On Perfect Privacy and Maximal Correlation

Borzoo Rassouli and Deniz Gündüz

Abstract

The problem of private data disclosure is studied from an information theoretic perspective. Considering a pair of correlated random variables \((X, Y)\), where \(Y\) denotes the observed data while \(X\) denotes the private latent variables, the following problem is addressed: What is the maximum information that can be revealed about \(Y\), while disclosing no information about \(X\)? Assuming that a Markov kernel maps \(Y\) to the revealed information \(U\), it is shown that the maximum mutual information between \(Y\) and \(U\), i.e., \(I(Y; U)\), can be obtained as the solution of a standard linear program, when \(X\) and \(U\) are required to be independent, called perfect privacy. This solution is shown to be greater than or equal to the non-private information about \(X\) carried by \(Y\). Maximal information disclosure under perfect privacy is shown to be the solution of a linear program also when the utility is measured by the reduction in the mean square error, \(E[(Y - U)^2]\), or the probability of error, \(Pr\{Y \neq U\}\). For jointly Gaussian \((X, Y)\), it is shown that perfect privacy is not possible if the kernel is applied to only \(Y\); whereas perfect privacy can be achieved if the mapping is from both \(X\) and \(Y\); that is, if the private latent variables can also be observed at the encoder.

Next, measuring the utility and privacy by \(I(Y; U)\) and \(I(X; U)\), respectively, the slope of the optimal utility-privacy trade-off curve is studied when \(I(X; U) = 0\). Finally, through a similar but independent analysis, an alternative characterization of the maximal correlation between two random variables is provided.

Index Terms

Privacy, perfect privacy, non-private information, maximal correlation, mutual information.

I. INTRODUCTION

With the explosion of machine learning algorithms, and their applications in many areas of science, technology, and governance, data is becoming an extremely valuable asset. However, with the growing power of machine learning algorithms in learning individual behavioral patterns from diverse data sources, privacy is becoming a major concern, calling for strict regulations on data ownership and distribution. On the other hand, many recent examples of de-anonymization attacks on publicly available anonymized data ([1], [2]) show that regulation on its own will not be sufficient to limit access to private data. An alternative approach, also considered in this paper, is to process the data at the time of release, such that no private information is leaked, called perfect privacy. Assuming that the joint distribution of the observed data and the latent variables that should be kept private is known, an information-theoretic study is carried out in this paper to characterize the fundamental limits on perfect privacy.

Consider a situation in which Alice wants to release some useful information about herself to Bob, represented by random variable \(Y\), and she receives some utility from this disclosure of information. This may represent some

Borzoo Rassouli and Deniz Gündüz are with the Information Processing and Communications Lab at the Department of Electrical and Electronics Engineering, Imperial College London, United Kingdom. emails: \{b.rassouli12, d.gunduz\}@imperial.ac.uk.
data measured by a health monitoring system [3], her smart meter measurements [4], or the sequence of a portion of her DNA to detect potential illnesses [5]. At the same time, she wishes to conceal from Bob some private information which depends on $Y$, represented by $X$. To this end, instead of letting Bob have a direct access to $Y$, a privacy-preserving mapping is applied, whereby a distorted version of $Y$, denoted by $U$, is revealed to Bob. In this context, privacy and utility are competing goals: The more distorted version of $Y$ is revealed by the privacy mapping, the less information can Bob infer about $X$, while the less utility can be obtained. This trade-off is the very result of the dependencies between $X$ and $Y$. An extreme point of this trade-off is the scenario termed as perfect privacy, which refers to the situation where nothing is allowed to be inferred about $X$ by Bob through the disclosure of $U$. This condition is modelled by the statistical independence of $X$ and $U$.

The concern of privacy and the design of privacy-preserving mappings has been the focus of a broad area of research, e.g., [6]–[9], while the information-theoretic view of privacy has gained increasing attention more recently. In [10], a general statistical inference framework is proposed to capture the loss of privacy in legitimate transactions of data. In [11], the privacy-utility trade-off under the self-information cost function (log-loss) is considered and called the privacy funnel, which is closely related to the information bottleneck introduced in [12]. In [13], sharp bounds on the optimal privacy-utility trade-off for the privacy funnel are derived, and an alternative characterization of the perfect privacy condition (see [14]) is proposed. Measuring both the privacy and the utility in terms of mutual information, perfect privacy is fully characterized in [15] for the binary case. Furthermore, a new quantity is introduced to capture the amount of private information about the latent variable $X$ carried by the observable data $Y$.

We study the information theoretic perfect privacy in this paper, and our main contributions can be summarized as follows:

- Adopting mutual information as the utility measure, i.e., $I(Y;U)$, we show that the maximum utility under perfect privacy is the solution to a standard linear program (LP). We obtain similar results when other measures of utility, e.g., the minimum mean-square error or the probability of error, are considered.

- We show that when $(X,Y)$ is a jointly Gaussian pair with non-zero correlation coefficient, for the privacy mapping $p_{U|Y}$, perfect privacy is not feasible. In other words, $U$ is independent of $X$ if and only if it is also independent of $Y$, i.e., maximum privacy is obtained at the expense of zero utility. This, however, is not the case when the mapping is of the form $p_{U|X,Y}$; that is, when the encoder has access to the private latent variables as well as the data.

- Denoting the maximum $I(Y;U)$ under perfect privacy by $g_0(X,Y)$, we characterize the relationship between the non-private information about $X$ carried by $Y$, $D_X(Y)$ as defined in [15], and $g_0(X,Y)$.

- Considering mutual information as both the privacy and the utility measure, the optimal utility-privacy trade-off curve, characterized by the supremum of $I(Y;U)$ over $p_{U|Y}$ vs. $I(X;U)$, is not a straightforward problem. Instead, we investigate the slope of this curve when $I(X;U) = 0$. This linear approximation to the trade-off curve provides the maximum utility rate when a small amount of private data leakage is allowed. We obtain this slope when perfect privacy is not feasible, i.e., $g_0(X,Y) = 0$, and propose a lower bound on it when perfect privacy is feasible, i.e., $g_0(X,Y) > 0$. 


• As a by-product of this slope analysis, we provide an alternative characterization of the maximal correlation between two random variables [16]–[18].

Notations. Random variables are denoted by capital letters, their realizations by lower case letters, and their alphabets by capital letters in calligraphic font. Matrices and vectors are denoted by bold capital and bold lower case letters, respectively. For integers \( m \leq n \), we have the discrete interval \([m : n] \triangleq \{m, m + 1, \ldots, n\}\), and the tuple \((a_m, a_{m+1}, \ldots, a_n)\) is written in short as \(a_{[m:n]}\). For an integer \( n \geq 1 \), \(1_n\) denotes an \(n\)-dimensional all-one column vector. For a random variable \(X \in \mathcal{X}\), with finite \(|\mathcal{X}|\), the probability simplex \(\mathcal{P}(\mathcal{X})\) is the standard \((|\mathcal{X}| - 1)\)-simplex given by

\[
\mathcal{P}(\mathcal{X}) = \left\{ \mathbf{v} \in \mathbb{R}^{|\mathcal{X}|} \middle| \mathbf{1}_{|\mathcal{X}|}^T \mathbf{v} = 1, \; v_i \geq 0, \; \forall i \in [1 : |\mathcal{X}|] \right\}.
\]

Furthermore, to each probability mass function (pmf) on \(X\), denoted by \(p_X(\cdot)\), corresponds a matrix \(\mathbf{P}_X = \text{diag}(p_X)\), where \(p_X\) is a probability vector in \(\mathcal{P}(\mathcal{X})\), whose \(i\)-th element is \(p_X(x_i)\) \((i \in [1 : |\mathcal{X}|])\). For a pair of random variables \((X, Y)\) with joint pmf \(p_{X,Y}\), \(\mathbf{P}_{X,Y}\) is an \(|\mathcal{X}| \times |\mathcal{Y}|\) matrix with \((i,j)\)-th entry equal to \(p_{X,Y}(i,j)\). Likewise, \(\mathbf{P}_{X|Y}\) is an \(|\mathcal{X}| \times |\mathcal{Y}|\) matrix with \((i,j)\)-th entry equal to \(p_{X|Y}(i|j)\). \(F_Y(\cdot)\) denotes the cumulative distribution function (CDF) of random variable \(Y\), and if it admits a density, its probability density function (pdf) is denoted by \(f_Y(\cdot)\). For \(0 \leq t \leq 1\), \(H_b(t) \triangleq -t \log_2 t - (1 - t) \log_2(1 - t)\) denotes the binary entropy function. The unit-step function is denoted by \(s(\cdot)\). Throughout the paper, for a random variable \(Y\) with the corresponding probability vector \(\mathbf{p}_Y\), \(H(Y)\) and \(H(\mathbf{p}_Y)\) are written interchangeably, and so are the quantities \(D(p_Y(\cdot)||q_Y(\cdot))\) and \(D(\mathbf{p}_Y||\mathbf{q}_Y)\).

II. System model and preliminaries

Consider a pair of random variables \((X, Y) \in \mathcal{X} \times \mathcal{Y}\) \((|\mathcal{X}|, |\mathcal{Y}| < \infty)\) distributed according to the joint distribution \(p_{X,Y}\). We assume that \(p_Y(y) > 0, \forall y \in \mathcal{Y}\) and \(p_X(x) > 0, \forall x \in \mathcal{X}\), since otherwise the supports \(\mathcal{Y}\) or/and \(\mathcal{X}\) could have been modified accordingly. This equivalently means that the probability vectors \(\mathbf{p}_Y\) and \(\mathbf{p}_X\) are in the interior of their corresponding probability simplices, i.e. \(\mathcal{P}(\mathcal{Y})\) and \(\mathcal{P}(\mathcal{X})\). Let \(X\) denote the private/sensitive data that the user wants to conceal and \(Y\) denote the useful data the user wishes to disclose. Assume that the privacy mapping/data release mechanism takes \(Y\) as input and maps it to the released data denoted by \(U\). In this scenario, \(X - Y - U\) form a Markov chain, and the privacy mapping is captured by the conditional distribution \(p_{U|Y}\).

Let \(g_\epsilon(X, Y)\) be defined\(^1\) as \([15]\)

\[
g_\epsilon(X, Y) \triangleq \sup_{p_{U|Y}: X \rightarrow Y \rightarrow U} I(Y; U). \tag{1}
\]

In other words, when mutual information is adopted as a measure of both utility and privacy, \((1)\) gives the best utility that can be obtained by privacy mappings which keep the sensitive data \((X)\) private within a certain level of \(\epsilon\).

\(^1\)This is done with an abuse of notation, as \(g_\epsilon(X, Y)\) should be written as \(g_\epsilon(p_{X,Y})\).
Proposition 1. It is sufficient to have $|U| \leq |Y| + 1$. Also, we can write
\[ g_e(X, Y) = \max_{P_U|Y:} I(Y; U) = \max_{P_U|Y:} I(Y; U). \] (2)

Proof. The proof is provided in Appendix A.

Later, we show that it is sufficient to restrict our attention to $|U| \leq |Y|$, when $\epsilon = 0$.

III. Perfect Privacy

Definition. For a pair of random variables $(X, Y)$, we say that perfect privacy is feasible if there exists a random variable $U$ that satisfies the following conditions:

1) $X - Y - U$ forms a Markov chain,
2) $X \perp \perp U$, i.e., $X$ and $U$ are independent,
3) $Y \not\perp \perp U$, i.e., $Y$ and $U$ are not independent.

From the above definition, we can say that perfect privacy being feasible is equivalent to having $g_0(X, Y) > 0$.

Proposition 2. Perfect privacy is feasible if and only if
\[ \dim(\text{Null}(P_{X|Y})) \neq 0. \] (3)

Proof. In [14, Theorem 4], the authors showed that for a given pair of random variables $(X, Y) \in \mathcal{X} \times \mathcal{Y} (|\mathcal{X}|, |\mathcal{Y}| < \infty)$, there exists a random variable $U$ satisfying the conditions of perfect privacy if and only if the columns of $P_{X|Y}$ are linearly dependent. Equivalently, there must exist a non-zero vector $v$, such that $P_{X|Y}v = 0$, which is equivalent to (3).

Proposition 3. For the null space of $P_{X|Y}$, we have
\[ z \in \text{Null}(P_{X|Y}) \implies 1^T_{|Y|} \cdot z = 0. \]

Therefore, for any $z \in \text{Null}(P_{X|Y})$, there exists a positive real number $\alpha$, such that $p_Y + \alpha z \in \mathcal{P}(\mathcal{Y})$.

Proof. We have
\[ 1^T_{|Y|} \cdot z = 1^T_{|X|} P_{X|Y} z = 0, \] (4)
where (4) follows from the fact that $1^T_{|X|} P_{X|Y} = 1^T_{|Y|}$, and (5) from the assumption that $z$ belongs to Null$(P_{X|Y})$.

The last claim of the proposition is due to the fact that $p_Y$ is in the interior of $\mathcal{P}(\mathcal{Y})$.

Theorem 1. For a pair of random variables $(X, Y) \in \mathcal{X} \times \mathcal{Y} (|\mathcal{X}|, |\mathcal{Y}| < \infty)$, $g_0(X, Y)$ is the solution to a standard linear program (LP) as given in (13).

Proof. Let $P_{X|Y}$ be an $|\mathcal{X}| \times |\mathcal{Y}|$ matrix with $(i, j)$-th entry equal to $p_{X|Y}(i|j)$. 


From the singular value decomposition\(^2\) of \(P_{X|Y}\), we have
\[
P_{X|Y} = U\Sigma V^T,
\]
where the matrix of right eigenvectors is
\[
V = \begin{bmatrix} v_1 & v_2 & \ldots & v_{|Y|} \end{bmatrix}_{|Y| \times |Y|}.
\]
\(^6\) is equivalent to having the null space of \(P_{X|Y}\) written as
\[
\text{Null}(P_{X|Y}) = \text{Span}\{v_m, v_{m+1}, \ldots, v_{|Y|}\}, \text{ for some } m \leq |Y|.
\]
The random variables \(X\) and \(U\) are independent if and only if
\[
p_X(\cdot) = p_X(U(\cdot|u)), \forall u \in U \iff p_X = p_{X|u}, \forall u \in U.
\]
Furthermore, if \(X - Y - U\) form a Markov chain, \(^8\) is equivalent to
\[
P_{X|Y}(p_Y - p_{Y|u}) = 0, \forall u \in U \iff (p_Y - p_{Y|u}) \in \text{Null}(P_{X|Y}), \forall u \in U.
\]
From the column vectors in \(^6\) and the definition of index \(m\) afterwards, construct the matrix \(A\) as
\[
A \triangleq \begin{bmatrix} v_1 & v_2 & \ldots & v_{m-1} \end{bmatrix}^T.
\]
From \(^7\), we can write
\[
(p_Y - p_{Y|u}) \in \text{Null}(P_{X|Y}), \forall u \in U \iff A(p_Y - p_{Y|u}) = 0, \forall u \in U.
\]
Therefore, for the triplet \((X, Y, U)\), if \(X - Y - U\) forms a Markov chain and \(X \perp U\), we must have \(p_{Y|u} \in S, \forall u \in U\), where \(S\) is a convex polytope defined as
\[
S \triangleq \left\{ x \in \mathbb{R}^{|Y|} \middle| Ax = Ap_Y, x \geq 0 \right\}.
\]
Note that any element of \(S\) is a probability vector according to Proposition 3.

On the other hand, for any pair \((Y, U)\), for which \(p_{Y|u} \in S, \forall u \in U\), we can simply have \(X - Y - U\) and \(X \perp U\). Therefore, we can write
\[
X - Y - U, \ X \perp U \iff p_{Y|u} \in S, \forall u \in U.
\]
This leads us to
\[
g_0(X, Y) = \max_{p_U(Y): \ I(X;U) \leq 0} I(Y;U) = \max_{p_U(Y): \ I(X;U) = 0} I(Y;U) = \max_{p_U(Y): \ p_{Y\mid u} \in S, \forall u \in U} I(Y;U) = \min_{p_U(Y): \ p_{Y\mid u} \in S, \forall u \in U} H(Y|U),
\]
\(^2\)We assume, without loss of generality, that the singular values are arranged in a descending order.
where in [12], since the minimization is over \( p_{Y|u} \) rather than \( p_{U|Y} \), a constraint was added to preserve the marginal distribution \( p_Y \).

**Proposition 4.** In minimizing \( H(Y|U) \) over \( \{ p_{Y|u} \in S, \ \forall u \in U \mid \sum_u p_U(u)p_{Y|u} = p_Y \} \), it is sufficient to consider only \(|Y|\) extreme points of \( S \).

**Proof.** The proof is provided in Appendix B.

From Proposition 4, the problem in (12) can be divided into two phases: in phase one, the extreme points of set \( S \) are identified, while in phase two, proper weights over these extreme points are obtained to minimize the objective function.

For the first phase, we proceed as follows. The extreme points of \( S \) are the basic feasible solutions (see [19], [20]) of it, i.e., the basic feasible solutions of the set

\[
\left\{ x \in \mathbb{R}^{|Y|} \bigg| Ax = b, \ x \geq 0 \right\},
\]

where \( b = Ap_Y \).

The procedure of finding the extreme points of \( S \) is as follows. Pick a set \( B \subset [1 : |Y|] \) of indices that correspond to \( m-1 \) linearly independent columns of matrix \( A \) in (9). Since matrix \( A_{(m-1) \times |Y|} \) is full rank (note that its rows are mutually orthonormal and \( \text{rank}(A) = m-1 \)), there are at most \( \binom{|Y|}{m-1} \) ways of choosing \( m-1 \) linearly independent columns of \( A \). Let \( A_B \) be an \( (m-1) \times (m-1) \) matrix whose columns are the columns of \( A \) indexed by the indices in \( B \). Also, for any \( x \in S \), let \( \tilde{x} = \begin{bmatrix} x_B^T & x_N^T \end{bmatrix}^T \), where \( x_B \) and \( x_N \) are \((m-1)\)-dimensional and \((|Y|-m+1)\)-dimensional vectors whose elements are the elements of \( x \) indexed by the indices in \( B \) and \([1 : |Y|]\setminus B \), respectively.

For any basic feasible solution \( x^* \), there exists a set \( B \subset [1 : |Y|] \) of indices that correspond to a set of linearly independent columns of \( A \), such that the corresponding vector of \( x^* \), i.e. \( \tilde{x}^* = \begin{bmatrix} x_B^{*T} & x_N^{*T} \end{bmatrix}^T \), satisfies the following

\[
x_N^* = 0, \quad x_B^* = A_B^{-1}b, \quad x_B^* \geq 0.
\]

On the other hand, for any set \( B \subset [1 : |Y|] \) of indices that correspond to a set of linearly independent columns of \( A \), if \( A_B^{-1}b \geq 0 \), then \( \begin{bmatrix} A_B^{-1}b \\ 0 \end{bmatrix} \) is the corresponding vector of a basic feasible solution. Hence, the extreme points of \( S \) are obtained as mentioned above, and their number is at most \( \binom{|Y|}{m-1} \).

For the second phase, we proceed as follows. Assume that the extreme points of \( S \), found in the previous phase, are denoted by \( p_1, p_2, \ldots, p_K \). Then (12) is equivalent to

\[
g_0(X,Y) = H(Y) - \min_{w \geq 0} \begin{bmatrix} H(p_1) & H(p_2) & \cdots & H(p_K) \end{bmatrix} \cdot w
\]

s.t. \( \begin{bmatrix} p_1 & p_2 & \cdots & p_K \end{bmatrix} w = p_Y \),

(13)

where \( w \) is a \( K \)-dimensional weight vector, and it can be verified that the constraint \( \sum_{i=1}^{K} w_i = 1 \) is satisfied if the constraint in (13) is met. The problem in (13) is a standard linear program (LP), which can be efficiently solved. 

\[ \square \]
The following example clarifies the optimization procedure in the proof of Theorem 1.

**Example 1.** Consider the pair \((X, Y) \in [1:2] \times [1:4]\) joint distribution is specified by the following matrix:

\[
P_{X,Y} = \begin{bmatrix}
0.15 & 0.2 & 0.0625 & 0.05 \\
0.35 & 0.05 & 0.0625 & 0.075
\end{bmatrix}.
\]

This results in

\[
p_Y = \begin{bmatrix}
\frac{1}{2} \\
\frac{1}{4} \\
\frac{1}{4} \\
\frac{1}{8}
\end{bmatrix}, \quad P_{X|Y} = \begin{bmatrix}
0.3 & 0.8 & 0.5 & 0.4 \\
0.7 & 0.2 & 0.5 & 0.6
\end{bmatrix}.
\]

Since \(|Y| > |X|\), we have \(\text{dim}(\text{Null}(P_{X|Y})) \neq 0\); and therefore, \(g_0(X, Y) > 0\). The singular value decomposition of \(P_{X|Y}\) is

\[
P_{X|Y} = \begin{bmatrix}
-0.7071 & -0.7071 \\
-0.7071 & 0.7071
\end{bmatrix} \begin{bmatrix}
1.4142 & 0 & 0 & 0 \\
0 & 0.5292 & 0 & 0
\end{bmatrix} \begin{bmatrix}
-0.5 & 0.5345 & -0.4163 & -0.5394 \\
-0.5 & -0.8018 & -0.3154 & -0.0876 \\
-0.5 & 0 & 0.8452 & -0.1889 \\
-0.5 & 0.2673 & -0.1135 & 0.8159
\end{bmatrix},
\]

where it is obvious that columns 3 and 4 of the matrix of the right eigenvectors span the null space of \(P_{X|Y}\).

Hence, the matrix \(A\) in (9) is given by

\[
A = \begin{bmatrix}
-0.5 & -0.5 & -0.5 & -0.5 \\
0.5345 & -0.8018 & 0 & 0.2673
\end{bmatrix}.
\]

For the first phase, i.e., finding the extreme points of \(\mathcal{S}\), it is clear that there are \(\binom{4}{2}\) possible ways of choosing 2 linearly independent columns of \(A\). Hence, the index set \(\mathcal{B}\) can be \(\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}\) or \(\{3, 4\}\).

From \(x_B = A_B^{-1}b\), we get

\[
x_{\{1,2\}} = \begin{bmatrix} 0.675 \\ 0.325 \end{bmatrix}, x_{\{1,3\}} = \begin{bmatrix} 0.1875 \\ 0.8125 \end{bmatrix}, x_{\{1,4\}} = \begin{bmatrix} -0.625 \\ 1.625 \end{bmatrix}, x_{\{2,3\}} = \begin{bmatrix} -0.125 \\ 1.125 \end{bmatrix}, x_{\{2,4\}} = \begin{bmatrix} 0.1563 \\ 0.8437 \end{bmatrix}, x_{\{3,4\}} = \begin{bmatrix} 0.625 \\ 0.375 \end{bmatrix}.
\]

It is obvious that \(x_{\{1,4\}}\) and \(x_{\{2,4\}}\) are not feasible, since they do not satisfy \(x_B \geq 0\). Therefore, the extreme points of \(\mathcal{S}\) are obtained as

\[
P_1 = \begin{bmatrix} 0.675 \\ 0.325 \\ 0 \end{bmatrix}, P_2 = \begin{bmatrix} 0.1875 \\ 0.8125 \\ 0 \end{bmatrix}, P_3 = \begin{bmatrix} 0 \\ 0.1563 \\ 0.8437 \end{bmatrix}, P_4 = \begin{bmatrix} 0 \\ 0 \\ 0.625 \end{bmatrix}.
\]
Now, for the second phase, the standard LP in (13) is written as
\[
\begin{align*}
\min_{\mathbf{w} \geq 0} & \begin{bmatrix} 0.9097 & 0.6962 & 0.6254 & 0.9544 \end{bmatrix} \cdot \mathbf{w} \\
\text{S.t.} & \begin{bmatrix} 0.675 & 0.1875 & 0 & 0 \\
0.325 & 0 & 0.1563 & 0 \\
0 & 0.8125 & 0 & 0.625 \\
0 & 0 & 0.8437 & 0.375 \end{bmatrix} \mathbf{w} = \begin{bmatrix} 1/2 \\
1/4 \\
1/8 \\
1/8 \end{bmatrix},
\end{align*}
\]
where the minimum value of the objective function is 0.8437 bits, which is achieved by
\[
\mathbf{w}^* = \begin{bmatrix} 0.698 \\
0.1538 \\
0.1481 \\
0 \end{bmatrix}.
\]
Therefore, \(g_0(X,Y) = H(Y) - 0.8437 = 0.9063\) (note that \(H(Y|X) = 1.3061\)), \(U = \{u_1, u_2, u_3\}\), \(p_U = \begin{bmatrix} 0.698 & 0.1538 & 0.1481 \end{bmatrix}^T\) and \(p_{Y|u_i} = p_i, \forall i \in [1:3]\). Finally, \(p_{U|Y}^*\) corresponds to the matrix \(P_{U|Y}^*\) given as
\[
P_{U|Y}^* = \begin{bmatrix} 0.9423 & 0.9074 & 0 & 0 \\
0.0577 & 0 & 1 & 0 \\
0 & 0.0926 & 0 & 1 \end{bmatrix}.
\]

**Remark 1.** It can be verified that in the degenerate case of \(X \perp \perp Y\), we have \(\text{Null}(P_{X|Y}) = \text{Span}\{v_1, v_2, \ldots, v_{|Y|}\}\), or equivalently, \(S = \mathcal{P}(Y)\). In this case, the extreme points of \(S\) have zero entropy. Therefore, the minimum value of \(H(Y|U)\) is zero, and \(U = \{u_1, u_2, \ldots, u_{|Y|}\}\) with \(p_U(u_i) = p_Y(y_i), \forall i \in [1:|Y|]\) and \(p_{Y|u_i} = e_i\), which is the \(i\)th extreme point of \(\mathcal{P}(Y)\). As a result, \(g_0(X,Y) = H(Y)\), which is also consistent with the fact that \(U = Y\) is independent of \(X\) and maximizes \(I(Y;U)\).

**A. MMSE under perfect privacy**

Assume that instead of \(I(Y;U)\), the goal is to minimize \(E[(Y - U)^2]\) under the perfect privacy constraint. This can be formulated as follows:
\[
\min_{p_{U|Y}} E[(Y - U)^2],
\]
where the expectation is according to the joint distribution \(p_{Y,U}\). Obviously, an upperbound for (14) is \(\text{Var}[Y]\), as one could choose \(U = E[Y]\). In what follows, we show that (14) has a similar solution to that of \(g_0(X,Y)\). The only difference is that the realizations of \(U\), i.e., the particular values of the elements in \(U\), are irrelevant for the
solution of $g_0(X,Y)$, since only their mass probabilities have a role when evaluating $I(Y;U)$, while the objective function in (14) takes into account both the pmf and the realizations of $U$. We can write

$$E_{U,Y}[(Y - U)^2] = E_U \left[ E_{Y|U}[(Y - U)^2|U] \right]$$

$$\geq E_U \left[ E_{Y|U} [(Y - E_{Y|U})^2|U] \right]$$

$$= \int \text{Var}[Y|U = u] dF(u),$$

where (15) is a classical result from MMSE estimation [21]; and in (16), we have used $\text{Var}[Y|U = u] = E_{Y|U} \left( (Y - E[Y|U])^2 | U = u \right)$. Therefore, from (11) and (16), we have

$$\min_{p_U|Y, X - Y - U, X \perp \perp U} \left\{ \frac{1}{w} \right\} = \min_{p_U|Y, p_Y|u \in S} \int \text{Var}[Y|U = u] dF(u),$$

where the equality holds if and only if $E[Y|U = u] = u$, $\forall u \in U$.

**Proposition 5.** $\text{Var}[Y|U = u]$ is a strictly concave function of $p_Y|u$.

*Proof.* The proof is provided in Appendix C.

From the concavity of $\text{Var}[Y|U = u]$ in Proposition 5, we can apply the reasoning in the proof of Proposition 4, and conclude that in (17), it is sufficient to consider only the extreme points of $S$. Hence, the problem has two phases, where in phase one, the extreme points of $S$ are found. For the second phase, denoting the extreme points of $S$ by $p_1, p_2, \ldots, p_K$, (17) boils down to a standard linear program as follows.

$$\min_{w \geq 0} \begin{bmatrix} \text{Var}_1 & \text{Var}_2 & \ldots & \text{Var}_K \end{bmatrix} \cdot w$$

$$\text{S.t.} \begin{bmatrix} p_1 & p_2 & \ldots & p_K \end{bmatrix} w = p_Y,$$

where $\text{Var}_i (\forall i \in [1:K])$ denotes $\text{Var}[Y|U = u]$ under $p_i$, i.e., when $p_Y|u = p_i$. Finally, once the LP in (18) is solved, the realizations of the random variable $U$ are set to equalize the expectations of $Y$ under the corresponding distributions of those extreme points ($p_i$) of $S$ with non-zero mass probability. For example, the problem in (14) for the pair $(X,Y)$ given in Example 1 is

$$\min_{w \geq 0} \begin{bmatrix} 0.2194 & 0.6094 & 0.52754 & 0.2344 \end{bmatrix} \cdot w$$

$$\text{s.t.} \begin{bmatrix} 0.675 & 0.1875 & 0 & 0 \\ 0.325 & 0 & 0.1563 & 0 \\ 0 & 0.8125 & 0 & 0.625 \\ 0 & 0 & 0.8437 & 0.375 \end{bmatrix} w = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix},$$

(19)
where the extreme points of $S$, i.e., $p_i$ ($i \in [1 : 4]$), are already known from Example 1. The minimum value of the standard LP in $\mathbf{24}$ is 0.2406, which is achieved by

$$
\mathbf{w}^* = \begin{bmatrix}
0.7407 \\
0 \\
0.0593 \\
0.2
\end{bmatrix}.
$$

Therefore, $\text{MMSE} = 0.2406$ ($\leq \text{Var}[Y] = 1.1094$), $p_U = \begin{bmatrix}
0.7407 \\
0.0593 \\
0.2
\end{bmatrix}^T$ and $p_{Y|u_1} = p_1$, $p_{Y|u_i} = p_{i+1}$, $i = 2, 3$. By setting $u = \mathbb{E}[Y|U = u]$, we obtain $\mathcal{U} = \{u_1, u_2, u_3\} = \{1.325, 3.6874, 3.375\}$.

**B. Minimum probability of error under perfect privacy**

The objective of the optimization can be the error probability as

$$
\min_{p_U|Y} \Pr\{Y \neq U\}. \quad (20)
$$

Obviously, an upper bound for $\mathbf{20}$ is $1 - \max_y p_{Y}(y)$ as one could choose $U = \arg \max_y p_{Y}(y)$. For an arbitrary joint distribution on $(Y, U)$, we can write

$$
\Pr\{Y \neq U\} = 1 - \Pr\{Y = U\} \\
= 1 - \int_{\mathcal{U}} \Pr\{Y = u|U = u\}dF_U(u) \\
\geq 1 - \int_{\mathcal{U}} \max_y p_{Y|U}(y|u)dF_U(u), \quad (21)
$$

where $\mathbf{21}$ holds with equality when $u = \arg \max_y p_{Y|U}(y|u)$. Then,

$$
\min_{p_U|Y} \Pr\{Y \neq U\} \geq 1 - \max_{F_U(\cdot), \ p_{Y|u} \in \mathbb{S}} \int_{\mathcal{U}} \max_y p_{Y|U}(y|u)dF_U(u). \quad (22)
$$

It can be verified that $\max_y p_{Y}(y)$ is convex in $p_{Y}(\cdot)$. Hence, following the same reasoning in the proof of Proposition 4, it is sufficient to consider only the extreme points of $S$ in the optimization in $\mathbf{22}$. Therefore, the problem has two phases: in phase one, the extreme points of $S$ are identified. For the second phase, denoting the extreme points of $S$ by $p_1, p_2, \ldots, p_K$, the problem boils down to a standard linear program as follows:

$$
1 - \max_{\mathbf{w} \geq 0} \begin{bmatrix}
p_{m_1} & p_{m_2} & \ldots & p_{m_K}
\end{bmatrix} \cdot \mathbf{w} \\
\text{s.t.} \begin{bmatrix}
p_1 & p_2 & \ldots & p_K
\end{bmatrix} \mathbf{w} = \mathbf{p_Y}, \quad (23)
$$

where $p_{m_i}$ is the maximum element of the vector $p_i$, $i \in [1 : K]$. Once the LP is solved, the optimal conditionals $p_{Y|u}$ and the optimal mass probabilities of $U$ are obtained. Finally, the realizations of the random variable $U$ are
set to equalize $u = \arg \max_y p_{Y|u}(y|u)$. For example, the problem in (20) for the pair $(X, Y)$ given in Example 1 is

$$1 - \max_{w \geq 0} \begin{bmatrix} 0.675 & 0.8125 & 0.8437 & 0.625 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ 0.675 & 0.1875 & 0 & 0 \\ 0.325 & 0 & 0.1563 & 0 \\ 0 & 0.8125 & 0 & 0.625 \\ 0 & 0 & 0.8437 & 0.375 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad (24)$$

where the extreme points of $S$, i.e. $p_i$ $(i \in [1 : 4])$, are already known from Example 1. The minimum probability of error is obtained as $0.2789$ ($\leq 1 - \max_y p_{Y}(y) = \frac{1}{2}$) achieved by

$$w^* = \begin{bmatrix} 0.698 \\ 0.1538 \\ 0.1482 \\ 0 \end{bmatrix}.$$ 

Hence, $p_U = \begin{bmatrix} 0.698 & 0.1538 & 0.1482 \end{bmatrix}^T$ with $U = \{u_1, u_2, u_3\} = \{1, 3, 4\}$, and $p_{Y|u_i} = p_i$, $i \in [1 : 3]$.

Thus far, we have investigated the constraint of perfect privacy when $|X|, |Y| < \infty$. The next theorem and its succeeding example consider two cases in which at least one of $|X|$ and $|Y|$ is infinite. The following theorem shows that perfect privacy is not feasible for the (correlated) jointly Gaussian pair.

**Theorem 2.** Let $(X, Y) \sim \mathcal{N}(\mu, \Sigma)$ be a pair of jointly Gaussian random variables, where

$$\mu = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_X^2 & \rho \sigma_X \sigma_Y \\ \rho \sigma_X \sigma_Y & \sigma_Y^2 \end{bmatrix},$$

(25)
in which $\rho \neq 0$, since otherwise $X \perp \perp Y$. We have $g_0(X, Y) = 0$ for the above pair.

**Proof.** If there exists a random variable $U$ such that $X - Y - U$ form a Markov chain and $X \perp \perp U$, we must have $F_X(\cdot) = F_{X|U}(\cdot|u), \forall u \in U$, and hence, $f_X(\cdot) = f_{X|U}(\cdot|u), \forall u \in U$, since $X$ has a density. Equivalently, we must have

$$f_X(\cdot) = \int f_{X|Y}(\cdot|y)dF_{Y|U}(y|u), \forall u \in U. \quad (26)$$

Also, to have $g_0(X, Y) > 0$, there must exist at least $u_1, u_2 \in U$, such that

$$F_{Y|U}(\cdot|u_1) \neq F_{Y|U}(\cdot|u_2). \quad (27)$$

In what follows we show that if (26) holds, (27) cannot be satisfied; and therefore, perfect privacy is not feasible for a jointly Gaussian $(X, Y)$ pair.

It is known that $X$ conditioned on $\{Y = y\}$ is also Gaussian, given by

$$X|\{Y = y\} \sim \mathcal{N}\left(\frac{\rho \sigma_X}{\sigma_Y}(y - \mu_Y) + \mu_X, \frac{(1 - \rho^2)\sigma_X^2}{\sigma^2}\right).$$

(28)
From (26), (28), and for \( u_1, u_2 \in \mathcal{U} \), we have

\[
    f_X(x) = \int e^{-\frac{(x-y-\beta)^2}{2\sigma^2}} dF_{Y|U}(y|u_1)
    = \int e^{-\frac{(x-y-\beta)^2}{2\sigma^2}} dF_{Y|U}(y|u_2), \quad \forall x \in \mathbb{R},
\]
or, equivalently,

\[
    \int e^{-\frac{(x-y-\beta)^2}{2\sigma^2}} d\left( F_{Y|U}(y|u_1) - F_{Y|U}(y|u_2) \right) = 0, \quad \forall x \in \mathbb{R}. \tag{29}
\]

Multiplying both sides of (29) by \( e^{j\omega x} \), and taking the integral with respect to \( x \), we obtain

\[
    \int e^{j\omega x} \left[ \int e^{-\frac{(x-y-\beta)^2}{2\sigma^2}} d\left( F_{Y|U}(y|u_1) - F_{Y|U}(y|u_2) \right) \right] dx = 0.
\]

By Fubini’s theorem\(^1\) we can write

\[
    \int \left[ \int e^{j\omega x} e^{-\frac{(x-y-\beta)^2}{2\sigma^2}} dx \right] d\left( F_{Y|U}(y|u_1) - F_{Y|U}(y|u_2) \right) = 0.
\]

After some manipulations, we get

\[
    \int e^{j\omega \rho y} d\left( F_{Y|U}(y|u_1) - F_{Y|U}(y|u_2) \right) = 0. \tag{30}
\]

Since \( \rho \neq 0 \), from (28), we have \( \alpha \neq 0 \). Hence, the LHS of (30) is a Fourier transform. Due to the invertibility of the Fourier transform, i.e. \( \int e^{j\omega t} dg(t) = 0 \iff dg(t) = 0 \), we must have \( F_{Y|U}(\cdot|u_1) = F_{Y|U}(\cdot|u_2) \). Therefore, (27) does not hold and perfect privacy is not feasible for the (correlated) jointly Gaussian pair \((X,Y)\). \(\blacksquare\)

In the following example, we consider \(|\mathcal{X}| = 2\) and \(|\mathcal{Y}| = \infty\). We observe that we can have bounded \( I(X;Y) \) and \( h(Y) \), while an unbounded \( g_0(X,Y) \) without even revealing \( Y \) undistorted. This renders the usage of mutual information as a measure of dependence counterintuitive for continuous alphabets. This is related to the fact that differential entropy cannot be interpreted as a measure of the information content in a random variable, as it can take negative values.

**Example 2.** Let \( \mathcal{X} = \{x_0, x_1\} \) with \( p_X(x_0) = p \). Let \( Y|\{X = x_0\} \sim \text{Uniform}[0,1] \) and \( Y|\{X = x_1\} \sim \text{Uniform}[0,2] \). It can be verified that the probability density function (pdf) of \( Y \)

\[
    f_Y(y) = \begin{cases} 
        \frac{1+p}{2} & y \in [0,1] \\
        \frac{1-p}{2} & y \in [1,2] \\
        0 & y \notin [0,2] 
    \end{cases}, \tag{31}
\]

and the conditional pmf of \( X \) conditioned on \( Y \) is given by

\[
    p_{X|Y}(x_0|y) = \begin{cases} 
        \frac{2p}{1+p} & y \in [0,1] \\
        0 & y \notin [0,1] 
    \end{cases}, \quad p_{X|Y}(x_1|y) = \begin{cases} 
        \frac{1-p}{1+p} & y \in [0,1] \\
        1 & y \in [1,2] \\
        0 & y \notin [0,2] 
    \end{cases}.
\]

\(^3\)Note that \( \int |f_{X|U}(x|u_1) - f_{X|U}(x|u_2)| dx \leq \int (|f_{X|U}(x|u_1)| + |f_{X|U}(x|u_2)|) dx = 2 < +\infty \).
Since the support of Y is the interval [0, 2], the support of \(Y\{|U = u\}\) must be a subset of [0, 2], \(\forall u \in \mathcal{U}\). Also, the independence of X and U in \(-Y - U\) implies 
\[
p_{X|U}(x|u) = p, \ \forall u \in \mathcal{U} \implies \Pr\{0 \leq Y \leq 1|U = u\} = \int_0^1 dF_{Y|U}(y|u) = \frac{1 + p}{2}, \ \forall u \in \mathcal{U}.
\]
Finally, in order to preserve the pdf of Y in (31), the conditional CDF \(F_{Y|U}(\cdot|u)\) must satisfy the following 
\[
\int_{\mathcal{U}} F_{Y|U}(\cdot|u)dF_U(u) = F_Y(\cdot),
\] (32)
where \(F_Y(\cdot)\) is the CDF corresponding to the pdf in (31).

Let \(\mathcal{F}\) be the set of all CDFs defined on [0, 2] and \(\mathcal{F}\) be defined as 
\[
\mathcal{F} \triangleq \left\{ F_Y \in \mathcal{F} \left| \int_0^1 dF_Y(y) = \frac{1 + p}{2} \right. \right\}.
\]
We can write 
\[
g_0(X, Y) = \sup_{F_U(\cdot), \ F_{Y|U}(\cdot|u) \in \mathcal{F}, \ \forall u \in \mathcal{U}, \ \int_0^1 F_{Y|U}(\cdot|u)dF_U(u)=F_Y(\cdot)} I(Y; U).
\] (33)
In what follows, we show that the supremum in (33) is unbounded. Let \(M\) be an arbitrary positive integer. Construct the joint distribution \(p_{Y,U}\) as follows. Let \(\mathcal{U} = \{u_1, u_2, \ldots, u_M\}\), with \(p_U(u_i) = \frac{1}{M}, \forall i \in [1 : M]\). Also, let 
\[
f_{Y|U}(y|u_i) = \begin{cases} 
\frac{1 + p}{2}M & y \in \left[\frac{i - 1}{M}, \frac{i}{M}\right] \\
\frac{1 - p}{2}M & y \in \left[1 + \frac{i - 1}{M}, 1 + \frac{i}{M}\right), \ \forall i \in [1 : M]. \\
0 & \text{otherwise}
\end{cases}
\]
It can be verified that \(F_{Y|U}(\cdot|u_i) \in \mathcal{F}, \forall i \in [1 : M]\), and (32) is also satisfied. We have \(h(Y) = H_b(\frac{1 + p}{2})\) and \(h(Y|U) = H_b(\frac{1 + p}{2}) - \log_2 M\). Hence, with this construction, we have  
\[
I(Y; U) = \log_2 M.
\]
Since this is true for any positive integer \(M\), we conclude that \(g_0 = +\infty\), while \(I(X; Y) = H_b(\frac{1 + p}{2}) - 1 + p\) is finite. Therefore, letting \(U\) be a quantized version of \(Y\) results in an unbounded \(g_0\).

When we consider MMSE as the utility function, we have to solve the following problem 
\[
\min_{F_U(\cdot), \ F_{Y|U}(\cdot|u) \in \mathcal{F}, \ \forall u \in \mathcal{U}, \ \int_0^1 F_{Y|U}(\cdot|u)dF_U(u)=F_Y(\cdot)} \var{Var}[Y|U = u]dF(u).
\]
Since \(\var{Var}[Y|U = u]\) is a strictly concave function of \(F_{Y|U}(\cdot|u)\), it is sufficient to consider the optimization over the extreme points of \(\mathcal{F}\), which are the distributions concentrated at two mass points; one in the interval [0, 1] with mass probability \(\frac{1 + p}{2}\), and the other one in the interval [1, 2] with mass probability \(\frac{1 - p}{2}\). In other words, we have \(F_{Y|U}(y|u) = \frac{1 + p}{2} s(y - y_1(u)) + \frac{1 - p}{2} s(y - y_2(u))\), where \(s(\cdot)\) denotes the unit step function, and \(y_1(u) \in [0, 1], y_2(u) \in [1, 2]\) denote the two aforementioned mass points as a function of \(u\).

A simple analysis shows that 
\[
\var{Var}[Y|U = u] = \frac{1 - p^2}{4} (y_1(u) - y_2(u))^2.
\]
Therefore,
\[
\min_{F_U(\cdot), F_{Y|U}(\cdot|u) \in \mathbb{F}, \forall u \in U} \int \text{Var}[Y|U = u] dF(u)
\]
\[
= \min_{F_U(\cdot), y_1(u) \in [0,1], y_2(u) \in [1,2], \forall u \in U}
\frac{1 - p^2}{4} \int \left( y_1(u) - y_2(u) \right)^2 dF(u)
\]
\[
\geq \min_{F_U(\cdot), y_1(u) \in [0,1], y_2(u) \in [1,2], \forall u \in U}
\frac{1 - p^2}{4} \int \left( |y_1(u) - y_2(u)| dF(u) \right)^2
\]
\[
= \min_{F_U(\cdot), F_{Y|U}(\cdot|u) \in \mathbb{F}, \forall u \in U}
\frac{1 - p^2}{4} \left( \int \left[ \int_0^1 \frac{2y}{1+p} dF_{Y|U}(y|u) - \int_1^2 \frac{2y}{1-p} dF_{Y|U}(y|u) \right] dF(u) \right)^2
\]
\[
= \frac{1 - p^2}{4} \left( \text{Var}[Y|0 \leq Y \leq 1] - \text{Var}[Y|1 \leq Y \leq 2] \right)^2
\]
\[
= \frac{1 - p^2}{4} \left( \frac{1}{2} - \frac{3}{2} \right)^2
\]
\[
= \frac{1 - p^2}{4}.
\]  
(36)

where (34) is due to the convexity of \(x^2\) and Jensen’s inequality, and (35) is from the fact that \(\int_0^1 y dF_{Y|U}(y|u) = \frac{1+p}{2} y_1(u)\) and \(\int_1^2 y dF_{Y|U}(y|u) = \frac{1-p}{2} y_2(u)\).

In order to achieve the minimum in (36), we proceed as follows. Let \(U \sim \text{Uniform} \left[ \left[ \frac{1-p}{2}, \frac{3-p}{2} \right] \right]\), and
\[
F_{Y|U}(y|u) = \frac{1+p}{2} s(y - u + \frac{1-p}{2}) + \frac{1-p}{2} s(y - u - \frac{1-p}{2}),
\]
which means that \(Y|\{U = u\}\) has two mass points at \(u - \frac{1-p}{2} \in [0,1]\) and \(u + \frac{1+p}{2} \in [1,2]\) with probabilities \(\frac{1+p}{2}\) and \(\frac{1-p}{2}\), respectively. It can then be verified that (37) satisfies (32). Finally, the following equalities hold.
\[
\text{Var}[Y|U = u] = \frac{1 - p^2}{4}
\]
\[
\text{E}[Y|U = u] = u, \forall u \in \left[ \left[ \frac{1-p}{2}, \frac{3-p}{2} \right] \right].
\]

Therefore, we have \(\text{MMSE} = \frac{1 - p^2}{4} < \frac{1}{3} - \frac{p^2}{4} = \text{Var}[Y]\).

When we consider the probability of error as the utility function, we have to solve the following problem
\[
1 - \max_{F_U(\cdot), F_{Y|U}(\cdot|u) \in \mathbb{F}, \forall u \in U}
\text{Pr}\{Y = U\}.
\]
It can be verified that similarly to the analysis when MMSE is the utility function, we can restrict our attention to the extreme points of \(\mathbb{F}\) and obtain the minimum error probability of \(\frac{1-p}{2}\), which is achieved by \(U \sim \text{Uniform}[0,1]\) and \(F_{Y|U}(y|u) = \frac{1+p}{2} s(y - u) + \frac{1-p}{2} s(y - u - 1)\).

IV. NON-PRIVATE INFORMATION VS. \(g_0(X,Y)\)

For a pair of random variables \((X,Y) \in \mathcal{X} \times \mathcal{Y}\) \((|\mathcal{X}|,|\mathcal{Y}| < \infty\), the private information about \(X\) carried by \(Y\) is defined in (15) as
\[
C_X(Y) \triangleq \min_{W: \mathcal{X} \rightarrow W \rightarrow Y, \text{H}(W) \rightarrow \text{H}(W|Y) = 0} \text{H}(W). \tag{38}
\]
Since $H(W|Y) = 0$ implies that $W$ is a deterministic function of $Y$, Eq. (38) means that among all the functions of $Y$ that make $X$ and $Y$ conditionally independent, we want to find the one with the lowest entropy.

It can be verified that

$$I(X; Y) \leq C_X(Y) \leq H(Y),$$

where the first inequality is due to the data processing inequality applied on the Markov chain $X - W - Y$, i.e., $I(W; Y) \geq I(X; Y)$, and the second inequality is a direct result of the fact that $W = Y$ satisfies the constraints in Eq. (38).

The non-private information about $X$ carried by $Y$ is defined in [15] as

$$D_X(Y) \triangleq H(Y) - C_X(Y).$$

Let $T^X : \mathcal{Y} \to \mathcal{P}(\mathcal{X})$ be a mapping from $\mathcal{Y}$ to the probability simplex on $\mathcal{X}$ defined by $y \to p_{X|Y}(\cdot|y)$. It was shown in [15, Theorem 3] that the minimizer in Eq. (38) is $W^* = T^X(Y)$; and hence,

$$D_X(Y) = H(Y) - H(T^X(Y)).$$

Furthermore, it was proved in [15, Lemma 5] that $C_X(Y) = H(Y)$, i.e., $D_X(Y) = 0$, if and only if there do not exist $y_1, y_2 \in \mathcal{Y}$ such that $p_{X|Y}(\cdot|y_1) = p_{X|Y}(\cdot|y_2)$.

In [15], $g_0(X, Y)$ and $D_X(Y)$ were loosely connected to each other, as the latter represents roughly the amount of information contained in $Y$ and not correlated with $X$. Three examples were provided, where in two of them $g_0(X, Y) = D_X(Y)$, while in the last one $g_0(X, Y) > D_X(Y)$. Finally a question was raised regarding the condition on the joint distribution $p_{X,Y}$ under which $g_0(X, Y) = D_X(Y)$ holds. In what follows, we characterize the relation between $D_X(Y)$ and $g_0(X, Y)$.

Let $P_{X|Y}$ denote the matrix corresponding to the conditional distribution $p_{X|Y}$. From [15, Lemma 5], if $D_X(Y) > 0$, we must have at least $y_1, y_2 \in \mathcal{Y}$ such that their corresponding columns in $P_{X|Y}$ are equal. Let $\mathcal{E}_1 \subset [1 : |\mathcal{Y}|]$ be a set of (at least two) indices corresponding to the columns in $P_{X|Y}$ that are equal, i.e.,

$$p_{X|y_1} = p_{X|y_2}, \quad \forall i, j \in \mathcal{E}_1, i \neq j, \quad \text{and} \quad p_{X|y_1} \neq p_{X|y_k}, \quad \forall i \in \mathcal{E}_1, \forall k \in [1 : |\mathcal{Y}|] \setminus \mathcal{E}_1.$$

We can generalize this definition if the matrix $P_{X|Y}$ has several subsets of identical columns. Hence, the corresponding index sets are denoted by $\mathcal{E}_m, \forall m \in [1 : B]$, for some $B \geq 1$. In other words,

$$p_{X|y_1} = p_{X|y_2}, \quad \forall i, j \in \mathcal{E}_m, i \neq j, \quad \text{and} \quad p_{X|y_1} \neq p_{X|y_k}, \quad \forall i \in \mathcal{E}_m, \forall k \in [1 : |\mathcal{Y}|] \setminus \mathcal{E}_m, \forall m \in [1 : B].$$

Let $G \triangleq \sum_{i=1}^{B} |\mathcal{E}_i|$. For the $|\mathcal{X}| \times |\mathcal{Y}|$-dimensional matrix $P_{X|Y}$ that has the index sets $\mathcal{E}_m, \forall m \in [1 : B]$, we construct a corresponding $|\mathcal{X}| \times (|\mathcal{Y}| - G + B)$-dimensional matrix $\hat{P}_{X|Y}$ from $P_{X|Y}$ by eliminating all the columns in each $\mathcal{E}_m$, except one. For example, we have the following pair

$$P_{X|Y} = \begin{bmatrix} 0.3 & 0.3 & 0.4 & 0.5 & 0.4 \\ 0.2 & 0.2 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \hat{P}_{X|Y} = \begin{bmatrix} 0.3 & 0.4 & 0.5 \\ 0.2 & 0.5 & 0.5 \\ 0.5 & 0.1 & 0 \end{bmatrix}, \quad (41)$$

where $B = 2, G = 4, \mathcal{E}_1 = \{1, 2\}$, and $\mathcal{E}_2 = \{3, 5\}$. 
Theorem 3. For a pair of random variables $(X,Y) \in \mathcal{X} \times \mathcal{Y}$ ($|\mathcal{X}|, |\mathcal{Y}| < \infty$) distributed according to $p_{X,Y}$, we have
\[ g_0(X,Y) \geq D_X(Y), \] (42)
where the equality holds if and only if either of the following holds:
1) Perfect privacy is not feasible, i.e. $\dim(\text{Null}(P_{X|Y})) = 0$,
2) Perfect privacy is feasible, and $\dim(\text{Null}(P_{X|Y})) = 0$. In other words,
\[ |\mathcal{Y}| - |\mathcal{X}| \leq G - B, \text{ and } \sigma_i(P_{X|Y}) \neq 0, \forall i \in \left[1:|\mathcal{Y}| - G + B\right]. \]
Proof. It is obvious that when there exist no $y_1, y_2 \in \mathcal{Y}$ such that $p_{X|Y}(\cdot | y_1) = p_{X|Y}(\cdot | y_2)$, we have $D_X(Y) = 0$, and (42) holds from the non-negativity of $g_0$. Assume that there exist index sets $E_m, \forall m \in [1:B]$, corresponding to equal columns of $P_{X|Y}$, as defined before. Since $p_{X|y} = p_{X|y', \forall i, j \in E_m, \forall m \in [1:B]}$, we have
\[ T^X(y_i) = T^X(y_j), \forall i, j \in E_m, \forall m \in [1:B]. \]
Hence, $T^X(Y)$ is a random variable whose support has the cardinality $|\mathcal{Y}| - G + B$ and whose mass probabilities are the elements of the following set
\[ \left\{ \sum_{i \in E_1} p_Y(y_i), \sum_{i \in E_2} p_Y(y_i), \ldots, \sum_{i \in E_B} p_Y(y_i) \right\} \cup \left\{ p_Y(y_i) \left| i \notin \bigcup_{m=1}^B E_m \right. \right\}. \] (43)
Let $\mathcal{S}'$ be a set of $\prod_{i=1}^B |E_i|$ probability vectors on the simplex $\mathcal{P}(\mathcal{Y})$ given by
\[ \mathcal{S}' = \left\{ s_{m|[1:B]} \left| \forall m|[1:B] \in \prod_{i=1}^B E_i \right. \right\}, \] (44)
where the tuple $(m_1, m_2, \ldots, m_B)$ is written in short as $m|[1:B]$ and the probability vectors $s_{m|[1:B]}$ are defined element-wise as
\[ s_{m|[1:B]}(k) = \begin{cases} \sum_{i \in E_k} p_Y(y_i) & k = m_i, i \in [1:B] \\ p_Y(y_k) & k \notin \bigcup_{m=1}^B E_k, \forall k \in [1:|\mathcal{Y}|], \forall m|[1:B] \in \prod_{i=1}^B E_i. \end{cases} \] (45)

Proposition 6. For the set $\mathcal{S}'$ in (44) and the set $\mathcal{S}$ in (10), we have
\[ \mathcal{S}' \subseteq \mathcal{S}, \text{ and } H(s) = H(T^X(Y)), \forall s \in \mathcal{S}'. \]
Furthermore, the probability vector $p_Y$ can be written as a convex combination of the points in $\mathcal{S}'$, i.e.
\[ p_Y = \sum_{m|[1:B] \in \prod_{i=1}^B E_i} \alpha_{m|[1:B]} s_{m|[1:B]}, \] (46)
where $\alpha_{m|[1:B]} \geq 0, \forall m|[1:B] \in \prod_{i=1}^B E_i$ and
\[ \sum_{m|[1:B] \in \prod_{i=1}^B E_i} \alpha_{m|[1:B]} = 1. \]
Proof. The proof is provided in Appendix D.
Finally, we can write

$$g_0(X, Y) = H(Y) - \min_{F_U(\cdot), \ p_Y|U \in \mathbb{S}, \ \forall u \in U: \ \int_U p_Y|u \ dF(u) = p_Y} H(Y|U)$$  \hspace{1cm} (47)

$$\geq H(Y) - \sum_{m_{1:B} \in \mathbb{P}^B} \alpha_{m_{1:B}} H(s_{m_{1:B}})$$  \hspace{1cm} (48)

$$= H(Y) - \sum_{m_{1:B} \in \mathbb{P}^B} \alpha_{m_{1:B}} H(T^X(Y))$$  \hspace{1cm} (49)

$$= H(Y) - H(T^X(Y))$$

$$= D_X(Y),$$  \hspace{1cm} (50)

where (47) is from (12); (48) is justified as follows. According to Proposition 6, \(S' \subseteq S\), and \(p_Y\) is preserved from (46). Hence, the vectors in \(S'\) belong to the constraint of the minimization in (47), and the inequality follows. (49) is from Proposition 6, and (50) is due to (40). This proves (42).

For the proof of the necessary and sufficient condition of equality in (42), first, we prove the second direction, i.e. the sufficient condition. If perfect privacy is not feasible, we have \(g_0(X, Y) = 0\), and the equality is immediate. Assume that perfect privacy is feasible and there exist index sets \(E_m, \ \forall m \in [1:B]\), corresponding to equal columns of \(P_{X|Y}\), as defined before.

**Proposition 7.** If \(\dim(\text{Null}(\hat{P}_{X|Y})) = 0\), the extreme points of the convex polytope \(S\), defined in (10), are the elements of \(S'\). If \(\dim(\text{Null}(\hat{P}_{X|Y})) \neq 0\), none of the elements in \(S'\) is an extreme point of \(S\).

**Proof.** The proof is provided in Appendix E. \(\square\)

Now, if \(\dim(\text{Null}(\hat{P}_{X|Y})) = 0\), from Proposition 7, we can say that for any vector \(s\) that is an extreme point of \(S\), we have \(H(s) = H(T^X(Y))\), which means, from Proposition 4, that

$$\min_{F_U(\cdot), \ p_Y|U \in \mathbb{S}, \ \forall u \in U: \ \int_U p_Y|u \ dF(u) = p_Y} H(Y|U) = H(T^X(Y)).$$

This is equivalent to \(g_0(X, Y) = D_X(Y)\), from (12) and (40).

For the first direction, i.e. the necessary condition of equality in (42), assume that \(g_0(X, Y) = D_X(Y)\). If perfect privacy is not feasible, there is nothing to prove, i.e. \(g_0(X, Y) = D_X(Y) = 0\). If perfect privacy is feasible, i.e. \(g_0(X, Y) > 0\), we must have \(D_X(Y) > 0\), which from [15], means that there must exist index sets \(E_m, \ \forall m \in [1:B]\) , corresponding to equal columns of \(P_{X|Y}\). We prove that \(\dim(\text{Null}(\hat{P}_{X|Y})) = 0\) by contradiction. Assume that

For example, assume that in the example in (41), we have \(p_Y = \begin{bmatrix} 0.1 & 0.2 & 0.15 & 0.25 & 0.3 \end{bmatrix}^T\). We can write \(p_Y = \frac{1}{2}s_{1,3} + \frac{1}{2}s_{1,5} + \frac{1}{2}s_{2,3} + \frac{3}{2}s_{2,5}\), where

\[
\begin{align*}
\mathbf{s}_{1,3} &= \begin{bmatrix} 0.3 \\ 0 \\ 0.25 \\ 0 \end{bmatrix}, \\
\mathbf{s}_{1,5} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \\
\mathbf{s}_{2,3} &= \begin{bmatrix} 0.3 \\ 0.25 \\ 0 \end{bmatrix}, \\
\mathbf{s}_{2,5} &= \begin{bmatrix} 0.3 \\ 0.25 \end{bmatrix}.
\end{align*}
\]
\( \dim(\text{Null}(\hat{P}_{X|Y})) \neq 0 \). From Proposition 7, we conclude that none of the elements in \( S' \) is an extreme point of \( S \). In other words, for any \( s \) in \( S' \), we can find the triplet \((s', s'', \beta)\), such that \( s = \beta s' + (1 - \beta) s'' \), where \( s', s'' \in S \) \((s' \neq s'')\) and \( \beta \in (0, 1) \). Therefore,

\[
H(T^X(Y)) = \sum_{m_{1:B} \in \prod_{i=1}^B \mathcal{E}_i} \alpha_{m_{1:B}} H(s_{m_{1:B}})
\]

\[
= \sum_{m_{1:B} \in \prod_{i=1}^B \mathcal{E}_i} \alpha_{m_{1:B}} H(\beta s_{m_{1:B}} + (1 - \beta) s_{m_{1:B}}')
\]

\[
> \sum_{m_{1:B} \in \prod_{i=1}^B \mathcal{E}_i} \beta s_{m_{1:B}} \alpha_{m_{1:B}} H(s_{m_{1:B}}') + \sum_{m_{1:B} \in \prod_{i=1}^B \mathcal{E}_i} (1 - \beta) s_{m_{1:B}} \alpha_{m_{1:B}} H(s_{m_{1:B}}')
\]

(51)

\[
\geq \min_{F_Y(\cdot), \mathbf{p}_Y \mid u \in \mathbb{S}, \forall u \in \mathcal{U}, \int_{\mathcal{U}} \mathbf{p}_Y | u \cdot dF(u) = p_Y} H(Y | U),
\]

(52)

where (51) is due to the strict concavity of the entropy; (52) comes from the fact that \( s_{m_{1:B}}' \) and \( s_{m_{1:B}}'' \) with corresponding mass probabilities \( \beta s_{m_{1:B}} \alpha_{m_{1:B}} \) and \( (1 - \beta) s_{m_{1:B}} \alpha_{m_{1:B}} \), \( \forall m_{1:B} \in \prod_{i=1}^B \mathcal{E}_i \), belong to the constraints of minimization in (52). This results in \( g_0(X, Y) > D_X(Y) \), which is a contradiction. Hence, we must have \( \dim(\text{Null}(\hat{P}_{X|Y})) = 0 \).

\[\Box\]

V. FULL DATA OBSERVATION VS. OUTPUT PERTURBATION

Thus far, we have assumed that the privacy mapping/data release mechanism takes \( Y \) as input and maps it to the released data denoted by \( U \), where \( X - Y - U \) form a Markov chain and the privacy mapping is captured by the conditional distribution \( p_U(Y) \). In a more general scenario, the privacy mapping can take a noisy version \( W \) of \( (X, Y) \) as input, as in [22]. In this case, the privacy mapping is denoted by \( p_U(W) \), and \( (X, Y) \rightarrow W \rightarrow U \) form a Markov chain, where the triplet \((X, Y, W) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{W} \) \((X, Y, W < \infty)\) is distributed according to some given joint distribution \( p_{X,Y,W} \). In this model, perfect privacy is feasible for the triplet \((X, Y, W) \) if there exists a privacy mapping \( p_U(W) \) whose output \( U \) depends on the useful data \( Y \), while being independent of the private data \( X \); that is, \( I(Y; U) > 0 \) and \( I(X; U) = 0 \) as before.

**Proposition 8.** Perfect privacy is feasible for \((X, Y, W) \) if and only if

\[
\dim \left( \text{Null}(P_{X|W}) \setminus \text{Null}(P_{Y|W}) \right) \neq 0.
\]

(53)

**Proof.** The proof follows similarly to that of Proposition 2 by noting that both \( X - (X, Y) - W - U \) and \( Y - (X, Y) - W - U \) form Markov chains. In other words, there must exist a vector in \( \mathcal{P}(\mathcal{W}) \), such that a change in \( p_W \) along that vector changes \( p_Y \), while keeps \( p_X \) unchanged. \( \Box \)

It can be verified that for the general scenario of \((X, Y) - W - U \), where the mapping is denoted by \( p_U(W) \), perfect privacy can be obtained through a similar LP as in Theorem 1 with the following modifications: The convex
polytope \( S \) is modified as the set of probability vectors \( x \) in \( P(W) \), such that \( (p_W - x) \in \text{Null}(P_{X|W}) \); denoting

the the extreme points of \( S \) by \( p_1, p_2, \ldots, p_K \), (13) changes to

\[
\max_{p_U: (X,Y) - W - U} I(Y;U) = H(Y) - \min_{w \geq 0} \left[ H(P_{Y|W} p_1) - H(P_{Y|W} p_2) - \ldots - H(P_{Y|W} p_K) \right] \cdot w
\]

s.t. \[ p_1 p_2 \ldots p_K w = p_W. \]

The special cases of full data observation and output perturbation (22) refer to scenarios where the privacy mapping has direct access to both the private and useful data \((W = (X,Y))\) and only the useful data \((W = Y)\), respectively. With these definitions, Sections II to IV consider the particular case of output perturbation. In what follows, we consider the full data observation scenario briefly.

**Proposition 9.** If \( Y \) is not a deterministic function of \( X \), perfect privacy is always feasible in the full data observation model.

**Proof.** If \( Y \) is not a deterministic function of \( X \), there must exist \( x_1 \in X \) and \( y_1, y_2 \in Y (y_1 \neq y_2) \) such that \( p_{X,Y}(x_1, y_1) > 0 \) and \( p_{X,Y}(x_1, y_2) > 0 \). Let \( U = \{u_1, u_2\} \) and \( p_U(u_1) = \frac{1}{2} \). Choose a sufficiently small \( \epsilon > 0 \) and let

\[
p_{X,Y|U}(x,y|u_1) = \begin{cases} 
 p_{X,Y}(x_1, y_1) + \epsilon & (x,y) = (x_1, y_1) \\
 p_{X,Y}(x_1, y_2) - \epsilon & (x,y) = (x_1, y_2) \\
 p_{X,Y}(x,y) & \text{otherwise}
\end{cases}
\]

\( p_{X,Y|U}(x,y|u_2) = 2p_{X,Y}(x,y) - p_{X,Y|U}(x,y|u_1), \forall (x,y) \in X \times Y. \) (54)

It can be verified that \( p_{X,Y} \) is preserved in \( p_{X,Y|U} \). Also, \( p_{X|U} (\cdot | u) = p_X (\cdot), \forall u \in U \), and \( p_{Y|U} (y_1 | u_1) \neq p_Y(y_1) \), where the former indicates that \( X \perp \perp U \), and the latter shows that \( Y \not\perp \perp U \).

Considering the output perturbation model, Theorem 2 proved that perfect privacy is not feasible for the (correlated) jointly Gaussian pair. The following theorem states the opposite for the full data model.

**Theorem 5.** For a jointly Gaussian pair \((X,Y)\) with correlation coefficient \( \rho (\neq 0) \), perfect privacy is feasible for the full data observation model, and we have

\[
\sup_{P_{U|X,Y}: X \perp \perp U} I(Y;U) \geq -\log \rho.
\]

**Proof.** Denoting the variances of \( X \) and \( Y \) by \( \sigma_X^2 \) and \( \sigma_Y^2 \), respectively, it is already known that we can write

\[
Y = \frac{\rho \sigma_Y}{\sigma_X} X + \sigma_Y \sqrt{1 - \rho^2} N,
\]

(56)
where \( N \sim \mathcal{N}(0, 1) \) is independent of \( X \). By letting 
\[
U = \frac{1}{\sigma_Y \sqrt{1 - \rho^2}} (Y - \frac{\rho \sigma_Y}{\sigma_X} X),
\]
i.e., \( U = N \), we have \( X \perp U \), and

\[
I(Y; U) = h(Y) - h(Y|U)
= \frac{1}{2} \log_2 2\pi e \sigma_Y^2 - h\left( \frac{\rho \sigma_Y}{\sigma_X} X + \sigma_Y \sqrt{1 - \rho^2} N \right)
= \frac{1}{2} \log_2 2\pi e \sigma_Y^2 - h\left( \frac{\rho \sigma_Y}{\sigma_X} X \right)
= \frac{1}{2} \log_2 2\pi e \sigma_Y^2 - \frac{1}{2} \log_2 2\pi e \rho^2 \sigma_Y^2
= -\log_2 \rho,
\]
where (56) is used in (57), and (58) follows from the fact that \( X \perp N \).

\section*{VI. Maximal correlation}

Consider a pair of random variables \((X, Y) \in \mathcal{X} \times \mathcal{Y}\) distributed according to \( p_{X,Y} \), with \(|\mathcal{X}|, |\mathcal{Y}| < \infty \). Let \( \tilde{\mathcal{F}} \) denote the set of all real-valued functions of \( X \), and define

\[
\mathcal{F} \triangleq \{ f(\cdot) \in \tilde{\mathcal{F}} \mid E[f(X)] = 0, \ E[f^2(X)] = 1 \}.
\]

Let \( \tilde{\mathcal{G}} \) and \( \mathcal{G} \) be defined similarly for the random variable \( Y \). The maximal correlation of \((X, Y)\) is defined as ([16], [17], [18]):

\[
\rho_m(X; Y) = \max_{f \in \mathcal{F}, g \in \mathcal{G}} E[f(X)g(Y)].
\]

If \( \mathcal{F} \) (and/or \( \mathcal{G} \)) is empty\(^4\), then \( \rho_m \) is defined to be zero.

An alternative characterization of the maximal correlation is given by Witsenhausen in [23] as follows. Let the matrix \( Q \) be defined as

\[
Q \triangleq P_X^{-\frac{1}{2}} P_{X,Y} P_Y^{-\frac{1}{2}},
\]
with singular values \( \sigma_1 \geq \sigma_2 \geq \cdots \). It is shown in [23] that \( \sigma_1 = 1 \), and the maximal correlation of \((X, Y)\), i.e., \( \rho_m(X; Y) \), is equal to the second largest singular value of matrix \( Q \), i.e., \( \sigma_2 \).

In what follows, we propose an alternative characterization of the maximal correlation, which also helps us interprete the other singular values of matrix \( Q \). The following preliminaries from [25] are needed in the sequel.

\section*{A. Preliminaries}

Assume that \( \mathbf{R} \) is an \( n \)-by-\( n \) real symmetric matrix, and \( \mathbf{c} \) is an \( n \)-dimensional vector satisfying \( \| \mathbf{c} \|_2 = 1 \). Assume that we are interested in finding the stationary values of

\[
x^T \mathbf{R} x,
\]

\(^4\)When \( X \) (and/or \( Y \)) is constant almost surely.

\(^5\)For other characterizations, see [24].
subject to the constraints
\[ c^T x = 0, \]
\[ \|x\|_2 = 1. \] (62)

Letting \( \lambda \) and \( \mu \) be the Lagrange multipliers, we have
\[ L(x, \lambda, \mu) = x^T Rx - \lambda (x^T x - 1) + 2\mu x^T c. \]

Differentiating the Lagrangian with respect to \( x \), we obtain
\[ Rx - \lambda x + \mu c = 0, \] (63)
which results in \( \mu = -c^T Rx \), after multiplying both sides by \( c^T \) and noting that \( \|c\|_2 = 1 \). By substituting this value of \( \mu \) in (63), we are led to
\[ PRx = \lambda x, \]
where \( P = I - cc^T \). Since \( P \) is a projection matrix, i.e. \( P^2 = P \), the stationary values of \( x^T Rx \) are the singular values of the matrix \( PR \) that occur at the corresponding eigenvectors.

Finally, assume that the vector \( c \) in the constraints is replaced with an \( n \times r \) matrix \( C \) with \( r \leq n \). Also, assume that the columns of matrix \( C \) are orthonormal. It can be verified that the results remain the same after having \( P \) modified as \( P = I - CC^T \).

**B. Alternative characterization of \( \rho_m(X;Y) \)**

Consider a pair of random variables \( (X, Y) \in \mathcal{X} \times \mathcal{Y} \) \( (|\mathcal{X}|, |\mathcal{Y}| < \infty) \) distributed according to \( P_{X,Y} \), with the marginals \( p_X \) and \( p_Y \). The matrix \( P_{X|Y} \) can be viewed as a channel with input \( Y \) and output \( X \). When the input of this channel is distributed according to \( q_Y \), the output is distributed according to \( q_X = P_{X|Y} q_Y \).

Let \( r : \mathcal{P}(\mathcal{Y}) \backslash \{p_Y\} \to [0, 1] \) be defined as
\[ r(q_Y) \triangleq \frac{D(q_X || p_X)}{D(q_Y || p_Y)}, \] (64)

We write \( q_Y \to p_Y \) when \( \|q_Y - p_Y\|_2 \to 0 \) and \( q_Y \neq p_Y \). We are interested in finding the stationary values of \( r(q_Y) \) when \( q_Y \to p_Y \).

**Theorem 6.** The stationary values of (64), when \( q_Y \to p_Y \), are the squared singular values of matrix \( Q \triangleq P_X^{-\frac{1}{2}} P_{X|Y} P_Y^{-\frac{1}{2}} \), and in particular,
\[ \rho_m^2(X;Y) = \lim_{\eta \to 0} \sup_{\eta < \|q_Y - p_Y\|_2 \leq \eta} r(q_Y). \]

**Proof.** Having \( q_Y \to p_Y \), we can write
\[ q_Y = p_Y + \epsilon, \quad 1_{[|\mathcal{Y}|]} \cdot \epsilon = 0, \quad \|\epsilon\|_2 \to 0, \epsilon \neq 0, \]
where \( \epsilon \) is an auxiliary vector. From the relationship \( q_X = P_{X|Y} q_Y \), we have
\[ r(q_Y) = \frac{D(p_X + P_{X|Y} \epsilon || p_X)}{D(p_Y + \epsilon || p_Y)}. \] (65)
Assume that \( p_0 \) and \( p \) are two probability vectors in the interior of \( \mathcal{P}(\mathcal{Y}) \). Let \( p_0(\cdot) \) and \( p(\cdot) \) denote their corresponding probability mass functions. We can write the Taylor series expansion of the relative entropy as

\[
D(p_0 + \epsilon || p) = D(p_0 || p) + \epsilon^T \nabla D |_{p_0} + \frac{1}{2} \epsilon^T \nabla^2 D |_{p_0} \epsilon + \cdots,
\]

where

\[
\nabla D |_{p_0} = \begin{bmatrix} \log \frac{p_0(y_1)}{p(y_1)} + 1 & \log \frac{p_0(y_2)}{p(y_2)} + 1 & \cdots & \log \frac{p_0(y_1)}{p(y_1)} + 1 \end{bmatrix}^T,
\]

\[
\nabla^2 D |_{p_0} = \text{diag} \left( \frac{1}{p_0(y_1)} \frac{1}{p_0(y_2)} \cdots \frac{1}{p_0(y_1)} \right),
\]

are the gradient and the Hessian of \( D(\cdot || p) \) at \( p_0 \), respectively, and the higher order terms of \( \epsilon \) are denoted by dots in (66). Therefore, (65) boils down to

\[
r(q_Y) = \frac{D(p_X || p_X) + \epsilon^T p_X^T Y Y 1_{|X|} + \epsilon^T P X^T Y Y P X^T Y Y \epsilon + \cdots}{D(p_Y || p_Y) + \epsilon^T 1_{|Y|} + \epsilon^T P Y^{-1} \epsilon + \cdots} = \frac{\epsilon^T P X^T Y Y \epsilon + \cdots}{\epsilon^T P Y^{-1} \epsilon + \cdots},
\]

where we have used the facts that \( D(p || p) = 0 \), \( P X^T Y Y 1_{|X|} = 1_{|Y|} \) and \( \epsilon^T 1_{|Y|} = 0 \). When \( \|\epsilon\|_2 \to 0 \), the higher order terms of \( \epsilon \) in (67), shown with dots, can be ignored. Hence, we are interested in finding the stationary values of

\[
\frac{\epsilon^T P X^T Y Y \epsilon}{\epsilon^T P Y^{-1} \epsilon},
\]

when \( 1_{|Y|} \cdot \epsilon = 0 \), \( \epsilon \neq 0 \). Note that the condition \( \|\epsilon\|_2 \to 0 \) is redundant as the norm \( \|\epsilon\|_2 \) cancels out from both the numerator and the denominator of (68). We can equivalently write (68) as

\[
\frac{v^T \epsilon^T P X^T Y Y \epsilon}{v^T \epsilon^T P Y^{-1} \epsilon},
\]

where \( v = P Y^{-\frac{1}{2}} \epsilon \), \( v \neq 0 \), \( c^T v = 0 \) with \( c = P Y^{\frac{1}{2}} 1_{|Y|} \), and it is obvious that \( \|c\|_2 = 1 \). Without loss of generality, we assume that \( \|v\|_2 = 1 \). Therefore, we are led to finding the stationary values of

\[
v^T R v,
\]

where \( R = P Y^{\frac{1}{2}} P X^T Y Y P^{-1} X^T Y P X^T Y Y P Y^{\frac{1}{2}} \), subject to the constraints

\[
\epsilon^T \cdot v = 0,
\]

\[
\|v\|_2 = 1.
\]

Note that \( R \) is a \(|Y|\)-by-\(|Y|\) real symmetric matrix, and \( c \) is a \(|Y|\)-dimensional vector satisfying \( \|c\|_2 = 1 \). Therefore, (69) is the same problem as in (61) whose stationary values are the eigenvalues of the matrix \((I - cc^T)R\), which occur at their corresponding eigenvectors.

We have

\[
R = P Y^{\frac{1}{2}} P X^T Y Y P^{-1} X^T Y P Y^{\frac{1}{2}} = \left( P X^{\frac{1}{2}} P X^T Y Y P Y^{\frac{1}{2}} \right)^T \left( P X^{\frac{1}{2}} P X^T Y Y P Y^{\frac{1}{2}} \right) = Q^T Q,
\]
where \( Q \) is defined in (60). Also, \( c \) is the eigenvector of \( R \) corresponding to the eigenvalue of 1, which follows from:

\[
Rc = P^T_{\chi}P_{\chi y}^{-1}P_{\chi y}^{-1}P_{\chi |y}P^T_{\chi}c = P^T_{\chi}P_{\chi y}^{-1}P_{\chi y}p_y = P^T_{\chi}P_{\chi y} P_{\chi y}|y|
\]

Therefore, the eigenvalues of the matrix \((I - cc^T)R\) are \( \lambda_1 = \sigma^2 \geq \lambda_2 = \sigma^2 \geq \cdots \) and 0, where \( \sigma_i \)'s are the singular values of matrix \( Q \), and hence, \( \lambda_1 = \rho^2_{m} \). This leads us to the following equality for the maximal correlation

\[
\rho^2_{m}(X, Y) = \sup_{q_Y \rightarrow p_Y} \frac{D(q_X || p_X)}{D(q_Y || p_Y)}.
\]

The other eigenvalues of \((I - cc^T)R\) (or equivalently, the other singular values of matrix \( Q \), except the largest one) can be interpreted in a similar way. Assume that \( v_1 \) is the maximizer of (69), i.e., \( v_1 \) is the eigenvector of \((I - cc^T)R\) that corresponds to the eigenvalue \( \lambda_1 = \rho^2_{m} \). Equivalently, when \( q_Y \rightarrow p_Y \) (\( q_Y \neq p_Y \)), the ratio in (64) is maximized if \( q_Y \) converges to \( p_Y \) in the direction of \( \epsilon_1 = P^T_{\chi}v_2 \). If besides the constraints in (70), we also impose the constraint that \( v \) should be orthogonal to \( v_1 \), i.e., replacing \( c \) by matrix \( \hat{c} \) whose first and second columns are, respectively, \( c \) and \( v_1 \), the maximum of (69) would be \( \lambda_2 = \sigma^2_{m} \), achieved by its corresponding eigenvector \( v_2 \). Equivalently, when \( q_Y \rightarrow p_Y \) (\( q_Y \neq p_Y \)) and \((q_Y - p_Y) \perp P^T_{\chi}v_1 \), the ratio in (64) is maximized if \( q_Y \) converges to \( p_Y \) in the direction of \( \epsilon_2 = P^T_{\hat{c}}v_2 \). This procedure can be continued to cover all the singular values of matrix \( Q \), from the second largest to the smallest.

**Remark 2.** A natural question arises whether the largest singular value, which is one, has a similar interpretation. In (70), the constraint \( c^T \cdot v = 0 \) is omitted, the maximum of (69) would be 1, which occurs at \( v = c \). The constraint \( c^T \cdot v = 0 \) is due to \( 1_{|y|} \cdot e = 0 \), which in turn results from the fact that \( q_Y \) is a probability vector. Therefore, if the definition of relative entropy is extended to the vectors with positive elements, when \( q_Y \rightarrow p_Y \) (\( q_Y \neq p_Y \)) and \( q_Y \) can be any vector with positive elements, the ratio in (64) is maximized if \( q_Y \) converges to \( p_Y \) in the direction of \( \epsilon_0 = P^T_{\hat{c}}c = p_Y \).

In Sections III and IV of this paper, the problem of perfect privacy, i.e., the quantity \( g_0(X, Y) \), is studied. Although the optimal utility-privacy trade-off curve is not straightforward analytically, it is of interest to check the behaviour of this curve when a small amount of private data leakage is allowed. To this end, we study the slope of this curve in the following section.
VII. THE SLOPE OF THE UTILITY-PRIVACY TRADE-OFF REGION

In this section, we consider the trade-off region \( g_0(X, Y) \) vs. \( \epsilon \) as defined in (1). We are interested in evaluating the rate of increase in the utility when \( I(X; U) = 0 \). In other words, we focus on the slope of this trade-off curve at \((0,0)\). For the case when \( g_0 = 0 \), this slope is obtained, while for the case \( g_0 > 0 \), a lower bound is proposed.

Let \( V^* \in [1, +\infty) \) be defined as

\[
V^* \triangleq \sup_{q_Y \neq P_Y} \frac{D(q_Y \| P_Y)}{D(q_X \| P_X)},
\]

with the convention that if for some \( q_Y(\neq P_Y) \), we have \( q_X = P_X \), then \( V^* = +\infty \).

Proposition 10. We have \( g_0(X; Y) = 0 \) if and only if \( V^* < +\infty \).

\[ \exists q_Y^* \neq P_Y : V^* - \delta < \frac{D(q_Y^* \| P_Y)}{D(q_X^* \| P_X)} \leq V^*, \]

where \( q_X^* \) is induced by \( q_Y^* \), i.e. \( q_X^* = P_{X|Y} q_Y^* \). Let \( U = \{u_1, u_2\} \). For sufficiently small \( \zeta > 0 \), let

\[
p_{U}(u_1) = \zeta, \quad p_{Y|u_1} = q_Y^*, \quad p_{Y|u_2} = \frac{1}{1-\zeta}(p_Y - \zeta q_Y^*),
\]

where this sufficiently small \( \zeta \) makes \( p_{Y|u_2} \) a probability vector. We have

\[
\frac{I(Y; U)}{I(X; U)} = \frac{\sum_{u\in\mathcal{U}} p_{U}(u)D(p_{Y|u} \| P_Y)}{\sum_{u\in\mathcal{U}} p_{U}(u)D(p_{X|u} \| P_X)} = \frac{(1 - \zeta)\frac{1}{1-\zeta}(p_Y - \zeta q_Y^*) \| P_Y) + \zeta D(q_Y^* \| P_Y)}{(1 - \zeta)\frac{1}{1-\zeta}(p_X - \zeta q_X^*) \| P_X) + \zeta D(q_X^* \| P_X)}. \tag{73}
\]

We are interested in inspecting the behaviour of \( (73) \), when \( \zeta \to 0 \). From the Taylor series expansion, we have

\[
D\left(\frac{1}{1-\zeta}(p_Y - \zeta q_Y^*) \| P_Y\right) = D\left(p_Y + \frac{\zeta}{1-\zeta}(p_Y - q_Y^*) \| P_Y\right) = 0 + \frac{\zeta^2}{2(1-\zeta)} (p_Y - q_Y^*)^T P_Y^{-1}(p_Y - q_Y^*) + \ldots, \tag{74}
\]
where $K_y$ is a constant and the dots denote the higher order terms in the Taylor series expansion. A similar expansion can be written for the first term in the denominator of (73) with the corresponding constant denoted by $K_x$. Hence, we have

$$
\lim_{\epsilon \to 0} \frac{I(Y;U)}{I(X;U)} = \lim_{\epsilon \to 0} \frac{\epsilon^2 K_y + \zeta D(q_Y^* || p_Y)}{\epsilon^2 K_x + \zeta D(q_X^* || p_X)} = \frac{D(q_Y^* || p_Y)}{D(q_X^* || p_X)} > V^* - \delta.
$$

(75)

(76)

(77)

Since $\delta$ is chosen arbitrarily, we can write

$$
\lim_{\epsilon \to 0} \frac{q_e(X;Y)}{\epsilon} \geq V^*.
$$

(78)

On the other hand,

$$
\lim_{\epsilon \to 0} \frac{q_e(X;Y)}{\epsilon} \leq \sup_{U: X \rightarrow Y - U \mid I(X;U) > 0} \frac{I(Y;U)}{I(X;U)} = \sup_{U: X \rightarrow Y - U \mid I(X;U) > 0} \frac{\sum_{u \in U} p_U(u) D(p_Y^* || p_Y)}{\sum_{u \in U} p_U(u) D(p_X^* || p_X)} \leq \sup_{q_Y \neq p_Y} \frac{D(q_Y^* || p_Y)}{D(q_X^* || p_X)} = V^*.
$$

(79)

(80)

where the assumption of $p_{Y|U} \neq p_Y$ in the summations of (79) causes no loss of generality, as it only excludes the zero terms. From (78) and (80), (72) is proved.

\[\square\]

**Remark 3.** In the general observation model of Section V, i.e., $(X, Y) - W - U$, where the privacy-preserving mapping is $p_U|W$, it can be similarly verified that the slope at $(0, 0)$ of the optimal utility-privacy trade-off curve, characterized by $I(Y; U)$ vs. $I(X; U)$, is given by

$$
\sup_{q_W \notin \text{Null}(p_{Y|W})} \frac{D(p_Y|W q_W || p_Y)}{D(p_X|W q_W || p_X)}.
$$

(81)

where the same convention of the definition in (71) holds; the probability vectors $p_W$, $p_X$, $p_Y$, and the matrices $P_{Y|W}$, $P_{X|W}$ are all derived from the joint distribution $p_{X,Y,W}$.

**Remark 4.** Let $X \sim \text{Bernoulli}(p_x)$ ($p_x \in (0, \frac{1}{2})$) and $Y$ is connected to $X$ through a binary symmetric channel (BSC) with crossover probability $\alpha \in (0, \frac{1}{2}]$. Hence, $Y$ is distributed as $Y \sim \text{Bernoulli}(p_y)$ with $p_y = p_x * \alpha$, where $p * q = p(1 - q) + q(1 - p)$ ($p, q \in [0, 1]$) denotes the binary convolution. From [15, Lemma 1],

$$
g_e(X, Y) \leq H(Y) - H_b(\alpha * H_b^{-1}(H(X) - \epsilon)),
$$

(82)
where $H_b(t) = -t \log t - (1 - t) \log(1 - t)$ is the binary entropy function, and $H_b^{-1} : [0, 1] \rightarrow [0, \frac{1}{2}]$ is its inverse. From (82), we have

$$\lim_{\epsilon \to 0} \frac{g_\epsilon(X; Y)}{\epsilon} \leq \lim_{\epsilon \to 0} \frac{H(Y) - H_b(\alpha * H_b^{-1}(H(X) - \epsilon))}{\epsilon}$$

$$= \left. \frac{d}{d\epsilon} \left( H(Y) - H_b(\alpha * H_b^{-1}(H(X) - \epsilon)) \right) \right|_{\epsilon = 0}$$

$$= (1 - 2\alpha) \frac{\log \frac{1 - p_x}{p_x}}{\log \frac{1 - p_x}{p_x}}, \quad (84)$$

where (83) is from the application of L'Hospital's rule. According to [15], perfect privacy is not feasible for this $(X, Y)$, and we have $g_0(X, Y) = 0$. Hence,

$$\lim_{\epsilon \to 0} \frac{g_\epsilon(X; Y)}{\epsilon} = \sup_{q_Y \neq p_Y} \frac{D(q_Y \| p_Y)}{D(q_X \| p_X)}$$

$$= 1$$

$$\geq \inf_{q_Y \neq p_Y} \frac{D(q_Y \| p_Y)}{D(q_X \| p_Y)}$$

$$\geq \frac{1}{\rho_m}, \quad (85)$$

where (85) is from Proposition 11; (86) is permissible, since the ratios involved are bounded away from zero and infinity; (87) is a direct result of adding constraint to the infimum, and (88) is from the analysis after (64). Note that, in the specific case of this example where $|Y| = 2$, the inequality in (88) can be replaced by equality, as the standard 1-simplex has only one dimension. In other words, when $q_Y \rightarrow p_Y$, the infimum and the supremum of the ratio are the same, i.e., $\frac{1}{\rho_m}$.

From (88) and (84), we must have

$$\frac{1}{\rho_m} \leq (1 - 2\alpha) \frac{\log \frac{1 - p_x}{p_x}}{\log \frac{1 - p_x}{p_x}}. \quad (89)$$

By a simple calculation of $\rho_m$, which is the second largest singular value of the matrix $P_X^{-\frac{1}{2}} P_{X,Y} P_Y^{-\frac{1}{2}}$, we get

$$\rho_m^2 = \alpha^2 + (1 - 2\alpha)(p_x + p_y - 2p_x p_y) - 1. \quad (90)$$

It is obvious that for a fixed $p_x$, when $\alpha \rightarrow \frac{1}{2}$; we have $p_y \rightarrow \frac{1}{2}$, and therefore, from (90), $\rho_m \rightarrow 0$. This is intuitive, since having $\alpha \rightarrow \frac{1}{2}$, the pair $(X, Y)$ moves towards independence; and therefore, any correlation between them vanishes. As a result, the left hand side (LHS) of (89) becomes unbounded, while its right hand side (RHS) tends to zero. For example, with $p_x = 0.6$ and $\alpha = 0.45$, the LHS is approximately 104.12, while its RHS is 0.0099, which makes (89) invalid. The reason for this phenomenon is that the upperbound in (82) does not hold in general, which is due to a subtle error in employing Mrs. Gerber's lemma in [15, Lemma 1], in which, the conditional entropy $H(Y|U)$ is bounded below as

$$H(Y|U) = H(X \oplus N|U)$$

$$\geq H_b(\alpha * H_b^{-1}(H(X|U))), \quad (91)$$
where the crossover probability $\alpha$ is captured in the additive binary noise $N \sim \text{Bernoulli}(\alpha)$, and (91) is due to Mrs. Gerber’s lemma. Then, $H(X|U)$ is replaced with $H(X) - \epsilon$, since $I(X;U) = \epsilon$, to obtain the bound in (82). However, in the statement of Mrs. Gerber’s lemma [26], $N$ must be independent of the pair $(X,U)$, while this is not necessarily the case here. Assume that, in the Markov chain $X - Y - U$, $U$ is obtained by passing $Y$ through a $Z$ channel. Then, $N$ is not independent of the pair $(X,U)$. Actually, it can be verified that for one realization of $U$, $N$ becomes a deterministic function of $X$. Therefore, the application of Mrs. Gerber’s lemma is not permissible.

**Remark 5.** It can be readily verified from Theorem 6 and Proposition 11 that

$$\frac{1}{V^*} = \lim_{\delta \to 0} \sup_{U:X \to Y-U} \frac{I(X;U)}{I(Y;U)},$$

(92)

$$\rho_m^2(X;Y) = \lim_{\delta \to 0} \sup_{U:X \to Y-U} \frac{I(X;U)}{I(Y;U)} = 0,$$

(93)

where in (92), we use the convention $\frac{1}{0} = 0$; in (92), the term $\mathbb{E}_U[D(P_{X|U}(\cdot|U)||P_X(\cdot))]$ is equal to $I(X;U)$, and it is written like this to be comparable with the constraint of supremization in (93), i.e., an average constraint in (92) versus a per-realization constraint in (93).

**B. The slope at $(0, g_0)$**

In the previous subsection this slope was obtained when $g_0 = 0$. In what follows, we consider the case $g_0 > 0$, and we are interested in finding

$$\lim_{\epsilon \to 0} \frac{g_0(X;Y) - g_0}{\epsilon}.$$

(94)

In the sequel, we propose a lower bound for (94). Assume that $g_0$, as obtained through an LP formulation in Theorem 1, is achieved by

$$U' \in U' = \{ u'_1, u'_2, \ldots, u'_{|U'|}\}, \quad p_{Y|u'}, \forall u' \in U',$$

where the vectors $p_{Y|u'}, \forall u' \in U'$ belong to the extreme points of the set $\mathbb{S}$, defined in (10). Define

$$\psi(u') \triangleq \sup_{q_Y: 0 < D(q_Y||p_{Y|u'}) < +\infty} \frac{D(q_Y||p_{Y|u'})}{D(q_Y||p_X)}, \quad \forall u' \in U',$$

(95)

and if for some $u'$, there is no $q_Y$ for which $0 < D(q_Y||p_{Y|u'}) < +\infty$ (which happens exactly when $p_{Y|u'}$ is a corner point of the probability simplex), then let $\psi(u') = 0$.

**Proposition 12.** We have $\psi(u') < +\infty, \forall u' \in U'$.

**Proof.** The proof is provided in Appendix H.

Let

$$L \triangleq \max_{u' \in U'} \psi(u'),$$

(96)

6The Z channel has binary input and output alphabets, and conditional pmf $p(0|0) = 1, p(1|1) = p(0|1) = \frac{1}{2}$. 

and denote a/the maximizer of (96) by \( u'_j \) for some \( j \in [1 : |\mathcal{U}'|] \). From (95) and (96), for a fixed \( \delta \), we have

\[
\exists q_Y \neq p_{Y|u'_j}, \quad D(q_Y \| p_{Y|u'_j}) < +\infty: \quad L - \delta < \frac{D(q_Y \| p_{Y|u'_j})}{D(p_X \| p_X)} \leq L
\]

(97)

Construct the pair \((Y, U)\) as follows. Let \( \mathcal{U} = \{u_1, u_2, \ldots, u_{|\mathcal{U}'|}, \hat{u}_j\} \), and

\[
p_U(u_i) = p_U(U'_{i}), \quad p_{Y|u_i} = p_{Y|u'_i}, \quad \forall i \in [1 : |\mathcal{U}'|], i \neq j
\]

(98)

\[
p_U(u_j) = \epsilon p_{U'}(u'_j), \quad p_U(\hat{u}_j) = (1 - \epsilon)p_{U'}(u'_j), \quad p_{Y|u_j} = q_Y, \quad p_{Y|\hat{u}_j} = \frac{1}{1 - \epsilon}(p_{Y|u'_j} - \epsilon q_Y).
\]

(99)

Note that for sufficiently small \( \epsilon > 0 \), \( p_{Y|\hat{u}_j} \) is a probability vector, since we have \( D(q_Y \| p_{Y|u'_j}) < +\infty \). In other words, for any entry of the vector \( p_{Y|u'_j} \) that is zero (note that it is an extreme point of \( \mathcal{S} \)), the corresponding entry in \( q_Y \) is also zero. Finally, it can be verified that from (98) and (99), the marginal probability vector \( p_Y \) is also preserved.

With this construction, we can have the Markov chain \( X - Y - U \) and

\[
\frac{I(Y; U) - g_0}{I(X; U)} = \frac{\sum_{u \in \mathcal{U}} p_U(u)D(p_{Y|u} \| p_Y) - \sum_{q \in \mathcal{Q}} qD(p_{Y|q} \| p_Y)}{\sum_{u \in \mathcal{U}} p_U(u)D(p_X \| p_X)}
\]

\[
= \sum_{u \in \{(u_i, \hat{u}_j)\}} p_U(u)D(p_{Y|u} \| p_Y) - p_U(u')D(p_{Y|u'} \| p_Y)
\]

\[
= \epsilon D(q_Y \| p_Y) + (1 - \epsilon)D\left(\frac{1}{1 - \epsilon}(p_{Y|u'_j} - \epsilon q_Y) \biggm\| p_Y\right) - D(p_{Y|u'_j} \| p_Y)
\]

(100)

\[
= \epsilon D(q_Y \| p_Y) + (1 - \epsilon)D\left(\frac{1}{1 - \epsilon}(p_{Y|u'_j} - \epsilon q_Y) \biggm\| p_Y\right) - D(p_{Y|u'_j} \| p_Y)
\]

(101)

where the numerator in (100) is from (98), the denominator in (100) is from the fact that \( p_{X|u_i} = p_{X|u'_i}, \forall i \in [1 : |\mathcal{U}'|], i \neq j \) and \( p_{X|u'} = p_{X|Y}p_{Y|u'} = p_X, \forall u' \in \mathcal{U}' \); (101) is due to (99). We can write the Taylor series expansion for the second term in the numerator of (101) as

\[
D\left(\frac{1}{1 - \epsilon}(p_{Y|u'_j} - \epsilon q_Y) \biggm\| p_Y\right) = D(p_{Y|u'_j} \| p_Y) + \frac{\epsilon}{1 - \epsilon}(p_{Y|u'_j} - q_Y) + \frac{\epsilon^2}{2(1 - \epsilon)^2}(p_{Y|u'_j} - q_Y)^T \mathbf{P}_{Y|u'_j}(p_{Y|u'_j} - q_Y) + \ldots
\]

where \( \mathbf{P}_{Y|u'_j} = \text{diag}(p_{Y|u'_j}) \). Replacing the above in (101), the numerator becomes

\[
\epsilon D(q_Y \| p_Y) + (1 - \epsilon)D(p_{Y|u'_j} \| p_Y) + \epsilon(p_{Y|u'_j} - q_Y)^T \mathbf{P}_{Y|u'_j}(p_{Y|u'_j} - q_Y)
\]

\[
+ \frac{\epsilon^2}{2(1 - \epsilon)^2}(p_{Y|u'_j} - q_Y)^T \mathbf{P}_{Y|u'_j}(p_{Y|u'_j} - q_Y) + \ldots
\]

which, after some manipulations, becomes equal to

\[
\epsilon D(q_Y \| p_Y) + \frac{\epsilon^2}{2(1 - \epsilon)} K_y + \ldots
\]
Following similar steps, an expansion for the second term in the denominator of (101) can be obtained. Letting $\epsilon \to 0$, and ignoring the higher order terms (denoted by the dots), we get

$$
\frac{I(Y; U) - g_0}{I(X; U)} = \lim_{\epsilon \to 0} \frac{\epsilon D(q_Y \| p_{Y|u'}) + \frac{\epsilon^2}{2(1-\epsilon)} K_y}{\epsilon D(q_Y \| p_X) + \frac{\epsilon^2}{2(1-\epsilon)} K_x}
$$

$$
= \frac{D(q_Y \| p_{Y|u'})}{D(q_Y \| p_X)} > L - \delta,
$$

(102)

where (102) is due to (97). Since $\delta > 0$ was chosen arbitrarily, we get

$$
\lim_{\epsilon \to 0} g_0 \frac{I(Y; U) - g_0}{\epsilon} \geq L.
$$

(103)

On the other hand, construct the pair $(Y, U)$ as follows. Let $U = \{u_1, u_2, \ldots, u_{|u'|} \}$, $\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_{|Y|}$, and

$$
p_U(u_i) = (1 - \epsilon)p_{U|(u')_i}, \quad p_{Y|u_i} = p_{Y|u'_i}, \quad \forall i \in [1 : |u'|]
$$

where $e_i$ is an $|Y|$-dimensional vector denoting the $i$th extreme point of the probability simplex, i.e. $e_1 = \left[\begin{array}{c}1 \\ 0 \\ 0 \\ \ldots \end{array}\right]^T$, $e_2 = \left[\begin{array}{c}0 \\ 1 \\ 0 \\ \ldots \end{array}\right]^T$, and so on. It can be verified that the marginal probability vector $p_Y$ is preserved.

With this construction, we can have the Markov chain $X - Y - U$ and

$$
\frac{I(Y; U) - g_0}{I(X; U)} = \frac{(1 - \epsilon)g_0 + \sum_{i=1}^{|Y|} p_U(\tilde{u}_i)D(p_{Y|\tilde{u}_i} \| p_Y) - g_0}{\sum_{i=1}^{|Y|} p_U(\tilde{u}_i)D(p_X|\tilde{u}_i \| p_X)}
$$

$$
= \frac{\epsilon \sum_{i=1}^{|Y|} p_V(y_i)D(e_i \| p_Y) - g_0}{\epsilon \sum_{i=1}^{|Y|} p_V(y_i)D(p_X|y_i \| p_X)}
$$

$$
= \frac{H(Y) - g_0}{I(X; Y)},
$$

(104)

where we have used the facts that $D(e_i \| p_Y) = -\log p_Y(y_i)$ and $p_X|\tilde{u}_i = p_X|y_i$. Therefore, combining (103) and (104), we have

$$
\lim_{\epsilon \to 0} g_0 \frac{I(Y; U) - g_0}{\epsilon} \geq \max \left\{ L, \frac{H(Y) - g_0}{I(X; Y)} \right\}.
$$

(105)

Example 3. Assume $X = \{1, 2\}, Y = \{1, 2, 3\},$ and

$$
p_Y = \left[\begin{array}{c}0.2 \\ 0.5 \\ 0.3 \end{array}\right], \quad p_{X|Y} = \left[\begin{array}{ccc}0.5 & 0.3 & 0.6 \\ 0.5 & 0.7 & 0.4 \end{array}\right] \implies p_X = \left[\begin{array}{c}0.43 \\ 0.57 \end{array}\right].
$$

We have

$$
H(Y) = 1.4855, \quad I(X; Y) = 0.0539, \quad g_0(X; Y) = 0.5147,
$$

where $g_0$ is achieved, through an LP formulation, by $U' = \{u'_1, u'_2\}$ with $p_{U'|u'_1} = 0.3077$ and

$$
p_{Y|u'_1} = \left[\begin{array}{c}0.65 \\ 0.35 \\ 0 \end{array}\right], \quad p_{Y|u'_2} = \left[\begin{array}{c}0 \\ 0.5667 \\ 0.4333 \end{array}\right].$$
We can simply obtain $\psi(\cdot)$ as

$$
\psi(u') = \sup_{q_Y \neq p_Y|u'} \frac{D(q_Y||p_Y|u')} {D(q_X||p_X)} = \begin{cases} 
43.52 & u' = u'_1 \\
15.86 & u' = u'_2 
\end{cases},
$$

where $\psi(u'_1)$ and $\psi(u'_2)$ are achieved at $q_Y = e_1$ and $q_Y = e_2$, respectively. This results in $L = 43.52$, and from (105), we have

$$
\lim_{\epsilon \to 0} \frac{g_\epsilon(X;Y) - g_0} {\epsilon} \geq \max \{43.52, 18.011\} = 43.52.
$$

Figure 1 illustrates this example, where the triangle represents the probability simplex with the corner points $e_i$, $i \in [1:3]$. $v_3$ is the third right eigenvector of $P_{X|Y}$, whose span is the null space of $P_{X|Y}$. The line segment passing through $p_Y$ in the direction of $v_3$ is the convex polytope $S$ with the two extreme points $p_{Y|u'_1}$ and $p_{Y|u'_2}$.

VIII. CONCLUSIONS

This paper addresses the problem of perfect privacy, where the goal is to find the maximum utility obtained through a (distorted) disclosure of available data $Y$, while guaranteeing maximum privacy for the private latent variable $X$. It is shown that this problem boils down to a standard linear program when the utility is measured by the mutual information between $Y$ and its disclosed (distorted) version $U$. Similar results are obtained for other utility measures, in particular mean-square error and probability of error. It is shown that when the private variable and the observed data form a jointly Gaussian pair, utility can be obtained only at the expense of privacy when the data release mechanism has access only to the observed data $Y$. On the other hand, it is shown that, when
the privacy mapping has direct access to both the data $Y$ and the latent variable $X$, perfect privacy is feasible. Measuring both the utility and privacy by mutual information, we have then investigated the slope of the optimal utility-privacy trade-off curve when the revealed and private data are independent, i.e., $I(X; U) = 0$. Finally, we have proposed an alternative characterization of the maximal correlation between two random variables.

**APPENDIX A**

Let $U$ be an arbitrary set. Let $\mathcal{P}$ be the set of probability mass functions (pmf) on $\mathcal{Y}$. Let $r : \mathcal{P} \to \mathbb{R}^{\mathcal{Y}+1}$ be a vector-valued mapping defined element-wise as

$$
\begin{align*}
    r_i(p|Y|U(u|\cdot|u)) &= p_{Y|U}(y|u), \quad i \in [1 : |\mathcal{Y}| - 1], \\
    r_{|\mathcal{Y}|}(p|Y|U(u|\cdot|u)) &= H(Y|U = u), \\
    r_{|\mathcal{Y}|+1}(p|Y|U(u|\cdot|u)) &= H(X|U = u).
\end{align*}
$$

(106)

Since $\mathcal{P}$ corresponds to the standard $(|\mathcal{Y}| - 1)$-simplex, which is a closed and bounded subset of $\mathbb{R}^{|\mathcal{Y}|}$, it is compact. Also, $r$ is a continuous mapping from $\mathcal{P}$ to $\mathbb{R}^{\mathcal{Y}+1}$. Therefore, from the support lemma [26], for every $U \sim F(u)$ defined on $U$, there exists a random variable $U' \sim p(u')$ with $|U'| \leq |\mathcal{Y}| + 1$ and a collection of conditional pmfs $p_{Y|U'}(\cdot|u') \in \mathcal{P}$ indexed by $u' \in U'$, such that

$$
\int_\mathcal{U} r_i(p(y|u))dF(u) = \sum_{u' \in U'} r_i(p(y|u'))p(u'), \quad i \in [1 : |\mathcal{Y}| + 1].
$$

This means that for an arbitrary $U$ with $X - Y - U$, the terms $p_{Y}(\cdot)$, $I(Y; U)$ and $I(X; U)$ are preserved if $U$ is replaced with $U'$. Since we can simply have $X - Y - U'$, there is no loss of optimality in considering $|U| \leq |\mathcal{Y}| + 1$.

Let $\mathcal{P}_y \triangleq \left\{p_{U|Y}(\cdot|y) \bigg| U \in \mathcal{U}, \ Y \in \mathcal{Y}, \ |U| \leq |\mathcal{Y}| + 1 \right\}$ and $\mathcal{P}_y \triangleq \left\{p_{U|Y}(\cdot|y) \bigg| U \in \mathcal{U}, \ |U| \leq |\mathcal{Y}| + 1 \right\}$, $\forall y \in \mathcal{Y}$. The set $\mathcal{P}_y$ is the standard $(|\mathcal{U}| - 1)$-simplex, and therefore compact (since $|U| \leq |\mathcal{Y}| + 1 < \infty$). The set $\mathcal{P} = \bigcup_y \mathcal{P}_y$ is a finite $(|\mathcal{Y}| < \infty)$ union of these compact sets, which is still compact. Finally, the set $\mathcal{P}' = \left\{p_{U|Y}(\cdot|y) \in \mathcal{P} \bigg| X - Y - U', \ I(X; U) \leq \epsilon \right\}$ is a closed subset of $\mathcal{P}$ (due to the continuity of mutual information and closedness of the interval $[0, \epsilon]$), and therefore, it is also compact. Since $I(Y; U)$ is a continuous mapping over $\mathcal{P}'$, its supremum is achieved; and therefore, it is a maximum. This proves the first equality in (2). The second equality in (2) follows from the convexity of the objective function on $\mathcal{P}'$, and the maximum occurs at an extreme point of $\mathcal{P}'$, for which $I(X; U) = \epsilon$.

**APPENDIX B**

It is already known that when $g_0 > 0$, we have $p_{Y|U} \in \mathcal{S}, \forall u \in \mathcal{U}$. The reasoning in Appendix A can be modified as follows. $\mathcal{P}$ can be replaced with $\mathcal{S}$, and the mapping $r$ is modified as $r : \mathcal{S} \to \mathbb{R}^{\mathcal{Y}}$, where the last constraint in (106) is removed, as for any element in $\mathcal{S}$, we have $H(X|U = u) = H(X)$. Therefore, we obtain the sufficiency of $|U| \leq |\mathcal{Y}|$. 
Now, assume that the minimum in (12) is achieved by $K(\leq |\mathcal{Y}|)$ points in $\mathcal{S}$. We prove that all of these $K$ points must belong to the extreme points of $\mathcal{S}$. Let $p$ be an arbitrary point among these $K$ points. $p$ can be written as

$$p = \sum_{i=1}^{|\mathcal{Y}|} \alpha_i p_i,$$

(107)

where $\alpha_i \geq 0 \ (\forall i \in [1 : |\mathcal{Y}|])$ and $\sum_{i=1}^{|\mathcal{Y}|} \alpha_i = 1$; points $p_i \ (\forall i \in [1 : |\mathcal{Y}|])$ belong to the extreme points of $\mathcal{S}$ and $p_i \neq p_j \ (i \neq j)$. From the concavity of entropy, we have

$$H(p) \geq \sum_{i=1}^{|\mathcal{Y}|} \alpha_i H(p_i),$$

(108)

where the equality holds if and only if all of the $\alpha_i$s but one are zero. From the definition of an extreme point, if $p$ is not an extreme point of $\mathcal{S}$, it can be written as in (107) with at least two non-zero $\alpha_i$s, which makes the inequality in (108) strict. However, this violates the assumption that the $K$ points achieve the minimum. Hence, all of the $K$ points of the minimizer must belong to the set of extreme points of $\mathcal{S}$.

**APPENDIX C**

Let $p_{Y|u}$ be given as $p_{Y|u} = \lambda p_{Y|u_1} + (1 - \lambda)p_{Y|u_2}$, where $\lambda \in [0, 1]$. It is obvious that for an arbitrary function $b(\cdot)$,

$$E[b(Y)|U = u] = \lambda E[b(Y)|U = u_1] + (1 - \lambda)E[b(Y)|U = u_2].$$

(109)

Therefore,

$$\text{Var}[Y|U = u] = E\left[\left(Y - E[Y|U = u]\right)^2 \bigg| U = u\right]$$

$$= E[Y^2|U = u] - \left(E[Y|U = u]\right)^2$$

$$= \lambda E[Y^2|U = u_1] + (1 - \lambda)E[Y^2|U = u_2] - \left(\lambda E[Y|U = u_1] + (1 - \lambda)E[Y|U = u_2]\right)^2$$

(110)

$$\geq \lambda E[Y^2|U = u_1] + (1 - \lambda)E[Y^2|U = u_2] - \lambda \left(E[Y|U = u_1]\right)^2 - (1 - \lambda) \left(E[Y|U = u_2]\right)^2$$

(111)

$$= \lambda E\left[\left(Y - E[Y|U = u_1]\right)^2 \bigg| U = u_1\right] + (1 - \lambda)E\left[\left(Y - E[Y|U = u_2]\right)^2 \bigg| U = u_2\right]$$

$$= \lambda \text{Var}[Y|U = u_1] + (1 - \lambda)\text{Var}[Y|U = u_2],$$

where (110) is due to (109); and (111) is due to the (strict) convexity of $x^2$.

7The convex polytope $\mathcal{S}$ is an at most $(|\mathcal{Y}| - 1)$-dimensional convex subset of $\mathbb{R}^{|\mathcal{Y}|}$. Therefore, any point in $\mathcal{S}$ can be written as a convex combination of at most $|\mathcal{Y}|$ extreme points of $\mathcal{S}$. 
APPENDIX D

From the construction in (45), we have \( \forall m_{1:B} \in \prod_{i=1}^B E_i \),

\[
P_{X|Y}(s_{m_{1:B}}) = \sum_{k=1}^{|\mathcal{Y}|} P_X(y_k) s_{m_{1:B}}(k)
\]

\[
= \sum_{k \in \cup_{i=1}^B E_i} P_X(y_k) s_{m_{1:B}}(k) + \sum_{k \notin \cup_{i=1}^B E_i} P_X(y_k) s_{m_{1:B}}(k)
\]

\[
= \sum_{k \in [1:B]} P_X(y_k) \left( \sum_{t \in E_k} p_Y(y_t) \right) + \sum_{k \notin \cup_{i=1}^B E_i} P_X(y_k) s_{m_{1:B}}(k)
\]

\[
= \sum_{k \in \mathcal{Y}} P_X(y_k) \sum_{t \in E_k} p_Y(y_t) + \sum_{k \notin \cup_{i=1}^B E_i} P_X(y_k) s_{m_{1:B}}(k)
\]

\[
= \sum_{k=1}^{p_X} P_X(y_k) s_{m_{1:B}}(k)
\]

Hence, \( s \in \mathcal{S}, \forall s \in \mathcal{S}' \), which means that \( \mathcal{S}' \subseteq \mathcal{S} \).

From (45), the non-zero entries of any probability vector \( s \in \mathcal{S}' \) are the same as the elements of the set in (43).

Therefore, \( H(s) = H(T_X(Y)) \), \( \forall s \in \mathcal{S}' \).

Finally, let the set of \( |\mathcal{Y}| \)-dimensional probability vectors \( \{s_{m_{1}}\}_{m_{1} \in \mathcal{E}_{1}} \) on \( \mathcal{Y} \) be defined element-wise as

\[
s_{m_{1}}(y_k) = \begin{cases} 
p_Y(y_k) & \forall k \notin \mathcal{E}_{1} \\
\sum_{j \in \mathcal{E}_{1}} p_Y(y_j) & k = m_{1}, \forall m_{1} \in \mathcal{E}_{1}, \forall k \in [1 : |\mathcal{Y}|]. \\
0 & k \neq m_{1}, k \in \mathcal{E}_{1}
\end{cases}
\]  \hspace{1cm} (112)

By induction, define

\[
s_{m_{1:n-1}}(y_k) = \begin{cases} 
s_{m_{1:n-1}}(y_k) & \forall k \notin E_n \\
\sum_{j \in E_n} p_Y(y_j) & k = m_n, \forall m_{1:n} \in \prod_{i=1}^{n} E_i, \forall n \in [2 : B], \forall k \in [1 : |\mathcal{Y}|], \\
0 & k \neq m_n, k \in E_n
\end{cases}
\]  \hspace{1cm} (113)

where it can be verified that (113) and (45) are equivalent for \( n = B \). By constructions in (112) and (113), we can, respectively, write

\[
p_Y = \sum_{m_{1} \in \mathcal{E}_{1}} \sum_{k \in \mathcal{E}_{1}} p_Y(y_k) s_{m_{1}}
\]  \hspace{1cm} (114)

and

\[
s_{m_{1:n-1}} = \sum_{m_{n} \in E_n} \sum_{k \in E_n} p_Y(y_k) s_{m_{1:n-1}}
\]  \hspace{1cm} (115)

Therefore, \( p_Y \) can be written from (114) and (115) as

\[
p_Y = \sum_{m_{1:B} \in \prod_{i=1}^B E_i} \frac{p_Y(m_1)p_Y(m_2) \ldots p_Y(m_B)}{s_{m_{1:B}}}
\]  \hspace{1cm} (116)
By letting
\[ \alpha_{m[1:B]} = \frac{p_Y(m_1)p_Y(m_2)\ldots p_Y(m_B)}{\sum_{k\in E_1} p_Y(y_k) \sum_{k\in E_2} p_Y(y_k) \ldots \sum_{k\in E_B} p_Y(y_k)}, \forall m[1:B] \in \prod_{i=1}^{B} E_i, \]
we are led to (46).

**APPENDIX E**

Without loss of generality, by an appropriate labelling of the elements in \(Y\), we can assume that \(E_1 = [1 : |E_1|]\), \(E_2 = [E_1 + 1 : |E_1| + |E_2|]\), and so on. Let the common column vector corresponding to \(E_m\) be denoted by \(p_m\), i.e. \(p_m \triangleq p_X|_{y_m}\), \(\forall i \in E_m, \forall m \in [1 : B]\). We can write
\[ P_X|_Y = \left[ p_1, \ldots, p_1, p_2, \ldots, p_2, \ldots, p_B, \ldots, p_B, p_X|_{y_{G+1}}, p_X|_{y_{G+2}}, \ldots, p_X|_{y_{|Y|}} \right], \]
and
\[ \hat{P}_X|_Y = \left[ p_1, p_2, \ldots, p_B, p_X|_{y_{G+1}}, p_X|_{y_{G+2}}, \ldots, p_X|_{y_{|Y|}} \right]. \]

Define the vectors
\[ e_m^1 = \begin{bmatrix} 0_{\sum_{i=1}^{m-1} |E_i|} \\ 1 \\ -1 \\ 0_{(|Y|-\sum_{i=1}^{m-1} |E_i|-2)} \end{bmatrix}, e_m^2 = \begin{bmatrix} 0_{\sum_{i=1}^{m-1} |E_i|} \\ \frac{1}{2} \\ \frac{1}{2} \\ -1 \\ 0_{(|Y|-\sum_{i=1}^{m-1} |E_i|-3)} \end{bmatrix}, \ldots, e_m^{|E_m|-1} = \begin{bmatrix} 0_{\sum_{i=1}^{m-1} |E_i|} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ -1 \\ 0_{(|Y|-\sum_{i=1}^{m-1} |E_i|)} \end{bmatrix}, \forall m \in [1 : B], \]
where \(\sum_{i=1}^{m-1} |E_i| = \emptyset\) when \(m = 1\).

**Proposition 13.** Let
\[ \mathbb{N} \triangleq \text{Span}\left\{ e_m^i \mid \forall i \in [1 : |E_m| - 1], \forall m \in [1 : B] \right\}. \]
We have
\[ \mathbb{N} \subseteq \text{Null}(P_X|_Y), \]
where \(118\) holds with equality if and only if \(\dim(\text{Null}(\hat{P}_X|_Y)) = 0\).

**Proof.** The proof is provided in Appendix [E] \(\square\)

If \(\dim(\text{Null}(\hat{P}_X|_Y)) = 0\), any element in \(S'\) is an extreme point of \(S\). The reasoning is as follows. Note that \(S' \subseteq S\). Hence, it remains to show that no point of \(S'\) can be written as a convex combination of two different points of \(S\). Pick an arbitrary point \(s\) in \(S'\). It can be verified that no \(\epsilon > 0\) and \(e \in \mathbb{N}\) exist such that both \(s + \epsilon e\) and \(s - \epsilon e\) remain a probability vector. This is due to having a negative element in either or both of them. From Proposition 13, we have \(\mathbb{N} = \text{Null}(P_X|_Y)\) which in turn means that \(s\) cannot be written as a convex combination of two different points of \(S\). Therefore, the elements in \(S'\) belong to the extreme points of \(S\).

On the other hand, when \(\dim(\text{Null}(\hat{P}_X|_Y)) = 0\), assume that there exists \(s^* \notin S'\) that is an extreme point of \(S\). We show that this leads to a contradiction. Firstly, note that among the elements of \(s^*\) that correspond to
$\mathcal{E}_m, \forall m \in [1 : B]$, there must be at most one non-zero element. This is justified as follows. Assume that $s^*_i$ and $s^*_j$ ($i \neq j$) are two non-zero elements of $s^*$, where $i, j \in \mathcal{E}_m$ for some $m \in [1 : B]$. Construct the $|Y|$-dimensional vector $f$ where $f_i = -f_j = 1$, and the remaining terms are zero. Obviously, $f \in \text{Null}(P_{X|Y})$, as $P_{X|Y}f = 0$. Let $\epsilon = \min\{s^*_i, s^*_j\}$. It is obvious that $s^*$ can be written as a convex combination of the vectors $s^* + \epsilon f$ and $s^* - \epsilon f$, where both are in $S$. However, this contradicts the assumption of $s^*$ being an extreme point of $S$. Hence, among the elements of $s^*$ that correspond to $\mathcal{E}_m, \forall m \in [1 : B]$, at most one element is non-zero. As a result, we can find a point $s \in S'$ whose positions of its non-zero elements in $\bigcup_{i=1}^{|E|} \mathcal{E}_i$ matches those of $s^*$. Since $s^* \notin S'$, $s^*$ must differ with this $s$ in at least one element. Assume that for some $m \in [1 : B]$, there exists $j \in \mathcal{E}_m$ such that $s^*_j \neq s_j$.

Then, the elements of $\Delta s = s^* - s$ that correspond to $\mathcal{E}_m$ are all zero, except $\Delta s_j = s^*_j - s_j \neq 0$. It can then be verified that $\Delta s$ cannot be written as a linear combination of the vectors in $\mathbb{N}$, as no linear combination of the vectors $e_m^i, \forall i \in [1 : |E_m| - 1]$ can produce a vector whose all the elements corresponding to $\mathcal{E}_m$ are zero except one. Since $\text{dim}(\text{Null}(P_{X|Y})) = 0$, we have from Proposition 13 that $\mathbb{N} = \text{Null}(P_{X|Y})$, which in turn means that $s^* - s \notin \text{Null}(P_{X|Y})$. Therefore, $s^* \notin S$, which is a contradiction. If $s^*_j = s_j, \forall j \in \bigcup_{i=1}^{|E|} \mathcal{E}_i$, then we must have $s^*_j \neq s_j$ for some $j \in [G + 1 : |Y|]$. Still, $\Delta s$ cannot be written as a linear combination of the vectors in $\mathbb{N}$, as for any vector $n \in \mathbb{N}$, we have $n_k = 0, \forall k \in [G + 1 : |Y|]$. This leads us to $s^* \notin S$, which is again a contradiction. Therefore, we conclude that when $\text{dim}(\text{Null}(P_{X|Y})) = 0$, the extreme points of $S$ are the elements of $S'$.

If $\text{dim}(\text{Null}(P_{X|Y})) \neq 0$, from Proposition 13, there must exist a non-zero vector $v$ such that $v \in \text{Null}(P_{X|Y})$ and $v \notin \mathbb{N}$. Pick an arbitrary point of $S'$. In order to make the analysis simple, let the picked vector be $v_0$, which is

$$v_0 = \begin{bmatrix} \sum_{i \in \mathcal{E}_1} P_Y(y_i) \\ 0_{|\mathcal{E}_1| - 1} \\ \sum_{i \in \mathcal{E}_2} P_Y(y_i) \\ 0_{|\mathcal{E}_2| - 1} \\ \vdots \\ \sum_{i \in \mathcal{E}_{|E|}} P_Y(y_i) \\ 0_{|\mathcal{E}_{|E|}| - 1} \\ \vdots \\ P_Y(y_{|Y|}) \end{bmatrix}$$

From $v$, we can construct a non-zero vector $\tilde{v} \in \text{Null}(P_{X|Y})$, as done in (119). Then, it is obvious that for sufficiently small $\epsilon > 0$, $\tilde{s}_0$ can be written as a convex combination of $\tilde{s}_0 + \epsilon \tilde{v}$ and $\tilde{s}_0 - \epsilon \tilde{v}$, where both are in $S$.

This shows that $\tilde{s}_0$ cannot be an extreme point of $S$. A similar approach can be applied to show that the other.

---

8The only difference is in constructing a vector $\hat{v}$, such that when a point of $S'$ is perturbed along the direction of $\hat{v}$, it still lies in $S$. This can be done by noting that it is sufficient to construct a $\hat{v}$ whose position of zero elements in $[1 : G]$ matches that of the arbitrary point from $S$. In a similar way that $\tilde{v} \in \text{Null}(P_{X|Y})$ was constructed from $v$ in (119), by using other orthogonal vectors in $\mathbb{N}$, instead of $e_m^i$, a new $\tilde{v}$ can be constructed whose position of zero elements in $[1 : G]$ matches that of the arbitrary point from $S$.
points of $\mathcal{S}'$ do not belong to the set of extreme points of $\mathcal{S}$. Hence, from $\dim(\text{Null}(\hat{P}_{X|Y})) \neq 0$, we conclude that none of the elements in $\mathcal{S}'$ is an extreme point of $\mathcal{S}$.

**APPENDIX F**

The fact that $\mathbb{N} \subseteq \text{Null}(P_{X|Y})$ can be verified by observing that $P_{X|Y}e_m = 0, \forall i \in [1 : |E_m| - 1], \forall m \in [1 : B]$. If $\dim(\text{Null}(\hat{P}_{X|Y})) = 0$, we must have $\mathbb{N} = \text{Null}(P_{X|Y})$. If this is not true, from (118), there must exist a non-zero vector $v$ such that $v \in \text{Null}(P_{X|Y})$ and $v \notin \mathbb{N}$. Let $v_i$ denote the $i$th element in $v$. We can write

$$v + \sum_{m=1}^{B} \sum_{i=1}^{|E_m|-1} \alpha^i_m e^i_m = \begin{bmatrix} \sum_{i \in E_1} v_i \\ 0_{|E_1|-1} \\ \vdots \\ \sum_{i \in E_B} v_i \\ 0_{|E_B|-1} \\ v_{G+1} \\ \vdots \\ v_{|Y|} \end{bmatrix} = \hat{v},$$

(119)

where it can be verified that the coefficients $\alpha^i_m, \forall i \in [1 : |E_m| - 1], \forall m \in [1 : B]$ are obtained uniquely, as the vectors $e^i_m$ are mutually orthogonal. Since $v, e^i_m \in \text{Null}(P_{X|Y}), \forall i \in [1 : |E_m| - 1], \forall m \in [1 : B]$, we have $\hat{v} \in \text{Null}(P_{X|Y})$. Also, note that $\hat{v}$ is a non-zero vector, since otherwise (119) would result in $v \in \mathbb{N}$. Finally, from the structure of $\hat{v}$ and $P_{X|Y}$, we observe that $P_{X|Y}\hat{v}' = P_{X|Y}\hat{v} = 0$, where $\hat{v}'$ is obtained from eliminating the zero vectors of $\hat{v}$, denoted by $0_{|E_i|-1}, \forall i \in [1 : B]$, in (119). Since $\hat{v}$ is a non-zero vector, so must be $\hat{v}'$. Hence, $\dim(\text{Null}(\hat{P}_{X|Y})) \neq 0$, which is a contradiction. Therefore, we must have $\mathbb{N} = \text{Null}(P_{X|Y})$.

If $\mathbb{N} = \text{Null}(P_{X|Y})$, we must have $\dim(\text{Null}(\hat{P}_{X|Y})) = 0$. If this is not true, there exists a non-zero vector $\hat{r}'$ such that $P_{X|Y}\hat{r}' = 0$, and correspondingly a non-zero vector $\hat{r}$ such that $P_{X|Y}\hat{r} = 0$, where the relation between $\hat{r}'$ and $\hat{r}$ is similar to that between $\hat{v}'$ and $\hat{v}$ in the previous paragraph. Therefore, we have $\hat{r} \in \text{Null}(P_{X|Y})$. However, it can be verified that due to the structure of the vectors $e^i_m$, i.e. the positions of the zero and non-zero elements in $[1 : G]$, $\hat{r}$ cannot be written as a linear combination of the vectors $e^i_m$. This results in $\hat{r} \notin \mathbb{N}$, which is a contradiction, as we assumed $\mathbb{N} = \text{Null}(P_{X|Y})$. This proves that $\dim(\text{Null}(\hat{P}_{X|Y})) = 0$.

**APPENDIX G**

When $g_0(X, Y) = 0$, there is no $q_Y \neq p_Y$, such that $q_X = p_X$, since otherwise we could have constructed a random variable $U \in \{u_1, u_2\}$, and a sufficiently small $\alpha > 0$, such that

$$p_U(u_1) = \alpha, \quad p_Y|u_1 = q_Y, \quad p_Y|u_2 = \frac{1}{1-\alpha}(p_Y - \alpha q_Y),$$

where the sufficiently small $\alpha$ makes $p_Y|u_2$ still a probability vector. With this construction, it can be verified that $X - Y - U, X \perp U$ and $Y \not\perp U$ which contradicts $g_0(X, Y) = 0$. Hence, the only way to have $V^* > M, \forall M \in \mathbb{R}$
is through the existence of a sequence of distributions, i.e. \( \{q^n_Y\}_n \), where \( q^n_Y \neq p_Y, \forall n \) and \( q^n_X \rightarrow p_X \). Since perfect privacy is not feasible, we must have \( |\mathcal{Y}| \leq |\mathcal{X}| \) and \( \sigma_i(P_{X|Y}) \neq 0, \forall i \in \left[ 1 : \min(|\mathcal{X}|,|\mathcal{Y}|) \right] \). This means that in order to have \( q^n_X \rightarrow p_X \), we must have \( q^n_Y \rightarrow p_Y \). We know that when \( q_Y \rightarrow p_Y (q_Y \neq p_Y) \), the ratio in (64) is bounded below by the minimum eigenvalue of the matrix \( Q^T Q \). If for an arbitrary non-zero vector \( e \), we have \( Q e = 0 \), then we must have
\[
Q e = P_X^{-\frac{1}{2}} P_{X|Y} P_Y^{-\frac{1}{2}} e = P_X^{-\frac{1}{2}} P_{X|Y} \underbrace{P_Y^{\frac{1}{2}} e}_{e'} = 0,
\]
which is not possible, since \( e' \) is a non-zero vector, and so is \( P_{X|Y} e' \) due to the fact that \( \sigma_i(P_{X|Y}) \neq 0, \forall i \in \left[ 1 : \min(|\mathcal{X}|,|\mathcal{Y}|) \right] \) and \( |\mathcal{Y}| \leq |\mathcal{X}| \), i.e. the null space of \( P_{X|Y} \) is only the all-zero vector. Therefore, the minimum eigenvalue of the matrix \( Q^T Q \) is bounded away from zero. Equivalently, the inverse of (64) is bounded above by the inverse of the minimum eigenvalue of \( Q^T Q \). Hence, \( V^* < +\infty \).

The proof of the second direction is immediate, since having \( g_0(X, Y) > 0 \) leads to the existence of \( q_Y \neq p_Y \), such that \( q_X = p_X \), which in turn violates \( V^* < +\infty \).

**APPENDIX H**

Firstly, note that for any point \( q_Y \) that satisfies \( 0 < D(q_Y||p_{Y|u'}) < +\infty \), we have \( q_Y \notin \mathcal{S} \) (i.e., \( q_X \neq p_X \)), where \( \mathcal{S} \) is defined in (10), since otherwise from the fact that \( D(q_Y||p_{Y|u'}) < +\infty \), for sufficiently small \( \epsilon \), we can make the probability vector \( q_Y = \frac{1}{1-\epsilon}(p_{Y|u'} - \epsilon q_Y) \) which also belongs to \( \mathcal{S} \), as \( P_{X|Y} q_Y = p_X \). However, this violates the fact that \( p_{Y|u'} \) is an extreme point of \( \mathcal{S} \), since it can be written as a convex combination of two points of \( \mathcal{S} \), i.e. \( q_Y \) and \( q_Y' \) (\( q_Y \neq q_Y' \)). Alternatively, we can say that for any \( q_Y \) that satisfies \( 0 < D(q_Y||p_{Y|u'}) < +\infty \), we have \( (q_Y - p_{Y|u'}) \notin \text{Null}(P_{X|Y}) \). Hence, the only way to have \( \psi(u') \) possibly unbounded is through the existence of a sequence of distributions, i.e. \( \{q^n_Y\}_n \), where \( 0 < D(q^n_Y||p_{Y|u'}) < +\infty, \forall n \) and \( q^n_X \rightarrow p_X \), which requires \( q^n_Y \) converging to a point of \( \mathcal{S} \). Let \( \mathcal{I}_{u'} \) denote the set of indices corresponding to the zero elements of \( p_{Y|u'} \). Let \( \mathbb{T} \) denote the set of probability vectors \( p \), such that \( D(p||p_{Y|u'}) < +\infty \). In other words, \( \mathbb{T} \) is the set of probability vectors whose elements corresponding to the indices in \( \mathcal{I}_{u'} \) are zero. Since \( \mathbb{T} \) is a closed set, we conclude that if \( q^n_Y \in \mathbb{T} \) converges to \( p_0 \) (a point of \( \mathcal{S} \), \( p_0 \) must also be in \( \mathbb{T} \), i.e. it satisfies \( D(p_0||p_{Y|u'}) < +\infty \). If \( D(p_0||p_{Y|u'}) > 0 \), from what mentioned before, we have \( (p_0 - p_{Y|u'}) \notin \text{Null}(P_{X|Y}) \), which contradicts the fact that \( p_0 \in \mathcal{S} \), hence, we must have \( p_0 = p_{Y|u'} \). Therefore, it suffices to consider the following problem
\[
\liminf_{0 < D(q_Y||p_{Y|u'}) < +\infty} \frac{D(q_X||p_X)}{D(q_Y||p_{Y|u'})}.
\]
Similarly to the analysis after (64), the above becomes equal to the minimum eigenvalue of \( \tilde{Q}^T \tilde{Q} \), where
\[
\tilde{Q} = P_X^{-\frac{1}{2}} P_{X|Y} P_{Y|u'}^{-\frac{1}{2}},
\]
and \( \tilde{P}_{X|Y} \) is an \( |\mathcal{X}| \times (|\mathcal{Y}| - |\mathcal{I}_{u'}|) \)-dimensional matrix obtained by eliminating the columns of \( P_{X|Y} \) that correspond to the indices in \( \mathcal{I}_{u'} \); \( P_{Y|u'} \) is a \( (|\mathcal{Y}| - |\mathcal{I}_{u'}|) \times (|\mathcal{Y}| - |\mathcal{I}_{u'}|) \) diagonal matrix whose diagonal elements are the corresponding non-zero elements of \( p_{Y|u'} \).
In what follows, we show that the minimum eigenvalue of the matrix \( \bar{Q}^T \bar{Q} \) is bounded away from zero, since otherwise there must exists a non-zero \((|\mathcal{Y}|-|\mathcal{I}_{u'}|)\)-dimensional vector \( \bar{e} \), such that \( \bar{Q} \bar{e} = 0 \). Let \( \bar{e}' \triangleq \bar{P}^{\frac{1}{2}}_{X'_{u'}} \bar{e} \). From Proposition 3, \( 1_{|\mathcal{Y}|-|\mathcal{I}_{u'}|} \bar{e}' = 0 \). Construct the \(|\mathcal{Y}|-\) dimensional vector \( e' \) as follows. Let its elements corresponding to the indices in \( \mathcal{I}_{u'} \) be zero, and its other terms be equal to the elements of \( e' \). It is obvious that \( e' \in \text{Null}(\bar{P}_{X|Y}) \), since the elements of \( e' \) corresponding to the columns of \( \bar{P}_{X|Y} \) that are not in \( \hat{\bar{P}}_{X|Y} \) are zero and we have \( \bar{e}' \in \text{Null}(\hat{\bar{P}}_{X|Y}) \). From Proposition 3, having \( e' \in \text{Null}(\bar{P}_{X|Y}) \) results in \( 1_{|\mathcal{Y}|} e' = 0 \). For sufficiently small \( \epsilon > 0 \), let \( q_Y = p_{Y|u'} + \epsilon e' \). Since \( \epsilon \neq 0 \) and \( e' \neq 0 \), we have \( D(q_Y \| p_{Y|u'}) > 0 \). Moreover, since the elements in \( q_Y \) corresponding to the indices in \( \mathcal{I}_{u'} \) are zero, we have \( D(q_Y \| p_{Y|u'}) < +\infty \). Therefore, from the reasoning at the beginning of this Appendix, we have \( q_Y \notin \mathcal{S} \). However, since \( e' \in \text{Null}(\bar{P}_{X|Y}) \) and \( \bar{P}_{X|Y} p_{Y|u'} = p_X \), we have \( q_Y \in \mathcal{S} \), which is a contradiction. Therefore, the minimum eigenvalue of the matrix \( \bar{Q}^T \bar{Q} \) is bounded away from zero. This in turn means that the inverse of \( \psi(u') \) is bounded away from zero, and therefore, \( \psi(u') < +\infty \), \( \forall u' \).

REFERENCES

[1] A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets,” in IEEE Symp. on Security and Privacy (SP), 2008, pp. 111–125.
[2] X. Ding, L. Zhang, and W. Zhiguo, “A brief survey on de-anonymization attacks in online social networks,” in International Conf. on Computational Aspects of Social Networks (CASoN), 2010, pp. 611–615.
[3] S. Kumar, W. Nilsen, M. Pavel, and M. Srivastava, “Mobile health: Revolutionizing healthcare through transdisciplinary research,” Computer, vol. 46, pp. 28–35, 2013.
[4] J. Gomez-Vilardebo and D. Gündüz, “Smart meter privacy for multiple users in the presence of an alternative energy source,” IEEE Trans. on Information Forensics and Security, pp. 132–141, 2015.
[5] A. Motahari, G. Bresler, and D. Tse, “Information theory of DNA shotgun sequencing,” IEEE Trans. on Information Theory, vol. 59, no. 10, pp. 6273–6289, Oct. 2013.
[6] C. Dwork, F. McSherry, K. Nissim, and A. Smith, “Calibrating noise to sensitivity in private data analysis,” Theory of Cryptography, Springer, pp. 265–284, 2006.
[7] L. Sweeney, “k-anonymity: A model for protecting privacy,” Intl. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 5, pp. 557–570, 2002.
[8] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, “l-diversity: Privacy beyond k-anonymity,” ACM Trans. on Knowledge Discovery from Data, vol. 1.
[9] N. Li, T. Li, and S. Venkatasubramaniam, “t-closeness: Privacy beyond k-anonymity and l-diversity,” IEEE Intl. Conf. on Data Eng., 2007.
[10] F. Calmon and N. Fawaz, “Privacy against statistical inference,” in 50th Annual Allerton Conference, Illinois, USA, Oct. 2012, pp. 1401–1407.
[11] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Médard, “From the information bottleneck to the privacy funnel,” in IEEE Information Theory Workshop (ITW), 2014, pp. 501–505.
[12] N. Tishby, F. Pereira, and W. Bialek, “The information bottleneck method,” arXiv preprint physics/0004057, 2000.
[13] F. Calmon, A. Makhdoumi, and M. Médard, “Fundamental limits of perfect privacy,” in IEEE Int. Symp. Inf. Theory (ISIT), 2015, pp. 1796–1800.
[14] T. Berger and R. Yeung, “Multiterminal source encoding with encoder breakdown,” IEEE Trans. Inf. Theory, pp. 237–244, 1989.
[15] S. Assoodeh, F. Alajaji, and T. Linder, “Notes on information-theoretic privacy,” in 52nd Annual Allerton Conference, Illinois, USA, Oct. 2014, pp. 1272–1278.
[16] H. Hirschfeld, “A connection between correlation and contingency,” in Proc. Cambridge Philosophical Soc. 31, 1935, pp. 520–524.
[17] H. Gebelein, “Das statistische problem der korrelation als variations- und eigenwert-problem und sein zusammenhang mit der ausgleichungsrechnung,” Zeitschrift f"ur angew. Math. und Mech. 21, pp. 364–379, 1941.
[18] A. Rényi, “On measures of dependence,” *Acta Math. Hung.*, vol. 10, pp. 539–550, 1959.

[19] D. Bertsimas and J. N. Tsitsiklis, *Introduction to linear optimization*. Athena Scientific, 1997.

[20] K. G. Murty, *Linear Programming*. John Wiley and Sons, 1983.

[21] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. Springer, 2008.

[22] Y. Wang, Y. Basciftci, and P. Ishwar, “Privacy-utility tradeoffs under constrained data release mechanisms,” https://arxiv.org/pdf/1710.09295.pdf, Oct. 2017.

[23] H. Witsenhausen, “On sequences of pairs of dependent random variables,” *SIAM Journal on Applied Mathematics*, vol. 28, no. 1, pp. 100–113, Jan. 1975.

[24] V. Anantharam, A. Gohari, S. Kamath, and C. Nair, “On maximal correlation, hypercontractivity, and the data processing inequality studied by Erkip and Cover,” https://arxiv.org/pdf/1304.6133.pdf, Apr. 2013.

[25] G. Golub, “Some modified matrix eigenvalue problems,” *SIAM Review*, vol. 15, no. 2, pp. 318–334, Apr. 1973.

[26] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2012.