Information channels to channel operators

Yuichiro Kakihara

October 19, 2010
References


References (continued)


Functional Analysis

- Analysis on infinite dimensional spaces
- Analysis on the structure of subsets of spaces
- Fun Analysis

Objects = Spaces and their subsets such as
- $M(X) =$ the space of measures
- $P(X) =$ the set of probability measures $\subset M(X)$
- $C(X, Y) =$ the set of information channels
- $\mathcal{O}(X, Y) =$ the set of channel operators
Measure spaces

$X$ is a nonempty set.

$\mathcal{X}$ is a $\sigma$-algebra of subsets of $X$, i.e.,

1) $X \in \mathcal{X}$
2) $A \in \mathcal{X} \Rightarrow A^c \in \mathcal{X}$
3) $A_1, A_2, \ldots \in \mathcal{X} \Rightarrow A_1 \cup A_2 \cup \cdots \in \mathcal{X}$

Any set in $\mathcal{X}$ is said to be measurable.

$(X, \mathcal{X})$ is a measurable space.

$\mu$ is a (positive) measure on $\mathcal{X}$, i.e., $\mu : \mathcal{X} \rightarrow [0, \infty)$ such that

$A_i \in \mathcal{X}, A_i \cap A_j = \emptyset \ (i \neq j)$$\Rightarrow \mu(A_1 \cup A_2 \cup \cdots) = \mu(A_1) + \mu(A_2) + \cdots$

$(X, \mathcal{X}, \mu)$ is a measure space.

$S : X \rightarrow X$ is a measurable transformation if

$A \in \mathcal{X} \Rightarrow S^{-1}A = \{x \in X : Sx \in A\} \in \mathcal{X}$
Integration

(1) \((X, \mathcal{X}, \mu)\) is a measure space.

(2) A real valued function \(f : X \rightarrow (-\infty, \infty) = \mathbb{R}\) is said to be \textbf{measurable} if, for any \(c \in \mathbb{R}\), \(\{f \leq c\} = \{x \in X : f(x) \leq c\} \in \mathcal{X}\).

(3) For \(A \in \mathcal{X}\), \(1_A(x)\) stands for the \textbf{indicator function} of \(A\), i.e., \(1_A(x) = 1\) if \(x \in A\) and \(= 0\) otherwise.

(4) A function of the form \(\sum_{i=1}^{n} \alpha_i 1_{A_i}(x)\) is called a \textbf{simple function}, where \(\alpha_1, \ldots, \alpha_n \in \mathbb{R}\).

(5) \(L(X)\) stands for the set of all \(\mathbb{R}\)-valued simple functions on \(X\).
Integration (continued)

(6) $\int_X 1_A(x) \mu(dx) = \mu(A)$ for $A \in \mathcal{X}$.

(7) $\int_X \sum_{i=1}^n \alpha_i 1_{A_i}(x) \mu(dx) = \sum_{i=1}^n \alpha_i \mu(A_i)$

(8) If $f$ is a nonnegative measurable function on $X$, then

$$\int_X f(x) \mu(dx) = \sup \left\{ \int_X g(x) \mu(dx) : 0 \leq g \leq f, g \in L(X) \right\}.$$

(9) If $f$ is a measurable function, then write $f = f^+ - f^-$ with $f^+, f^- \geq 0$ and define

$$\int_X f(x) \mu(dx) = \int_X f^+(x) \mu(dx) - \int_X f^-(x) \mu(dx).$$
Linear spaces

Let $\mathcal{X}$ be a set. Then $\mathcal{X}$ is called a **linear space** over $\mathbb{R}$ (or a real linear space) if the addition and the scalar multiplication of elements in $\mathcal{X}$ are defined, i.e.,

$$x, y \in \mathcal{X}, \quad \alpha \in \mathbb{R} \implies x + y, \alpha x \in \mathcal{X}.$$ 

If $\mathbb{R}$ is replaced by $\mathbb{C}$ (the set of complex numbers), then $\mathcal{X}$ is called a complex linear space.
Examples

(1) $\mathbb{R}^2 = \{(x_1, x_2) : x_1, x_2 \in \mathbb{R}\}$ is a real linear space, where the addition and scalar multiplication are defined by

$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2),$$

$$\alpha(x_1, x_2) = (\alpha x_1, \alpha x_2).$$

(2) Similarly, $\mathbb{R}^n = \{(x_1, \ldots, x_n) : x_i \in \mathbb{R}, 1 \leq i \leq n\}$ is a real linear space.

(3) Let $C([0, 1])$ be the set of all real valued continuous functions on $[0, 1]$. Then, $C([0,1])$ is a real linear space, where the addition and scalar multiplication are defined by

$$(f + g)(t) = f(t) + g(t), \quad (\alpha f)(t) = \alpha f(t)$$

for $f, g \in C([0, 1]), \alpha \in \mathbb{R}$ and $t \in [0, 1]$. 
Banach spaces

Let $\mathcal{X}$ be a linear space over $\mathbb{R}$. That is, for $x, y \in \mathcal{X}$ and $\alpha, \beta \in \mathbb{R}$ we have $\alpha x + \beta y \in \mathcal{X}$.

A **norm** on $\mathcal{X}$ is a function $\| \cdot \| : \mathcal{X} \to [0, \infty)$ such that

1. $\|x\| = 0$ if and only if $x = 0$;
2. $\|\alpha x\| = |\alpha|\|x\|$ for $\alpha \in \mathbb{R}$ and $x \in \mathcal{X}$;
3. $\|x + y\| \leq \|x\| + \|y\|$ for $x, y \in \mathcal{X}$.

If a norm is defined on $\mathcal{X}$, then $\mathcal{X}$ is a **normed linear space**.

A normed linear space $\mathcal{X}$ is called a **Banach space** if it is **complete**, i.e., if $\{x_n\} \subset \mathcal{X}$ is such that $\|x_m - x_n\| \to 0$ as $m, n \to \infty$, then there exists an $x \in \mathcal{X}$ such that $\|x_n - x\| \to 0$ as $n \to \infty$. 
Examples

(1) Let $C([0, 1])$ be the set of all real-valued continuous functions on $[0, 1]$. This is a real linear space. Also $C([0, 1])$ is a Banach space, where a norm, the sup norm, is defined by

$$
\|f\| = \sup \{|f(x)| : x \in [0, 1]\} = \max \{|f(x)| : x \in [0, 1]\}
$$

for $f \in C([0, 1])$.

(2) The space $C(\mathbb{R})$ consisting of all bounded continuous functions on $\mathbb{R}$ is a Banach space with the sup norm.
Examples (continued)

(3) Let \((X, \mathcal{X})\) be a measurable space. Let \(M(X)\) be the set of all \(\mathbb{R}\)-valued measures on \(\mathcal{X}\). \(M(X)\) becomes a Banach space with the total variation norm, which is given by

\[
\|\mu\| = \sup \left\{ \sum_{i=1}^{n} |\mu(A_i)| : A_i \cap A_j = \emptyset, i \neq j \right\}, \quad \mu \in M(X).
\]

Let \(P(X)\) be the set of all probability measures in \(M(X)\). That is, \(P(X) = \{\mu \in M(X) : \mu \geq 0, \mu(X) = 1\}\). Hence, \(\|\mu\| = 1\) for \(\mu \in P(X)\). \(P(X)\) is a closed convex subset of \(M(X)\), where \(\mu, \eta \in P(X)\) and \(0 \leq \alpha \leq 1\) imply \(\alpha \mu + (1 - \alpha)\eta \in P(X)\).
A mathematical model of an input space

$$\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$$ is the set of integers.

$$\mathcal{A} = \{a_1, \ldots, a_\ell\}$$ is an alphabet.

$$X = \mathcal{A}^\mathbb{Z} = \{(x_k) = (\ldots, x_{-1}, x_0, x_1, \ldots) : x_k \in \mathcal{A}, k \in \mathbb{Z}\}$$

$$[x_i^0 \ldots x_j^0] = \{x = (x_k) \in X : x_k = x_k^0, i \leq k \leq j\}$$ is a message.

$$\mathcal{M}$$ is the set of all messages in $$X$$ and $$\mathcal{X}$$ be the $$\sigma$$-algebra generated by $$\mathcal{M}$$, i.e., the smallest $$\sigma$$-algebra containing $$\mathcal{M}$$.

$$S : X \rightarrow X$$ is a shift on $$X$$, where, for $$x = (x_k) \in X$$,

$$Sx = x' = (x'_k)$$ is defined by $$x'_k = x_{k+1}$$, $$k \in \mathbb{Z}$$.

$$S$$ is a measurable transformation on $$X$$.

$$(X, \mathcal{X}, S)$$ is called an alphabet message space that is taken to be the input space or input.

Each $$\mu \in P(X)$$ is called an input source.
A mathematical model of an information channel

Let \((Y, \mathcal{Y}, T)\) be another alphabet message space and take this to be the output space or output.

An information channel or channel with input \(X\) and output \(Y\) is a function \(\nu : X \times \mathcal{Y} \rightarrow [0, 1]\) such that

1. \(\nu(x, \cdot) \in P(Y) \) for every \(x \in X\);
2. \(\nu(\cdot, C) \in B(X) \) for every \(C \in \mathcal{Y}\),

where \(B(X)\) is the set of all \(\mathbb{R}\)-valued measurable functions on \(X\).

\(\mathcal{C} = C(X, Y)\) is the set of all information channels with input \(X\) and output \(Y\).
Input, output and compound sources

An **input source** \( \mu \in P(X) \) and a channel \( \nu \in C \) induce an **output source** \( \mu \nu \in P(Y) \) and a **compound source** \( \mu \otimes \nu \in P(X \times Y) \) that are given by

\[
\mu \nu(C) = \int_X \nu(x, C) \mu(dx), \quad C \in \mathcal{Y}.
\]

\[
\mu \otimes \nu(A \times C) = \int_A \nu(x, C) \mu(dx), \quad A \in \mathcal{X}, \ C \in \mathcal{Y}.
\]

Here, \((X \times Y, \mathcal{X} \otimes \mathcal{Y}, S \times T)\) is the **compound space**.
Stationarity

(1) An input source $\mu \in P(X)$ is said to be stationary if $\mu(S^{-1}A) = \mu(A)$ for every $A \in \mathcal{X}$. $P_s(X)$ stands for the set of all stationary sources. We use $P_s(Y)$ and $P_s(X \times Y)$ to denote the set of all stationary output and compound sources, respectively.

(2) A channel $\nu \in \mathcal{C}$ is said to be stationary if

\[(c3) \nu(Sx, C) = \nu(x, T^{-1}C) \text{ for every } x \in X \text{ and } C \in \mathcal{C}.\]

$\mathcal{C}_s = \mathcal{C}_s(X, Y)$ is the set of all stationary channels in $\mathcal{C}$. 

Proposition 1. A stationary channel preserves stationarity:

\[ \mu \in P_s(X), \, \nu \in C_s \implies \mu \nu \in P_s(Y), \, \mu \otimes \nu \in P_s(X \times Y), \]

Proof. Let \( \mu \in P_s(X), \, \nu \in C_s, \, A \in \mathcal{X} \text{ and } C \in \mathcal{Y}. \) Then, we want to show that

\[ \mu \otimes \nu((S \times T)^{-1}(A \times C)) = \mu \otimes \nu(A \times C). \]
Proof (continued). Now we have

\[
\mu \otimes \nu((S \times T)^{-1}(A \times C))
\]

\[
= \mu \otimes \nu(S^{-1}A \times T^{-1}C)
\]

\[
= \int_{S^{-1}A} \nu(x, T^{-1}C) \mu(dx)
\]

\[
= \int_{S^{-1}A} \nu(Sx, C) \mu(dx), \quad \text{by (c3)},
\]

\[
= \int_{A} \nu(x', C) \mu(dx'), \quad \text{by } x' = Sx \text{ and } \mu(dx') = \mu(dS^{-1}x')
\]

\[
= \mu \otimes \nu(A \times C).
\]

Thus, \( \mu \otimes \nu \in P_s(X \times Y) \).

Similarly, \( \mu \nu \in P_s(Y) \). \qed
Ergodicity, extremality and absolute continuity

(1) A stationary input source $\mu \in P_s(X)$ is said to be ergodic if $S^{-1}A = A$ implies $\mu(A) = 0$ or 1.

Let $P_{se}(X)$, $P_{se}(Y)$, $P_{se}(X \times Y)$ be the sets of all stationary ergodic sources on $X$, $Y$ and $X \times Y$, respectively.

(2) Let $K$ be a convex subset of a linear space $L$. Then, $c \in K$ is an extreme point of $K$, if $c = (a + b)/2$ with $a, b \in K$, then $a = b = c$. $\text{ex } K$ denotes the set of all extreme points of $K$.

(3) Let $\mu, \eta \in P(X)$. Then, $\mu$ is absolutely continuous with respect to $\eta$, denoted $\mu \ll \eta$, if $\eta(A) = 0$ implies $\mu(A) = 0$. 
Proposition 2. For a stationary input source $\mu \in P_s(X)$ the following statements are equivalent:

(1) $\mu \in P_{se}(X)$, i.e., $\mu$ is ergodic;
(2) $\mu \in \text{ex } P_s(X)$, i.e., $\mu$ is an extreme point of $P_s(X)$;
(3) There is some ergodic source $\eta \in P_{se}(X)$ such that $\mu \ll \eta$;
(4) $\xi \in P_s(X)$, $\xi \ll \mu \implies \xi = \mu$. 
Ergodicity and identicalness

(1) A stationary channel $\nu \in \mathcal{C}_s$ is said to be **ergodic** if

\[(c4) \quad \mu \in P_{se}(X) \implies \mu \otimes \nu \in P_{se}(X \times Y).\]

$\mathcal{C}_{se} = \mathcal{C}_{se}(X, Y)$ is the set of all ergodic channels in $\mathcal{C}_s$.

(2) Let $\mathcal{P} \subseteq P(X)$. Two channels $\nu_1, \nu_2 \in \mathcal{C}$ are said to be **identical** mod $\mathcal{P}$, denoted $\nu_1 \equiv \nu_2 \ (\text{mod } \mathcal{P})$, if

$$\nu_1(x, \cdot) = \nu_2(x, \cdot) \quad \mu\text{-a.e. } x \in X \quad \text{for every } \mu \in \mathcal{P}. $$

(If $A = \{x \in X : \nu_1(x, \cdot) \neq \nu_2(x, \cdot)\}$, then $\mu(A) = 0$.)
Proposition 3 (Umegaki, Nakamura 1969). For a stationary channel $\nu \in C_s$ the following statements are equivalent:

(1) $\nu \in C_{se}$, i.e., $\nu$ is ergodic;

(2) $\nu$ is extremal in $C_s$ in the sense that if $\nu \equiv \alpha \nu_1 + (1 - \alpha)\nu_2$ (mod $P_{se}(X)$) with $\nu_1, \nu_2 \in C_s$ and $0 < \alpha < 1$, then $\nu \equiv \nu_1 \equiv \nu_2$ (mod $P_{se}(X)$);

(3) There is some ergodic channel $\nu_1 \in C_{se}$ such that $\nu(x, \cdot) \ll \nu_1(x, \cdot) \ P_{se}(X)$-a.e. $x$;

(4) $\nu_0 \in C_s$, $\nu_0(x, \cdot) \ll \nu(x, \cdot) \ P_{se}(X)$-a.e. $x$ $\implies \nu_0 \equiv \nu$ (mod $P_{se}(X)$).
Induced channel operators

A channel $\nu \in \mathcal{C}$ induces channel operators $F_\nu : M(X) \to M(Y)$ and $G_\nu : M(X) \to M(X \times Y)$ given by

$$F_\nu \mu(C) = \mu \nu(C) = \int_X \nu(x, C) \mu(dx),$$

$$G_\nu \mu(A \times C) = \mu \otimes \nu(A \times C) = \int_A \nu(x, C) \mu(dx)$$

for $\mu \in M(X)$, $A \in \mathcal{X}$ and $C \in \mathcal{Y}$. 
Channel operators

A channel operator is a linear operator $\mathcal{G}: M(X) \to M(X \times Y)$ that satisfies the following conditions:

(o1) $\mathcal{G}$ is a positive linear operator of norm one such that $\mathcal{G}: P(X) \to P(X \times Y)$;
(o2) $\mu(\cdot) = \mathcal{G}\mu(\cdot \times Y)$ for every $\mu \in P(X)$;
(o3) $\mathcal{G}\mu \leq \mu \times F\mu$ for every $\mu \in P(X)$, where $F\mu(\cdot) = \mathcal{G}\mu(X \times \cdot)$;
(o4) $\mu_1 \ll \mu_2 \implies \mathcal{G}\mu_1 \ll \mathcal{G}\mu_2$ for $\mu_1, \mu_2 \in P(X)$.

$\mathcal{O} = \mathcal{O}(X, Y)$ is the set of all channel operators.

($\mathcal{G}$ is linear if $\mathcal{G}(\alpha \mu + \beta \eta) = \alpha \mathcal{G}\mu + \beta \mathcal{G}\eta$ for $\alpha, \beta \in \mathbb{R}$ and $\mu, \eta \in M(X)$.)
Stationarity and ergodicity

A channel operator $G \in \mathcal{O}$ is said to be **stationary** if

$$(o5) \mu \in P_s(X) \implies G\mu \in P_s(X \times Y)$$

and to be **ergodic** if

$$(o6) \mu \in P_{se}(X) \implies G\mu \in P_{se}(X \times Y).$$

$\mathcal{O}_s = \mathcal{O}_s(X, Y)$ is the set of all stationary channel operators in $\mathcal{O}$ and $\mathcal{O}_{se} = \mathcal{O}_{se}(X, Y)$ is the set of all ergodic channel operators in $\mathcal{O}_s$.

If we identify a channel $\nu \in \mathcal{C}$ with the channel operator $G_{\nu} \in \mathcal{O}$, then we can regard $\mathcal{C} \subseteq \mathcal{O}$, $\mathcal{C}_s \subseteq \mathcal{O}_s$ and $\mathcal{C}_{se} \subseteq \mathcal{O}_{se}$.
Identicalness, extremality and absolute continuity

(1) Two channel operators \( G_1, G_2 \in \mathcal{O} \) are said to be **identical** (mod \( P_{se}(X) \)), denoted \( G_1 \equiv G_2 \) (mod \( P_{se}(X) \)), if \( G_1 \mu = G_2 \mu \) for every \( \mu \in P_{se}(X) \).

(2) A stationary channel operator \( G \in \mathcal{O}_s \) is said to be **extremal** in \( \mathcal{O}_s \) mod \( P_{se}(X) \), denoted \( G \in \text{ex} \mathcal{O}_s \) (mod \( P_{se}(X) \)), provided that if \( G \equiv \alpha G_1 + (1 - \alpha) G_2 \) (mod \( P_{se}(X) \)) with \( G_1, G_2 \in \mathcal{O}_s \) and \( 0 < \alpha < 1 \), then \( G \equiv G_1 \equiv G_2 \) (mod \( P_{se}(X) \)).

(3) Let \( G_1, G_2 \in \mathcal{O} \). Then, \( G_1 \) is said to be **absolutely continuous** with respect to \( G_2 \) mod \( P_{se}(X) \), denoted

\[
G_1 \ll G_2 \quad \text{(mod } P_{se}(X) \text{)},
\]

if \( G_1 \mu \ll G_2 \mu \) for every \( \mu \in P_{se}(X) \).
Proposition 4. For a stationary channel operator $G \in O_s$ the following statements are equivalent:

(1) $G \in O_{se}$, i.e., $G$ is ergodic;

(2) $G \in \text{ex } O_s \pmod{P_{se}(X)}$;

(3) There is some stationary ergodic channel operator $G_1 \in O_{se}$ such that $G \ll G_1 \pmod{P_{se}(X)}$;

(4) If a stationary channel operator $G_1 \in O_s$ is such that $G_1 \ll G \pmod{P_{se}(X)}$, then $G_1 \equiv G \pmod{P_{se}(X)}$. 
**Proposition 5.** In the alphabet message space formulation, we have $C = \mathcal{O}$, $C_s = \mathcal{O}_s$ and $C_{se} = \mathcal{O}_{se}$. That is, for any channel operator $G \in \mathcal{O}$ there exists a unique channel $\nu \in \mathcal{C}$ such that $G = G_\nu$.

**Remark.** The proof of the above is based on the fact that $X$ is a totally disconnected compact Hausdorff space.
Topology

Let $X$ be a nonempty set. A **topology** on $X$ is a set $\mathcal{X}$ of subsets of $X$ such that

1. $\emptyset, X \in \mathcal{X}$;
2. $O_1, O_2 \in \mathcal{X} \implies O_1 \cap O_2 \in \mathcal{X}$;
3. $O_j \in \mathcal{X}, j \in J \implies \bigcup_{j \in J} O_j \in \mathcal{X}$, where $J$ is any set.

Each set $O \in \mathcal{X}$ is called an **open set**. A complement of an open set is a **closed set**.

A subset $Y \subset X$ is said to be **compact** if $Y \subset \bigcup_{j \in J} O_j$ with $O_j \in \mathcal{X}, j \in J$ imply that for some finite sets $j_1, \ldots, j_m \in J$ such that $Y \subset \bigcup_{i=1}^m O_{j_i}$. 
Example

(1) In $\mathbb{R}$, any interval of the form $(a, b)$ is an open set. Let $\mathcal{X}$ be the topology determined by the set of all such intervals. A subset $Y \subset \mathbb{R}$ is compact if and only if $Y$ is bounded and closed.

(2) Let $X$ be a nonempty set and $\mathcal{X} = 2^X$, the power set. That is, $\mathcal{X}$ is the set of all subsets of $X$. $\mathcal{X}$ satisfies the conditions of a topology and is called the \textbf{discrete topology}.
**Pointwise weak* topology**

The **pointwise weak* topology** on the set of all channel operators \( \mathcal{O} = \mathcal{O}(X, Y) \) is defined by the system of neighborhoods given by

\[
W(G_0 : \mu_1, \ldots, \mu_m ; \varepsilon ; f_1, \ldots, f_n) = \bigcap_{i=1}^{m} \bigcap_{j=1}^{n} \left\{ G \in \mathcal{O} : |G\mu_i(f_j) - G_0\mu_i(f_j)| < \varepsilon \right\},
\]

where \( G_0 \in \mathcal{O}, \varepsilon > 0, \mu_1, \ldots, \mu_m \in P(X) \), and \( f_1, \ldots, f_n \in C(X \times Y) \). Here, we write

\[
G\mu(f) = \int_{X \times Y} f(x, y) G\mu(dx, dy), \quad \mu \in P(X), f \in C(X \times Y).
\]
Proposition 6.

(1) $\mathcal{O}$ and $\mathcal{O}_s$ are compact in the pointwise weak* topology.
(2) $\text{ex} \mathcal{O}_s = \mathcal{O}_{se}$ is nonempty.
(3) $\mathcal{O}_s$ is the closure of the convex hull of $\mathcal{O}_{se}$ in the pointwise weak* topology.
Open problems

Consider general input and output spaces. That is, we take compact Hausdorff spaces $X$ and $Y$ be the input and output, respectively. Then,

(1) Find sufficient conditions for $C = \mathcal{O}$. Note that $C \subseteq \mathcal{O}$ is always true. If possible, find necessary and sufficient conditions.

(2) What is the pointwise weak* closure of $C$ in $\mathcal{O}$? Can we approximate a channel operator by a net (or sequence) of information channels in some sense?

(3) What happens if we consider topologies other than the pointwise weak* topology on $\mathcal{O}$?