An Algorithm for Computing the Stratonovich’s Value of Information

Akira Kamatsuka
Shonan Institute of Technology
Email: kamatsuka@info.shonan-it.ac.jp

Takahiro Yoshida
Nihon University
Email: yoshida.takahiro@nihon-u.ac.jp

Koki Kazama, Toshiyasu Matsushima
Waseda University
Email: kokikazama@aoni.waseda.jp, toshimat@waseda.jp

Abstract—We propose an algorithm for computing Stratonovich’s value of information (VoI) that can be regarded as an analog of the distortion-rate function. We construct an alternating optimization algorithm for VoI under a general information leakage constraint and derive a convergence condition. Furthermore, we discuss algorithms for computing VoI under specific information leakage constraints, such as Shannon’s mutual information (MI), f-leakage, Arimoto’s MI, Sibson’s MI, and Csiszár’s MI.

A full version of this paper [1] is accessible at: https://arxiv.org/abs/2205.02778

I. INTRODUCTION

Decision-making based on noisy data has recently been studied extensively in the field of machine learning and information theory. Examples include the information disclosure problem with privacy protection (e.g. [2]) and the classification problem in the presence of label noise (e.g. [3]).

Such research can be traced back to the theory of value of information (VoI) pioneered by Stratonovich in the 1960s [4]. He analyzed the inference gain when using the noisy data Y containing at most R [bits] of mutual information I(X; Y) about the original data X and derived a theoretical result on the fundamental trade-off between the amount of mutual information and the inferential gain. Recently, this result has been extended by us to a general information leakage measure that is not limited to mutual information [6]. In [6], we also gave an interpretation of the result in terms of optimal privacy mechanism in the privacy-utility trade-off (PUT) problem.

Rate-distortion theory developed by Shannon [7], on the other hand, is a well-known trade-off problem in information theory, and various theoretical studies have been conducted. In particular, an alternating optimization algorithm was proposed by Blahut as a computational algorithm for the rate-distortion function [8], which is now called the Arimoto-Blahut algorithm. Later, a simple sufficient convergence condition was derived by Yueng [9], [10].

Inspired by these results in the rate-distortion theory, we consider applying the alternating optimization algorithm for computing VoI under the general information leakage measure constraint. This algorithm allows us to construct an optimal privacy mechanism in the PUT problem.

Our main contributions are as follows:

- We provide an alternating optimization algorithm framework for computing VoI under a general information leakage constraint (Algorithm 1) and derive a convergence condition to a globally optimal solution (Theorem 2, Corollary 2).
- We consider alternating optimization algorithms for VoI under Shannon’s MI, f-leakage, Arimoto’s MI, Sibson’s MI, and Csiszár’s MI constraints. Then we derive and discuss the KKT conditions for them (Section IV).

II. PRELIMINARY

We first review the theory of the Value of Information (VoI) [4], [6] on the system model in Figure 1 and the alternating optimization problem [9], [10].

A. Notations

Let X, Y and A be discrete random variables on finite alphabets \( \mathcal{X}, \mathcal{Y} \) and \( \mathcal{A} \). X and Y represent the original data and the noisy data, respectively, while A represents an action. Let \( p_{X,Y} = p_X \times p_{Y|X} \) be a given joint distribution of \( (X,Y) \). Let \( \delta : \mathcal{Y} \rightarrow \mathcal{A} \) be a deterministic decision rule and \( \ell(x,a)(\geq 0) \) be a non-negative loss function which represents a loss for making an action \( A = a \) when the true state is \( X = x \). We use \( E_X[f(X)] \) and \( E_X[f(X)|Y = y] \) to represent expectation on \( f(X) \) and conditional expectation on \( f(X) \) given \( Y = y \), respectively, where \( f(X) \) is a function of \( X \). We also use \( E_{X|Y}[f(X)] \) to emphasize that we are taking expectations in \( p_X \). Finally, we use \( \log \) to represent the natural logarithm.

\[
\begin{align*}
X & \rightarrow p_{Y|X} \rightarrow Y \rightarrow \delta \rightarrow A
\end{align*}
\]

Fig. 1. System Model

Acknowledgment

This research is supported in part by Grant-in-Aid JP17K06446 for Scientific Research (C).

1Recently, his book containing this theory has been translated into English [5].
B. Stratonovich’s Value of Information (VoI)

In [6], we introduced a general information leakage measure in an axiomatic way as follows.

**Definition 1** ([6, Def. 3]). The information leakage \( \mathcal{L}(X \rightarrow Y) = \mathcal{L}(p_{X,Y} | p_Y | X) \) is defined as a functional of \( p_X \) and \( p_{Y|X} \) that satisfies the following properties:

1) Non-negativity:
   \[
   \mathcal{L}(X \rightarrow Y) \geq 0. \tag{1}
   \]

2) Data Processing Inequality (DPI):
   If \( X \rightarrow Y \rightarrow Z \) forms a Markov chain, then
   \[
   \mathcal{L}(X \rightarrow Z) \leq \mathcal{L}(X \rightarrow Y). \tag{2}
   \]

3) Independence:
   \[
   \mathcal{L}(X \rightarrow Y) = 0 \iff X \perp Y. \tag{3}
   \]

We also assume that the information leakage \( \mathcal{L}(X \rightarrow Y) \) is bounded above, i.e., there exists an upper bound \( K(X) \) that can depend on \( p_X \) such that for all \( p_{Y|X}, \mathcal{L}(X \rightarrow Y) \leq K(X). \)

**Example 1**. Some examples of the information leakage are Shannon’s mutual information (MI) \( I(X; Y) := H(X) - H(X | Y) \), f-leakage \( \mathcal{L}_f(X \rightarrow Y) := \min_{q_Y > 0} D_f(p_{X,Y} | p_X \times q_Y) \), Arimoto’s MI of order \( \alpha I^\alpha(X; Y) := H_\alpha(X) - H_\alpha(X | Y) \), Sibson’s MI of order \( \alpha I^\alpha(X; Y) := \min_{q_Y > 0} D_\alpha(p_{X,Y} | p_X \times q_Y) \), and Csiszár’s MI of order \( \alpha I^\alpha(X; Y) := \min_{q_Y > 0} E_X \left[ D_\alpha(p_{Y|X}(\cdot | X)) q_Y \right] \), where \( \alpha \in (0,1) \cup (1,\infty) \), the minimums are over all distributions \( q_Y \) that satisfy \( q_Y(y) > 0 \) for all \( y \in \mathcal{Y} \), \( H(X) := -\sum_x p_X(x) \log p_X(x) \) is the Shannon entropy of \( X \), \( H(X | Y) := -\sum_{x,y} p_{X,Y}(x,y) \log p_{X,Y}(x,y) \) is the conditional entropy of \( X \) given \( Y \), and \( H_\alpha(X) := \frac{\alpha}{1-\alpha} \log (\sum_x p_X(x)^\alpha)^{\frac{1}{\alpha}} \) is the Rényi entropy of \( X \) of order \( \alpha \), \( H_\alpha(X | Y) := \frac{\alpha}{1-\alpha} \log (\sum_y p_{X,Y}(x,y)^\alpha)^{\frac{1}{\alpha}} \) is Arimoto’s conditional entropy of \( X \) given \( Y \), and \( D_f(p|q) := \sum_{x \in \mathcal{X}} q(x) f \left( \frac{p(x)}{q(x)} \right) \) is the f-divergence, where \( f: [0, \infty) \rightarrow \mathbb{R} \) is a convex function such that \( f(1) = 0 \), strictly convex at \( t = 1 \). Table I shows a list of f-divergence.

Stratonovich introduced the following quantity, which we term as average gain, to quantify the inferential gain of using the noisy data \( Y \) for a decision-making as the largest reduction of the minimal expected loss compared to a no-data situation.

**Definition 2** (Average gain [6, Def. 5]). The average gain of using \( Y \) on \( X \) for making an action \( A \) with a loss function \( \ell(x,a) \) is defined as

\[
\text{gain}_f^\ell(X; Y) := \inf_a \mathbb{E}_X [\ell(X,a)] - \mathbb{E}_Y \left[ \inf_a \mathbb{E}_X [\ell(X,a) | Y] \right]. \tag{4}
\]

Then VoI is defined as follows.

**Definition 3** (VoI [6, Def. 7]). VoI for a loss function \( \ell(x,a) \) and an information leakage measure \( \mathcal{L}(X \rightarrow Y) \) is given as

\[
\mathcal{V}_f^\ell(R; \mathcal{Y}) := \sup_{\mathcal{L}(X \rightarrow Y) \leq R} \text{gain}_f^\ell(X; Y) \tag{5}
\]

\[
= \inf_a \mathbb{E}_X [\ell(X,a)] - \inf_{\mathcal{L}(X \rightarrow Y) \leq R} \mathbb{E}_Y \left[ \inf_a \mathbb{E}_X [\ell(X,a) | Y] \right]. \tag{6}
\]

Stratonovich first proved the fundamental trade-off between the amount of information leakage and inferential gain, showing the following achievable upper bound \( \mathcal{V}_f^\ell(R; \mathcal{Y}) \) [5, Chapter 9.7], which is extended by us [6, Thm. 1].

**Theorem 1** ([6, Thm. 1]). For a loss function \( \ell(x,a) \), define a function as follows:

\[
\mathcal{V}_f^\ell(R) := \inf_a \mathbb{E}_X [\ell(X,a)] - \inf_{\mathcal{L}(X \rightarrow Y) \leq R} \mathbb{E}_X,A [\ell(X,A)]. \tag{7}
\]

Then \( \mathcal{V}_f^\ell(0) = 0 \) and for \( 0 \leq R \leq K(X) \) and for arbitrary alphabet \( \mathcal{Y} \),

\[
\mathcal{V}_f^\ell(R; \mathcal{Y}) \leq \mathcal{V}_f^\ell(R). \tag{8}
\]

Moreover, let \( t(A) \) be a sufficient statistic of \( X \) and \( t(A) \) be a set of all values of the statistic. Then the equality in the inequality (8) holds when \( \mathcal{Y} = t(A) \) and the optimal conditional distribution is given by

\[
p_{A|X}^* (y | x) := \sum_a p_{A|X}(a | x) \mathbb{I}_{\{y = t(a)\}}, \tag{9}
\]

where \( p_{A|X}^* = \arg\inf_{p_{A|X}}: \mathcal{L}(X \rightarrow A) \leq R \mathbb{E}_{X,A} [\ell(X,A)] \).

**Remark 1.** Stratonovich call \( \mathcal{V}_f^\ell(R) \) as Value of Shannon’s Information in [5, Chapter 9.3]. Thus we call \( \mathcal{V}_f^\ell(R) \) (resp. \( \mathcal{V}_f^\ell(R), \mathcal{V}_f^\ell(R) \)) and \( \mathcal{V}_f^\ell(R) \) as Value of Arimoto’s (resp. Sibson’s, Csiszár’s) Information and Value of f-leakage.

**Remark 2.** In [6], we gave an interpretation of the Theorem 1 in terms of optimal privacy mechanism \( p_{A|X}^* \) in the privacy-utility trade-off problem. To construct the optimal mechanism \( p_{A|X}^* \), we need to construct \( p_{A|X}^* = \arg\inf_{p_{A|X}}: \mathcal{L}(X \rightarrow A) \leq R \mathbb{E}_{X,A} [\ell(X,A)] \). The algorithm for computing \( \mathcal{V}_f^\ell(R) \) in Section III allows us to obtain this distribution.

**Proposition 1** ([6, Prop. 7]).
1) \( \mathcal{V}_f^\ell(R) \) is increasing in \( R \).
2) \( \mathcal{V}_f^\ell(R) \) is concave (resp. quasi-concave) if \( \mathcal{L}(X \rightarrow A) \) is convex (resp. quasi-convex) in \( p_{A|X} \).

**Corollary 1** ([6, Cor. 1]). From the property 2) above, the following hold:

- \( \mathcal{V}_f^\ell(R) \) is concave (see Figure 2) since \( I(X; A) \) is convex in \( p_{A|X} \) for fixed \( p_X \).
| $f(t)$ | Name of $D_f(p||q)$ | $q_A^* = \arg\min_{q_A} D_f(p_X|p_{A|X}|p_X|q_A)$ |
|-------|------------------|-------------------------------------------------|
| $t \log t$ | KL-divergence | $\sum_{x \in \mathcal{X}} p_X(x) \log \frac{p_{A|X}(a|x)}{p_X(x)}$ |
| $- \log t$ | reverse KL-divergence | $\frac{\exp(\sum_{x \in \mathcal{X}} p_X(x) \log p_{A|X}(a|x))}{\sum_{a} \exp(\sum_{x \in \mathcal{X}} p_X(x) \log p_{A|X}(a|x))}$ |
| $2(\sqrt{t} - 1)$ | squared Hellinger distance | $\sum_{x} \sqrt{p_X(x)^2 p_{A|X}(a|x)}$ |
| $(t-1)^2$ | Pearson $\chi^2$-divergence | $\frac{\sum_{x} p_X(x) p_{A|X}(a|x)^2}{\sum_{a} \sum_{x} p_X(x) p_{A|X}(a|x)^2}$ |
| $1/t - 1$ | Neyman $\chi^2$-divergence | $\frac{\sum_{x} \sum_{a} p_X(x) p_{A|X}(a|x) (a|x)^{-1}}{\sum_{a} \sum_{x} p_X(x) p_{A|X}(a|x) (a|x)^{-1}}$ |
| $(t^{\alpha} - 1)/(\alpha - 1)$ | Hellinger divergence of order $\alpha$ | $\frac{\sum_{x} \sum_{a} p_X(x) p_{A|X}(a|x)^{\alpha}(a|x)^{-1}}{\sum_{a} \sum_{x} p_X(x) p_{A|X}(a|x)^{\alpha}(a|x)^{-1}}$ |
| $4(1 - t^{(\alpha+1)/2})/(1 - \alpha^2)$ | $\alpha$-divergence | $\frac{\sum_{x} \sum_{a} p_X(x) p_{A|X}(a|x)(\alpha+1)/2}{\sum_{a} \sum_{x} p_X(x) p_{A|X}(a|x)(\alpha+1)/2}$ |

**Table 1**

List of $f$-divergence and its minimizer

- $V_{f}^{L}(R)$ is concave since $L_{f}(X \to A)$ is convex in $p_{A|X}$ for fixed $p_X$.
- For $\alpha > 0$, $V_{f}^{L}(R)$ is quasi-concave since $I_{f}^{\alpha}(X;A)$ is quasi-convex in $p_{A|X}$ for fixed $p_X$. For $0 < \alpha \leq 1$, $V_{f}^{L}(R)$ is concave since $I_{f}^{\alpha}(X;A)$ is convex in $p_{A|X}$ for fixed $p_X$ (see [1, Prop. 9]).
- For $0 < \alpha \leq 1$, $V_{f}^{L}(R)$ is concave since $I_{f}^{\alpha}(X;A)$ is convex in $p_{A|X}$ for fixed $p_X$.
- For $0 < \alpha \leq 1$, $V_{f}^{L}(R)$ is concave since $I_{f}^{\alpha}(X;A)$ is convex in $p_{A|X}$ for fixed $p_X$.

![Fig. 2. Value of Shannon’s information](image)

**Proposition 2.** Assume that $L(A \to X) = L(p_X,p_{A|X})$ is convex in $p_{A|X}$ for fixed $p_X$. Then the inequality constraint in (7) for $V_{f}^{L}(R)$ can be replaced by an equality constraint, i.e., the following holds:

$$V_{f}^{L}(R) = \inf_{\alpha} \mathbb{E}_{X}[\ell(X,\alpha)] - \inf_{p_{A|X} : L(X \to A) = R} \mathbb{E}_{X,A}[\ell(X, A)].$$

**C. Alternating Optimization**

Let $B_i$ be a convex subset of $\mathbb{R}^{u_i}$ for $i = 1, 2$. Let $f: B_1 \times B_2 \to \mathbb{R}$ be a continuous function defined on $B_1 \times B_2$ that are bounded from below and has continuous partial derivatives $\nabla f = (\partial f / \partial u_1, \partial f / \partial u_2)$ on $B_1 \times B_2$. Then consider the double infimum

$$f^* := \inf_{u_1 \in B_1} \inf_{u_2 \in B_2} f(u_1, u_2).$$

Assume that for all $u_2 \in B_2$ there exists a unique $c_1(u_2) \in B_1$ such that

$$f(c_1(u_2), u_2) = \min_{u_1} f(u_1, u_2).$$

Similarly, assume that for all $u_1 \in B_1$ there exists a unique $c_2(u_1) \in B_2$ such that

$$f(u_1, c_2(u_1)) = \min_{u_2} f(u_1, u_2).$$

Let $u_{1(0)}$ be an arbitrarily chosen vector in $B_1$. Then define a sequence $\{(u_{1(k)}, u_{2(k)})\}_{k=0}^{\infty}$ as follows:

$$u_{1(k)} := c_1(u_{2(k-1)}),$$

$$u_{2(k)} := c_2(u_{1(k)}).$$

**Proposition 3.** ([10, Thm. 10.5]) If $f$ is convex on $B_1 \times B_2$, then $f(u_{1(k)}, u_{2(k)}) \to f^*$ as $k \to \infty$.

**III. Computation of the Value of Information**

Note that $V_{f}^{L}(R)$ can be represented as

$$V_{f}^{L}(R) = U_{f}^{L}(0) - U_{f}^{L}(R),$$

where $U_{f}^{L}(R) := \inf_{p_{A|X} : L(X \to A) \leq R} \mathbb{E}_{X,A}[\ell(X, A)]$. Thus the computation of $V_{f}^{L}(R)$ results in the computation of $U_{f}^{L}(R)$. In this section, we provide an alternating optimization

2Note that, instead of the double infimum problem (12), a double supremum problem is considered in [10].
Algorithm framework for computing $U^f_{\mathbf{L}}(R)$ and derive a convergence condition to a globally optimal solution. Due to the space limitation, some of the proofs and examples will be omitted. See the full version [1].

**Theorem 2.** Assume that there exists a non-negative functional $G(p_{A|X}, q_A) \geq 0$ such that
\begin{equation}
\mathbf{L}(p_X, p_{A|X}) = \min_{q_A > 0} G(p_{A|X}, q_A),
\end{equation}
where the minimum is over all distributions $q_A$ that satisfies $q_A(a) > 0$ for all $a \in A$. For $\beta \geq 0$, define a function $F_\beta(p_{A|X}, q_A)$ and distributions $(p^*_A, q^*_A)$ as follows:
\begin{align}
F_\beta(p_{A|X}, q_A) &:= E_{X,A} [\ell(X, A)] + \beta G(p_{A|X}, q_A), \\
F_\beta(p^*_A, q^*_A) &:= \inf_{p_{A|X} \in B_1} \inf_{q_A \in B_2} F_\beta(p_{A|X}, q_A),
\end{align}
where
\begin{align}
B_1 &= \left\{ p_{A|X} \mid \forall a, x : p_{A|X}(a \mid x) > 0, \sum_a p_{A|X}(a \mid x) = 1 \right\}, \\
B_2 &= \left\{ q_A \mid \forall a > 0 : q_A(a) > 0, \sum_a q_A(a) = 1 \right\}.
\end{align}
Then, the following holds:
\begin{equation}
U^f_{\mathbf{L}}(R_\beta) + \beta R_\beta = F_\beta(p^*_A, q^*_A),
\end{equation}
where $R_\beta := \mathbf{L}(p_X, p^*_A|X)$ and
\begin{equation}
U^f_{\mathbf{L}}(R_\beta) := \inf_{p_{A|X} \in B_1} \inf_{\mathbf{L}(X \rightarrow A) = R_\beta} \mathbb{E}_{X,A} [\ell(X, A)].
\end{equation}

**Corollary 2.** Assume that for all $p_{A|X} \in B_1$ there exists a unique $c_2(p_{A|X}) \in B_2$ such that $F_\beta(p_{A|X}, c_2(p_{A|X})) = \min_{q_A \in B_2} F(p_{A|X}, q_A)$. Similarly, assume that for all $q_A \in B_2$ there exists a unique $c_1(q_A) \in B_1$ such that $F_\beta(c_1(q_A), q_A) = \min_{p_{A|X} \in B_1} F(p_{A|X}, q_A)$. Let $p^*_{A|X} \in B_1$ be an arbitrary probability distribution on $A$ and define sequences $\{q^{(k)}_A, p^{(k)}_{A|X}\}_{k=0}^\infty, \{F^{(k)}\}_{k=0}^\infty$ as follows:
\begin{align}
q^{(k)}_A &:= \arg\min_{q_A \in B_2} F(p^{(k-1)}_{A|X}, q_A), \\
p^{(k)}_{A|X} &:= \arg\min_{p_{A|X} \in B_1} F(p^{(k)}_{A|X}, q^{(k)}_A), \\
F^{(k)} &:= F(p^{(k)}_{A|X}, q^{(k)}_A). \tag{26}
\end{align}
If $G(p_{A|X}, q_A)$ is jointly convex on $B_1 \times B_2$, then
\begin{equation}
F^{(k)} \rightarrow U^f_{\mathbf{L}}(R_\beta) + \beta R_\beta, \quad \text{as } k \rightarrow \infty. \tag{27}
\end{equation}

**Proof.** First, $F_\beta(p_{A|X}, q_A)$ is bounded from below since $\ell(x, a) \geq 0, \beta \geq 0$ and $G(p_{A|X}, q_A) \geq 0$. Moreover, since $\mathbb{E}_{X,A} [\ell(X, A)]$ is linear (thus both convex and concave) on $B_1 \times B_2$, the joint convexity of $F_\beta(p_{A|X}, q_A)$ is equivalent to the joint convexity of $G(p_{A|X}, q_A)$. Therefore, the proof is complete by applying Proposition 3 to $F_\beta(p_{A|X}, q_A)$.

From Theorem 2 and Corollary 2, the following Arimoto-Blahut-like alternating optimization algorithm is derived.

**Remark 3.** $U^f_{\mathbf{L}}(R)$ corresponds to the distortion-rate function $D(R) := \inf_{\ell(x,y) \leq R} E_{X,Y} [d(X,Y)]$, where $d(x, y)$ is a distortion function. Therefore, by replacing $\ell(x, a)$ with $d(x, y)$, the results above also hold for the generalized distortion-rate function defined as $D_{\mathbf{L}}(R) := \inf_{\ell(x,y) \leq R} E_{X,Y} [d(X,Y)]$.

**Algorithm 1** Arimoto-Blahut-like algorithm

**Input:**
$\epsilon > 0, \beta \geq 0$
$p_{A|X} \in B_1$

**Output:**
$U^f_{\mathbf{L}}(R_\beta)$

1: **Initialization:**
$F^{(-1)} \leftarrow 0$
$q^{(0)}_A \leftarrow \arg\min_{q_A > 0} F_\beta(p^{(0)}_{A|X}, q_A)$
$F^{(0)} \leftarrow F_\beta(p^{(0)}_{A|X}, q^{(0)}_A), k \leftarrow 1$
2: **while** $|F^{(k)} - F^{(k-1)}| > \epsilon$
3: $p^{(k)}_{A|X} \leftarrow \arg\min_{p_{A|X} > 0} F_\beta(p^{(k)}_{A|X}, q^{(k)}_A)$
4: $q^{(k)}_A \leftarrow \arg\min_{q_A > 0} F_\beta(p^{(k)}_{A|X}, q^{(k)}_A)$
5: $F^{(k)} \leftarrow F_\beta(p^{(k)}_{A|X}, q^{(k)}_A)$
6: $k \leftarrow k + 1$
7: **end while**
8: **return** $F^{(k)} - \beta \mathbf{L}(p_X, p^{(k)}_{A|X})$

IV. APPLICATIONS

In this section, we discuss alternating optimization algorithms for computing Vol under the constraint of a specific information leakage measure that includes Shannon’s MI, f-leakage, Arimoto’s MI, Sibson’s MI and Csiszár’s MI. Due to the space limitation, discussion on algorithms for value of Arimoto’s MI, Sibson’s MI and Csiszár’s MI will be omitted. See the full version [1].

A. Computation of the Value of Shannon’s Information

**Proposition 4** ([11, Lem. 10.8.11].
\begin{equation}
I(X; A) = \min_{q_A > 0} D(p_{X|A|X} || p_{X|A}), \tag{28}
\end{equation}
where the minimum is achieved at
\begin{equation}
q^*_A(a) := \sum_x p_{X}(x)p_{A|X}(a \mid x). \tag{29}
\end{equation}

**Proposition 5.** Let
\begin{equation}
F_\beta(p_{A|X}, q_A) := \mathbb{E}_{X,A} [\ell(X, A)] + \beta D(p_{X|A|X} || p_{X|A}), \tag{30}
\end{equation}
Then
1) For fixed $p_{A|X}$, $F_{\beta}(p_{A|X}, q_A)$ is minimized by

$$q_A^*(a) = \sum_x p_X(x)p_{A|X}(a | x).$$

(31)

2) For fixed $q_A$, $F_{\beta}(p_{A|X}, q_A)$ is minimized by

$$p_{A|X}^*(a | x) = \frac{q_A(a)e^{\frac{1}{\beta} f(x,a)}}{\sum_a q_A(a)e^{\frac{1}{\beta} f(x,a)}}.$$  

(32)

Since $G(p_{A|X}, q_A) := D(p_{A|X}||q_A)$ is jointly convex in $(p_{A|X}, q_A)$ (see, e.g., [11, Thm. 2.7.2]), the following holds from Corollary 2.

**Corollary 3.** The alternating optimization algorithm corresponding to this problem converges to $U^*_1(R_\beta)$.

**B. Computation of Value of f-leakage**

The minimizer of the $f$-leakage $\mathcal{L}_f(X \rightarrow A)$ depends on the function $f$. Since $G_f(p_{A|X}, q_A) := D(p_{XPA}||p_Xq_A)$ is jointly convex in $(p_{A|X}, q_A)$ (see [12, Lem. 4.1]), the following propositions follow from the KKT condition.

**Proposition 6.** Let $f : [0, \infty) \rightarrow \mathbb{R}$ be a differentiable convex function such that $f(1) = 0$ and strictly convex at $t = 1$. Any minimizers $q_A^*$ of the $f$-leakage $\mathcal{L}_f(X \rightarrow A) = \min_{q_A > 0} D_f(p_{XPA}||p_Xq_A)$ satisfy the following equation for some $\lambda$

$$\sum_x p_X(x) \left\{ f \left( \frac{p_{A|X}(a | x)}{q_A^*(a)} \right) - \left( \frac{p_{A|X}(a | x)}{q_A^*(a)} \right) \cdot f' \left( \frac{p_{A|X}(a | x)}{q_A^*(a)} \right) \right\} + \lambda = 0.$$  

(33)

**Example 2.** By solving this equation (33) for $q_A^*$, the minimizers for each $f$ are obtained as shown in the third column of Table 1.

**Proposition 7.** Let $f : [0, \infty) \rightarrow \mathbb{R}$ be a differentiable convex function such that $f(1) = 0$ and strictly convex at $t = 1$. Define

$$F_{\beta}^f(p_{A|X}, q_A) := \mathbb{E}_{X,A} [\ell(X, A)] + \beta D_f(p_{XPA}||p_Xq_A).$$  

(34)

Then

1) For fixed $p_{A|X}$, any minimizers $q_A^*$ of $\min_{q_A > 0} F_{\beta}^f(p_{A|X}, q_A)$ satisfy (33) for some $\lambda$.

2) For fixed $q_A$, any minimizers $p_{A|X}^*$ of $\min_{p_{A|X} > 0} F_{\beta}^f(p_{A|X}, q_A)$ satisfy the following equation for all $x \in \mathcal{X}$ and for some $\lambda_x$:

$$\ell(x, a) + \beta f' \left( \frac{p_{A|X}^*(a | x)}{q_A^*(a)} \right) + \lambda_x = 0.$$  

(35)

**Remark 4.** Unfortunately, the equation (35) does not necessarily have an analytic solution or a unique solution. For example, when $f(t) = 2(1 - \sqrt{t})$, by solving (35) we have

$$p_{A|X}^*(a | x) = \frac{\beta^2 q_A^*(a)}{(\ell(x, a) + \lambda_x)^2},$$  

(36)

$$1 = \sum_a \frac{\beta^2 q_A^*(a)}{(\ell(x, a) + \lambda_x)^2},$$  

(37)

which does not have an analytic solution or a unique solution in general.

**V. CONCLUSION**

In this study, we proposed an alternating optimization algorithm for computing the Stratonovich’s value of information $V_{\beta}^f(R)$ under a general information leakage constraint. We also derived a convergence condition to globally optimal solution. Future work includes constructing algorithms for solving the equations for the KKT conditions in value of $f$-leakage, Arimoto information, Sibson information, Csiszár information numerically and conducting numerical experiments.

**REFERENCES**

[1] A. Kamatsuka, T. Yoshida, K. Kazama, and T. Matsushima, “An algorithm for computing the stratonovich’s value of information,” 2022. [Online]. Available: https://arxiv.org/abs/2205.02278

[2] J. Liao, O. Kosut, L. Sankar, and F. du Pin Calmon, “Tunable measures for information leakage and applications to privacy-utility tradeoffs,” IEEE Transactions on Information Theory, vol. 65, no. 12, pp. 8043–8066, 2019.

[3] B. Frenay and M. Verleysen, “Classification in the presence of label noise: A survey,” IEEE Transactions on Neural Networks and Learning Systems, vol. 25, no. 5, pp. 845–869, 2014.

[4] R. Stratonovich, “On value of information,” Izvestiya of USSR Academy of Sciences, Technical Cybernetics, vol. 5, pp. 3–12, 1965.

[5] R. Stratonovich, R. Belavkin, P. Pardalos, and J. Principe, Theory of Information and its Value. Springer International Publishing, 2020.

[6] A. Kamatsuka, T. Yoshida, and T. Matsushima, “A generalization of the stratonovich’s value of information and application to privacy-utility tradeoff,” 2022. [Online]. Available: https://arxiv.org/abs/2201.11449

[7] C. E. Shannon, “Coding theorems for a discrete source with a fidelity criterion,” Institute of Radio Engineers, International Convention Record, vol. 7 (part 4), pp. 142–163, 1959.

[8] R. Blahut, “Computation of channel capacity and rate-distortion functions,” IEEE Transactions on Information Theory, vol. 18, no. 4, pp. 460–473, 1972.

[9] R. W. Yeung and T. Berger, “Multi-way alternating minimization,” in Proceedings of 1995 IEEE International Symposium on Information Theory, 1995, pp. 74–.

[10] R. W. Yeung, A First Course in Information Theory (Information Technology: Transmission, Processing and Storage). Berlin, Heidelberg: Springer-Verlag, 2006.

[11] T. M. Cover and J. A. Thomas, Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience, 2006.

[12] I. Csiszár and P. C. Shields, “Information theory and statistics: A tutorial,” Commun. Inf. Theory, vol. 1, no. 4, pp. 417–528, Dec. 2004. [Online]. Available: http://dx.doi.org/10.1561/0100000004