Data Processing Inequalities Based on a Certain Structured Class of Information Measures With Application to Estimation Theory

Neri Merhav

Department of Electrical Engineering

Technion—Israel Institute of Technology

Haifa 32000, Israel

ITA 2012, San Diego, CA, February 5–10, 2012.

Background

Classical joint source–channel data processing inequality (DPI) for $U \to X \to Y \to V$:

$$R(D) \leq I(U;V) \leq I(X;Y) \leq C \quad \Rightarrow \quad D \geq R^{-1}(C).$$

Ziv and Zakai (1973) generalized to:

$$R_Q(D) \leq I_Q(U;V) \leq I_Q(X;Y) \leq C_Q \quad \Rightarrow \quad D \geq R_Q^{-1}(C_Q),$$

where

$$I_Q(A;B) = \mathbf{E}\left\{\log Q\left(\frac{P(A)P(B)}{P(A,B)}\right)\right\}$$

for a general convex function Q (see also Csiszár's f–divergence, 1972). Further generalization (Zakai & Ziv, 1975) to multivariate convex functions

$$I_Q(A;B) = \mathbf{E}\left\{\log Q\left(\frac{\mu_1(A,B)}{P(A,B)},\ldots,\frac{\mu_k(A,B)}{P(A,B)}\right)\right\}.$$

Background (Cont'd)

Gurantz (1974) examined

$$G(Y|x, x_1, \ldots, x_k) = \int_{\mathcal{Y}} \mathrm{d}y \cdot P_{Y|X}(y|x) \times$$

$$Q_1 \left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x)} \cdot Q_2 \left(\frac{P_{Y|X}(y|x_2)}{P_{Y|X}(y|x_1)} \cdot Q_3 \left(\ldots Q_k \left(\frac{P_{Y|X}(y|x_k)}{P_{Y|X}(y|x_{k-1})} \right) \ldots \right) \right) \right),$$

and showed that for $X \to Y \to Z$,

$$G(Y, x, x_1, \ldots, x_k) \geq G(Z, x, x_1, \ldots, x_k).$$

This yields $R_G(U;V) \leq C_G$ w.r.t. $I_G(A;B) = \mathbf{E}\{G(B|A, A_1, \ldots, A_k)\}$, where $E\{\cdot\}$ is w.r.t. $P_{AB}(a,b) \times P_A(a_1) \times \ldots \times P_A(a_k)$.

While I_G can be shown to be a special case of the ZZ75 information measure, it has an interesting structure that calls for further study.

Choice of the Convex Functions

Consider the functions

$$
\begin{aligned}
Q_1(t) &= -t^{a_1} \quad 0 \le a_1 \le 1 \\
Q_i(t) &= t^{a_i} \quad 0 \le a_i \le 1, \quad 2 \le i \le k
\end{aligned}
$$

leading to

$$
G(Y|x_0, x_1, \ldots, x_k) = -\int_{\mathcal{Y}} \mathrm{d}y P_{Y|X}(y|x_0) \times
$$

$$
\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_0)} \left(\frac{P_{Y|X}(y|x_2)}{P_{Y|X}(y|x_1)} \left(\cdots \left(\frac{P_{Y|X}(y|x_k)}{P_{Y|X}(y|x_{k-1})} \right)^{a_k} \right)^{a_{k-1}} \cdots \right)^{a_2} \right)^{a_1}
$$

$$
= \quad -\int_{\mathcal{Y}} \mathrm{d}y \prod_{i=0}^{k} P_{Y|X}^{b_i}(y|x_i)
$$

where $b_i \ge 0$ for all i and $\sum_{i=0}^{k} b_i = 1$.

Choice of the Convex Functions (Cont'd)

Choosing $b_i = 1/(k+1)$ for all i yields

$$I_G(X;Y) = -\int_{\mathcal{Y}} \mathrm{d}y \left[\int_{\mathcal{X}} \mathrm{d}x P_X(x) P_{Y|X}^{1/(k+1)}(y|x)\right]^{k+1} = -\exp\{-E_0(\rho, P_X)\}\Big|_{\rho=k}.$$

Comments:

- Gallager's function E_0 indeed satisfies a DPI (Kaplan & Shamai 1993).
- Choice of integer ρ ($\rho = k$) is relatively easy:
  Square brackets $\to$ multidimensional integral $\to$ swapping with $\int \mathrm{d}y$.
- Generalizing from the Bhattacharyya distance ($k = 1$) to a general k.

Questions:

- Zakai & Ziv (1975) examined the choice $k = 1$ in signal parameter estimation. Is $k = 1$ the best choice or can it be improved?
- How does the best bound of this type compare to other bounds from estimation theory?

Application to Estimation Theory

Consider the model

$$y(t) = x(t, u) + n(t), \quad 0 \leq t < T,$$

where $x(t, u)$ is an arbitrary waveform, parameterized by u, with

$$\int_0^T \mathrm{d}t \cdot x^2(t, u) = E$$

and $n(t)$ is AWGN with spectral density $N_0/2$.
It is assumed that u is realization of $U \sim \text{Unif}[-1/2, +1/2)$. We are interested in lower bounds on

$$\overline{\epsilon^2} = \mathbf{E}(\hat{U} - U)^2$$

in the high–SNR regime $E/N_0 \gg 1$.
We focus on universal lower bounds (fundamental limits), that are independent of the waveform. No bandwidth constraints are imposed.

Calculation of $R_G(D)$

The high–res behavior of $R_G(D)$ is as follows:

$$R_G(D) \sim \begin{cases} -4c\sqrt{D} & k = 1 \\ -4\left(\frac{k}{k-2}\right)^k \cdot D & k > 2 \end{cases}$$

where

$$c = \int_{-\infty}^{+\infty} \frac{\mathrm{d}t}{(1+t^2)^2}.$$

For $k = 2$, we have

$$\log[-R_G(D)] \sim \log D$$

in the sense that

$$\lim_{D \to 0} \frac{\log[-R_G(D)]}{\log D} = 1.$$

Calculation of $I_G(U;Y)$

For the AWGN channel

$$P(y|u) \propto \exp\left\{-\frac{1}{N_0}\int_0^T [y(t)-x(t,u)]^2 \mathrm{d}t\right\},$$

we have

$$I_G(U;Y) \leq -\exp\left\{-\frac{E}{N_0}\cdot\frac{k}{(k+1)}\cdot(1-\varrho)\right\},$$

where

$$\varrho = \frac{1}{E}\mathbf{E}\left\{\int_0^T \mathrm{d}t\cdot x(t,U)x(t,U')\right\} = \frac{1}{E}\int_0^T \mathrm{d}t\cdot[\overline{x}(t)]^2,$$

and

$$\overline{x}(t) = \mathbf{E}\{x(t,U)\} = \int_{-1/2}^{+1/2} \mathrm{d}u\cdot x(t,u).$$

Note that

$$E(1-\varrho) = \int_0^T \mathrm{d}t\cdot \mathrm{Var}\{x(t,U)\}.$$

DPI Estimation Error Bounds

Applying the DPI, $R_G(D) \le I_G(U;Y)$, we get

$$\overline{\epsilon^2} \ge \begin{cases} \frac{1}{16c^2}\exp\{-(1-\varrho)E/N_0\} & k=1 \quad \text{(Zakai \& Ziv '75)} \\ \frac{1}{4}\left(1-\frac{2}{k}\right)^k \exp\left\{-(1-\varrho)\frac{k}{k+1}\cdot\frac{E}{N_0}\right\} & k>2 \end{cases}$$

and for $k=2$

$$\liminf_{E/N_0\to\infty} \frac{\log \overline{\epsilon^2}}{E/N_0} \ge -\frac{2}{3}\cdot(1-\rho).$$

Discussion:

- $k=2$ is the best choice of k for high SNR.
- The bounds are minimized by signals with $\varrho=0$.
- Upon setting $\varrho=0$, the bounds are independent of the modulation.
- For the bounds to be tight, $\rho(U,U')=\int_0^T \mathrm{d}t\cdot x(t,U)x(t,U')/E$ should be nearly zero with high probability – rapidly vanishing correlation.
- It is possible to achieve $\overline{\epsilon^2}\sim e^{-E/(3N_0)}$, e.g., by PPM. The gap is 3dB.

Comparison to Other Bounds

The Weiss–Weinstein bound (WWB) for a given modulation is

$$\text{WWB} = \sup_{h \neq 0} \frac{h^2 \exp\{-[1 - r(h)]E/(2N_0)\}}{2(1 - \exp\{-[1 - r(2h)]E/(2N_0)\})},$$

where

$$r(h) = \rho(u, u+h) = \frac{1}{E} \int_0^T x(t,u)x(t,u+h)\mathrm{d}t.$$

To derive a universal lower bound, this should be minimized over all feasible correlation functions $r(\cdot)$ – not a trivial minimax problem.
One can lower bound by solving the maximin problem, yielding

$$\text{WWB} = \frac{e^{-E/N_0}}{2(1 - e^{-E/N_0})}.$$

But this is inferior to our earlier bounds for $k > 1$.

Comparison to Other Bounds (Cont'd)

A simple consideration of M–ary signal detection yields

$$\overline{\epsilon^2} \geq \frac{1}{8M^2} \cdot Q\left(\sqrt{\frac{E}{N_0} \cdot \frac{M}{M-2}}\right),$$

where $M = 4, 6, 8, \ldots$. For high SNR, this is exponentially equivalent to

$$\exp\left\{-\frac{E}{2N_0} \cdot \frac{M}{M-2}\right\},$$

which, for large enough M, is arbitrarily close to $e^{-E/(2N_0)}$. This is better than our best bound $e^{-2E/(3N_0)}$.

Q: In what situations is the DPI bound superior to other bounds?

Channels with Uncertainty – AWGN with Fading

Suppose that there is an unknown nuisance parameter A (e.g., fading), independent of U and

$$P_{Y|U}(y|u) = \int_{-\infty}^{+\infty} \mathrm{d}a \cdot P_A(a) P_{Y|U,A}(y|u,a).$$

Think of $I_G(U;Y)$ as a functional of $P_{Y|U}$, denoted $\mathcal{I}(P_{Y|U}(\cdot|u))$, then it is a convex functional, namely,

$$\begin{aligned}\mathcal{I}(P_{Y|U}(\cdot|u)) &= \overbrace{\mathcal{I}\left(\int_{-\infty}^{+\infty} \mathrm{d}a P_A(a) P_{Y|U,A}(\cdot|u,a)\right)}^{\text{unknown } A} \\ &\leq \underbrace{\int_{-\infty}^{+\infty} \mathrm{d}a P_A(a) \mathcal{I}(P_{Y|U,A}(\cdot|u,a))}_{\text{known } A}\end{aligned}$$

The AWGN Channel With Fading

Consider the channel

$$y(t) = a \cdot x(t,u) + n(t), \quad 0 \le t < T,$$

where a and u are realizations of A and U, respectively.
Assume that $A \sim \mathcal{N}(0, \sigma^2)$ is independent of U.

$$P_{Y|U}(y|u) \propto \int_{-\infty}^{+\infty} \mathsf{d}a \cdot \frac{e^{a^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{ -\frac{1}{N_0} \int_0^T [y(t) - a \cdot x(t,u)]^2 \mathsf{d}t \right\}$$
$$\propto \exp\left\{ \theta \left[\int_0^T y(t) x(t,u) \mathsf{d}t \right]^2 \right\}$$

where

$$\theta \triangleq \frac{2\sigma^2}{N_0^2(1 + 2\sigma^2 E/N_0)}.$$

Estimation Error Bounds for AWGN With Fading

Upon calculating $I_G(U;Y)$ for the AWGN channel with fading (under the rapidly vanishing correlation assumption), we obtain the high–SNR bounds

$$\overline{\epsilon^2} \geq \frac{g_k}{\sigma} \cdot \sqrt{\frac{N_0}{E}}$$

with

$$g_k = \frac{1}{4\sqrt{2}} \left(1 - \frac{2}{k}\right)^k \left(1 + \frac{1}{k}\right)^{(k+1)/2}, \qquad k = 1, 2, \ldots$$

The tightest bound is obtained with $k \to \infty$. Let

$$g_\infty = \lim_{k \to \infty} g_k = \frac{1}{4\sqrt{2}e^{3/2}} = 0.03944.$$

Thus, our asymptotic lower bound for high SNR is

$$\liminf_{E/N_0 \to \infty} \sqrt{\frac{E}{N_0}} \cdot \overline{\epsilon^2} \geq \frac{0.03944}{\sigma}.$$

Comparison with Other Bounds

The Weiss–Weinstein bound:

$$\text{WWB} \propto \frac{N_0}{\sigma^2 E}.$$

The M–ary signal detection bound:

$$\liminf_{E/N_0 \to \infty} \sqrt{\frac{E}{N_0}} \cdot \overline{\epsilon^2} \geq \frac{0.001758}{\sigma}.$$

The Chazan–Zakai–Ziv bound:

$$\liminf_{E/N_0 \to \infty} \sqrt{\frac{E}{N_0}} \cdot \overline{\epsilon^2} \geq \frac{0.00716}{\sigma},$$

a factor of 5.5 (7.4dB) smaller than the DPI bound.

Conclusion and Future Work

- We examined a family of information measures with a certain structure (Gurantz, 1974).
- For a specific choice of the the convex functions – equivalent to $E_0(\rho, P_X)|_{\rho=k}$ – an extension of the Bhattacharyya distance.
- Best choice of k: $k = 2$ for AWGN; $k \to \infty$ – for AWGN with fading.
- Bounds compete favorably with existing bounds, especially in situations of uncertainty. Explanation: convexity of $\mathcal{I}_G(P_{Y|U})$.
- Future work: Trying to close the gap between upper bound and universal lower bound of $\lim_{E/N_0 \to \infty} N_0 \log \overline{\epsilon^2}/E$.