LETTER *Special Section on Information Theory and Its Applications*

Upper Bound on Privacy-Utility Tradeoff Allowing Positive Excess Distortion Probability

Shota SAITO[†a)], ***Member and*** **Toshiyasu MATSUSHIMA**[††], ***Fellow***

SUMMARY This letter investigates the information-theoretic privacy-utility tradeoff. We analyze the minimum information leakage (f-leakage) under the utility constraint that the excess distortion probability is allowed up to $\epsilon \in [0, 1)$. The derived upper bound is characterized by the ϵ-cutoff random transformation and a distortion ball.

key words: *distortion function, f-leakage, privacy-utility tradeoff*

1. Introduction

Information-theoretic analysis of privacy-utility tradeoff is one of the research topics in Shannon theory. Many studies have investigated privacy-utility tradeoffs in various settings (see, e.g., Section I of [2] and references therein.) Among these studies, Liao et al. [2] considered the following setup.

Let X denote a private *original data*, where X is a random variable taking values in a finite set $\mathcal{X}$. The probability distribution of X is denoted by P_X. A *released data* is denoted by Y, which takes values in a finite set $\mathcal{Y}$. A *privacy mechanism* $P_{Y|X} : \mathcal{X} \to \mathcal{Y}$ is a random mapping which transforms X to Y. As a utility measure, Liao et al. [2] adopted a *distortion function* between X and Y, i.e., a function $d : \mathcal{X} \times \mathcal{Y} \to [0, +\infty)$ which measures differences between X and Y. On the other hand, as a privacy measure, they considered the following *f-leakage*:

Definition 1 ([2]): Given a joint distribution $P_{X,Y} = P_X P_{Y|X}$ and a convex function $f : \mathbb{R}_+ \to \mathbb{R}$ such that $f(1) = 0$, the f-leakage is defined as

$$\mathcal{L}_f(X \to Y) := \inf_{Q_Y} D_f(P_{X,Y} \| P_X \times Q_Y),$$

where the infimum is taken over all probability distributions on $\mathcal{Y}$ and $D_f(P_{X,Y} \| P_X \times Q_Y)$ is the f-divergence given by

$$D_f(P_{X,Y} \| P_X \times Q_Y) := \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P_X(x) Q_Y(y) f\left(\frac{P_{X,Y}(x,y)}{P_X(x) Q_Y(y)}\right).$$

Setting $f(t) = t \log t$*, we have $\mathcal{L}_f(X \to Y) = I(X;Y)$, where $I(X;Y)$ is the mutual information between X and Y. Hence, the f-leakage $\mathcal{L}_f(X \to Y)$ can be seen as a generalized notion of the mutual information between an original data X and a released data Y.

As an analysis of privacy-utility tradeoff, Liao et al. [2] investigated the quantity

$$\inf_{\substack{P_{Y|X}: \\ \mathbb{P}[d(X,Y)>D]=0}} \mathcal{L}_f(X \to Y),$$

where $D \geq 0$ is a maximal permitted distortion. In other words, they analyzed the minimum information leakage under the utility constraint $d(X,Y) \leq D$ with probability one.

In this letter, we consider generalization of [2] and investigate the quantity

$$\inf_{\substack{P_{Y|X}: \\ \mathbb{P}[d(X,Y)>D]\leq\epsilon}} \mathcal{L}_f(X \to Y) \quad \text{for } \epsilon \in [0,1), \tag{1}$$

i.e., the minimum information leakage under the utility constraint that the excess distortion probability is allowed up to $\epsilon \in [0, 1)$. As our main result, we derive an upper bound of (1) by using the ϵ-cutoff random transformation [1] and a distortion ball. Our bound is sharp in the sense that we will describe in Remark 1. Further, the result is closely related to lossy source coding, and thus it gives insight to lossy source coding (see Remark 2).

2. Result of Prior Work

The following theorem shows the minimum information leakage under the utility constraint that $\mathbb{P}[d(X,Y) > D] = 0$.

Theorem 1 ([2]): For any probability distribution P_X, distortion function $d : \mathcal{X}\times\mathcal{Y} \to [0, +\infty)$, distortion level $D \geq 0$, and convex function $f : \mathbb{R}_+ \to \mathbb{R}$ such that $f(1) = 0$,

$$\inf_{\substack{P_{Y|X}: \\ \mathbb{P}[d(X,Y)>D]=0}} \mathcal{L}_f(X \to Y) = f(0) + \inf_{Q_Y} \mathbb{E}_{P_X}\left[Q_Y(B_D(X))\left(f\left(\frac{1}{Q_Y(B_D(X))}\right) - f(0)\right)\right],$$

where the infimum on the right-hand side is taken over all distributions on $\mathcal{Y}$, $B_D(x)$ is a distortion ball centered at x, i.e., $B_D(x) := \{y \in \mathcal{Y} : d(x,y) \leq D\}$ and $Q_Y(B_D(x)) := \sum_{y \in B_D(x)} Q_Y(y)$.

Manuscript received February 6, 2021.
Manuscript revised June 4, 2021.
Manuscript publicized July 14, 2021.

[†]The author is with the Faculty of Informatics, Gunma University, Maebashi-shi, 371-8510 Japan.
[††]The author is with the Department of Applied Mathematics, School of Fundamental Science and Engineering, Waseda University, Tokyo, 169-8555 Japan.
a) E-mail: shota.s@gunma-u.ac.jp

DOI: 10.1587/transfun.2021TAL0002

*All logarithms are of base 2 throughout this letter.

3. Main Results

Before showing our result, we need to define the notion of the ϵ-cutoff random transformation:

Definition 2 ([1]): For $\epsilon \in [0,1)$ and a random variable Z taking values in a set $\mathcal{Z}$, the ϵ-cutoff random transformation $\langle\cdot\rangle_\epsilon$ is defined as

$$\langle Z\rangle_\epsilon := \begin{cases} Z & \text{if } Z < \eta, \\ \eta & \text{if } Z = \eta \text{ (with prob. } 1-\alpha), \\ 0 & \text{if } Z = \eta \text{ (with prob. } \alpha), \\ 0 & \text{otherwise,} \end{cases}$$

where $\eta \in \mathbb{R}$ and $\alpha \in [0,1)$ are determined by $\mathbb{P}[Z > \eta] + \alpha\mathbb{P}[Z = \eta] = \epsilon^\dagger$.

Now, the following theorem gives the upper bound of the minimum information leakage under the utility constraint that $\mathbb{P}[d(X,Y) > D] \le \epsilon$. The proof is in Sect. 4.1.

Theorem 2: For any probability distribution P_X, distortion function $d : \mathcal{X} \times \mathcal{Y} \to [0,+\infty)$, distortion level $D \ge 0$, convex function $f : \mathbb{R}_+ \to \mathbb{R}$ such that $f(1) = 0$, and $\epsilon \in [0,1)$, we have

$$\inf_{\substack{P_{Y|X}:\\ \mathbb{P}[d(X,Y)>D]\le\epsilon}} \mathcal{L}_f(X \to Y) \le \inf_{Q_Y} \sum_{\substack{x\in\mathcal{X}:\\ \left\langle \log\frac{1}{Q_Y(B_D(x))}\right\rangle_\epsilon > 0}} P_X(x)\mathbb{E}_{Q_Y}\left[f\left(\frac{\mathbf{1}\{d(x,Y)\le D\}}{Q_Y(B_D(x))}\right)\right], \quad (2)$$

where $\mathbf{1}\{\cdot\}$ is the indicator function and the infimum on the right-hand side is taken over all distributions on $\mathcal{Y}$.

Setting $f(t) = t\log t$ in Theorem 2, we have the following corollary (the proof is in Sect. 4.2):

Corollary 1: Under the same conditions of Theorem 2,

$$\inf_{\substack{P_{Y|X}:\\ \mathbb{P}[d(X,Y)>D]\le\epsilon}} I(X;Y) \le \inf_{Q_Y} \mathbb{E}_{P_X}\left[\left\langle \log\frac{1}{Q_Y(B_D(X))}\right\rangle_\epsilon\right], \quad (3)$$

where the infimum on the right-hand side is taken over all distributions on $\mathcal{Y}$.

Remark 1: Set $\epsilon = 0$ in (2). Then, some calculation yields

$$\inf_{\substack{P_{Y|X}:\\ \mathbb{P}[d(X,Y)>D]=0}} \mathcal{L}_f(X \to Y) \le f(0) + \inf_{Q_Y} \mathbb{E}_{P_X}\left[Q_Y(B_D(X))\left(f\left(\frac{1}{Q_Y(B_D(X))}\right) - f(0)\right)\right].$$

Due to Theorem 1, the above inequality is actually equality. Thus, the bound in Theorem 2 is sharp in the special setting.

Remark 2: The inequality (3) is the same inequality obtained by Kostina et al. [1] in the context of lossy source coding (see the inequality (132) of Theorem 8 in [1]).

Remark 3: Theorem 2 shows only the upper bound of (1). It is a future work to derive the lower bound of (1).

4. Proofs

4.1 Proof of Theorem 2

First, we fix an arbitrary probability distribution $Q_{\overline{Y}}$ on $\mathcal{Y}$ and define the privacy mechanism $P_{Y|X}$ as follows:

$$P_{Y|X}(y|x) := \begin{cases} \frac{\mathbf{1}\{d(x,y)\le D\}Q_{\overline{Y}}(y)}{Q_{\overline{Y}}(B_D(x))} & \text{if } \left\langle \log\frac{1}{Q_{\overline{Y}}(B_D(x))}\right\rangle_\epsilon > 0, \\ Q_{\overline{Y}}(y) & \text{otherwise,} \end{cases} \quad (4)$$

Then, it is easy to see that the excess distortion probability under $P_XP_{Y|X}$ is $\mathbb{P}[d(X,Y) > D] \le \epsilon$. Therefore, we obtain the following chain:

$$\begin{aligned} \inf_{\substack{P_{Y|X}:\\ \mathbb{P}[d(X,Y)>D]\le\epsilon}} \mathcal{L}_f(X \to Y) &\overset{(a)}{\le} D_f(P_XP_{Y|X} \| P_X \times Q_{\overline{Y}}) \\ &= \sum_{y\in\mathcal{Y}} Q_{\overline{Y}}(y) \sum_{x\in\mathcal{X}} P_X(x) f\left(\frac{P_{Y|X}(y|x)}{Q_{\overline{Y}}(y)}\right) \\ &= \sum_{y\in\mathcal{Y}} Q_{\overline{Y}}(y) \left\{ \sum_{\substack{x\in\mathcal{X}:\\ \left\langle \log\frac{1}{Q_{\overline{Y}}(B_D(x))}\right\rangle_\epsilon = 0}} P_X(x) f\left(\frac{P_{Y|X}(y|x)}{Q_{\overline{Y}}(y)}\right) \right. \\ &\qquad \left. + \sum_{\substack{x\in\mathcal{X}:\\ \left\langle \log\frac{1}{Q_{\overline{Y}}(B_D(x))}\right\rangle_\epsilon > 0}} P_X(x) f\left(\frac{P_{Y|X}(y|x)}{Q_{\overline{Y}}(y)}\right) \right\} \\ &\overset{(b)}{=} \sum_{y\in\mathcal{Y}} Q_{\overline{Y}}(y) \left\{ \sum_{\substack{x\in\mathcal{X}:\\ \left\langle \log\frac{1}{Q_{\overline{Y}}(B_D(x))}\right\rangle_\epsilon = 0}} P_X(x) f(1) \right. \\ &\qquad \left. + \sum_{\substack{x\in\mathcal{X}:\\ \left\langle \log\frac{1}{Q_{\overline{Y}}(B_D(x))}\right\rangle_\epsilon > 0}} P_X(x) f\left(\frac{\mathbf{1}\{d(x,y)\le D\}}{Q_{\overline{Y}}(B_D(x))}\right) \right\} \end{aligned}$$

†As pointed out in [1], the values of η and α satisfying this equality are not uniquely determined in general, but any pair (η,α) satisfying this equality defines the same $\langle Z\rangle_\epsilon$ up to almost-sure equivalence. See the description just after (14) in [1].

$$\stackrel{(c)}{=} \sum_{\substack{x\in\mathcal{X}: \\ \left\langle \log \frac{1}{Q_{\overline{Y}}(B_D(x))} \right\rangle_\epsilon > 0}} P_X(x)\mathbb{E}_{Q_{\overline{Y}}}\left[f\left(\frac{\mathbf{1}\{d(x,Y)\le D\}}{Q_{\overline{Y}}(B_D(x))} \right) \right],$$

where (a) and (b) follows from (4), and (c) is due to $f(1)=0$.

Since $Q_{\overline{Y}}$ is a fixed arbitrary distribution on $\mathcal{Y}$, we take infimum over all distributions on $\mathcal{Y}$ and complete the proof.

4.2 Proof of Corollary 1

Setting $f(t)=t\log t$ in Theorem 2, we see that the left-hand side of (2) reduces to

$$\inf_{\substack{P_{Y|X}: \\ \mathbb{P}[d(X,Y)>D]\le\epsilon}} I(X;Y).$$

On the other hand, when $f(t)=t\log t$, the right-hand side of (2) is calculated as

$$\begin{aligned}
&\inf_{Q_Y} \sum_{\substack{x\in\mathcal{X}: \\ \left\langle \log \frac{1}{Q_Y(B_D(x))} \right\rangle_\epsilon > 0}} P_X(x)\mathbb{E}_{Q_Y}\left[f\left(\frac{\mathbf{1}\{d(x,Y)\le D\}}{Q_Y(B_D(x))} \right) \right] \\
&= \inf_{Q_Y} \sum_{\substack{x\in\mathcal{X}: \\ \left\langle \log \frac{1}{Q_Y(B_D(x))} \right\rangle_\epsilon > 0}} P_X(x) \\
&\quad \cdot \left\{ \sum_{y\in B_D(x)} Q_Y(y)\frac{\mathbf{1}\{d(x,y)\le D\}}{Q_Y(B_D(x))}\log\frac{\mathbf{1}\{d(x,y)\le D\}}{Q_Y(B_D(x))} \right. \\
&\quad \left. + \sum_{y\in B_D(x)^c} Q_Y(y)\frac{\mathbf{1}\{d(x,y)\le D\}}{Q_Y(B_D(x))}\log\frac{\mathbf{1}\{d(x,y)\le D\}}{Q_Y(B_D(x))} \right\} \\
&\stackrel{(a)}{=} \inf_{Q_Y} \sum_{\substack{x\in\mathcal{X}: \\ \left\langle \log \frac{1}{Q_Y(B_D(x))} \right\rangle_\epsilon > 0}} P_X(x) \\
&\quad \cdot \sum_{y\in B_D(x)} Q_Y(y)\frac{1}{Q_Y(B_D(x))}\log\frac{1}{Q_Y(B_D(x))} \\
&= \inf_{Q_Y} \sum_{\substack{x\in\mathcal{X}: \\ \left\langle \log \frac{1}{Q_Y(B_D(x))} \right\rangle_\epsilon > 0}} P_X(x)\log\frac{1}{Q_Y(B_D(x))} \\
&\stackrel{(b)}{=} \inf_{Q_Y} \mathbb{E}_{P_X}\left[\left\langle \log\frac{1}{Q_Y(B_D(X))} \right\rangle_\epsilon \right],
\end{aligned}$$

where $B_D(x)^c$ denotes the complement of $B_D(x)$, (a) follows from the definition of the distortion ball $B_D(x)$ and the convention that $0\log 0=0$, and (b) follows from the definition of the ϵ-cutoff random transformation.

References

[1] V. Kostina, Y. Polyanskiy, and S. Verdú, "Variable-length compression allowing errors," IEEE Trans. Inf. Theory, vol.61, no.8, pp.4316–4330, Aug. 2015.

[2] J. Liao, O. Kosut, L. Sankar, and F.P. Calmon, "Tunable measures for information leakage and applications to privacy-utility tradeoffs," IEEE Trans. Inf. Theory, vol.65, no.12, pp.8043–8066, Dec. 2019.