Strong data-processing of mutual information: beyond Ahlswede and Gács

Yury Polyanskiy

Department of EECS
MIT

yp@mit.edu

Joint work with Yihong Wu (UIUC)

Feb 2015

Preview of main result

Consider Markov chain:

$$W \to X \to Y$$

- $X, Y \in \mathbb{R}$
- $X \to Y$ is additive (over $\mathbb{R}$) noise channel $Y = X + Z$
- moment-constraint: $\mathbb{E}|X|^p \le \gamma$

- Data-processing tells us:

$$I(W;Y) \le I(W;X)$$

Preview of main result

Consider Markov chain:

$$W \to X \to Y$$

- $X, Y \in \mathbb{R}$
- $X \to Y$ is additive (over $\mathbb{R}$) noise channel $Y = X + Z$
- moment-constraint: $\mathbb{E}|X|^p \le \gamma$

- Data-processing tells us:

$$I(W;Y) \le I(W;X)$$

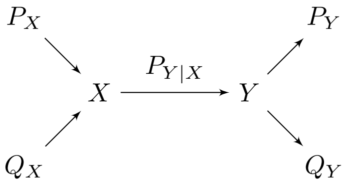
- This work:

$$I(W;Y) \le F(I(W;X)) \qquad \text{and } F(t) < t$$

Discrete channels

Data processing inequality

- KL divergence



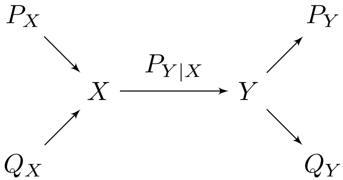
$$\Rightarrow D(P_Y||Q_Y) \le D(P_X||Q_X)$$

- mutual information

$$U \to X \to Y \Rightarrow I(U;Y) \le I(U;X)$$

Strong data processing inequality

- KL divergence



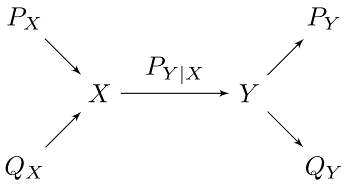
$$\Rightarrow D(P_Y||Q_Y) \le \eta_{\mathrm{KL}} D(P_X||Q_X)$$

- mutual information

$$U \to X \to Y \Rightarrow I(U;Y) \le \eta_{\mathrm{KL}} I(U;X)$$

[Ahlswede-Gács'76], [Anantharam-Gohari-Kamath-Nair'13]

Strong data processing inequality



- For fixed $P_{Y|X}$, KL contraction ratio:

$$\eta_{\mathrm{KL}} = \sup_{P_X \neq Q_X} \frac{D(P_Y || Q_Y)}{D(P_X || Q_X)} = \sup_{U \to X \to Y} \frac{I(U;Y)}{I(U;X)}$$

Theorem (Ahlswede-Gács)

$\eta_{\mathrm{KL}} < 1$ *iff zero-error capacity* $C_0 = 0$.

Dobrushin coefficient

- total variation

$$\|P - Q\|_{\mathrm{TV}} = \frac{1}{2}\int |\mathrm{d}P - \mathrm{d}Q|$$

- Dobrushin coefficient

$$\eta_{\mathrm{TV}} = \sup_{P_X \neq Q_X} \frac{\|P_Y - Q_Y\|_{\mathrm{TV}}}{\|P_X - Q_X\|_{\mathrm{TV}}} = \sup_{x,x'} \|P_{Y|X=x} - P_{Y|X=x'}\|_{\mathrm{TV}}.$$

Dobrushin coefficient

- total variation

$$\|P - Q\|_{\mathrm{TV}} = \frac{1}{2}\int |\mathrm{d}P - \mathrm{d}Q|$$

- Dobrushin coefficient

$$\eta_{\mathrm{TV}} = \sup_{P_X \neq Q_X} \frac{\|P_Y - Q_Y\|_{\mathrm{TV}}}{\|P_X - Q_X\|_{\mathrm{TV}}} = \sup_{x,x'} \|P_{Y|X=x} - P_{Y|X=x'}\|_{\mathrm{TV}}.$$

Theorem (Cohen-Iwasa-Rautu-Ruskai-Seneta-Zbăganu'93)

$$\eta_{\mathrm{KL}} \le \eta_{\mathrm{TV}}$$

- Not tight, but easy to compute. Example:

$$BSC(\delta): \quad \begin{cases} \eta_{\mathrm{KL}} = (1-2\delta)^2 \\ \eta_{\mathrm{TV}} = |1-2\delta| \end{cases}$$

Application 1: mixing of Markov chains

Consider a M.C. with invariant dist. P^*:

$$X_0 \to X_1 \to X_2 \to \cdots$$

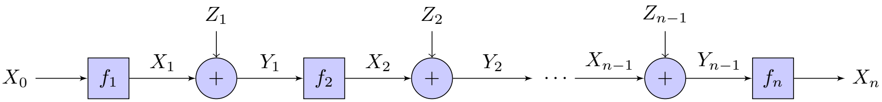
Contraction:

$$\|P_{X_n} - P^*\|_{\mathrm{TV}} \le (\eta_{\mathrm{TV}})^n \qquad \text{(Dobrushin)}$$

$$\chi^2(P_{X_n} \| P^*) \le (\eta_{\chi^2})^n \cdot \chi^2(P_{X_0} \| P^*) \qquad \text{(spectral gap)}$$

$$D(P_{X_n} \| P^*) \le (\eta_{\mathrm{KL}})^n \cdot D(P_{X_0} \| P^*) \qquad \text{(log-Sobolev)}$$

Application 2: Dissipation of information

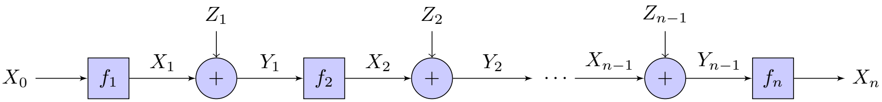


If $\eta_{\mathrm{KL}} < 1$, then

$$I(X_0; X_n) \le I(X_0; Y_{n-1}) \le \eta_{\mathrm{KL}} I(X_0; X_{n-1}) \le \cdots \le \eta_{\mathrm{KL}}^{n-1} I(X_0; X_1)$$

$\to 0$ exponentially

Application 2: Dissipation of information



If $\eta_{\mathrm{KL}} < 1$, then

$$I(X_0; X_n) \le I(X_0; Y_{n-1}) \le \eta_{\mathrm{KL}} I(X_0; X_{n-1}) \le \cdots \le \eta_{\mathrm{KL}}^{n-1} I(X_0; X_1)$$

$$\to 0 \quad \text{exponentially}$$

Converse bound on computation via noisy gates

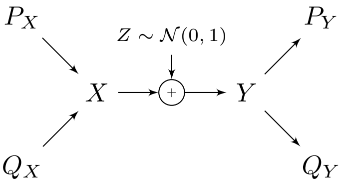
Theorem (Pippenger'88, Feder'89, Evans-Schulman'99)

Circuits of gates with fan-in k perturbed by BSC(δ) are not reliable if

$$\delta > \frac{1}{2} - \frac{1}{2\sqrt{k}}\,.$$

Non-discrete channels

Dobrushin coefficient for amplitude constraint



Two types of input constaints:

- Amplitude constraint: $X \in [-A, A]$ almost surely. This is easy:

$$\eta_{\mathrm{KL}} \le \eta_{\mathrm{TV}} = \sup_{x,x' \in [-A,A]} \|\mathcal{N}(x,1) - \mathcal{N}(x',1)\|_{\mathrm{TV}}$$

$$= 1 - 2Q(A) < 1 \qquad Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

- Power constraint:

$$\mathbb{E}[X^2] \le P\,.$$

No contraction under $\mathbb{E}[X^2] \le P$

- Consider:

$$\begin{cases} P_X = (1-t)\delta_0 + t\delta_{\sqrt{P/t}} \\ Q_X = (1-t)\delta_0 + t\delta_{-\sqrt{P/t}} \end{cases} \Rightarrow \frac{\|P_Y - Q_Y\|_{\mathrm{TV}}}{\|P_X - Q_X\|_{\mathrm{TV}}} \to 1$$

- Similarly, for KL

$$\sup_{P_X, Q_X} \frac{D(P_X * \mathcal{N} \| Q_X * \mathcal{N})}{D(P_X \| Q_X)} = 1$$

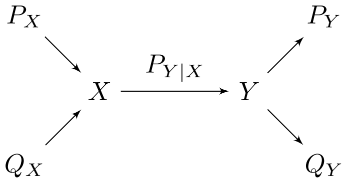
- . . . and for MI

$$\sup_{W-X-Y:\ \mathbb{E}[X^2] \le P} \frac{I(W; X+Z)}{I(W; X)} = 1$$

Punchline:

$$\boxed{\eta_{\mathrm{KL}} = \eta_I = \eta_{\mathrm{TV}} = 1}$$

Revisit strong data-processing

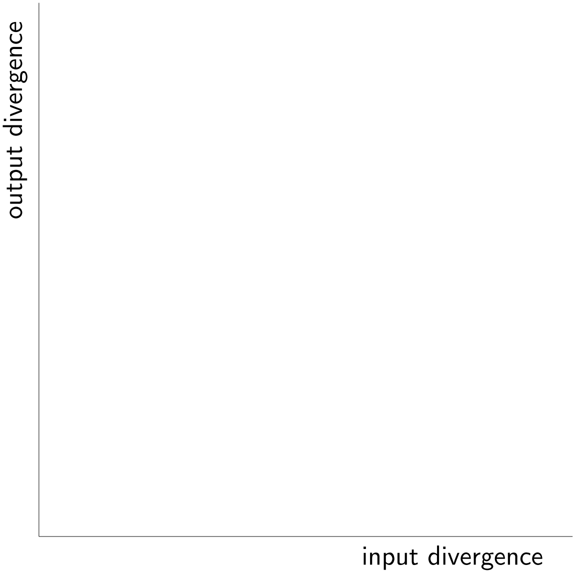


- Strong D.P.:

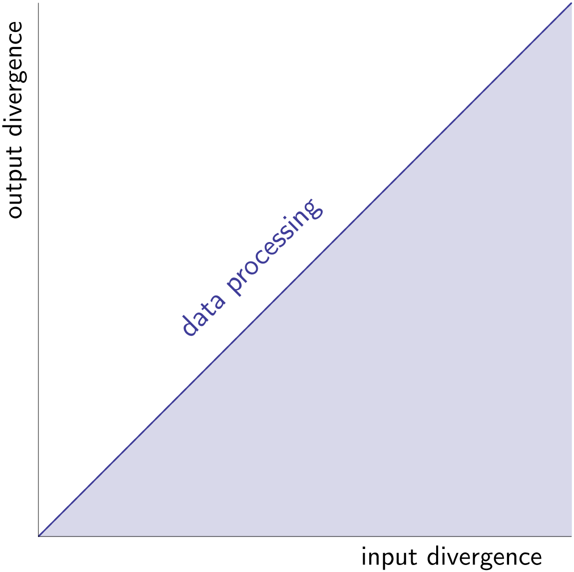
$$D(P_Y \| Q_Y) \le \eta_{KL} D(P_X \| Q_X)$$

- More precise characterization: joint range $(P_X, Q_X) \mapsto (D(P_X \| Q_X), D(P_Y \| Q_Y)) \in \mathbb{R}_+^2$

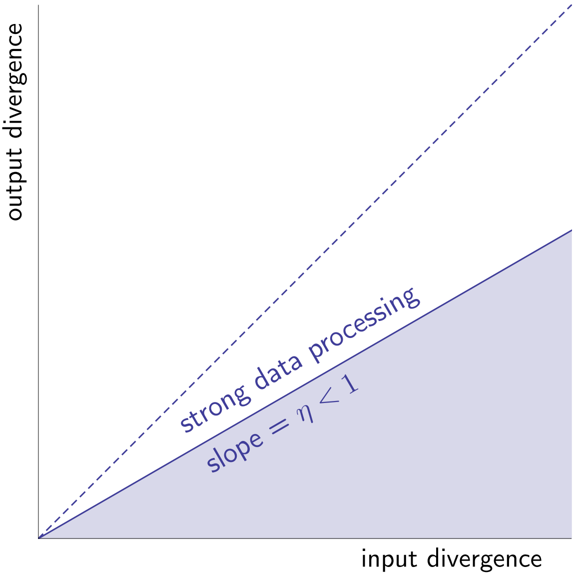
Joint range



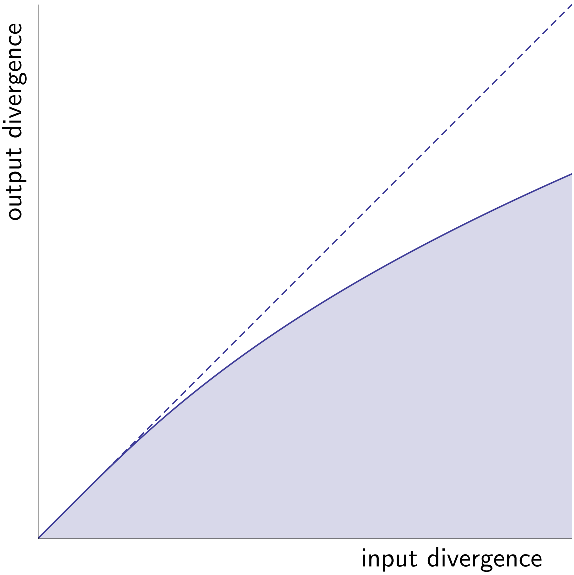
Joint range



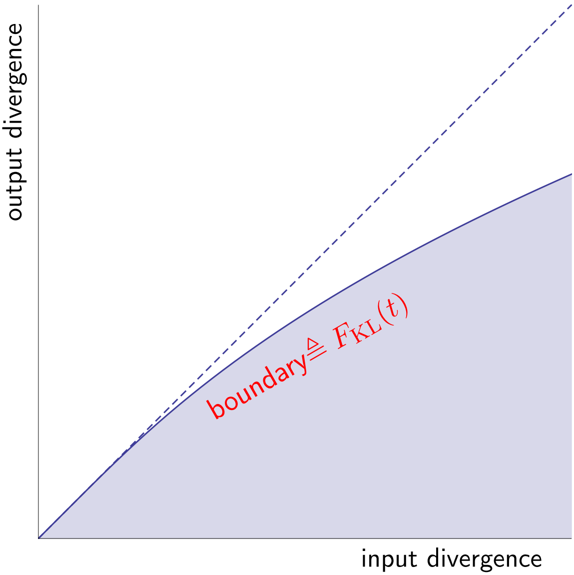
Joint range


output divergence
strong data processing
slope = $\eta < 1$
input divergence

Joint range

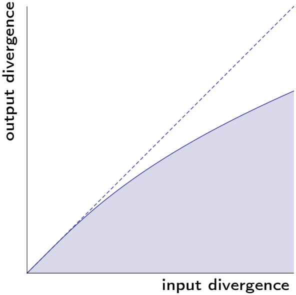


Joint range



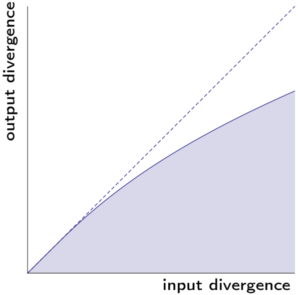
output divergence
boundary $\triangleq F_{\mathrm{KL}}(t)$
input divergence

Is joint range curved?



Want: joint range bounded away from diagonal.

Is joint range curved?

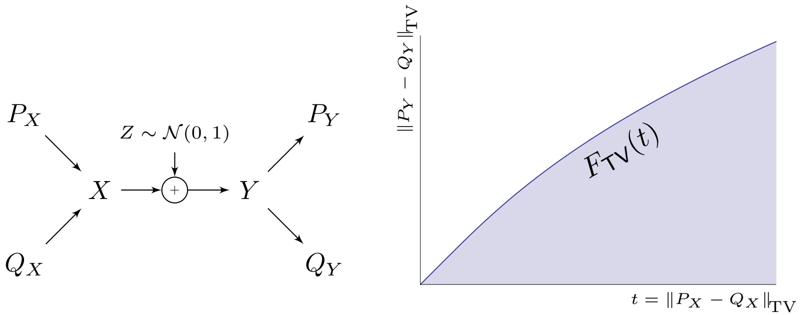


Want: joint range bounded away from diagonal.

- Sad news: For KL the boundary $F_{\mathrm{KL}}(t) = t$
- Good news: For TV the boundary $F_{\mathsf{TV}}(t) < t$ (!)

$F_{\mathsf{TV}}(t), t \in [0, 1]$ – Dobrushin curve of the channel

Dobrushin curve



- Upper boundary:

$$F_{\mathsf{TV}}(t) = \sup_{\|P_X - Q_X\|_{\mathrm{TV}} \le t} \|P_Y - Q_Y\|_{\mathrm{TV}}$$

with constraint:

$$\mathbb{E}_{P_X}|X|^2 + \mathbb{E}_{Q_X}|X|^2 \le 2P$$

- Dobrushin coefficient η_{TV} = maximal slope of Dobrushin curve F_{TV}

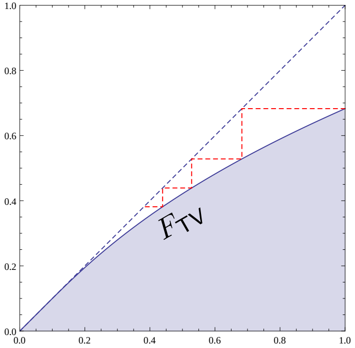
Dobrushin curve of AWGN

Theorem (P.-Wu'14)

Under power constraint $\mathbb{E}|X|^2 \le P$,

$$F_{\mathsf{TV}}(t) = t\left(1 - 2\mathrm{Q}\left(\sqrt{\frac{P}{t}}\right)\right)$$

where Q = *complementary normal CDF.*

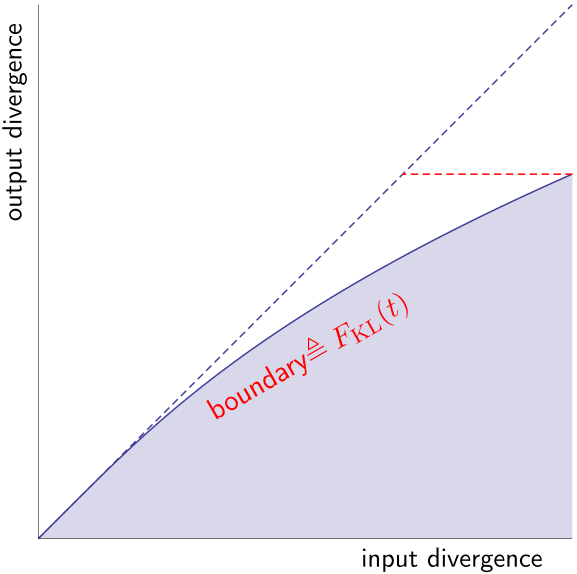


Note:

- F_{TV} smooth but not analytic: ${F_{\mathsf{TV}}}'(0) = 1$, ${F_{\mathsf{TV}}}^{(k)}(0) = 0$
- iterative mapping:

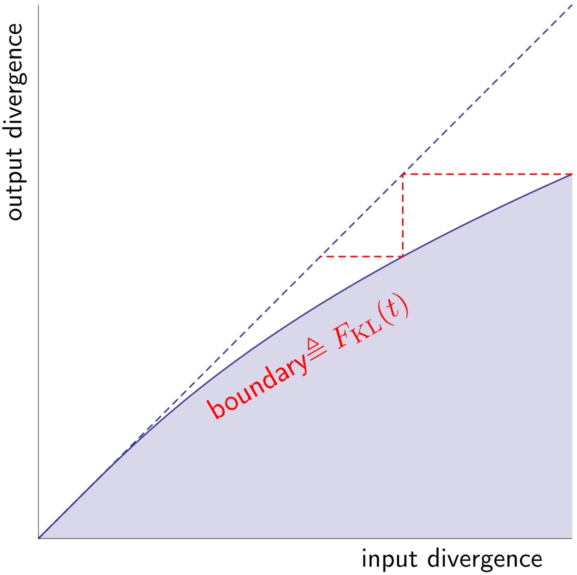
$$\|P_{X_n} - Q_{X_n}\|_{\mathrm{TV}} \le F_{\mathsf{TV}} \circ F_{\mathsf{TV}} \cdots \circ F_{\mathsf{TV}}(1) = O\left(\frac{1}{\log n}\right) \to 0$$

Joint range

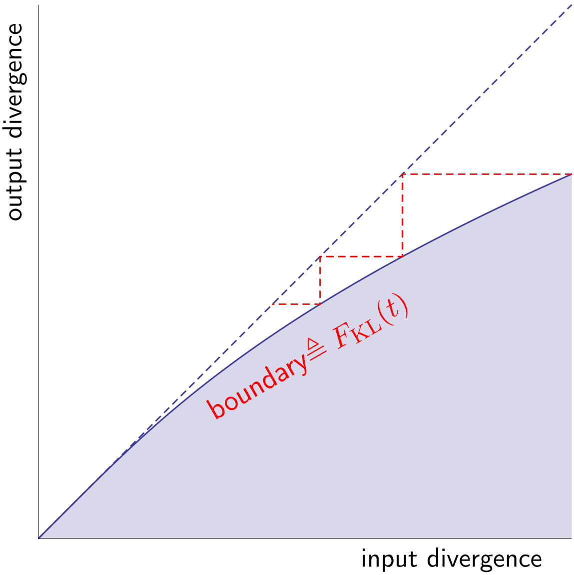


output divergence
boundary$\triangleq F_{\mathrm{KL}}(t)$
input divergence

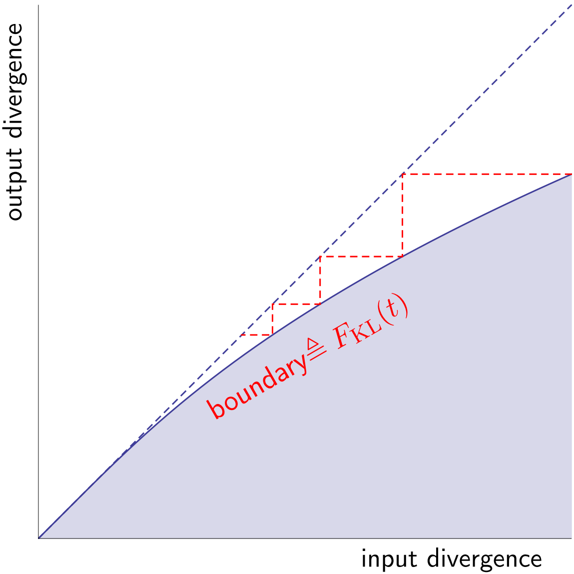
Joint range


output divergence
boundary $\triangleq F_{\mathrm{KL}}(t)$
input divergence

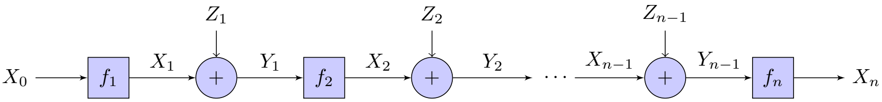
Joint range


output divergence
input divergence
boundary $\triangleq F_{\mathrm{KL}}(t)$

Joint range


output divergence
input divergence
boundary $\triangleq F_{\mathrm{KL}}(t)$

Application to chain of AWGN relays

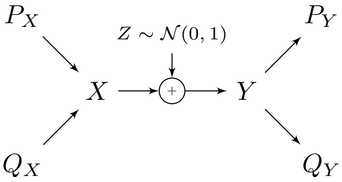


Theorem (P.-Wu'14)

For any processors $\{f_n\}$ *s.t.* $\mathbb{E}[X_n^2] \le P$*:*

$$I(X_0; X_n) \le \frac{CP \log\log n}{\log n} \to 0$$

F_I-curve of additive channels



$$\mathbb{E}|X|^2 \le P$$

For additive noise (non-discrete) channels:

- Contraction coeffs:

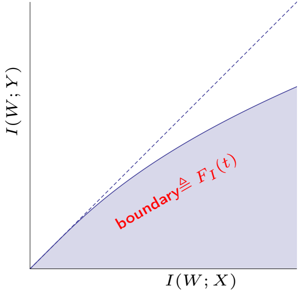
$$\eta_{\mathrm{TV}} = \eta_{\mathrm{KL}} = \eta_I = 1$$

- Joint range:

$$F_{\mathrm{KL}}(t) = t$$

$$F_{\mathsf{TV}}(t) < t$$

- Last question: Joint range for mutual info.



$$F_I(t) \triangleq \sup_{I(W;X)\le t} I(W;Y)$$

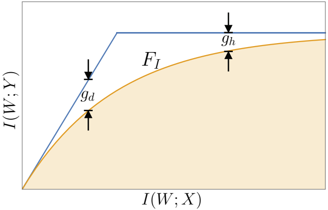
s.t. $W - X - Y$ and $\mathbb{E}|X^2| \le P$

Trivially:

$$F_I(t) \le t \qquad \text{(data-processing)}$$

$$F_I(t) \le C \qquad C \triangleq \max_{\mathbb{E}|X|^2 \le P} I(X;Y) \text{–capacity}$$

Main result

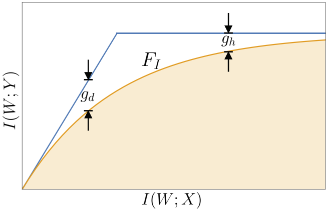


Theorem

There exist $g_d(t) > 0$ *and* $g_h(t) > 0$*:*

$$F_I(t) \le t - g_d(t)$$
$$F_I(t) \le C - g_h(t)$$

Main result



Theorem

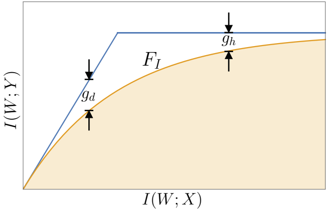
There exist $g_d(t) > 0$ *and* $g_h(t) > 0$*:*

$$F_I(t) \le t - g_d(t) \qquad (F_I \text{ bounded from diagonal})$$

$$F_I(t) \le C - g_h(t) \qquad (F_I \text{ bounded from capacity})$$

- Also holds for all P_Z s.t. $P_Z \not\perp P_{Z+a}$
- Also holds for other constraints $\mathbb{E}|X|^p \le \gamma$

Behavior of g_d and g_h



Theorem

There exist $g_d(t) > 0$ and $g_h(t) > 0$:

$$F_I(t) \le t - g_d(t)$$
$$F_I(t) \le C - g_h(t)$$

For the AWGN channel (w/ F. Calmon, cf. ISIT'15):

$$g_d(t) \approx e^{\frac{c \log t}{t}}, \qquad t \to 0$$
$$g_h(t) \approx e^{-c_1 \cdot \exp(c_2 t)}, \quad t \to \infty$$

Proof: Diagonal bound

Goal:

$$I(W;Y) < I(W;X) \qquad W \to X \to Y = X + Z$$

High level idea:

- small scale: local information in $I(W;X)$ is blurred away by noise
- large scale: $\mathbb{E}|X|$ restricts contribution to $I(W;X)$ from tails
- Hardest case: $I(W;X) \to 0$.

Proof: Diagonal bound

Goal:

$$I(W;Y) < I(W;X) \qquad W \to X \to Y = X + Z$$

High level idea:

- small scale: local information in $I(W;X)$ is blurred away by noise
- large scale: $\mathbb{E}|X|$ restricts contribution to $I(W;X)$ from tails
- Hardest case: $I(W;X) \to 0$.

Key definition:

- Let $\eta(A)$ = KL coeff. for the amplitude A constrained channel:

$$\eta(A) \triangleq \sup_{P,Q \text{ on } [-A,A]} \frac{D(P * P_Z \| Q * P_Z)}{D(P\|Q)}$$

- By Dobrushin for AWGN we have

$$\eta(A) \le 1 - 2Q(A) \approx 1 - e^{-A^2/2}$$

- . . . e.g. $\eta(\frac{1}{3}) < \frac{1}{3}$

Large scale bound (truncation)

Two simplifying assumptions:

- Assume $W \to X$ is determinstic. Want to prove:

$$I(X;Y) < H(X)$$

- Assume X takes values on some Δ-grid. (i.e. no "small scale" details in X.)

Steps:

- Introduce $E \triangleq 1\{|X| \geq A\}$ and $\epsilon = \mathbb{P}[|X| \geq A]$

$$I(X;Y) \leq I(X;Y,E)$$
$$\leq \quad H(E) + \epsilon I(X;Y|E=1) + \bar{\epsilon} \cdot I(X;Y|E=0)$$

Large scale bound (truncation)

Two simplifying assumptions:

- Assume $W \to X$ is determinstic. Want to prove:

$$I(X;Y) < H(X)$$

- Assume X takes values on some Δ-grid. (i.e. no "small scale" details in X.)

Steps:

- Introduce $E \triangleq 1\{|X| \geq A\}$ and $\epsilon = \mathbb{P}[|X| \geq A]$

$$\begin{aligned} I(X;Y) &\leq I(X;Y,E) \\ &\leq H(E) + \epsilon I(X;Y|E=1) + \bar{\epsilon} \cdot I(X;Y|E=0) \\ &\leq h(\epsilon) + \epsilon H(X|E=1) + \bar{\epsilon} \cdot \eta(A) H(X|E=0) \end{aligned}$$

contraction!

Large scale bound (truncation)

Two simplifying assumptions:

- Assume $W \to X$ is determinstic. Want to prove:

$$I(X;Y) < H(X)$$

- Assume X takes values on some Δ-grid. (i.e. no "small scale" details in X.)

Steps:

- Introduce $E \triangleq 1\{|X| \geq A\}$ and $\epsilon = \mathbb{P}[|X| \geq A]$

$$\begin{aligned} I(X;Y) &\leq I(X;Y,E) \\ &\leq H(E) + \epsilon I(X;Y|E=1) + \bar{\epsilon} \cdot I(X;Y|E=0) \\ &\leq h(\epsilon) + \epsilon H(X|E=1) + \bar{\epsilon} \cdot \eta(A) H(X|E=0) \quad \text{contraction!} \\ &= H(X) - \bar{\eta}(A)\big(H(X) - h(\epsilon) - \epsilon H(X|E=1)\big) \end{aligned}$$

- Entropy $H(X|E=1)$ is small by max-entropy exercise:

$$H(U) \leq (\mathbb{E}|U|+1) h\left(\frac{1}{1+\mathbb{E}|U|}\right) + \log 2 \,. \forall U \in \mathbb{Z}$$

Large scale bound (truncation)

Steps:

- With $E \triangleq 1\{|X| \geq A\}$ and $\epsilon = \mathbb{P}[|X| \geq A]$:

$$\begin{aligned} I(X;Y) &\leq I(X;Y,E) \\ &= H(X) - \bar{\eta}(A)\big(H(X) - h(\epsilon) - \epsilon H(X|E=1)\big) \end{aligned}$$

- $\mathbb{E}|X| \leq 1 \quad \Longrightarrow \quad \epsilon = \mathbb{P}[|X| \geq A] \leq \frac{1}{A}$
- ... plus max-entropy exercise:

$$h(\epsilon) + \epsilon H(X|E=1) \lesssim \frac{2 \log A}{A}, \qquad A \gg 1$$

- Take $A \gg 1$ s.t. $(\cdots) \geq \frac{H(X)}{2}$ and conclude

$$I(X;Y) \lesssim H(X) - e^{-\frac{A^2}{2}} H(X), \qquad A \approx \frac{1}{4H(X)} \log \frac{1}{H(X)}$$

Small scale bound (quantization)

- The truncation argument works also w/o assumptions:

$$I(W;Y) \le I(W;X) - \bar{\eta}(A)\big(I(W;X) - h(\epsilon) - \epsilon I(W;Y|E=1)\big)$$

- Problem: $I(W;Y|E=1) \le I(W;X|E=1)$ will not do.

Small scale bound (quantization)

- The truncation argument works also w/o assumptions:

$$I(W;Y) \le I(W;X) - \bar{\eta}(A)\big(I(W;X) - h(\epsilon) - \epsilon I(W;Y|E=1)\big)$$

- Problem: $I(W;Y|E=1) \le I(W;X|E=1)$ will not do.

Lemma

For any $W - X - Y$ over AWGN:

$$I(W;Y) \le H(\lfloor 3X \rfloor) + \frac{1}{3} I(W;X)$$

Proof: Let $Q = \lfloor 3X \rfloor$ be $\frac{1}{3}$-quantization of X. Then:

$$I(W;Y) \quad \le \quad I(Q;Y) + I(W;Y|Q)$$

Small scale bound (quantization)

- The truncation argument works also w/o assumptions:

$$I(W;Y) \le I(W;X) - \bar{\eta}(A)\big(I(W;X) - h(\epsilon) - \epsilon I(W;Y|E=1)\big)$$

- Problem: $I(W;Y|E=1) \le I(W;X|E=1)$ will not do.

Lemma

For any $W - X - Y$ over AWGN:

$$I(W;Y) \le H(\lfloor 3X \rfloor) + \frac{1}{3} I(W;X)$$

Proof: Let $Q = \lfloor 3X \rfloor$ be $\frac{1}{3}$-quantization of X. Then:

$$\begin{aligned} I(W;Y) &\le I(Q;Y) + I(W;Y|Q) \\ &\le I(Q;Y) + \frac{1}{3} I(W;X|Q) \qquad \text{contraction on } [-\tfrac{1}{6};\tfrac{1}{6}] \\ &\le H(Q) + \frac{1}{3} I(W;X) \end{aligned}$$

Small scale bound (quantization)

- …

$$I(W;Y) \le I(W;X) - \bar{\eta}(A)\big(I(W;X) - h(\epsilon) - \epsilon I(W;Y|E=1)\big)$$

- From Lemma

$$I(W;Y|E=1) \le H([3X]|E=1) + \frac{1}{3} I(W;X|E=1)$$

- Again, by max-entropy exercise

$$h(\epsilon) + \epsilon H([3X]|E=1) \lesssim \frac{2\log A}{A}, \qquad A \gg 1$$

- …and $\epsilon I(W;X|E=1) \le I(W;X)$:

$$I(W;Y) \lesssim I(W;X) - e^{-\frac{A^2}{2}} I(W;X) \qquad \text{if } \frac{\log A}{A} \le \frac{1}{6} I(W;X)$$

Proof: Horizontal bound

- We will prove:

$$I(W;Y) = C - \delta \implies I(W;X) \ge g(\delta)\,,$$

- Notation: P_X^*–caid, P_Y^*–caod. (Gaussian for the AWGNC)
- First:

$$\delta = C - I(W;Y) \ge D(P_Y \| P_Y^*) + I(X;Y|W)$$

- $\ldots \implies P_Y \overset{TV}{\approx} P_Y^*$ (Pinsker)
- and $P_{Y|W=w} \overset{TV}{\approx} P_{Z+x}$ since

$$I(X;Y|W) = \int dP_{X,W}(x,w) D(P_{Z+x} || P_{Y|W=w})$$

- Main idea: deconvolution

$$P * P_Z \approx Q * P_Z \quad \implies \quad P \approx Q$$

Then $P_X \approx P_X^*$ (diffuse) but $P_{X|W} \approx \delta_x$ (atomic) so

$$I(X;W) = D(P_{X|W} \| P_X | P_W) \gg 1$$

Lemma (Deconvolution lemma)

For every "regular" P_Z there is $\epsilon(\delta)$ s.t.

$$\|P * P_Z - Q * P_Z\|_{\mathrm{TV}} \le \delta \implies |P(B) - Q(B)| \lesssim \epsilon(\delta) \quad \forall \textit{small balls } B\,.$$

- Now go carefully. By Pinsker

$$\int dP_{X,W} \|P_{Z+x} - P_{Y|W=w}\|^2_{\mathrm{TV}} \lesssim \delta$$

so w.h.p. (over P_W):

$$\|P_{Y|W=w} - P_{Z+x_0(w)}\|_{\mathrm{TV}} \lesssim \sqrt{\delta}$$

- But

$$\begin{aligned} P_{Y|W=w} &= P_{X|W=w} * P_Z \\ P_{Z+x_0} &= \delta_{x_0} * P_Z \end{aligned}$$

- So by Lemma:

$$\exists B : P_{X|W=w}(B) \gtrsim 1 - \epsilon(\sqrt{\delta})$$

Lemma (Deconvolution lemma)

For every "regular" P_Z there is $\epsilon(\delta)$ s.t.

$$\|P * P_Z - Q * P_Z\|_{\mathrm{TV}} \le \delta \implies |P(B) - Q(B)| \lesssim \epsilon(\delta) \quad \forall \textit{small balls } B\,.$$

- On the other hand:

$$\begin{aligned} P_Y &= P_X * P_Z \\ P_Y^* &= P_X^* * P_Z \end{aligned}$$

- Assume P_X^* – diffuse (!) then

$$\lim_{|B|\to 0} \sup_B P_X^*(B) \to 0$$

(Levi's concentration function.)

- So by Lemma:

$$\forall B : P_X(B) \le P_X^*(B) + \epsilon(\sqrt{\delta}) \triangleq \epsilon'(\delta)$$

Proof: Horizontal bound (finish)

- So for P_W-many w we have a ball B_w s.t.

$$P_{X|W=w}(B_w) \gtrsim 1 - \epsilon \tag{1}$$

$$P_X(B_w) \lesssim \epsilon \tag{2}$$

- Thus, by data-processing:

$$\begin{aligned} I(X;W) &= D(P_{X|W} \| P_X | P_W) \\ &\gtrsim d(1-\epsilon \| \epsilon) \approx \log \frac{1}{\epsilon} \to \infty \end{aligned}$$

- The only assumption so far: caid P_X^* – has no atoms.

Deconvolution Lemma:

- Regularity of P_Z: (a) bounded density, (b) $\exists g_1 : \mathbb{R}^+ \to \mathbb{R}^+$ s.t.

$$\mathrm{Leb}\{\omega : |\Psi_Z(\omega)| \le \sqrt{u}, |\omega| \le g_1(u)\} \le \sqrt{g_1(u)}$$

  IOW: characteristic func. Ψ_Z is rarely zero.

- From boundedness of density and Plancherel:

(*) $$\int |\Psi_P(\omega) - \Psi_Q(\omega)|^2 |\Psi_Z(\omega)|^2 \lesssim \|P * P_Z - Q * P_Z\|_{\mathrm{TV}} .$$

Deconvolution Lemma:

- Regularity of P_Z: (a) bounded density, (b) $\exists g_1 : \mathbb{R}^+ \to \mathbb{R}^+$ s.t.

$$\text{Leb}\{\omega : |\Psi_Z(\omega)| \le \sqrt{u}, |\omega| \le g_1(u)\} \le \sqrt{g_1(u)}$$

  IOW: characteristic func. Ψ_Z is rarely zero.

- From boundedness of density and Plancherel:

(*) $$\int |\Psi_P(\omega) - \Psi_Q(\omega)|^2 |\Psi_Z(\omega)|^2 \lesssim \|P * P_Z - Q * P_Z\|_{\text{TV}} .$$

- Instead of ball B we consider "smooth" version:

$$P([x_0, x_0 + \delta]) \approx \mathbb{E}_P[v_\delta(X - x_0)] ,$$

$$v_\delta(x) \triangleq \frac{2\delta^2}{x^2}\left(1 - \cos\left(\frac{x}{\delta}\right)\right)$$

- Note that $|v_\delta| \le 1$ and by Fourier:

$$\int dP(x) v_\delta(x) = \delta \int_{-\frac{1}{\delta}}^{\frac{1}{\delta}} \Psi_P(\omega)\,(1 - \delta|\omega|)\, d\omega$$

Deconvolution Lemma:

- Regularity of P_Z: (a) bounded density, (b) $\exists g_1 : \mathbb{R}^+ \to \mathbb{R}^+$ s.t.

$$\mathrm{Leb}\{\omega : |\Psi_Z(\omega)| \le \sqrt{u}, |\omega| \le g_1(u)\} \le \sqrt{g_1(u)}$$

IOW: characteristic func. Ψ_Z is rarely zero.

- From boundedness of density and Plancherel:

(*) $$\int |\Psi_P(\omega) - \Psi_Q(\omega)|^2 |\Psi_Z(\omega)|^2 \lesssim \|P * P_Z - Q * P_Z\|_{\mathrm{TV}} .$$

- Instead of ball B we consider "smooth" version:

$$P([x_0, x_0 + \delta]) \approx \mathbb{E}_P[v_\delta(X - x_0)] ,$$

$$v_\delta(x) \triangleq \frac{2\delta^2}{x^2}\left(1 - \cos\left(\frac{x}{\delta}\right)\right)$$

- Note that $|v_\delta| \le 1$ and by Fourier:

$$|\mathbb{E}_P[v_\delta(X)] - \mathbb{E}_Q[v_\delta(X)]| \le \delta \int_{-\frac{1}{\delta}}^{\frac{1}{\delta}} |\Psi_P(\omega) - \Psi_Q(\omega)| \left(1 - \delta|\omega|\right) d\omega$$

Deconvolution Lemma:

- Regularity of P_Z: (a) bounded density, (b) $\exists g_1 : \mathbb{R}^+ \to \mathbb{R}^+$ s.t.

$$\mathrm{Leb}\{\omega : |\Psi_Z(\omega)| \le \sqrt{u}, |\omega| \le g_1(u)\} \le \sqrt{g_1(u)}$$

  IOW: characteristic func. Ψ_Z is rarely zero.

- From boundedness of density and Plancherel:

$$(*) \qquad \int |\Psi_P(\omega) - \Psi_Q(\omega)|^2 |\Psi_Z(\omega)|^2 \lesssim \|P * P_Z - Q * P_Z\|_{\mathrm{TV}}\,.$$

- Instead of ball B we consider "smooth" version:

$$P([x_0, x_0+\delta]) \approx \mathbb{E}_P[v_\delta(X - x_0)]\,,$$

$$v_\delta(x) \triangleq \frac{2\delta^2}{x^2}\left(1 - \cos\left(\frac{x}{\delta}\right)\right)$$

- Note that $|v_\delta| \le 1$ and by Fourier:

$$|\mathbb{E}_P[v_\delta(X)] - \mathbb{E}_Q[v_\delta(X)]| \le \delta \int_{-\frac{1}{\delta}}^{\frac{1}{\delta}} |\Psi_P(\omega) - \Psi_Q(\omega)|\,(1 - \delta|\omega|)\, d\omega$$

- ... Apply (*) on $\{|\Psi_Z(\omega)| > \sqrt{\delta}\}$

Take-away message

linear strong data processing inequality

nonlinear strong data processing inequality