Functional Methods for Information Theoretic Converes

Jingbo Liu

CISS 2018

March 21, 2018

Princeton University
Overview

- Parallel and independent developments in functional analysis and information theory.
- Information theoretic converses via functional formulations of information measures.
Many information measures have entropic and functional representations.
Example 1: strong data processing inequality

Fix $Q_{XY} = Q_X Q_{Y|X}$.

- Data processing:

  \[ D(P_Y \| Q_Y) - D(P_X \| Q_X) \leq 0, \quad \forall P_X. \quad (1) \]

- Strong data processing: for some $c \geq 1$,

  \[ cD(P_Y \| Q_Y) - D(P_X \| Q_X) \leq 0, \quad \forall P_X. \quad (2) \]

Ahlswede-Gács-Körner 1976:

\[
\sup_{P_X} \{ cD(P_Y \| Q_Y) - D(P_X \| Q_X) \} \\
= \sup_{f: \mathcal{Y} \rightarrow (0, \infty)} \{ \ln Q_X \left( e^{cQ_{Y|X}(\ln f)} \right) - c \ln Q_Y(f) \} \quad (3)
\]

Anantharam-Gohari-Kamath-Nair 2013: MI representation of SDPI.
Linear combinations of RE naturally arise in image-size problems

Ahlswede, Gács, Körner and Csiszár introduced image-size problem for strong converses in network information theory (NIT).

\[ x : Q_{Y|X=x}[A] \geq 1 - \epsilon \]

\[
\ln Q_X[\epsilon\text{-preimage of } A] - c(1 - \epsilon) \ln Q_Y[A] \\
\leq \sup_{P_X}\{cD(P_Y\|Q_Y) - D(P_X\|Q_X)\} + c \ln 2 \tag{4}
\]
Example 2: Brascamp-Lieb inequality and its reverse

[Brasamp, Lieb 76] Let $E, E_1, \ldots, E_m$ be Euclidean spaces, and $B_i : E \to E_i$ be linear maps. Let $(c_i)_{i=1}^m$ and $D$ be positive real numbers. Then

$$\int \prod_{i=1}^m f_{i}^{c_i}(B_i x) \, dx \leq D \prod_{i=1}^m \left( \int f_i(x_i) \, dx_i \right)^{c_i},$$

(5)

holds for all nonnegative $(f_i)$ if and only if it holds for all centered Gaussian $(f_i)$.

[Barthe 98] For $F$ a positive real number,

$$\int \sup_{(y_i) : \sum_{i=1}^m c_i B_i^* y_i = x} \prod_{i=1}^m f_{i}^{c_i}(y_i) \, dx \geq F \prod_{i=1}^m \left( \int f_i(y_i) \, dy_i \right)^{c_i},$$

(6)

holds for all nonnegative measurable $(f_i)$ if and only if it holds for all centered Gaussian $(f_i)$.

[Carlen, Cordero-Erausquin 09]: proof based on the entropic formulation.
Idea motivated by information theory: forward and reverse channels

How do we view the relation of the forward and the reverse Brascamp-Lieb inequalities?

- The image-size problem sometimes involves not only forward channels but also reverse channels (in the context of, e.g., degraded broadcast channels, zigzag networks, Gelfand-Pinsker).
- We can view each $B_i$ in the Brascamp-Lieb inequality as (deterministic) forward channels, and $B_i^*$ in the reverse Brascamp-Lieb inequality as (deterministic) reverse channels.
Our work: forward-reverse Brascamp-Lieb inequality

Theorem (L-Courtade-Cuff-Verdú 15)

Consider \( b_1, \ldots, b_l, c_1, \ldots, c_m, D \in (0, \infty) \). Let \( E_1, \ldots, E_l, E_1', \ldots, E_m' \) be Euclidean spaces, and let \( B_{ji} : E_i \to E_j' \) be a linear map for each \( i \in \{1, \ldots, l\} \) and \( j \in \{1, \ldots, m\} \). Then

\[
\prod_{i=1}^{l} \left( \int g_i \right)^{b_i} \leq D \prod_{j=1}^{m} \left( \int f_j \right)^{c_j}
\]

for all continuous functions \( f_j : E_j' \to [0, +\infty) \), \( g_i : E_i \to [0, \infty) \) satisfying

\[
\prod_{i=1}^{l} g_i^{b_i}(x_i) \leq \prod_{j=1}^{m} f_j^{c_j} \left( \sum_{i=1}^{l} B_{ji} x_i \right), \quad \forall x_1, \ldots, x_l
\]

if and only if the same holds for all centered Gaussian functions \( f_1, \ldots, f_m, g_1, \ldots, g_l \).

Proof by deriving an entropic formulation; use a technique in [Geng-Nair 14].
Forward-reverse Brascamp-Lieb implies other inequalities
A “geometric” forward-reverse Brascamp-Lieb inequality

Theorem (L-Courtade-Cuff-Verdú)

Let \( l \) be a positive integer, and let \( \mathbf{M} := (m_{ji})_{1 \leq j \leq l, 1 \leq i \leq l} \) be an orthogonal matrix. For any nonnegative continuous functions \((f_j)_{j=1}^l\) and \((g_i)_{i=1}^l\) on \( \mathbb{R} \) such that

\[
\prod_{i=1}^l g_i(x_i) \leq \prod_{j=1}^l f_j \left( \sum_{i=1}^l m_{ji} x_i \right), \quad \forall x^l \in \mathbb{R}^l, \tag{9}
\]

we have

\[
\prod_{i=1}^l \int g_i(x)dx \leq \prod_{i=1}^l \int f_j(x)dx. \tag{10}
\]
Part 2:
A technique of converse/impossibility bounds using functional representations of information measures
Review of applications in impossibility bounds

A partial list of examples in the context of estimation or common randomness generation:

- [Bogdanov, Mossel 11] distributed common randomness generation via hypercontractivity (without communication).
- [Zhang, Duchi, Jordan, Wainwright 13][Shamir 14][Braverman, Garg, Ma 14][Xu, Raginsky 15] distributed estimation via strong data processing.
- [L, Cuff, Verdú 15][Guruswami, Radhakrishnan 16] distributed common randomness generation via hypercontractivity (with communication).

Today: canonical approach to NIT converses (superseding blowing-up)
Source coding with compressed side information

Given: per-letter distribution $Q_{XY}$.

- $R_1 := \frac{1}{n} \log |\mathcal{W}_1|$, $R_2 := \frac{1}{n} \log |\mathcal{W}_2|$.
- Goal: $\hat{Y}^n = Y^n$ with high probability.
- Question: asymptotic behaviour of $\log |\mathcal{W}_1| + c \log |\mathcal{W}_2|$ for a nonvanishing error.
Achievability: well-developed.

Converse: more challenging!

*The study of second-order asymptotics for multi-terminal problems is at its infancy.... The primary difficulty is our inability to deal, in a systematic and principled way, with auxiliary random variables for the (strong) converse part. Thus, genuinely new non-asymptotic converses need to be developed....*

—Tan, F&T monograph, “Open problems and challenges.”
Source coding with compressed side information: results

\[ Y^n \rightarrow \text{Encoder 2} \rightarrow W_2 \rightarrow \text{Decoder} \rightarrow \hat{Y}^n \]

\[ X^n \rightarrow \text{Encoder 1} \]

\[ \phi_c(Q_{XY}) := \inf_{P_U|X} \{cH(Y|U) + I(U;X)\}. \] (11)

\[ \forall (n, \epsilon), \text{Second Order Rate} := \inf_{\text{codes}} \{\ln |W_1| + c \ln |W_2|\} - n\phi_c(Q_{XY}). \] (12)

Ahlswede, Gács, Körner 76
L, van Handel, Verdú 17

\[ \text{SOR}(n, \epsilon) = O_\epsilon(\sqrt{n \ln^{3/2} n}) \]

\[ \text{SOR}(n, \epsilon) = O_\epsilon(\sqrt{n}) \]

Dispersion

\[ \lim_{\epsilon \downarrow 0} \limsup_{n \to \infty} \frac{\text{SOR}^2(n, \epsilon)}{2n \ln \frac{1}{\epsilon}} = \text{Var} \left( \partial \phi_c|_{Q_{XY}} (X, Y) \right) \]
Challenge 1: connect the single-letter expression with a functional inequality

Recall that the minimum weighted sum of compression rates is

\[ \phi_c(Q_{XY}) := \inf_{P_{U|X}} \{cH(Y|U) + I(U; X)\}. \] (13)

However, the functional characterization is known [AGK76] for

\[ \inf_{P_X} \{cH(P_Y) + D(P_X \| Q_X)\}. \] (14)

How do we connect (13) and (14)?
Idea 1: fixed-composition analysis

Consider instead $P_{XY^n}$, the equiprobable distribution on sequences of $n$-type $P_{XY}$. Define

$$
\psi_{c,n}(P_{XY}) := \inf_{S_{X^n}} \{ cH(S_{Y^n}) + D(S_{X^n}||P_{X^n}) \}. 
$$

(15)

Lemma (L 18)

Given $Q_{XY}$ and $c \geq 1$, there exists $\lambda \in (0, 1)$ and $E > 0$ such that for any $n \geq 2$ and $n$-type $P_{XY}$:

$$
|P_{XY} - Q_{XY}| < \lambda,
$$

$$
\psi_{c,n}(P_{XY}) \geq n\phi_c(P_{XY}) - E \ln n. 
$$

(16)

Analogous to the single-letterization in the i.i.d. case, but pays extra $\ln n$ term.
Challenge 2: dealing with geometric average

For any $w_1$ (compression of $x^n$), let $B_{w_1} := \{y^n : \hat{y}^n(w_1, w_2(y^n)) = y^n\}$ be the recoverable set. Recall that

$$
\sup_{P_X} \{cD(P_Y \| Q_Y) - D(P_X \| Q_X)\} = \sup_{f : \mathcal{Y} \to (0, \infty)} \{\ln Q_X(e^{cQ_Y|X(\ln f)}) - c \ln Q_Y(f)\}. \tag{17}
$$

- Plugging in $f = 1_{B_{w_1}}$ is dismal, since the integral of $f$ becomes zero!
- If the geometric average $e^{cQ_Y|X(\ln f)}$ equals the arithmetic average $Q^c_{Y|X}(f)$, life would be much easier.
Idea 2: reverse hypercontractivity

Given $q < p < 1$, we say an operator $\Lambda$ is $(p, q)$-reverse hypercontractive if

$$\|\Lambda f\|_q \geq \|f\|_p, \quad \forall f : \mathcal{Y} \to [0, \infty).$$  \hspace{1cm} (18)

Note: $\|f\|_0 = \exp(\int \ln f)$ for probability measures.

Using the Markov semigroup theory, we construct the “magic operator” $\Lambda$ such that

- Geometric average of $\Lambda f$ roughly equals average of $f$ (w.r.t. $P_{Y^n|x^n}$ for any $x^n$) – solving Challenge 2!
- average of $\Lambda f$ roughly equals average of $f$ (w.r.t. $P_{Y^n}$).
Comparison with the blowing-up method

Figure: Schematic comparison of $1_A$, $1_{A_{nt}}$ and $T_t \otimes^n 1_A$, where $A$ is the indicator function of a Hamming ball.
High level description of the methodology

- Information measures
  - Integrals of functions
  - Measures of sets

Observables:
- Integrals of **functions**
  - convex duality
  - data processing
  - change of measure
  - take indicators (or the alike)

Observables:
- Measures of **sets**
  - data processing
  - change of measure
Thank you

Collaborators: Thomas Courtade, Paul Cuff, Ramon van Handel, Sergio Verdú (alphabetical)

Email: jingbo@princeton.edu

Personal page: http://www.princeton.edu/~jingbo/