z}$, we have:

$$\mathbb{E}\{\boldsymbol{Z}\} = (\mathbb{E}\{\boldsymbol{X}\}, \mathbb{E}\{f(\boldsymbol{X})\})$$

Therefore, $\boldsymbol{x}^* = \mathbb{E}\{\boldsymbol{X}\}$ and $y^* = \mathbb{E}\{f(\boldsymbol{X})\}$. From (2) it follows that $\mathbb{E}\{f(\boldsymbol{X})\} \geq f(\mathbb{E}\{\boldsymbol{X}\})$, proving the result.

D. Bounding Convex Functions by Linear Functions

Lemma 9: (from [1]) Let $f(\boldsymbol{x})$ be a convex function over $\mathbb{R}^N$. Then for any point $\hat{\boldsymbol{x}} \in \mathbb{R}^N$, there exists a vector $\hat{\boldsymbol{a}} \in \mathbb{R}^N$ such that:

$$f(\boldsymbol{x}) \geq f(\hat{\boldsymbol{x}}) + \hat{\boldsymbol{a}}^T(\boldsymbol{x} - \hat{\boldsymbol{x}}) \quad \text{for all } \boldsymbol{x} \in \mathcal{X}$$

where $\hat{\boldsymbol{a}}^T$ represents the transpose of $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{a}}^T\boldsymbol{y}$ represents the inner product between two N dimensional vectors $\boldsymbol{a}$ and $\boldsymbol{y}$, defined:

$$\hat{\boldsymbol{a}}^T\boldsymbol{y} \triangleq \sum_{i=1}^{N} \hat{a}_i y_i.$$

The above lemma says that for any convex function $f(\boldsymbol{x})$ over $\mathbb{R}^N$ and any point $\hat{\boldsymbol{x}} \in \mathbb{R}^N$, the function $f(\boldsymbol{x})$ can be lower bounded by a linear function[1], where that linear function shares the same function value as $f(\boldsymbol{x})$ at the point $\hat{\boldsymbol{x}}$. The linear function is sometimes called a *tangential hyperplane*, and the associated vector $\hat{\boldsymbol{a}}$ is called a *subgradient* of the function $f(\boldsymbol{x})$ at the point $\hat{\boldsymbol{x}}$. Lemma 9 can be proven using *hyperplane separation theory* [1]. For a simple example, consider the special case when $f(x)$ is a twice differentiable function of one real variable x (so that $N = 1$). Convexity in this case is equivalent to the property that $f''(x) \geq 0$ for all $x \in \mathbb{R}$. By Taylor's theorem, for any x and $\hat{x}$, we have:

$$f(x) = f(\hat{x}) + f'(\hat{x})(x - \hat{x}) + \frac{1}{2}f''(\tilde{x}) \geq f(\hat{x}) + f'(\hat{x})(x - \hat{x})$$

where $\tilde{x}$ is a value in the closed interval between x and $\hat{x}$. The linear function that bounds $f(x)$ in this case is the tangent of the function $f(x)$ at the point $\hat{x}$, and has slope given by $f'(\hat{x})$.

Below we use Lemma 9 to obtain an alternative proof of Jensen's inequality for the special case $\mathcal{X} = \mathbb{R}^N$. This alternative proof may provide some further insight.

Proof: (Alternate proof of Jensen's inequality for the special case $\mathcal{X} = \mathbb{R}^N$) Suppose $f(\boldsymbol{x})$ is convex over $\mathbb{R}^N$, and $\boldsymbol{X}$ is a random vector that takes values in $\mathbb{R}^N$. Define $\hat{\boldsymbol{x}} \triangleq \mathbb{E}\{\boldsymbol{X}\}$. Then by Lemma 9 there exists a vector $\hat{\boldsymbol{a}}$ such that:

$$f(\boldsymbol{x}) \geq f(\hat{\boldsymbol{x}}) + \hat{\boldsymbol{a}}^T(\boldsymbol{x} - \hat{\boldsymbol{x}}) \text{ for all } \boldsymbol{x} \in \mathbb{R}^N$$

Therefore, for any instantiation of the random variable $\boldsymbol{X}$, we have:

$$f(\boldsymbol{X}) \geq f(\hat{\boldsymbol{x}}) + \hat{\boldsymbol{a}}^T(\boldsymbol{X} - \hat{\boldsymbol{x}})$$

Taking expectations over both sides of the above inequality yields:

$$\mathbb{E}\{f(\boldsymbol{X})\} \geq f(\hat{\boldsymbol{x}}) + \boldsymbol{a}^T(\mathbb{E}\{\boldsymbol{X}\} - \hat{\boldsymbol{x}}) = f(\mathbb{E}\{\boldsymbol{X}\})$$

where the final equality follows from the definition of $\hat{\boldsymbol{x}}$. □

REFERENCES

[1] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar. *Convex Analysis and Optimization*. Boston: Athena Scientific, 2003.
[2] J. R. Munkres. *Topology*. NJ: Prentice Hall, Inc., 2000.
[3] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, 1995.
[4] L. Georgiadis, M. J. Neely, and L. Tassiulas. Resource allocation and cross-layer control in wireless networks. *Foundations and Trends in Networking*, vol. 1, no. 1, pp. 1-149, 2006.

[1]Strictly speaking, the function on the right hand side of the inequality in Lemma 9 is called an *affine* function of $\boldsymbol{x}$ (not a linear function of $\boldsymbol{x}$), as it is linear plus a constant.