ISIT JUNE 22-27 ANN ARBOR MICHIGAN

Approximation Guarantees for Minimum Rényi Entropy Functional Representations

Anuj K. Yadav
EPFL

Yanina Y. Shkel
EPFL

EPFL

Swiss National Science Foundation

EPFL Information Processing Group

Functional Representation Lemma

Given $(X, Y) \sim P_{XY}$, there exists Z $(Z \perp Y)$ and a function $g(\cdot, \cdot)$ such that $X = g(Y, Z)$
i.e.,

$$H(X \mid Y, Z) = 0$$
$$I(Y; Z) = 0$$

Minimum Rényi Entropy Functional Representation

Given : $(X, Y) \sim P_{XY}$

Find : Z (or $P_{Z|XY}$) …..

…with minimum $H_\alpha(Z)$ $(\forall \alpha \geq 0)$

Such that : $Y \perp Z$

$X = g(Y, Z)$

Equivalence to Minimum (Rényi) Entropy Coupling

Given : $(X, Y) \sim P_{XY}$

Find : Z (or $P_{Z|XY}$) …..

…with minimum $H_\alpha(Z)$ $(\forall \alpha \geq 0)$

Such that : $Y \perp Z$

$X = g(Y, Z)$.

Given : $|\mathcal{Y}|$ marginal PMFs $\{P_{X|Y=y}\}_{y\in\mathcal{Y}}$

Find : coupling $C(\{W_y\}_{y\in\mathcal{Y}})$…..

…with minimum $H_\alpha(C)$ $(\forall \alpha \geq 0)$

Such that : $W_y \sim P_{X|Y=y}$; $\forall y \in \mathcal{Y}$.

However …

- Computing $H_\alpha(Z^\star)$ or $H_\alpha(C^\star)$ is a **NP-hard** problem.

- **Lower bounds on** $H_\alpha(Z^\star)$ — Converse type results **[*Shkel-*Yadav '23]**
- **Upper bounds on** $H_\alpha(Z^\star)$ — Achievability type results

*Y. Y. Shkel, and *A. K. Yadav, "Information-spectrum converse for minimum entropy couplings and functional representations ," in *IEEE International Symposium on Information Theory (ISIT)*, 2023.

Prelude

Let X be a random variable such that $X \sim P_X$:

Information of X :

$$\imath_X(x) := \log\left(\frac{1}{P_X(x)}\right) \;;\text{ w. p. } P_X(x).$$

Information spectrum of X :

$$\mathbb{F}_{\imath_X(t)} = \mathbb{P}[\imath_X(X) \leq t] \quad ; \forall t \in [0,\infty)$$

Prelude

Let X be a random variable such that $X \sim P_X$:

Information of X :

$$\imath_X(x) := \log\left(\frac{1}{P_X(x)}\right) \text{ ; w. p. } P_X(x).$$

Information spectrum of X :

$$\mathbb{F}_{\imath_X(t)} = \mathbb{P}[\imath_X(X) \leq t] \quad ; \forall t \in [0,\infty)$$

Shannon entropy of X :

$$H(X) = \mathbb{E}[\imath_X(X)]$$
$$= \int_0^\infty \left(1 - \mathbb{F}_{\imath_X}(t)\right) dt$$

Rényi entropy of X :

$$H_\alpha(X) = \frac{1}{1-\alpha} \log\left(\mathbb{E}[2^{(1-\alpha)\imath_X(X)}]\right) ;$$
$$\forall \alpha \in [0,\infty)$$

Information-spectrum based Lower Bound

Theorem : Let $(X, Y) \sim P_{XY}$ be supported on countable $\mathcal{X}$ and countable $\mathcal{Y}$. Then, for any $\alpha \in [0,\infty)$ we have

$$H_\alpha(Z^\star) \geq K_\alpha(P_{XY})$$

where, $$K_\alpha(P_{XY}) = \begin{cases} \frac{1}{1-\alpha} \log \left[1 + \int_0^\infty J_\alpha(x)dx\right] & ; \text{if } \alpha \in [0,1) \cup (1,\infty) \\ \int_0^\infty G(x)dx & ; \alpha = 1 \end{cases}$$

such that : $$G(x) := \sup_{y \in \mathcal{Y}} \left(1 - \mathbb{F}_{\iota_{X|Y=y}}(x)\right)$$

$$J_\alpha(x) := (\ln 2)(1-\alpha)G(x)2^{(1-\alpha)x}$$

*Y. Y. Shkel, and *A. K. Yadav, "Information-spectrum converse for minimum entropy couplings and functional representations ," in *IEEE International Symposium on Information Theory (ISIT)*, 2023.

This Work …

- Concerned with **Upper Bounds on** $H_\alpha(Z^\star)$ i.e., Achievability type results.

- Approximation analysis based on the Greedy Coupling Algorithm **[Kocaoglu et al. '17]**
  - Let C_Z denote the output of the algorithm
  - $K_\alpha(P_{XY}) \leq H_\alpha(Z^\star)$ - - [from the Lower bound]
  - $K_\alpha(P_{XY}) \leq H_\alpha(Z^\star) \leq H_\alpha(C_Z)$ - - [problem's nature]

  - **Our work :** $H_\alpha(C_Z) \leq K_\alpha(P_{XY}) + Q$; (finding the smallest Q for every $\alpha \in [0,\infty)$).

Murat Kocaoglu, Alexandros G. Dimakis, Sriram Vishwanath, and Babak Hassibi, "Entropic causal inference", ***In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17),*** **AAAI Press, 1156–1162.**

This Work …

- Concerned with **Upper Bounds on** $H_\alpha(Z^\star)$ i.e., Achievability type results.

- Approximation analysis based on the Greedy Coupling Algorithm **[Kocaoglu et al. '17]**
  - Let C_Z denote the output of the algorithm
  - $K_\alpha(P_{XY}) \leq H_\alpha(Z^\star)$ - - [from the Lower bound]
  - $K_\alpha(P_{XY}) \leq H_\alpha(Z^\star) \leq H_\alpha(C_Z)$ - - [problem's nature]
  - **Our work :** $H_\alpha(C_Z) \leq K_\alpha(P_{XY}) + Q$; (finding the smallest Q for every $\alpha \in [0,\infty)$).
  - For the rest of the presentation : $m := |\mathscr{Y}|$ and $n := |\mathscr{X}|$

Murat Kocaoglu, Alexandros G. Dimakis, Sriram Vishwanath, and Babak Hassibi, "Entropic causal inference", ***In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17),*** **AAAI Press, 1156–1162.**

Greedy Coupling Algorithm

- **Input :** m PMFs $\{P_{X|Y=y_i}\}_{i=1}^{m}$, each with $\leq n$ states
- **Output :** Coupling $C_Z := (c_1, c_2, \ldots, c_T)$

Greedy Coupling Algorithm

- **Input :** m PMFs $\{P_{X|Y=y_i}\}_{i=1}^{m}$, each with $\leq n$ states
- **Output :** Coupling $C_Z := (c_1, c_2, \ldots, c_T)$
  - Sort each PMF in the non-increasing order
  - Find the minimum of maximum of each PMF i.e., $q = \min_i(P_{X|Y=y_i}(1))$
  - Append q as the next state of C_Z
  - Update the maximum state of each PMF

    i.e., $P_{X|Y=y_i}(1) = \left(P_{X|Y=y_i}(1) - q\right), \ \forall\, i \leq m$
  - Sort each PMF in non-increasing order
  - Find $q = \min_i(P_{X|Y=y_i}(1))$

Repeat until $q = 0$ (applies to the last four steps)

Greedy Coupling Algorithm : Example

- **Input :** $\{P_{X|Y=y_1} = (0.5,0.4,0.1)\,;\, P_{X|Y=y_2} = (0.6,0.2,0.2)\}$; $(m = 2\,,\, n = 3)$

Iteration (t)	Current PMFs	q	Updated PMFs	C_Z
1	(0.5, 0.4, 0.1) (0.6, 0.2, 0.2)	0.5	(0, 0.4, 0.1) (0.1, 0.2, 0.2)	(0.5)
2	(0.4, 0.1, 0) (0.2, 0.2, 0.1)	0.2	(0.2, 0.1, 0) (0, 0.2, 0.1)	(0.5, 0.2)
3	(0.2, 0.1, 0) (0.2, 0.1, 0)	0.2	(0, 0.1, 0) (0, 0.1, 0)	(0.5, 0.2, 0.2)
T = 4	(0.1, 0, 0) (0.1, 0, 0)	0.1	(0, 0, 0) (0, 0, 0)	(0.5, 0.2, 0.2, 0.1)
5	(0, 0, 0) (0, 0, 0)	0	(0, 0, 0) (0, 0, 0)	

- **Output :** Coupling $C_Z = (0.5,0.2,0.2,0.1)$

Connecting the dots…

- Recall, our goal : $H_\alpha(C_Z) \leq K_\alpha(P_{XY}) + Q$
- Also, recall that $K_\alpha(P_{XY})$ is a function of $G(x)$.

$$G(x) := \sup_{y \in \mathcal{Y}} \left(1 - \mathbb{F}_{\iota_{X|Y=y}}(x)\right)$$

$$J_\alpha(x) := (\ln 2)(1-\alpha)G(x)2^{(1-\alpha)x}$$

$$K_\alpha(P_{XY}) = \begin{cases} \frac{1}{1-\alpha} \log \left[1 + \int_0^\infty J_\alpha(x)dx\right] & ; \text{if } \alpha \in [0,1) \cup (1,\infty) \\ \int_0^\infty G(x)dx & ; \alpha = 1 \end{cases}$$

Connecting the dots…

- Recall, our goal : $H_\alpha(C_Z) \leq K_\alpha(P_{XY}) + Q$
- Also, recall that $K_\alpha(P_{XY})$ is a function of $G(x)$.
- Track the behavior of $G(x)$ at every iteration of the greedy algorithm.

$$G(x) := \sup_{y \in \mathcal{Y}} \left(1 - \mathbb{F}_{\imath_{X|Y=y}}(x)\right)$$

Connecting the dots…

- Recall, our goal : $H_\alpha(C_Z) \leq K_\alpha(P_{XY}) + Q$
- Also, recall that $K_\alpha(P_{XY})$ is a function of $G(x)$.
- Track the behavior of $G(x)$ at every iteration of the greedy algorithm.

$$G^{t+1}(x) \begin{cases} = G^t(x) - p_1^t(1) & ; x < a_1 \\ \leq G^t(x) + (p_m^t(1) - p_1^t(1)) & ; x \in [a_1, a_2) \\ \vdots & \\ \leq G^t(x) + (p_2^t(1) - p_1^t(1)) & ; x \in [a_{m-1}, a_m) \\ = G^t(x) & ; x \geq a_m \end{cases}$$

Where,

$$P_i := P_{X|Y=y_i} \quad \forall i \leq m$$

$$a_1 = \log \frac{1}{p_1^t(1)}$$

$$a_2 = \log \frac{1}{p_m^t(1) - p_1^t(1)}$$

$$\vdots$$

$$a_m = \log \frac{1}{p_2^t(1) - p_1^t(1)}$$

Connecting the dots…

- Recall, our goal : $H_\alpha(C_Z) \leq K_\alpha(P_{XY}) + Q$
- Also, recall that $K_\alpha(P_{XY})$ is a function of $G(x)$.

- Track the behavior of $G(x)$ at every iteration of the greedy algorithm.
- $J_\alpha(x)$ is a function of $G(x)$ i.e., $J_\alpha(x) = h(\alpha, x)G(x)$ where $h(\alpha, x) = \ln 2(1-\alpha)2^{(1-\alpha)x}$

$$J_\alpha^{t+1}(x) \begin{cases} = J_\alpha^t(x) - h(\alpha, x)p_1^t(1) & ; x < a_1 \\ \leq J_\alpha^t(x) + h(\alpha, x)(p_m^t(1) - p_1^t(1)) & ; x \in \left[a_1, a_2\right) \\ \vdots & \\ \leq J_\alpha^t(x) + h(\alpha, x)(p_2^t(1) - p_1^t(1)) & ; x \in \left[a_{m-1}, a_m\right) \\ = J_\alpha^t(x) & ; x \geq a_m \end{cases}$$

Connecting the dots…

- Track the behavior of $G(x)$ at every iteration of the greedy algorithm.
- $J_\alpha(x)$ is a function of $G(x)$ i.e., $J_\alpha(x) = h(\alpha, x)G(x)$ where $h(\alpha, x) = \ln 2(1-\alpha)2^{(1-\alpha)x}$
- Consequently,

$$\int_0^\infty J_\alpha^{t+1}(x)dx - \int_0^\infty J_\alpha^t(x)dx \le p_1^t(1) - (p_1^t(1))^\alpha\left[1 - \tilde{r}(\alpha, m)\right]$$

where, $\tilde{r}(\alpha, m) := \begin{cases} \max_{\substack{w_1 = 0; \\ w_{m+1} = 1; \\ w_1 < w_2 \le w_3 \le \ldots \le w_m < w_{m+1}.}} \sum_{k=2}^m w_k(w_k^{\alpha-1} - w_{k+1}^{\alpha-1}); \text{ for } \alpha \in [0,1), \\ \min_{\substack{w_1 = 0; \\ w_{m+1} = 1; \\ w_1 < w_2 \le w_3 \le \ldots \le w_m < w_{m+1}.}} \sum_{k=2}^m w_k(w_k^{\alpha-1} - w_{k+1}^{\alpha-1}); \text{ for } \alpha \in (1,\infty). \end{cases}$; and $w_k := \dfrac{p_k^t(1) - p_1^t(1)}{p_1^t(1)}$.

- Sum over all iterations of the greedy algorithm, $1 \le t \le T$.

Connecting the dots…

- Consequently,

$$\int_0^\infty J_\alpha^{t+1}(x)dx - \int_0^\infty J_\alpha^t(x)dx \leq p_1^t(1) - (p_1^t(1))^\alpha\left[1 - \tilde{r}(\alpha, m)\right]$$

- Sum over all iterations of the greedy algorithm, $1 \leq t \leq T$.

$$1 + \int_0^\infty J_\alpha^1(x)dx \geq \left[r(\alpha, m))\right] \sum_{t=1}^{T} (p_1^t(1))^\alpha$$

Where $r(\alpha, m) := \max(0, 1 - \tilde{r}(\alpha, m)$.

Connecting the dots…

- Sum over all iterations of the greedy algorithm, $1 \leq t \leq T$.

$$1 + \int_0^\infty J_\alpha^1(x)dx \geq \left[r(\alpha, m))\right] \sum_{t=1}^{T} (p_1^t(1))^\alpha$$

Where $r(\alpha, m) := \max(0, 1 - \tilde{r}(\alpha, m))$.

- On taking logarithm on both sides,

$$\frac{1}{1-\alpha} \log\left(1 + \int_0^\infty J_\alpha^1(x)dx\right) \geq \frac{1}{1-\alpha} \log\left[r(\alpha, m)\right] + \frac{1}{1-\alpha} \log\left(\sum_{t=1}^{T} (p_1^t(1))^\alpha\right)$$

Connecting the dots…

- Sum over all iterations of the greedy algorithm, $1 \leq t \leq T$.

$$1 + \int_0^\infty J_\alpha^1(x)dx \geq \left[r(\alpha, m))\right] \sum_{t=1}^{T} (p_1^t(1))^\alpha$$

Where $r(\alpha, m) := \max(0, 1 - \tilde{r}(\alpha, m)).$

- On taking logarithm on both sides,

$$\frac{1}{1-\alpha} \log \left(1 + \int_0^\infty J_\alpha^1(x)dx \right) \geq \frac{1}{1-\alpha} \log \left[r(\alpha, m)\right] + \frac{1}{1-\alpha} \log \left(\sum_{t=1}^{T} (p_1^t(1))^\alpha \right)$$

$K_\alpha(P_{XY})$

$H_\alpha(C_Z)$

Main Results

Theorem : Let $(X, Y) \sim P_{XY}$ be supported on countable $\mathcal{X}$ and countable $\mathcal{Y}$. Then, for any $\alpha \in [0,\infty)$ we have

$$H_\alpha(C_Z) \leq K_\alpha(P_{XY}) + F(\alpha, m)$$

where, $$F(\alpha, m) = \frac{1}{\alpha - 1} \log \left[r(\alpha, m) \right]$$
$$r(\alpha, m) = \max(0, 1 - \tilde{r}(\alpha, m)).$$

Main Results

Theorem : Let $(X, Y) \sim P_{XY}$ be supported on countable $\mathscr{X}$ and countable $\mathscr{Y}$. Then, for any $\alpha \in [0,\infty)$ we have

$$H_\alpha(C_Z) \leq K_\alpha(P_{XY}) + F(\alpha, m)$$

where, $$F(\alpha, m) = \frac{1}{\alpha - 1} \log \left[r(\alpha, m) \right]$$
$$r(\alpha, m) = \max(0, 1 - \tilde{r}(\alpha, m)).$$

Consequently,

$$K_\alpha(P_{XY}) \leq H^\star_\alpha(Z) \leq H_\alpha(C_Z) \leq K_\alpha(P_{XY}) + F(\alpha, m)$$
$$\leq H^\star_\alpha(Z) + F(\alpha, m)$$

Main Results

Corollary 1 : Let $(X, Y) \sim P_{XY}$ be supported on countable $\mathcal{X}$ and binary $\mathcal{Y}$ (i.e., $m = 2$). Then, for any $\alpha \in [0,\infty)$, we have

$$H_\alpha(C_Z) \le K_\alpha(P_{XY}) + F(\alpha,2)$$

where, $$F(\alpha,2) = \frac{1}{\alpha - 1} \log \left[1 + \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} - \left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} \right].$$

Main Results

Corollary 1 : Let $(X, Y) \sim P_{XY}$ be supported on countable $\mathscr{X}$ and binary $\mathscr{Y}$ (i.e., $m = 2$). Then, for any $\alpha \in [0,\infty)$, we have

$$H_\alpha(C_Z) \leq K_\alpha(P_{XY}) + F(\alpha,2)$$

where, $$F(\alpha,2) = \frac{1}{\alpha - 1} \log \left[1 + \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} - \left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} \right].$$

$F(\alpha, m)$ doesnot have a closed-form solution, in general!

Main Results

Recall that $F(\alpha, m) = \dfrac{1}{\alpha - 1} \log \left[r(\alpha, m) \right]$; where $r(\alpha, m) = \max(0, 1 - \tilde{r}(\alpha, m))$ such that

$$\tilde{r}(\alpha, m) := \begin{cases} \max_{\substack{w_1 = 0; \\ w_{m+1} = 1; \\ w_1 < w_2 \le w_3 \le \ldots \le w_m < w_{m+1}.}} \sum_{k=2}^{m} w_k (w_k^{\alpha-1} - w_{k+1}^{\alpha-1}); & \text{for } \alpha \in [0,1), \\ \min_{\substack{w_1 = 0; \\ w_{m+1} = 1; \\ w_1 < w_2 \le w_3 \le \ldots \le w_m < w_{m+1}.}} \sum_{k=2}^{m} w_k (w_k^{\alpha-1} - w_{k+1}^{\alpha-1}); & \text{for } \alpha \in (1,\infty). \end{cases}$$

Lemma : For every $\alpha \in [0,\infty)$, $F(\alpha, m)$ is an non-decreasing function of m.

As $m \to \infty$, $F(\alpha, m)$ approaches $\dfrac{1}{\alpha - 1} \log \left[\max \left(0, \dfrac{2\alpha - 1}{\alpha} \right) \right]$.

Main Results

Corollary 2 : Let $(X, Y) \sim P_{XY}$ be supported on countable $\mathcal{X}$ and countable $\mathcal{Y}$. Then, for any $\alpha \in [0,\infty)$, we have

$$H_\alpha(C_Z) \leq K_\alpha(P_{XY}) + \lim_{m\to\infty} F(\alpha, m)$$
$$= K_\alpha(P_{XY}) + \frac{1}{\alpha - 1} \log\left[\max\left(0, \frac{2\alpha - 1}{\alpha}\right)\right].$$

Comparison of Upper Bounds : ($m = 2$)

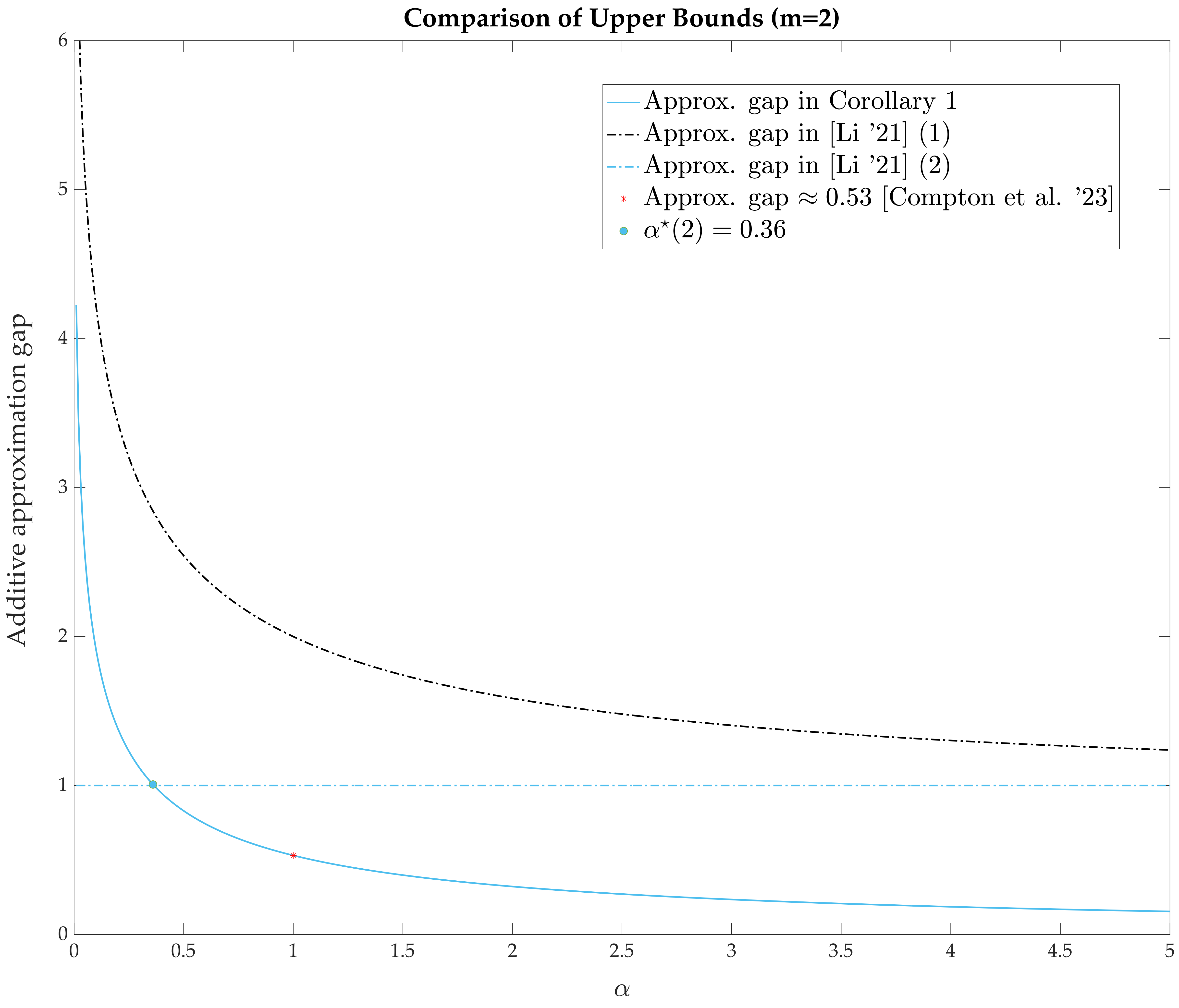
[Li, Trans. IT ’21] (1) : $H_\alpha(\tilde{Z}) \leq H_\alpha(Z^\star) + \begin{cases} \infty & ; \text{ if } \alpha = 0 \\ 2 & ; \text{ if } \alpha = 1 \\ 1 & ; \text{ if } \alpha = \infty \\ \frac{-\alpha - \log(1 - 2^{-\alpha})}{1 - \alpha} & ; \text{ otherwise} \end{cases}$

[Li, Trans. IT ’21] (2) : $H_\alpha(\tilde{Z}) \leq H_\alpha(Z^\star) + 1.$

[Compton et al., AISTATS ’23] : $H_1(C_Z) \leq H_1(Z^\star) + \frac{\log_2 e}{e} \approx 0.53$ **. (Only for Shannon Entropy)**

[Our Work] : $H_\alpha(C_Z) \leq H_\alpha(Z^\star) + F(\alpha,2).$

Comparison of Upper Bounds : ($m = 2$)



Comparison of Upper Bounds : (arbitrary m)

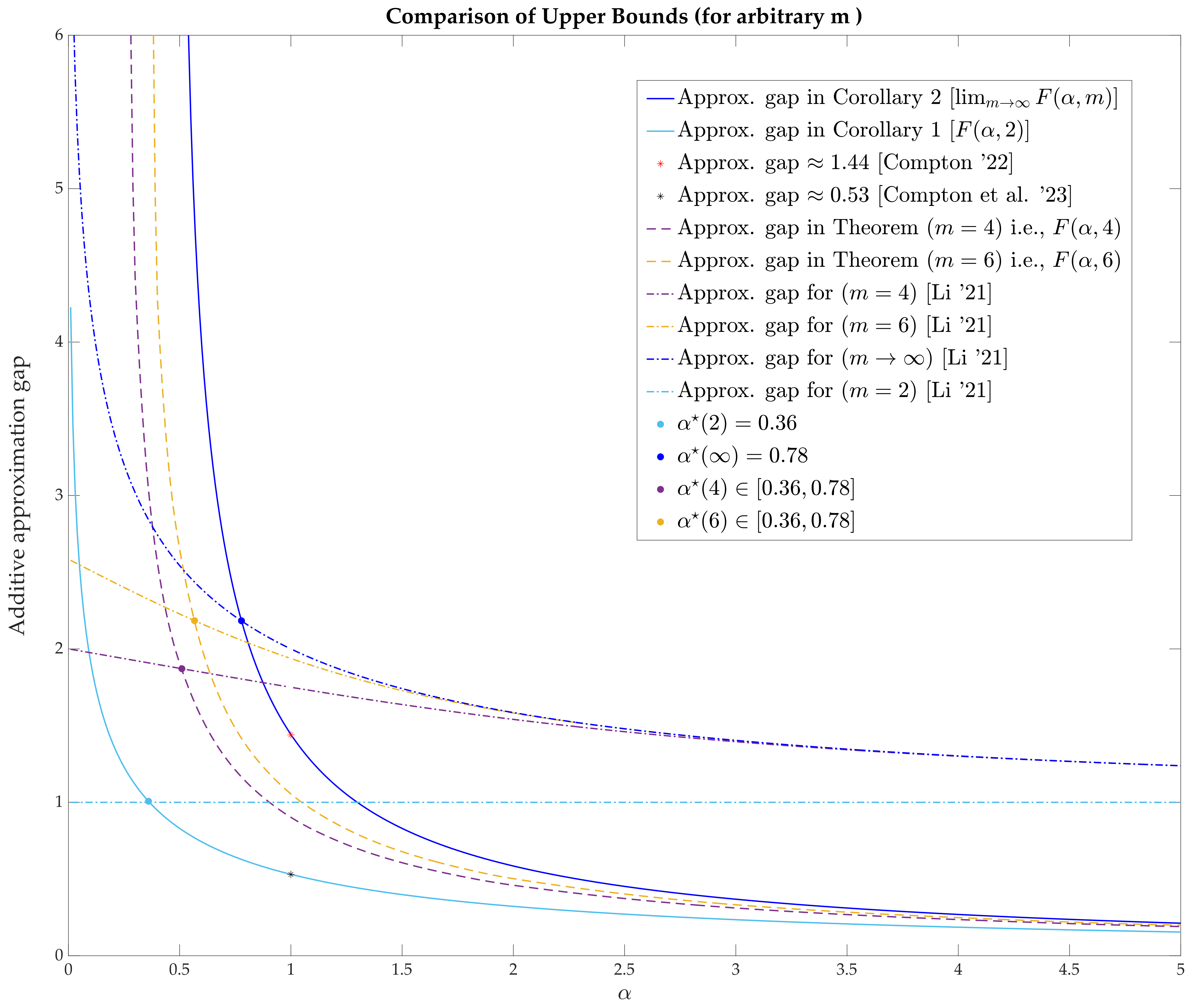
[Li, Trans. IT '21] (2) : $H_\alpha(\tilde{Z}) \leq H_\alpha(Z^\star) + \frac{1}{1-\alpha} \log \left(\frac{(2^\alpha - 2)2^{-\alpha m} + 2^{-\alpha}}{1 - 2^{-\alpha}} \right).$

[Compton, ISIT '22] : $H_1(C_Z) \leq H_1(Z^\star) + \log_2 e \approx 1.44$ **. (Only for Shannon Entropy)**

[Compton et al., AISTATS '23] : $H_1(C_Z) \leq H_1(Z^\star) + \frac{1 + \log_2 e}{e} \approx 1.22$ **. (Only for Shannon Entropy)**

[Our Work] : $H_\alpha(C_Z) \leq H_\alpha(Z^\star) + F(\alpha, m) \leq H_\alpha(Z^\star) + \frac{1}{\alpha - 1} \log \left[\max \left(0, \frac{2\alpha - 1}{\alpha} \right) \right].$

Comparison of Upper Bounds : (arbitrary m)



Summary

- Achievability type results **(Upper Bounds)** for Minimum Rényi Entropy Couplings and Functional Representations.
  - Approximation Analysis between the Rényi entropy of the 'output of the Greedy Coupling Algorithm' and the 'optimal coupling' i.e.,

$$H_\alpha(C_Z) \leq H_\alpha^\star(Z) + F(\alpha, m)$$

  - Our analysis is better for high values of α i.e., $\alpha \geq a^\star(m)$, where $\alpha^\star(m) \in [0.36,0.78]$ for every $m \geq 2$.
  - Greedy Coupling Algorithm is optimal for min-entropy i.e., $\alpha \to \infty$.