Forgot Your Password: Correlation Dilution

Ali Makhdoumi, Flavio P. Calmon, Muriel Médard

***Abstract*—We consider the problem of diluting common randomness from correlated observations by separated agents. This problem creates a new framework to study statistical privacy, in which a legitimate party, Alice, has access to a random variable X, whereas an attacker, Bob, has access to a random variable Y dependent on X drawn from a joint distribution $p_{X,Y}$. Alice's goal is to produce a non-trivial function of her available information that is uncorrelated with (has small correlation with) any function that Bob can produce based on his available information. This problem naturally admits a minimax formulation where Alice plays first and Bob follows her. We define dilution coefficient as the smallest value of correlation achieved by the best strategy available to Alice, and characterize it in terms of the minimum principal inertia components of the joint probability distribution $p_{X,Y}$. We then explicitly find the optimal function that Alice must choose to achieve this limit. We also establish a connection between differential privacy and dilution coefficient and show that if Y is ϵ-differentially private from X, then dilution coefficient can be upper bounded in terms of ϵ. Finally, we extend to the setting where Alice and Bob have access to i.i.d. copies of (X_i, Y_i), $i = 1, \dots, n$ and show that the dilution coefficient vanishes exponentially with n. In other words, Alice can achieve better privacy as the number of her observations grows.**

***Index Terms*—Statistical Privacy; Differential Privacy; Estimation; Principal Inertia Components.**

I. Introduction

We consider the setting where a legitimate party, Alice, has an observation of a random variable X, whereas an attacker, Bob, has access to a random variable Y dependent on X drawn from a joint distribution $p_{X,Y}$. Alice's goal is to produce a function $f(X)$ (non-trivial) that is uncorrelated with (has small correlation with) any function $g(Y)$ that Bob can produce. We call this general setup the *correlation dilution* problem, formally defined below.

Definition 1 (Correlation Dilution). Let X and Y be discrete random variables over finite support sets $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let

$$\mathcal{C}_X \triangleq \{f : \mathcal{X} \to \mathbb{R} \ : \ \mathbb{E}[f(X)] = 0, \ \mathbb{E}[f(X)^2] = 1\}$$

and

$$\mathcal{C}_Y \triangleq \{g : \mathcal{Y} \to \mathbb{R} \ : \ \mathbb{E}[g(Y)] = 0, \ \mathbb{E}[g(Y)^2] = 1\}$$

denote the set of normalized mean zero functions of X and Y. Alice and Bob choose functions $f \in \mathcal{C}_X$ and $g \in \mathcal{C}_Y$, respectively. Alice aims to minimize correlation between f and g, while Bob aims to maximize it.[1] We define *dilution coefficient* as

$$\delta(X;Y) = \min_{f \in \mathcal{C}_X} \max_{g \in \mathcal{C}_Y} \mathbb{E}[f(X)g(Y)],$$

where $\delta(X;Y)$ shows the extent to which Alice can decrease the correlation of her function with (worst-case) Bob's function.

We say that *full correlation dilution* between Alice and Bob is possible if $\delta(X;Y) = 0$. In other words, full correlation dilution is achieved if Alice can find a function $f \in \mathcal{C}_X$ that is uncorrelated with any function $g \in \mathcal{C}_Y$ that Bob can produce.

The correlation dilution problem, that we define, is a new framework to study statistical privacy (see [1] and references therein for a review of statistical privacy) that appears in a variety of security systems. Here, X plays the role of the secret information, Y is the information that leaks to an adversary or eavesdropper, and we wish to identify which functions of the secret information the adversary cannot determine reliably. Consider, for example, a password-restricted web service (e.g. email, online banking), where the user is asked to design a security question in case his or her password is forgotten. This situation is very common as studies have shown (see e.g. [2], [3]) that most users have at least one account for which they have forgotten their password, having to potentially resort to a security question. Choosing a pair of security question and its answer differs from selecting a password in that the selected secret string is usually a direct function of your personal information. Consequently, an attacker may have partial knowledge of the user's personal information (e.g. social network observations) which, in turn, could be correlated to the answer of certain security questions. The problem is then reduced to choosing a function of the personal data that bears little relation to any function that an attacker may compute from the data at his disposal.

In choosing a security question, we seek to find a function f of the personal data X that would still be hard to guess even if the adversary has gathered correlated side information Y from multiple sources. This example naturally motivates the questions studied in this paper: What is the optimal choice of function f (security question)? What is the fundamental limit of minimum correlation achievable? How does this fundamental value change as the amount of information available to both Alice and Bob grows? In particular, does the security risk increase as more observations of X and Y are available?

In this paper, we answer these questions by analyzing the *principal inertia components* ([4], [5]) of the joint distribution $p_{X,Y}$. In mathematical probability, the study of principal inertia components dates back to Hirschfeld [6], Gebelein [7], Sarmanov [8] and Rényi [9], and similar analysis have also recurrently appeared in the information theory and applied probability literature (see [6]–[13]). We present the formal definition of principal inertia components in the next section.

We prove that dilution coefficient, $\delta(X;Y)$, can be expressed in terms of the minimum principal inertia component of $p_{X,Y}$. We then characterize the dilution coefficient when Alice and Bob observe the sequences $X^n \triangleq (X_1, \dots, X_n)$ and $Y^n \triangleq (Y_1, \dots, Y_n)$, respectively, where (X_i, Y_i) are i.i.d. for $i = 1, \dots, n$ with joint distribution $p_{X,Y}$. We show that, even

A. Makhdoumi, F. P. Calmon and M. Médard are with the Massachusetts Institute of Technology, Cambridge, MA (e-mail: {makhdoum, flavio, medard}@mit.edu).

[1]This situation naturally creates a Stackelberg minimax game where Alice is the leader (plays first) and Bob is the follower.

though the mutual information between X^n and Y^n grows with n (i.e., $\lim_{n\to\infty} I(X^n;Y^n) = \lim_{n\to\infty} nI(X;Y) = \infty$ if X and Y are not independent), the value of $\delta(X^n;Y^n)$ vanishes exponentially with n and, in particular, $\delta(X^n;Y^n) = \delta(X;Y)^n$. This demonstrates that if $\delta(X;Y) < 1$, which we prove is equivalent to X not being deterministic mapping of Y, full correlation dilution becomes possible as n grows large.

Our results imply that, in general, mutual information $I(X;Y)$ does not characterize the extent to which Alice can hide her data from Bob when the data to be hidden is of Alice's choosing. The intuition behind this result is that if Alice has access to more observations X^n, then she can better exploit the properties of the distribution p_{X^n,Y^n} in order to determine her function $f(X^n)$. This supports the results of [14], showing the relevance of principal inertia components rather than mutual information in the context of secrecy. We also explicitly show how the optimal function f can be constructed in terms of the *principal inertia components decomposition*, explained in the next section. Finally, we show a connection between differential privacy and our measure $\delta(X;Y)$, proving that if Y is a differentially private mapping of X, then $\delta(X;Y)$ is small. This establishes the relevance of differential privacy in the context of correlation dilution.

One line of work in the literature concerns with the opposite problem of correlation dilution, i.e., extracting common randomness from correlated observations. In particular, Wyner [15] studied the problem of simulating a joint distribution from shared randomness while Gács and Körner [16] studied the problem of extracting common randomness from correlated observations. Non-Interactive correlation distillation, a setup in which separated agents have to each output a uniform random bit which agree with high probability, is studied in [10], [17] and a generalization of it is recently studied in [18].

The rest of the paper is organized as follows. In Section II, we present the notation and definitions used in this paper. In Section III, we formally define correlation dilution problem and characterize its fundamental limits. In Section IV, we establish a connection between differential privacy and our measure of dilution $\delta(X;Y)$. In Section V, we characterize correlation dilution of independent copies of (X_i,Y_i) for $i = 1,\dots,n$ as well as the optimal choice of functions, which leads to concluding remarks in Section VI.

II. Preliminaries

In this section, we define the principal inertia components and present the notations used in this paper.

A. Notation

Throughput the paper, X and Y denote discrete random variables with joint distribution $p_{X,Y}$, where $p_{X,Y}(x,y) = \mathbb{P}_{X,Y}[X = x, Y = y]$. The support of X and Y are finite sets $\mathcal{X} = \{1,\dots,|\mathcal{X}|\}$ and $\mathcal{Y} = \{1,\dots,|\mathcal{Y}|\}$, respectively. The joint distribution matrix P is a $|\mathcal{X}|\times|\mathcal{Y}|$ matrix with the (i,j)-th entry equal to $p_{X,Y}(i,j)$. We denote by $\mathbf{p}_X$ (respectively, $\mathbf{p}_Y$) the vector with i-th entry equal to $p_X(i)$ (respectively, $p_Y(i)$). For any vector $\mathbf{v}$, $\sqrt{\mathbf{v}}$ is a vector with i-th entry equal to $\sqrt{v_i}$. We define the Q matrix Q_{XY}, a $|\mathcal{X}|\times|\mathcal{Y}|$ matrix with the (i,j)-th entry equal to $\frac{p_{X,Y}(i,j)}{\sqrt{p_X(i)p_Y(j)}}$.[2]

[2]We suppose that p_X and p_Y are positive over their support set.

Let S and T be two finite sets. For two functions $f_1 : S \to \mathbb{R}$ and $f_2 : T \to \mathbb{R}$, we define $f = f_1 \otimes f_2$ as $f : (S,T) \to \mathbb{R}$, where $f(s,t) = f_1(s)f_2(t)$ for any $s \in S$ and $t \in T$.[3] We show the transpose of vector $\mathbf{v}$ and matrix Q by $\mathbf{v}'$ and Q', respectively. We denote the vector $(X_1,\dots,X_n)$ by X^n. For a given matrix Q, let Singular(Q) denote the set of singular values of Q.

B. Principal Inertia Components

The term "principal inertia" is borrowed from the correspondence analysis literature [4] and is used in recent works [5], [13]. Principal intertia components of the joint distribution of two random variables was studied in many works such as [6]–[13]. Next, we define the principal inertia components for the discrete setting considered here.

Definition 2. We call the singular value decomposition $Q_{XY} = U\Sigma V'$ the *principal inertia decomposition* of X and Y, where Σ is a diagonal matrix with $\sigma_1,\dots,\sigma_r$ on the diagonal and $r = \min\{|\mathcal{X}|,|\mathcal{Y}|\}$. The values σ_i^2, $i = 1,\dots,r$, are called the *principal inertia components* of X and Y. In particular, the second largest singular value is called maximal correlation between X and Y denoted by $\rho_m(X;Y) = \sigma_2$, where $\rho_m(X;Y)$ in turn, is given by

$$\rho_m(X;Y) \triangleq \sup\{\mathbb{E}[f(X)g(Y)] \ : \ f \in \mathcal{C}_X, g \in \mathcal{C}_Y\}.$$

We denote the columns of matrices U and V by $\mathbf{u}_1,\dots,\mathbf{u}_{|\mathcal{X}|}$ and $\mathbf{v}_1,\dots,\mathbf{v}_{|\mathcal{Y}|}$.

III. Correlation Dilution

A. Problem Statement

We now return to the setting presented in the introduction. Consider the scenario where Alice wishes to choose a function $f(X)$ of her observation X that is uncorrelated with (has small correlation with) any function $g(Y)$ that Bob can produce from his observation Y. This problem can be formulated as follows:

$$\text{Alice: } X \to f \in \mathcal{C}_X, \quad \text{Bob: } Y \to g \in \mathcal{C}_Y$$
$$\text{Objective: } \min_{f\in\mathcal{C}_X}\max_{g\in\mathcal{C}_Y}\mathbb{E}[f(X)g(Y)].$$

Note that we formulate the worst-case behavior of Bob, meaning that he acts in an adversarial manner in order to maximize the correlation after Alice chooses her function. We denote the optimal functions by

$$f^* \in \text{argmin}_{f\in\mathcal{C}_X}\max_{g\in\mathcal{C}_Y}\mathbb{E}[f(X)g(Y)],$$

and

$$g^* \in \text{argmax}_{g\in\mathcal{C}_Y}\mathbb{E}[f^*(X)g(Y)].$$

B. Characterization of Correlation Dilution

Definition 3. The *dilution coefficient* between X and Y is defined as

$$\delta(X;Y) = \min_{f\in\mathcal{C}_X}\max_{g\in\mathcal{C}_Y}\mathbb{E}[f(X)g(Y))]. \tag{1}$$

Next, we will characterize this quantity.

[3]In other words, if we treat functions f_1 and f_2 as vectors in $\mathbb{R}^{|S|}$ and $\mathbb{R}^{|T|}$, respectively, then $f = f_1 \otimes f_2$ is the Kronecker product of these two vectors.

Theorem 1. *For random variables X and Y with joint distribution $p_{X,Y}$, let the singular values of the corresponding Q_{XY} matrix be $\sigma_1 = 1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. We have that*

$$\delta(X;Y) = \begin{cases} \sigma_r, & \text{if } |\mathcal{X}| \leq |\mathcal{Y}|, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Proof: The proof is straightforward. We will present a proof based on singular value decomposition of the matrix Q_{XY}. This will help us to explicitly characterize the functions that achieve the dilution coefficient. Let $\mathcal{X} = \{1, \ldots, |\mathcal{X}|\}$ and $\mathcal{Y} = \{1, \ldots, |\mathcal{Y}|\}$. We consider the following basis for functions from $\mathcal{X}$ and $\mathcal{Y}$ to $\mathbb{R}$. For any $i \in \mathcal{X}$ let $\phi_i : \mathcal{X} \to \mathbb{R}$, where[4]

$$\phi_i(x) = \mathbf{1}\{x = i\} \frac{1}{\sqrt{p_X(i)}}.$$

For any $j \in \mathcal{Y}$ let $\psi_j : \mathcal{Y} \to \mathbb{R}$, where

$$\psi_j(y) = \mathbf{1}\{y = j\} \frac{1}{\sqrt{p_Y(j)}}.$$

The choice of basis is for convenience as it will simplify the analysis. We can write $f \in \mathcal{C}_X$ and $g \in \mathcal{C}_Y$ in terms of aforementioned basis as $f = \sum_i a_i \phi_i : \mathcal{X} \to \mathbb{R}$, and $g = \sum_i b_i \psi_i : \mathcal{Y} \to \mathbb{R}$, where $a_i = f(i)\sqrt{p_X(i)}$ and $b_i = g(i)\sqrt{p_Y(i)}$.
By definition of the basis, the expectation and variance constraints $f \in \mathcal{C}_X$ and $g \in \mathcal{C}_Y$ translate into $\mathbf{a} \perp \sqrt{\mathbf{p}_X}$, $\|\mathbf{a}\|_2 = 1$, $\mathbf{b} \perp \sqrt{\mathbf{p}_Y}$, and $\|\mathbf{b}\|_2 = 1$. Therefore, $\delta(X;Y)$ becomes

$$\delta(X;Y) = \min_{\mathbf{a}} \max_{\mathbf{b}} \mathbf{a}' Q_{XY} \mathbf{b}.$$

Now let $Q_{XY} = U \Sigma V^T$ be the singular value decomposition of Q_{XY}, and let $r = \min\{|\mathcal{X}|, |\mathcal{Y}|\}$. Since U and V are unitary matrices that span the column and row space of Q_{XY}, respectively, we can further write $\mathbf{a} = \sum_i c_i \mathbf{u}_i$ and $\mathbf{b} = \sum_i d_i \mathbf{v}_i$. Since $\sqrt{\mathbf{p}_X}$ and $\sqrt{\mathbf{p}_Y}$ are left and right singular vectors of Q_{XY} corresponding to the largest singular value 1, the constraints ($\mathbf{a} \perp \sqrt{\mathbf{p}_X}$, $||\mathbf{a}|| = 1$, $\mathbf{b} \perp \sqrt{\mathbf{p}_Y}$, and $||\mathbf{b}|| = 1$) translate into $||\mathbf{c}|| = ||\mathbf{d}|| = 1$ and $c_1 = d_1 = 0$. We have that

$$\max_{\mathbf{b}} \mathbf{a}' Q_{XY} \mathbf{b} = \max_{\mathbf{d}} \sum_{i=2}^{r} \sigma_i c_i d_i = \sqrt{\sum_{i=2}^{r} (c_i \sigma_i)^2},$$

where we used Cauchy-Schwartz inequality to obtain the last equality and maximum is achieved for $d_i = \frac{c_i \sigma_i}{\sqrt{\sum_{i=2}^{r} (c_i \sigma_i)^2}}$, $i = 2, \ldots, n$. Thus, the optimization problem simplifies to

$$\min_{\mathbf{c}} \sqrt{\sum_{i=2}^{r} (c_i \sigma_i)^2},$$

where $c_1 = 0$ and $||\mathbf{c}|| = 1$. The solution to this optimization problem is obtained by choosing $c_2, \ldots, c_{|\mathcal{X}|}$ such that $c_{|\mathcal{X}|} = 1$ and $c_2 = \ldots, = c_{|\mathcal{X}|-1} = 0$. ∎

Remark 1.

- Using the data processing inequality for principal inertia components (see [5], [19]), if $X' \to X \to Y$ form a Markov chain and $|\mathcal{X}'| \geq |\mathcal{X}|$, then $\delta(X';Y) \leq \delta(X;Y)$.
- Alice chooses function f such that $c_{|\mathcal{X}|} = 1$. This means that $\mathbf{a} = \mathbf{u}_{|\mathcal{X}|}$ and consequently $f^*(i) = \frac{\mathbf{u}_{|\mathcal{X}|}(i)}{\sqrt{p_X(i)}}$. Similarly, we have that $g^*(i) = \frac{\mathbf{v}_{|\mathcal{Y}|}(i)}{\sqrt{p_Y(i)}}$.
- If the minimum singular value σ_r of Q_{XY} is zero, then $\delta(X;Y) = 0$. This implies that Alice can achieve full correlation dilution, i.e., she can choose a function that is uncorrelated with any function that Bob can choose.
- In addition, if the size of the support set of X is larger than Y, then Alice can also achieve full correlation dilution. Intuitively, this is due to Alice having more degrees of freedom than Bob in the choice of function, and this asymmetry allows her to achieve full dilution.
- If the minimum singular value of Q_{XY} matrix is not unique, then the functions f^* and g^* are not unique. We have many choices for $\mathbf{a}$ and the corresponding f^*. In particular, $\mathbf{a} \in \text{span}\{\mathbf{u}_j : \sigma_j = \sigma_r\}$. In the rest of paper, we assume that f^* and g^* correspond to the smallest singular value, i.e., $\mathbf{a} = \mathbf{u}_r$ and $\mathbf{b} = \mathbf{v}_r$.

IV. Connection to Differential Privacy

We first define differential privacy and then study the connection between differential privacy and correlation dilution. More specifically, we investigate the following question: *if X and Y are differentially private, then is it correct that $\delta(X;Y)$ is small?* We show that the answer to this question is yes when a strong definition of differential privacy is used, and we establish a bound on $\delta(X;Y)$ when X and Y are differentially private. Differential privacy [20] is defined as follows:

Definition 4. For a given ϵ, Y is $\epsilon-$differentially private from X if $\sup_{j \in \mathcal{Y}, i, i' \in \mathcal{X}} \frac{p_{Y|X}(j|i)}{p_{Y|X}(j|i')} \leq e^{\epsilon}$, where we assume the random variables are discrete with finite support[5].

For a thorough explanation of differential privacy and its applications see [21]. The connection between differential privacy and other measures of privacy is studied in [22], [23]. Next, we show that if Y is differentially private from X, then $\delta(X;Y)$ can be bounded from above.

Theorem 2. *If Y is $\epsilon-$differentially private from X, then for $k-$th singular value of matrix Q_{XY}, we have*

$$\sigma_k \leq \frac{1}{\sqrt{k-1}} (e^{\epsilon} - 1) \sqrt{e^{\epsilon}}. \tag{3}$$

In particular, we have $\delta(X;Y) \leq \frac{1}{\sqrt{|\mathcal{X}|-1}} (e^{\epsilon} - 1) \sqrt{e^{\epsilon}}$.

Proof: For any $i, i' \in \mathcal{X}$ and $j \in \mathcal{Y}$, we have $p_{Y|X}(j|i) \leq e^{\epsilon} p_{Y|X}(j|i')$. We multiply both sides with $p_X(i')$ and take the summation over all $i' \in \mathcal{X}$ to obtain

$$p_{Y|X}(j|i) \leq e^{\epsilon} p_Y(j) \text{ for any } i \in \mathcal{X},\ j \in \mathcal{Y}. \tag{4}$$

We arbitrarily choose $i_0 \in \mathcal{X}$ and consider the matrix $\tilde{Q}$

[4]The indicator function $\mathbf{1}\{x = i\}$ is one if $x = i$ and zero otherwise.

[5]The original definition of differential privacy is that $\sup_{j \in \mathcal{Y}, i \sim i' \in \mathcal{X}} \frac{p_{Y|X}(j|i)}{p_{Y|X}(j|i')} \leq e^{\epsilon}$, where $i \sim i'$ denotes that i and i' are neighbors. The notion of neighboring can have multiple definitions as described in [20]. The definition presented here is *local differential privacy*.

defined as

$$\tilde{Q}(i,j) = \frac{\sqrt{p_X(i)}}{\sqrt{p_Y(j)}} p_{Y|X}(j|i_0).$$

Note that since $\frac{\tilde{Q}(i,j)}{\tilde{Q}(i_0,j)} = \frac{\sqrt{p_X(i)}}{\sqrt{p_X(i_0)}}$, all rows of the matrix $\tilde{Q}$ are a multiplicative of its i_0-th row, which results in $\text{rank}(\tilde{Q}) = 1$. On the other hand, since

$$\sum_{i\in\mathcal{X}}\sum_{j\in\mathcal{Y}} \sqrt{p_X(i)}\tilde{Q}(i,j)\sqrt{p_Y(j)} = \sum_{i\in\mathcal{X}}\sum_{j\in\mathcal{Y}} p_{Y|X}(j|i_0)p_X(i) = 1$$

the largest singular value of matrix $\tilde{Q}$ is one and the rest of singular values are zero. Next, we bound the Frobenius norm of the difference between Q and $\tilde{Q}$.

$$||Q_{XY} - \tilde{Q}||_{\text{F}}^2 = \sum_{i\in\mathcal{X},\ j\in\mathcal{Y}} \Big(Q_{XY}(i,j) - \tilde{Q}(i,j)\Big)^2$$

$$= \sum_{i\in\mathcal{X},\ j\in\mathcal{Y}} p_X(i)p_{Y|X}(j|i)\left(\frac{p_{Y|X}(j|i)}{p_Y(j)}\right)\left(\frac{p_{Y|X}(j|i_0)}{p_{Y|X}(j|i)} - 1\right)^2$$

$$\leq \sum_{i\in\mathcal{X},\ j\in\mathcal{Y}} p_X(i)p_{Y|X}(j|i)e^{\epsilon}\left(e^{\epsilon}-1\right)^2 = e^{\epsilon}\left(e^{\epsilon}-1\right)^2,$$

where we used the definition of differential privacy and (4) to obtain the last inequality. Using Hoffman-Wielandt inequality (see e.g. [24], Corollary 7.3.5) and the previous relation, we obtain

$$\sum_{i=2}^{r} \sigma_i^2 \leq e^{\epsilon}\left(e^{\epsilon}-1\right)^2.$$

For $k-$th singular value, we have $\sum_{i=2}^{r}\sigma_i^2 \geq (k-1)\sigma_k^2$. We combine the two previous relations to obtain $\sigma_k \leq \frac{1}{\sqrt{k-1}}\left(e^{\epsilon}-1\right)\sqrt{e^{\epsilon}}$. In particular, by Theorem 1, $\delta(X;Y) = \sigma_r \leq \frac{1}{\sqrt{|\mathcal{X}|-1}}\left(e^{\epsilon}-1\right)\sqrt{e^{\epsilon}}$. This completes the proof. ∎

V. Correlation Dilution with Multiple Observations

A. Problem Statement

Suppose Alice and Bob observe X^n and Y^n, respectively and $\{(X_i,Y_i)\}_{i=1}^n$ are independent. The formulation of correlation dilution becomes

$$\text{Alice: } X^n \to f \in \mathcal{C}_{X^n}, \quad \text{Bob: } Y^n \to g \in \mathcal{C}_{Y^n}$$
$$\text{Objective: } \min_f \max_g \mathbb{E}[f(X^n)g(Y^n)].$$

B. Characterization of Correlation Dilution with Multiple Observations

Proposition 1. *Let (X_1,Y_1) and (X_2,Y_2) be two independent random variables distributed drawn form p_{X_1,Y_1} and p_{X_2,Y_2}. We have*

$$\delta(X_1,X_2;Y_1,Y_2) = \delta(X_1;Y_1)\delta(X_2;Y_2). \tag{5}$$

Furthermore, if we let f_1^*, f_2^*, and f^* denote Alice's optimal choice of functions for random variables (X_1,Y_1), (X_2,Y_2), and (X_1X_2,Y_1Y_2). Similarly, if we let g_1^*, g_2^*, and g^* denote Bob's optimal choice of functions. We have $f^* = f_1^* \otimes f_2^*$ and $g^* = g_1^* \otimes g_2^*$.

Proof: If either $\delta(X_1;Y_1) = 0$ or $\delta(X_2;Y_2) = 0$, the result follows directly. Now assume that both $\delta(X_1;Y_1) > 0$ and $\delta(X_2;Y_2) > 0$. We will use tensorization of principal inertia components (see e.g. [19]):
Let (X_1,Y_1) and (X_2,Y_2) be independent random variables distributed drawn from p_{X_1,Y_1} and p_{X_2,Y_2}, respectively. Let $Q_{X_1Y_1}$, $Q_{X_2Y_2}$, and $Q_{X^2Y^2}$ denote the Q matrix of random variable (X_1,Y_1), (X_2,Y_2), and (X_1X_2,Y_1Y_2). We have $Q_{X^2Y^2} = Q_{X_1Y_1} \otimes Q_{X_2Y_2}$ and its set of singular values is

$$\{\sigma_i^{(1)}\sigma_i^{(2)} \ : \ \sigma_i^{(1)} \in \text{Singular}(Q_{X_1Y_1}), \sigma_i^{(2)} \in \text{Singular}(Q_{X_2Y_2})\}.$$

Using Theorem 1 and tensorization, we obtain

$$\delta(X_1,X_2;Y_1,Y_2) = \delta(X_1;Y_1)\delta(X_2;Y_2),$$

$f^* = f_1^* \otimes f_2^*$ and $g^* = g_1^* \otimes g_2^*$, which completes the proof. ∎

Corollary 1. *Let $(X_1,Y_1),\ldots,(X_n,Y_n)$ be n i.i.d random variables distributed drawn from $p_{X,Y}$. We have*

$$\delta(X^n;Y^n) = \delta(X;Y)^n \leq \sigma_r^n, \tag{6}$$

with equality if $\delta(X;Y) > 0$. Moreover, if (X_i,Y_i) are independent random variables distributed as $p_{X_iY_i}$ for $i = 1,\ldots,n$, then we obtain

$$\delta(X^n;Y^n) = \prod_{i=1}^{n} \delta(X_i;Y_i). \tag{7}$$

Proof: The proof follows by induction on n and using Theorem 1. ∎

Remark 2. If $\delta_r < 1$, we have that $\lim_{n\to\infty}\delta(X^n,Y^n) = \lim_{n\to\infty}\delta_r^n = 0$. Therefore, as n goes to infinity, Alice can achieve full dilution exponentially fast, meaning that she can choose a function that is uncorrelated with any function that Bob can choose. This may appear counter-intuitive at first, since a natural worry about security arises when the number of observations increases. Since the Mutual information between X^n and Y^n is higher than between X and Y, achieving privacy with (X^n,Y^n) seems to be harder than with (X,Y). However, we have $\delta(X^n,Y^n) \leq \delta(X;Y)$, meaning that Alice can dilute better when both Alice and Bob have n i.i.d. copies. The intuition behind this observation is that if Alice has more observations, then she can better exploit the inherent uncertainty of X^n given an observation of Y^n. In other words, even though $I(X^n;Y^n)$ grows, $H(X^n|Y^n)$ also grows, and Alice can find a mapping of X^n such that $f(X^n)$ cannot be reliably inferred from Y^n.

Next, we find a necessary and sufficient condition under which $\delta(X;Y) < 1$ holds.

Lemma 1. *For a given $p_{X,Y}$, $\delta(X;Y) < 1$ if and only if X is not a deterministic function of Y.*

Proof: Let $X = z(Y)$. Thus, letting $g(Y) = f(z(Y))$ we have $\delta(X;Y) = \min_f \mathbb{E}[f(X)f(z(Y))] = 1$. We now show the opposite direction. Suppose that $\delta(X;Y) = 1$. We show that either X or Y is a function of the other one. Suppose without loss of generality $|\mathcal{X}| \leq |\mathcal{Y}|$. Since $\sigma_1 = 1$, the equality $\delta(X;Y) = 1$ shows that all singular values of Q_{XY} are one. Therefore, all eigenvalues of $Q_{XY}Q'_{XY}$ are one. Matrix $Q_{XY}Q'_{XY}$ is symmetric, which shows that $Q_{XY}Q'_{XY}$ is equal to identity matrix. Next, we show that there exists

no $j \in \mathcal{Y}$ such that $p_{XY}(i,j) > 0$ and $p_{XY}(i',j) > 0$ for $i \neq i' \in \mathcal{X}$. Assume the contrary and consider the entry at (i,i') of $Q_{XY}Q'_{XY}$, which must be zero. We have that

$$[Q_{XY}Q'_{XY}]_{i,i'} = \sum_{y\in\mathcal{Y}} \frac{p_{XY}(i,y)p_{XY}(i',y)}{p_Y(y)\sqrt{p_X(i)p_X(i')}} \geq \frac{p_{XY}(i,j)p_{XY}(i',j)}{p_Y(j)\sqrt{p_X(i)p_X(i')}} > 0,$$

which is a contradiction. The fact that there exist no $j \in \mathcal{Y}$ such that $p_{X,Y}(i,j) > 0$ and $p_{X,Y}(i',j) > 0$ for $i \neq i' \in \mathcal{X}$ guarantees that for any j, there exist only one i with $p_{X,Y}(i,j) > 0$. This establishes that random variable X is a deterministic function of random variable Y. ∎

Example 1 (Discretion versus misinformation).

- **Misinformation:** Let $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$. For $i = 1, 2$, assume that X_1, X_2 are i.i.d. uniform random variables over $\{0, 1\}$, and Y_i is the result of passing X_i through a binary symmetric channel with cross-over probability $p < \frac{1}{2}$. We have $\delta(X;Y) = (1-2p)^2$ and

$$f^*(00) = f^*(11) = 1,\ f^*(01) = f^*(10) = -1,$$
$$g^*(00) = g^*(11) = 1,\ g^*(01) = g^*(10) = -1.$$

- **Discretion:** Let $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$. For $i = 1, 2$, assume now that X_i has an i.i.d. uniform distribution over $\{0, 1\}$ and Y_i is the result of passing X_i through a binary erasure channel with error probability $p < \frac{1}{2}$. We have $\delta(X;Y) = (1-p)$ and

$$f^*(00) = f^*(11) = 1,\ f^*(01) = f^*(10) = -1,$$
$$g^*(00) = g^*(11) = \frac{1}{1-p},\ g^*(01) = g^*(10) = -\frac{1}{1-p},$$
$$g^*(0e) = g^*(e0) = g^*(1e) = g^*(e1) = 0.$$

  Note that $(1-2p)^2 < (1-p)$ (for $p \leq \frac{1}{2}$). This shows that BSC better dilute Alice's function, comparing to BEC.
- In general, for n-fold product of the BSC and BEC with X_i uniformly distributed, the function f^* is the parity bit, i.e., we have that $f^*(x_1, \ldots, x_n) = (-1)^{\sum_{i=1}^n x_i}$.

Remark 3. The formulation of the *correlation distillation* problem studied in [10], [16], [17] is as follows:

$$\text{Alice: } X^n \to f \in \mathcal{C}_{X^n}, \quad \text{Bob: } Y^n \to g \in \mathcal{C}_{Y^n}$$
$$\text{Objective: } \max_{f,g} \mathbb{E}[f(X^n)g(Y^n)],$$

where, in contrast with correlation dilution problem, both Alice and Bob intend to maximize correlation without interaction with each other. The answer to this problem also relates to principal inertia components, and in particular, to the maximal correlation $\rho_m(X;Y)$.

VI. Conclusion

We considered a setting where a legitimate party, Alice, has an observation of random variable X, whereas an attacker, Bob, has access to a random variable Y dependent on X drawn from a joint distribution $p_{X,Y}$. Alice's goal is to produce a function of her data that is uncorrelated with (has small correlation with) any function that Bob can produce. We defined dilution coefficient, denoted by $\delta(X;Y)$, as the fundamental minimum correlation that Alice can achieve. We characterized dilution coefficient in terms of the minimum principal inertia component of p_{XY} and we explicitly found the optimal function to achieve it. We then established that if Y is ϵ-differentially private from X, then $\delta(X;Y)$ can be bounded in terms of ϵ. Finally, we considered the case where Alice and Bob have access to i.i.d. copies of $\{(X_i, Y_i)\}_{i=1}^n$, and showed that $\delta(X^n;Y^n) = \delta(X;Y)^n \to 0$ (if X is not a deterministic function of Y, then $\delta(X;Y) < 1$). This implies as n grows, dilution coefficient vanishes exponentially and Alice can achieve full correlation dilution.

References

[1] D. Kifer and B.-R. Lin, "An axiomatic view of statistical privacy and utility," *Journal of Privacy and Confidentiality*, 2012.

[2] S. Gaw and E. W. Felten, "Password management strategies for online accounts," in *Proceedings of the second symposium on Usable privacy and security*, 2006.

[3] D. Florencio and C. Herley, "A large-scale study of web password habits," in *Proceedings of the 16th international conference on World Wide Web*, 2007.

[4] M. J. Greenacre, *Theory and applications of correspondence analysis*, 1984.

[5] F. Calmon, M. Varia, M. Médard, M. Christiansen, K. Duffy, and S. Tessaro, "Bounds on inference," in *Allerton Conference on Communication, Control, and Computing*, 2013.

[6] H. O. Hirschfeld, "A connection between correlation and contingency," in *Mathematical Proceedings of the Cambridge Philosophical Society*. Cambridge Univ Press, 1935.

[7] H. Gebelein, "Das statistische problem der korrelation als variations-und eigenwertproblem und sein zusammenhang mit der ausgleichsrechnung," *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 1941.

[8] O. Sarmanov, "Maximum correlation coefficient (nonsymmetric case)," *Selected Translations in Mathematical Statistics and Probability*, 1962.

[9] A. Rényi, "On measures of dependence," *Acta mathematica hungarica*, 1959.

[10] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM Journal on Applied Mathematics*, 1975.

[11] V. Anantharam, A. Gohari, S. Kamath, and C. Nair, "On maximal correlation, hypercontractivity, and the data processing inequality studied by erkip and cover," *arXiv preprint arXiv:1304.6133*, 2013.

[12] Y. Polyanskiy, "Hypothesis testing via a comparator," in *International Symposium on Information Theory (ISIT)*, 2012.

[13] F. Calmon, M. Varia, and M. Médard, "An exploration of the role of principal inertia components in information theory," in *Information Theory Workshop (ITW)*, 2014.

[14] C. T. Li and A. E. Gamal, "Maximal correlation secrecy," *arXiv preprint arXiv:1412.5374*, 2014.

[15] A. D. Wyner, "The common information of two dependent random variables," *IEEE Transactions on Information Theory*, 1975.

[16] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, 1973.

[17] A. Bogdanov and E. Mossel, "On extracting common random bits from correlated sources," *IEEE Transactions on Information Theory*, 2011.

[18] S. Kamath and V. Anantharam, "Non-interactive simulation of joint distributions: The hirschfeld-gebelein-rényi maximal correlation and the hypercontractivity ribbon," in *Allerton Conference on Communication, Control, and Computing*, 2012.

[19] W. Kang and S. Ulukus, "A new data processing inequality and its applications in distributed source and channel coding," *IEEE Transactions on Information Theory*, 2011.

[20] C. Dwork, "Differential privacy," in *Automata, languages and programming*. Springer, 2006.

[21] C. Dwork, A. Roth *et al.*, "Foundations and trends® in theoretical computer science," *Foundations and Trends® in Theoretical Computer Science*, 2014.

[22] A. Makhdoumi and N. Fawaz, "Privacy-utility tradeoff under statistical uncertainty." in *Allerton Conference on Communication, Control, and Computing*, 2013.

[23] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Me'dard, "Form the information bottleneck to the privacy funnel," in *Information Theory Workshop (ITW)*, 2014.

[24] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.