Mathematical methods in communication July 26th, 2009

Appendix C : Introduction to analysis

Lecturer: Haim Permuter *Scribe: Iddo Naiss & Yonatan Yehezkeally*

I. Sets

Definition 1 (Set) A **Set** is a collection of distinct elements. Those elements are also referred to as the **Members** of the Set. No meaning is known to the order of elements in a Set, nor to repetition of elements. We notate $a \in A$ to mean than a is an element of the set A.
If every member of a Set A is also a member of a Set B, A is said to be **Contained** in B, or a **Subset** of B, which is notated $A \subset B$ or $A \subseteq B$. If in addition there exists an Element b of B which isn't a member of A, A is then said to be a **Proper Subset** of B, or $A \subsetneq B$.

Example 1 $\{3, 5, 8\}$ is a Set, whose only members are 3, 5 and 8.

Note the notation used to define our set. Another popular way to define a set is as follows:
$A = \{x|C\}$, where C is a condition on x, means that the members of A are exactly those which satisfy C.

We denote a special kind of subsets of the real line, referred to as Segments, which are simply $[a, b] \triangleq \{x|a \leq x \leq b\}$, for any two $a < b$ elements of the real line. We use rounded brackets to denote the fact that either a or b isn't a member of the Segment (e.g. $a \notin (a, b]$).

We define two binary operators on Sets, **Union** and **Intersection**.
The Union of two sets A, B is defined $A \cup B \triangleq \{x|x \in A \text{ or } x \in B\}$.
The Intersection of two sets A, B is defined $A \cap B \triangleq \{x|x \in A \text{ and } x \in B\}$.

Definition 2 (Ordered Pair) An **Ordered Pair** is a collection of exactly two objects, where importance is known to the order of those objects. Hence, unlike a set, repetition is allowed in an ordered pair. More formally, one can define an ordered pair using the concept of Sets as follows: $< a, b > \triangleq \{\{\{a\}, \emptyset\}, \{\{b\}\}\}$. This definition enables us to distinct between the first an last element of the pair.

Example 2 (Using the formal definition of an Ordered Pair) Notice that there is but one way to decipher the notation $\{\{\{\{3\}\}\}, \{\emptyset, \{3\}\}$, and that is as the ordered pair $< 3, \{3\} >$.

Definition 3 (Relation) Given two sets A, B we define $A \times B \triangleq \{< a, b > | a \in A \text{ and } b \in B\}$.
A **Binary Relation** between the two sets A, B is a subset $R \subset A \times B$. We sometime notate aRb to mean $< a, b > \in R$.

Example 3 (Familiar Relations) A very familiar relation (in every known context) is the **Equation** relation, i.e. we take the notation $x = y$ to mean $< x, y > \in =$ (in words, x, y are equal).
Another well known relation on the real line is $\leq$.

II. Functions and Sequences

A. Functions

Definition 4 (Function) A **Function** between two sets A, B, notated $F : A \rightarrow B$, is a binary relation between A, B which satisfies

- $\forall a \in A \exists b \in B$ s.t. aFb.
- Let $a \in A; b_1, b_2 \in B$. Then aFb_1 and $aFb_2 \Rightarrow b_1 = b_2$

This two properties justify the notation $F(a)$ to be that distinct $b \in B$ s.t. aFb. We shall use that notation when dealing with functions.
A is called the **Domain** of F, or $Dom(F)$, and B is called the **Range** of F, or $Rng(F)$.
The set $\{b | \exists a \in A \text{ s.t. } aFb\} \subset B$ is called the **Image** of F, or $Im(F)$.
If $\forall a_1, a_2 \in A; b \in B$ it holds that a_1Fb and $a_2Fb \Rightarrow a_1 = a_2$ we say that F is an **injection**, or that F is $1 - 1$.
If $Rng(F) = Im(F)$ we say that F is a **surjection**, or that F is onto B.
If F is both an injection and a surjection, F is called a **bijection** or a **1-1 Correlation** between A and B.

Example 4 (Information Measures) Let M be the set of Random Variables on a given Probability Space. Then the Entropy Measure defined in class is a function $H : M \rightarrow [0, \infty)$, and both the Divergence and Mutual Information are functions $D, I : M \times M \rightarrow [0, \infty)$.

B. Sequences

Definition 5 (Sequence) A **Sequence** is a collection of ordered elements. More formally, one can define a Sequence as a function from the Natural Numbers to some Set, called the Alphabet of the Sequence. We denote the elements of a Sequence as $(a_n)_{n \in \mathbb{N}}$.

Definition 6 (Convergence of Sequences) A sequence $(a_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ is said to **Converge** to $a \in \mathbb{R}$ if $\forall\epsilon > 0 \exists n_0 \in \mathbb{N}$ s.t. $n_0 < n \Rightarrow |a - a_n| < \epsilon$. a is then said to be the **Limit** of the sequence.
The sequence is said to converge to ∞ (or converge in the wide sence) if $\forall 0 < M \exists n_0 \in \mathbb{N}$ s.t. $n_0 < n \Rightarrow M < a_n$, and we define convergence to $-\infty$ in the same fashion.

Example 5 (A converging Real sequence) The sequence $(\frac{n}{n+1})_{n\in\mathbb{N}} \subset \mathbb{R}$, whose nth element is $\frac{n}{n+1}$, is easily shown to converge to 1 (since $|1 - \frac{n}{n+1}| = \frac{1}{n+1}$).

C. Properties of Functions

Definition 7 (Bounded Functions) A function $F : A \to B$ where $A, B \subset \mathbb{R}$ is said to be **Bounded from Above** if $\exists M > 0$ s.t. $b \in Im(F) \Rightarrow b < M$, and we similarly define when a function is **Bounded from Below**. If a function is said to be **Bounded**, we take that to mean it's bounded from both above and below.

Definition 8 (Continuous Function) A function $F : A \to B$ where $A, B \subset \mathbb{R}$ is said to be **Continuous** at $a \in A$ if $\forall 0 < \epsilon \exists 0 < \delta$ s.t. $|b - a| < \delta \Rightarrow |F(b) - F(a)| < \epsilon$.
The function is said simply to be continuous if it is continuous at all $a \in A$.

Also, recall the definition given in class for a **Convex** function of the real line.

III. Subsets of the Real line

Definition 9 (Closed Subset of $\mathbb{R}$) A subset $A \subseteq \mathbb{R}$ is said to be **Closed** if for all converging $(a_n)_{n\in\mathbb{N}} \subseteq A$, its limit a is a member of A.

Example 6 All segments $[a, b] \subseteq \mathbb{R}$ are closed, and so is $(-\infty, a]$, but not $(-\infty, a)$ (why?).

Definition 10 (Open Subset of $\mathbb{R}$) A subset $A \subseteq \mathbb{R}$ is said to be **Open** if $\mathbb{R}\backslash A$ is closed.

Example 7 All segments $(a, b) \subseteq \mathbb{R}$ are open, and so is $(-\infty, a)$.

It is interesting to note that an equivalent definition of an Open set:

Lemma 1 A set $A \subset \mathbb{R}$ is open if and only if $\forall x \in A \exists (a, b) \subseteq A$ s.t. $x \in (a, b)$.

We leave the proof as an exercise to the reader. [1]

Lemma 2 Every union of open sets is still open. In addition, a finite intersection of open sets is open as well. [2]

As a conclusion, derived by DeMorgan's laws from the preceding lemma, both infinite intersections and finite unions of closed sets are still closed.
Note, however, that an infinite intersection of open sets (or an infinite union of closed set) isn't necessarily open (or closed).

Example 8 (Even a countable union of closed sets may not be closed) Consider for $0 < n \in \mathbb{N}$ the segment $A_n \triangleq [\frac{1}{n}, 1]$. Notice that $\bigcup_{n\in\mathbb{N}} A_n = (0, 1]$, which of course isn't closed.

IV. Bounded Subsets of the Real Line

Definition 11 (upper and Lower Boundaries) Let A be a subset of the real line. $u \in \mathbb{R}$ is said to be an **Upper Boundary** of A if $x \in A \Rightarrow x \leq u$. Similarly, l is said to be a **Lower Boundary** of A if $x \in A \Rightarrow l \leq x$.

Definition 12 (Supremum and Infimum) An upper boundary $a \in \mathbb{R}$ of a set $A \subseteq \mathbb{R}$ is called the **Supremum** of A if $\forall \epsilon > 0 \exists x \in A$ s.t. $a - \epsilon < x$ (i.e. all open sets containing a have a non-empty intersection with A).
A Lower boundary $b \in \mathbb{R}$ of A is called the **Infimum** of A if $\forall \epsilon > 0 \exists x \in A$ s.t. $x < b + \epsilon$ (i.e. all open sets containing b have a non-empty intersection with A).

It is interesting to note that every $A \subseteq \mathbb{R}$ which has an upper boundary $a \in \mathbb{R}$ or a lower boundary $b \in \mathbb{R}$ also has a supremum or an infimum, respectively. That fact is sometimes referred to as the *Axiom of Completeness*, and is actually equivalent to the completeness of $\mathbb{R}$ as a metric space. Metric spaces will be shortly discussed at the end of this appendix.

Lemma 3 (Uniqueness of Supremum) Let $a \in \mathbb{R}$ be a supremum of $A \subseteq \mathbb{R}$. If $b \in \mathbb{R}$ is also a supremum of A, then $a = b$.

Proof: Suppose $a \neq b$. Without loss of generality, suppose $a < b$. Then $b \in (a, b + 1)$, and $x \in (a, b+1) \Rightarrow a < x \Rightarrow x \notin A$ (because a is in particular an upper boundary of A). Hence $A \cup (a, b+1) = \emptyset$, in contradiction. ■

Lemma 4 (Uniqueness of Infimum) Let $a \in \mathbb{R}$ be an infimum of $A \subseteq \mathbb{R}$. If $b \in \mathbb{R}$ is also an infimum of A, then $a = b$.

The proof can be done similarly to lemma 3, and will be left to the reader.

Definition 13 (Maximum and Minimum) Let $a, b \in \mathbb{R}$ be the supremum/infimum of a set $A \subseteq \mathbb{R}$ respectively. If $a \in A$ then a is called the **Maximum** of A, and similarly if $b \in A$ then b is called the **Minimum** of A.

Lemma 5 (Existence of Maximum for closed upper bounded Real subsets) Let $A \subseteq \mathbb{R}$ be closed, and let $b \in \mathbb{R}$ be an upper boundary of A. Then A has a Maximum.

Proof: As mentioned above, A has a supremum $a \in \mathbb{R}$. Then $\forall n \in \mathbb{N} \exists x_n \in A$ s.t. $a - \frac{1}{n} < x_n \leq a$. It follows that $|a - x_n| < \frac{1}{n}$, hence $(x_n)_{n\in\mathbb{N}}$ converges to a (why?). Remember that A is closed, and so $a \in A$, i.e. a is the maximum of A. ■

In much the same way, every closed and bounded from below subset of $\mathbb{R}$ has a minimum.

Theorem 1 (The set of upper bounds of a real subset is closed) Let $A \subseteq \mathbb{R}$.
Denote $U \triangleq \{x \in \mathbb{R} | x \text{ is an upper boundary of } A\}$, and suppose $U \neq \emptyset$. Then U is closed.

Proof: Let $(u_n)_{n\in\mathbb{N}} \subseteq U$, and suppose $(u_n)_{n\in\mathbb{N}}$ converges to $u \in \mathbb{R}$.
Suppose that there exists $a \in A$ s.t. $u < a$. Then $\exists n_0 \in \mathbb{N}$ s.t. $n_0 < n \Rightarrow |u - u_n| < (a - u) \Rightarrow u_n < a$, in contradiction. Hence u is an upper boundary of A. Thus U is closed. ■

Again, the set of lower boundaries of a subset of the real line is also closed.

Lemma 6 Let $A \subseteq \mathbb{R}$ be bounded from above, and define U to be the set of all upper boundaries of A. Then U has a minimum.
Similarly, if A is bounded from below, define L to be the set of all lower boundaries of A. Then L has a maximum.

Proof: This is a direct conclusion from Theorem 1, and the fact (which is easy enough to confirm) that U is bounded from below and L bounded from above. ■

Note that almost trivially, in the settings of Lemma 6, the supremum of A is the minimum of U, and the infimum of A is the maximum of L.

Example 9 (Capacity) Remember the operational definition of the Capacity of a channel
$C = \sup\{R|R \text{ is achievable}\}$. We can now prove that the capacity is well defined, i.e. that the set of achievable rates is bounded from above (hence it has a supremum).
Notice that for a finite Alphabet $|\mathcal{X}| < \infty$ and $\log_2(2|\mathcal{X}|) < R$ at least half of the messages $m \in \{1, ..., 2^{nR}\}$ can't be coded, hence the probability of an error can't converge to zero (whether we're using the mean criteria or the maximum criteria), rendering R unachievable.
Where the channel's Alphabet $\mathcal{X}$ is infinite, the set of achievable rates may very well be unbounded from above, and we take its supremum (hence, the Capacity of the channel) to be ∞.

Example 10 (Capacity, alternative definition) Now we can also note that the other definition of the Capacity of a DMC rightfully uses maximum of the Mutual Information rather than a supremum. Recall that given the transition probabilities $Pr(x|y)$ the Mutual Information $I(X;Y)$ is concave and continuous in the source distribution p_X (i.e. as a function $I : \{(p_i)_{i=1}^{|\mathcal{X}|} \in \mathbb{R}_+^{|\mathcal{X}|} | \sum_{i=1}^{|\mathcal{X}|} p_i = 1\} \to [0, \infty]$). We saw in class that it is concave, and the fact that it's continuous will go without proof for the moment (to proove this, consider the representation of the Mutual Information as a sum of Entropies, each a sum of elemntry functions of P_X). This follows from the next theorem.

Theorem 2 (Weierstrass's Extreme Value th.) [3] A continuous function from a closed and bounded subset of $\mathbb{R}^n$ to the real line has a maximum and a minimum.

Note that we didn't really defined what it means for a subset of $\mathbb{R}^n$ to be bounded. More on that later.

Example 11 (Rate Distortion definition) In much the same way, we note that in the Mutual Information definition of the Rate Distortion function, we indeed have a minimum over the transition probabilities, rather than an infimum. That is a result of the fact that the Mutual Information is also continuous as a function of the transition probabilities (and convex in it, where the source distribution is given), and the last theorem.

V. Basic Notions in Metric Spaces

Definition 14 (Metric) A **Metric** (also referred to as a **Distance Function**) on a non-empty set A is a function $d : A \times A \to [0, \infty]$ which satisfies:

- Positive-Definite: $d(a, b) = 0 \Leftrightarrow a = b$
- Symmetric: $\forall a, b \in A \; d(a, b) = d(b, a)$
- Triangle Inequality: $\forall a, b, c \in A \; d(a, c) \leq d(a, b) + d(b, c)$

The pair $< A, d >$ is then called a **Metric Space**

Example 12 (Metrics on $\mathbb{R}^n$) Perhaps the best known example of a metric is the distance between two real numbers, namely the metric defined on $\mathbb{R}$ by $d(x,y) \triangleq |x-y|$.

This metric is naturally generalized to $\mathbb{R}^n$ in a number of ways. We define for all $0 < p \in \mathbb{N}$

$d_p(\mathbf{x},\mathbf{y}) \triangleq \sqrt[p]{\sum_{i=1}^n |x_i - y_i|^p}$.

The actual proof that these functions are indeed distance functions will be left to the reader. For those who are interested, it requires the use of two inequalities known as Holder inequality and Minkowski inequality. [4]

Example 13 (Metrics on RVs) Define on $\{0,1\}^n$ the metric $d(\mathbf{x},\mathbf{y}) \triangleq \sum_{i=1}^n (x_i \oplus y_i) = \sum_{i=1}^n |x_i - y_i|$. This is in fact a reduction of d_1 from the last example to the subset $\{0,1\} \subset \mathbb{R}$, but in this case it is trivially a metric. This metric is known as the Hamming Metric, and is familiar to us from our discussion of Rate Distortion functions.

Another metric we discussed in class is the Mean Square Error metric on the set of Random Variables on a specific Probability Space, defined $d(X,Y) \triangleq \sqrt{\mathbb{E}((X-Y)^2)}$. This function is trivially positive-definite and symmetric. We can easily prove the triangle inequality using Cauchy-Schwarz inequality (stating $\mathbb{E}(XY) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$):

$$\begin{aligned}
\mathbb{E}((X-Z)^2) &= \mathbb{E}(((X-Y)+(Y-Z))^2) \\
&= \mathbb{E}((X-Y)^2) + 2\mathbb{E}((X-Y)(Y-Z)) + \mathbb{E}((Y-Z)^2) \\
&\leq \mathbb{E}((X-Y)^2) + 2\sqrt{\mathbb{E}((X-Y)^2)\mathbb{E}((Y-Z)^2)} + \mathbb{E}((Y-Z)^2) \\
&= (\sqrt{\mathbb{E}((X-Y)^2)} + \sqrt{\mathbb{E}((Y-Z)^2)})^2
\end{aligned}$$

Example 14 (Information Metrics) Divergence isn't a metric. As seen in one of the given homework, it is not symmetric.

Nevertheless, one can define a metric on the set of Random Variables on a given Probability Space (where we identify any two RVs X, Y if there exist $A, B \subset \mathbb{R}$ and an injection $f : A \overset{1-1}{\to} B$ s.t. $X = f(Y)$ with probability 1) using the Conditional Entropy measure in the following two ways:

$$d(X,Y) \triangleq H(X|Y) + H(Y|X) \tag{1}$$

$$d(X,Y) \triangleq \frac{H(X|Y) + H(Y|X)}{H(X,Y)} \tag{2}$$

In these cases, once more, it is trivially proven that these functions are positive-definite and symmetric, while proving the triangle inequality takes some effort.

For the purpose of the next three definition, let $< A, d >$ be a metric space.

Definition 15 (Bounded subsets) A subset $B \subset A$ is said to be bounded if $\exists D = sup\{d(b_1, b_2)|b_1, b_2 \in B\}$. D is then called the **Diameter** of B.

Definition 16 (Cauchy Sequence) A **Cauchy Sequence** is a sequence $(a_n)_{n\in\mathbb{N}} \subset A$ which satisfies $\forall 0 < \epsilon \exists n_0 \in \mathbb{N}$ s.t. $n_0 \leq n, m \Rightarrow d(a_n, a_m) < \epsilon$

It is easily seen that every converging sequence is a Cauchy Sequence.

Definition 17 (Complete Metric Space) $< A, d >$ is said to be **Complete** if every Cauchy Sequence on A is converging.

Example 15 (Completeness of $\mathbb{R}, \mathbb{Q}$) As mentioned before, the Real Line is complete. It is also trivial that the set of Rational Numbers isn't complete.

To finish this appendix, let us remark that it can be proven that each Metric Space is a subspace of a Complete Metric Space, and that the Real Line is the Completion of the Rational Numbers (this is actually one of the formal ways to construct $\mathbb{R}$).

NOTES

[1]Unknown author, "Topology/Metric Spaces", Wikibooks, `http://en.wikibooks.org/wiki/Topology/Metric_Spaces`

[2]See article no. 1

[3]James R. Munkres, "Topology; A First Course", Prentice Hall College Div, pp. 163-166 172-174 ,1974-06.

[4]Unknown author, "Holder's inequality", Wikipedia, `http://en.wikipedia.org/wiki/Holder's_inequality`

Unknown author, "Minkowski inequality", Wikipedia, `http://en.wikipedia.org/wiki/Minkowski_inequality`