Comparison of Noisy Channels and Reverse Data-Processing Theorems

Francesco Buscemi[1]

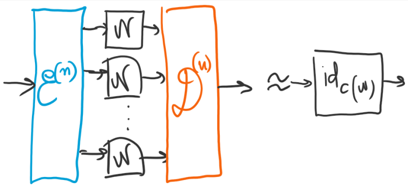
2017 IEEE Information Theory Workshop
Kaohsiung, 10 November 2017

[1]Dept. of Mathematical Informatics, Nagoya University, buscemi@i.nagoya-u.ac.jp

Summary

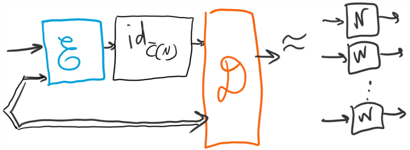
1. Partial orderings of communication channels (simulability orderings and coding orderings)
2. Reverse data-processing theorems
3. Degradability ordering: equivalent reformulations
4. Example application: characterization of memoryless stochastic processes

Direct Shannon Coding



direct capacity $C(\mathcal{N})$

Reverse Shannon Coding



reverse capacity $\overline{C}(\mathcal{N})$

Bennett, Devetak, Harrow, Shor, Winter (circa 2007-2014)

For a classical channel $\mathcal{N}$, when shared randomness is free,
$C(\mathcal{N}) = \overline{C}(\mathcal{N})$.

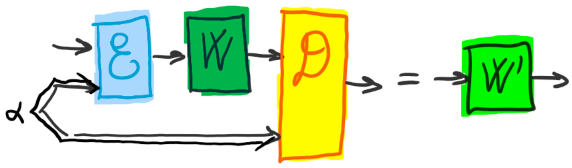
Shannon's noisy channel coding theorem is a statement about asymptotic simulability.

Shannon's "Channel Inclusion"

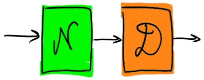
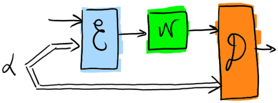
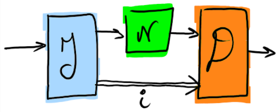
As a single-shot, zero-error analogue, Shannon, in *A Note on a Partial Ordering for Communication Channels* (1958), defines an exact form of simulability that he names "inclusion."

Definition (Inclusion Ordering)

Given two classical channels $\mathrm{W} : \mathcal{X} \to \mathcal{Y}$ and $\mathrm{W}' : \mathcal{X}' \to \mathcal{Y}'$, we write $\mathrm{W} \supseteq \mathrm{W}'$ if there exist encodings $\{\mathcal{E}_\alpha\}_\alpha$, decodings $\{\mathcal{D}_\alpha\}_\alpha$, and a probability distribution μ_α such that $\mathrm{W}' = \sum_\alpha \mu_\alpha (\mathcal{D}_\alpha \circ \mathrm{W} \circ \mathcal{E}_\alpha)$.



Three “Simulability” Orderings

		
Degradability	**Shannon’s Inclusion**	**Quantum Inclusion**
$\mathcal{N} \to \mathcal{N}'$	$\mathcal{N} \supseteq \mathcal{N}'$	$\mathcal{N} \supseteq_{\mathrm{q}} \mathcal{N}'$
$\exists \mathcal{D}$: CPTP such that $\mathcal{N}' = \mathcal{D} \circ \mathcal{N}$	$\exists \{\mathcal{E}_\alpha\}_\alpha, \{\mathcal{D}_\alpha\}_\alpha$: CPTP and μ_α : prob. dist. such that $\mathcal{N}' = \sum_\alpha \mu_\alpha (\mathcal{D}_\alpha \circ \mathcal{N} \circ \mathcal{E}_\alpha)$	$\exists \{\mathscr{I}^i\}_i$: CP instrument and $\{\mathcal{D}_i\}_i$: CPTP such that $\mathcal{N}' = \sum_i (\mathcal{D}_i \circ \mathcal{N} \circ \mathscr{I}^i)$

- for degradability, the two channels need to have the same input system; the two inclusion orderings allow to modify both input and output
- $\mathcal{N} \to \mathcal{N}' \implies \mathcal{N} \supseteq \mathcal{N}' \implies \mathcal{N} \supseteq_{\mathrm{q}} \mathcal{N}'$ (all strict implications)
- the “quantum inclusion” ordering $\supseteq_{\mathrm{q}}$ allows unlimited free classical forward communication: it is non-trivial only for quantum channels

Shannon's Coding Ordering

In the same paper, Shannon also introduces the following:

Definition (Coding Ordering)

Given two classical channels $\mathsf{W} : \mathcal{X} \to \mathcal{Y}$ and $\mathsf{W}' : \mathcal{X}' \to \mathcal{Y}'$, we write $\mathsf{W} \gg \mathsf{W}'$ if, for any (M, n) code for W' and any choice of prior distribution π_i on codewords, there exists an (M, n) code for W with average error probability $P_e = \sum_i \pi_i \lambda_i \le P'_e = \sum_i \pi_i \lambda'_i$.

Note: λ_i denotes the conditional probability of error, given that index i was sent.

Fact

$$\mathsf{W} \supseteq \mathsf{W}' \implies \mathsf{W} \gg \mathsf{W}' \implies C(\mathsf{W}) \ge C(\mathsf{W}')$$

The above definition and theorem can be directly extended to quantum channels and their classical capacity.

Other "Coding" Orderings

From: J. Körner and K. Marton, *The Comparison of Two Noisy Channels*. Topics in Information Theory, pp.411-423 (1977)

Definition (Capability and Noisiness Orderings)

Given two classical channels W : $\mathcal{X} \to \mathcal{Y}$ and W′ : $\mathcal{X} \to \mathcal{Z}$, we say that

1. W is more capable than W′ if, for any input random variable X, $H(X|Y) \leq H(X|Z)$
2. W is less noisy than W′ if, for any pair of jointly distributed random variables (U, X), $H(U|Y) \leq H(U|Z)$

Theorem (Körner and Marton, 1977)

It holds that

$$\textit{degradable} \implies \textit{less noisy} \implies \textit{more capable},$$

and all implications are strict.

Reverse Data-Processing Theorems

- two kinds of orderings: **simulability orderings** (degradability, Shannon inclusion, quantum inclusion) and **coding orderings** (Shannon coding ordering, noisiness and capability orderings)
- simulability orderings $\Longrightarrow$ coding orderings: data-processing theorems
- coding orderings $\Longrightarrow$ simulability orderings: reverse data-processing theorems (the problem discussed in this talk)

Why Reverse Data-Processing Theorems Are Relevant

- role in statistics: majorization, comparison of statistical models (Blackwell's sufficiency and Le Cam's deficiency), decision theory
- role in physics, esp. quantum theory: channels describe physical evolutions; hence, reverse-data processing theorems allow the reformulation of statistical physics in information-theoretic terms
- applications so far: quantum non-equilibrium thermodynamics; quantum resource theories; quantum entanglement and non-locality; stochastic processes and open quantum systems dynamics

Examples of Reverse Data-Processing Theorems: Equivalent Characterization of Degradability

A Classical Reverse Data-Processing Theorem...

Theorem

Given two classical channels $\mathsf{W} : \mathcal{X} \to \mathcal{Y}$ *and* $\mathsf{W}' : \mathcal{X} \to \mathcal{Z}$*, the following are equivalent:*

1. W *can be degraded to* W'*;*
2. *for any pair of jointly distributed random variables* (U, X)*,* $H_{\min}(U|Y) \leq H_{\min}(U|Z)$*.*

In fact, in point 2 it suffices to consider only random variables U *supported by* $\mathcal{Z}$ *and with uniform marginal distribution, i.e.,* $p(u) = \frac{1}{|\mathcal{Z}|}$*.*

Remarks

- condition (2) above is Körner's and Marton's noisiness ordering, with Shannon entropy replaced by $H_{\min}$
- by [König, Renner, Schaffner, 2009], W can be degraded to W' if and only if, for any initial joint pair (U, X), $P_{\text{guess}}(U|Y) \geq P_{\text{guess}}(U|Z)$

...and Its Quantum Version

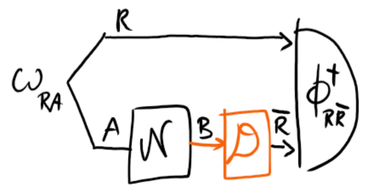
Theorem

Given two quantum channels $\mathcal{N} : A \to B$ and $\mathcal{N}' : A \to B'$, the following are equivalent:

1. *$\mathcal{N}$ can be degraded to $\mathcal{N}'$;*
2. *for any bipartite state ω_{RA},*
   $H_{\min}(R|B)_{(\mathrm{id}\otimes\mathcal{N})(\omega)} \leq H_{\min}(R|B')_{(\mathrm{id}\otimes\mathcal{N}')(\omega)}$.

In fact, in point 2 it suffices to consider only a system $R \cong B'$ and separable states ω_{RA} with maximally mixed marginal ω_R.

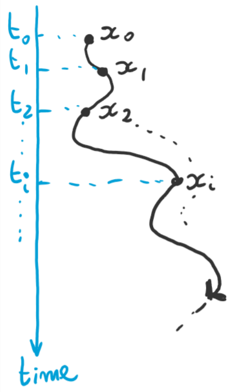
Remark. In words, for any initial bipartite state ω_{RA}, the maximal singlet fraction of $(\mathrm{id}_R \otimes \mathcal{N}_A)(\omega_{RA})$ is never smaller than that of $(\mathrm{id}_R \otimes \mathcal{N}'_A)(\omega_{RA})$.



An Application in Quantum Statistical Mechanics: Quantum Markov Processes

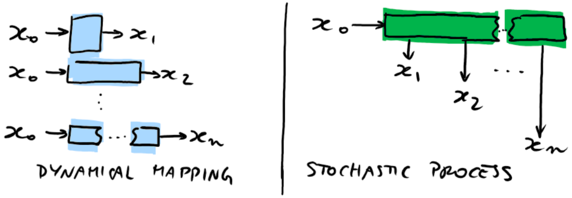
Discrete-Time Stochastic Processes

- Let x_i, for $i = 0, 1, \ldots,$ index the state of a system at time $t = t_i$
- Let $p(x_i)$ be the state distribution at time $t = t_i$
- The process is fully described by its joint distribution $p(x_N, x_{N-1}, \ldots, x_1, x_0)$
- If the system can be initialized at time $t = t_0$, it is convenient to identify the process with the conditional distribution $p(x_N, x_{N-1}, \ldots, x_1 | x_0)$



From Stochastic Processes to Dynamical Mappings

From a stochastic process $p(x_N, \ldots, x_1|x_0)$, we obtain a family of noisy channels $\{p(x_i|x_0)\}_{i\geq 0}$ by marginalization.



Definition (Dynamical Mappings)

A dynamical mapping is a family of channels $\{p(x_i|x_0)\}_{i\geq 1}$.

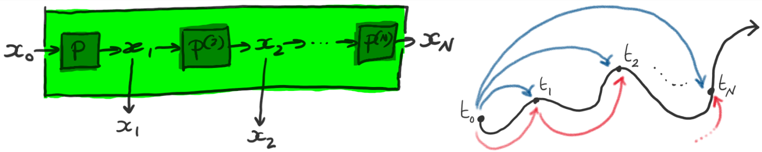
Remarks.

- Each stochastic process induces one dynamical mapping by marginalization; however, the same dynamical mapping can be "embedded" in many different stochastic processes.
- For quantum systems, dynamical mappings are okay, not so stochastic processes (no N-point time correlations).

Definition (Markovianity)

A stochastic process $p(x_N, \cdots, x_1|x_0)$ is said to be **Markovian** whenever

$$p(x_N, \cdots, x_1|x_0) = p^{(N)}(x_N|x_{N-1})p^{(N-1)}(x_{N-1}|x_{N-2}) \cdots p(x_1|x_0)$$



Definition (Divisibility)

A dynamical mapping $\{p(x_i|x_0)\}_{i\geq 1}$ is said to be **divisible** whenever

$$p(x_{i+1}|x_0) = \sum_{x_i} q^{(i+1)}(x_{i+1}|x_i)p(x_i|x_0) , \quad \forall i \geq 1 .$$

Hence, a divisible dynamical mapping can always be embedded in the Markovian process $q^{(N)}(x_N|x_{N-1}) \cdots q^{(2)}(x_2|x_1)p(x_1|x_0)$.

Divisibility as "Decreasing Information Flow"

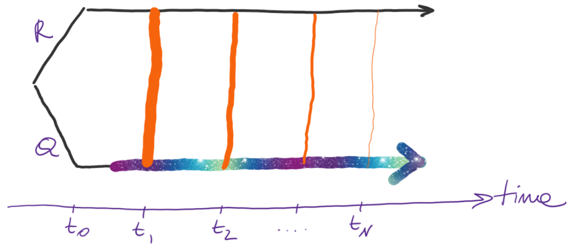
From the reverse data-processing theorems discussed before, we obtain:

Theorem

Given an initial open quantum system Q_0, a quantum dynamical mapping $\left\{\mathcal{N}^{(i)}_{Q_0\to Q_i}\right\}_{i\geq 1}$ is divisibile if and only if, for any initial state ω_{RQ_0},

$$H_{\min}(R|Q_1) \leq H_{\min}(R|Q_2) \leq \cdots \leq H_{\min}(R|Q_N)\ .$$

The same holds, mutatis mutandis, also for classical dynamical mappings.



Concluding Summary

Reverse data-processing theorems provide:

- a powerful framework to understand time-evolution in statistical physical systems
- complete (faithful) sets of monotones for generalized resource theories (including quantum non-equilibrium thermodynamics)
- new insights in the structure of noisy channels (e.g., new metrics, etc)

Applications to coding? Complexity theory?