On the Relation Between the Common Information Dimension and Wyner Common Information

Osama A. Hanna*, Xinlin Li*, Suhas Diggavi and Christina Fragouli
Electrical and Computer Engineering Department
University of California, Los Angeles
Los Angeles, CA, USA
Email: {ohanna, xinlinli, suhasdiggavi, christina.fragouli}@ucla.edu

2024 IEEE International Symposium on Information Theory (ISIT) | 979-8-3503-8284-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISIT57864.2024.10619484

***Abstract*—In this paper, we are interested in the regime where the common information between two Gaussian random vectors (X, Y) can be (or can approach) infinity. We ask two main questions: what is the rate of growth for common information from a finite to an infinite number of bits, as the dependency between the variables increases? and how well can we "approximately" simulate a pair of random variables (X, Y) with infinite common information using a finite number of shared bits? We analytically prove that the answer to both of these questions depends on the common information dimension $d(X, Y)$ between X and Y, that we introduced in our recent work [1]. Our work characterizes in a closed form the asymptotic behaviors, by building a connection to singular values associated with the covariance matrix $\mathbf{\Sigma}$ of (X, Y). We conclude the paper by providing numerical evaluation results that indicate fast convergence to the asymptotic regime.**

I. INTRODUCTION

Quantifying the common information between random variables is a problem with a long history in information theory [2]–[6], and has found application in diverse areas including source coding [7]–[9], cryptography [10]–[12] and learning [13]–[16]. A popular operational meaning comes from distributed simulation, where the common information captures the amount of shared randomness needed to simulate a joint target distribution [3]. In this paper, we promote our understanding in the regime where the common information between random variables can be (or can approach) infinity.

We will illustrate the scope of this paper through an example. Let $X = [X_1, V]^\top, Y = [Y_1, V]^\top$ be two Gaussian random vectors with X_1, Y_1, V independent scalar variables. The common information between X and Y is captured by V, which is a continuous scalar variable with infinite entropy - and thus the common information between X and Y, as calculated for instance in [17], is also infinite. To address this, in our recent work [1] we introduced the notion of common information dimension $d(X, Y)$, and showed that, for Gaussian variables, $d(X, Y)$ can be calculated as

$$d(X, Y) = \text{rank}(\mathbf{\Sigma}_X) + \text{rank}(\mathbf{\Sigma}_Y) - \text{rank}(\mathbf{\Sigma}),$$

where $\mathbf{\Sigma}_X = \mathbb{E}(XX^\top)$, $\mathbf{\Sigma}_Y = \mathbb{E}(YY^\top)$ and $\mathbf{\Sigma}$ is the joint covariance matrix of the vector $[X, Y]^\top$. Note the discontinuity in the space of common information: if $\text{rank}(\mathbf{\Sigma}_X) + \text{rank}(\mathbf{\Sigma}_Y) = \text{rank}(\mathbf{\Sigma})$ then $d(X, Y) = 0$ and the common information can be described using a finite number of bits; while if $\text{rank}(\mathbf{\Sigma}_X) + \text{rank}(\mathbf{\Sigma}_Y) > \text{rank}(\mathbf{\Sigma})$ the common information dimension takes discrete values and the common information measured in bits becomes infinite (in this second case, we say that X and Y are jointly singular). In our toy example, X, Y are jointly singular with $d(X, Y) = 1$.

In this paper, we ask two questions: (1) How fast does the common information grow, from a finite to an infinite number of bits, as the dependency between variables increases? and (2) Can we "approximately" simulate a pair of random variables (X, Y) using a finite number of shared bits, even though their common randomness is infinite? How large is the approximate common information for a certain approximation error?

We answer these questions for the case of Gaussian random vectors (of arbitrary dimension). To explore the first question, we consider a sequence of nearly singular Gaussian pairs with decreasing distances to a jointly singular target distribution. We show that the common information of the sequence grows as $\frac{1}{2}\log(1/\epsilon)d(X, Y)$, where $d(X, Y)$ is the common information dimension of the target distribution, and ϵ measures distance to the singular target distribution. For the second question, we define the ϵ-approximation common information as the minimum amount of common information between random variables that approximate a target distribution within a given error. We prove that, in this case as well, the approximate common information grows as $\frac{1}{2}\log(1/\epsilon)d(X, Y)$, where $d(X, Y)$ is the common information dimension of the target distribution, and ϵ is the approximation error (we comment on this similarity in Section III). Our proofs build on a new connection we make between the approximate common information and singular values associated with the covariance matrix $\mathbf{\Sigma}$ of (X, Y).

In summary, our results characterize the common information of Gaussian vectors in the nearly infinite regime, and establish a new link between common information and the common information dimension $d(X, Y)$. This offers a new interpretation for $d(X, Y)$. Our work also helps understand the quantity of common shared bits needed for a distributed simulation to achieve a desired level of accuracy. We illustrate this through numerical evaluations in Section IV, where we

$*$ Equal contribution.

The work was supported in part by NSF grants 2139304, 2007714, 2221871, 2146838, 1955632, and the Army Research Laboratory grant under Cooperative Agreement W911NF-17-2-0196.

show for instance that to simulate a target distribution with $d(X,Y)=5$ (as described in Section IV Example 2), with accuracy 2^{-5} we need to share 19 bits, while we can achieve a (very high) accuracy of 2^{-20} using around 53 bits.

Related work. One of the well-known classical notions of common information is Wyner's [3]. It is defined as the minimum amount of common randomness (in bits) that enables the distributed simulation of a pair of discrete random variables. [8], [18]–[20] generalize Wyner's common information to continuous sources. However, because of the complexity of the problem, the closed-form solution for continuous sources is only available for Gaussian random variables [8], [9], [17]. These works are useful when the common information only contains discrete variables (even when the sources are continuous); for scenarios where continuous variables are required for the distributed simulation, recently, the common information dimension [1] is defined as the minimum dimension of a random variable (within a certain class of functions) that enables distributed simulation of a set of random variables. A closed-form solution for the Gaussian vector was calculated based on the rank of the covariance matrices, when the common variables is a linear function of the sources. Unlike this work, both Wyner's notion and the common information dimension ask for an auxiliary variable that makes a given pair conditionally independent and they target the exact generation of a given distribution.

Wyner also describes two natural relaxations in [3]: (i) one replaces the conditional independence with a bounded conditional mutual information; (ii) the other allows a small distance between the generated and the target distributions, measured by Kullback–Leibler (KL) divergence. However, these were only analyzed in discrete settings. Recently, [9] studies the first relaxation in the case of Gaussian random variables. However, this version of the relaxed common information is still infinite when singular distributions are involved. In a separate study, [21] explores a related, but different, problem of exchanging a small number of bits to break/reduce the dependency between distributed source. On the other hand, this paper considers relaxation (ii) (with a different distance[1]) which allows an approximate generation when the sources can be continuous and the distributions may be singular.

Paper organization. We review preliminary results on the common information and the common information dimension of Gaussian vectors and introduce our problem formulations in Section II. We present our main results on the growth rate of the common information in Section III. We present the numerical evaluation in Section IV.

II. Preliminaries and Problem Statement

A. Notation

We use capital letters to represent (vectors of) random variables. We use d_X to denote the dimension of a random variable vector X. Since the quantity we are interested in (common information) is independent of the choice of mean values, we assume without loss of generality that all variables have zero mean. For a pair of zero-mean random variables (X,Y), we use $\mathbf{\Sigma}=\begin{bmatrix}\mathbf{\Sigma}_X & \mathbf{\Sigma}_{XY}^\top\\ \mathbf{\Sigma}_{XY} & \mathbf{\Sigma}_Y\end{bmatrix}$ to denote their covariance matrix, where $\mathbf{\Sigma}_X=\mathbb{E}(XX^\top)$ and $\mathbf{\Sigma}_Y=\mathbb{E}(YY^\top)$ are the marginal covariance matrices, and $\mathbf{\Sigma}_{XY}=\mathbb{E}(XY^\top)$ is the cross-covariance matrix. We say that X, Y are *jointly singular* if

$$\text{rank}(\mathbf{\Sigma})<\text{rank}(\mathbf{\Sigma}_X)+\text{rank}(\mathbf{\Sigma}_Y). \tag{1}$$

Our proofs show that we can assume without loss of generality that the marginal covariances are non-singular.

B. Common Information and Common Information Dimension

The (Wyner's) common information $C(X,Y)$ between random variables X and Y is defined in [3] as

$$C(X,Y):=\min_{X-W-Y}\quad I(X\,Y;W), \tag{2}$$

where $X-W-Y$ abbreviates a Markov chain; i.e., X and Y are conditionally independent given W. The general formula of common information between Gaussian vectors $X\in\mathbb{R}^n$ and $Y\in\mathbb{R}^n$ is given in [17] as

$$C(X,Y)=\frac{1}{2}\sum_{i=1}^{n}\log\frac{1+\rho_i}{1-\rho_i}, \tag{3}$$

where $\rho_1,\ldots,\rho_n$ are the singular values of the normalized cross-covariance matrix $\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}$, and $\mathbf{\Sigma}_X^{-1/2},\mathbf{\Sigma}_Y^{-1/2}$ are defined using pseudo-inverse when needed. Observe that when X and Y are jointly singular (i.e., (1) holds and thus $\rho_i=1$ for some i in (3)), the common information $C(X,Y)$ is infinite. In this case, a real-valued random variable W is needed to represent the common randomness and its complexity can be quantified by the common information dimension [1]. The common information dimension $d_{\mathcal{F}}(X,Y)$ of random variables X,Y with respect to a class of functions $\mathcal{F}$, is defined as[2]

$$d_{\mathcal{F}}(X,Y):=\min\{d_W|W\in\mathcal{W}_{\mathcal{F}}\}, \tag{4}$$

$$\text{where } \mathcal{W}_{\mathcal{F}}=\{W|\ \exists\, V,g:\ \mathbb{R}^{d_X}\to\mathbb{R}^{d_W}\in\mathcal{F},\text{such that } X\perp\!\!\!\perp Y|(V,W),\ H(V)<\infty,\ W=g(X,Y)\}.$$

That is, $d_{\mathcal{F}}(X,Y)$ measures the minimum dimension of an auxiliary random variable W that can break the dependency between X and Y, thus enabling distributed simulation. For Gaussian variables X,Y and with respect to the class of linear function, we can calculate it in closed form as [1]:

$$d(X,Y)=\text{rank}(\mathbf{\Sigma}_X)+\text{rank}(\mathbf{\Sigma}_Y)-\text{rank}(\mathbf{\Sigma}). \tag{5}$$

Remark 1. There are two variants of the common information dimension, the Wyner and the Gács-Körner (GK) version,

[1]Note that the KL divergence between any singular and non-singular distributions is always infinite, it is not suitable for the task of approximating a singular distribution with a non-singular one.

[2]W is restricted to be from a class of functions to avoid considering bijections between $\mathbb{R}^n$ and $\mathbb{R}$ which are unstable and not implementable [22].

which share the same solution in the case of two Gaussian variables and the class of linear functions. Details can be found in [1]. In this paper, we do not consider the GK-version of the approximation problem since the GK common information has an inherent discontinuity. In particular, it is easy to see that if the Gaussian sources are singular, then the GK common information is infinite (as is the case for the Wyner as well), however, if they are approximated by any non-singular Gaussian distribution, then the GK common information of the approximate distribution is zero. Hence, the GK version of the common information is not suitable for such approximations. Thus in this paper, we exclusively focus on the Wyner version.

C. Problem Statement

In this paper we ask the following two questions.

1) Common information of nearly singular sources: This formulation aims to study the growth rate of the common information for a sequence of pairs of random variables that approach joint singularity. In particular, let X, Y be Gaussian random variables with $\text{rank}(\mathbf{\Sigma}) < \text{rank}(\mathbf{\Sigma}_X) + \text{rank}(\mathbf{\Sigma}_Y)$, and hence, $d(X,Y) \geq 1$ and $C(X,Y) = \infty$. Let $\{(X_\epsilon, Y_\epsilon)\}_{\epsilon>0}$ be a sequence of Gaussian random variables satisfying

$$\mathbf{\Sigma}_{X_\epsilon} = \mathbf{\Sigma}_X, \mathbf{\Sigma}_{Y_\epsilon} = \mathbf{\Sigma}_Y, \text{ and } \forall\, i,\ \ |\rho_i(\epsilon) - \sigma_i| = \epsilon, \tag{6}$$

where $\{\sigma_i\}$ and $\{\rho_i(\epsilon)\}$ are the singular values of $\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}$ and $\mathbf{\Sigma}_{X_\epsilon}^{-1/2}\mathbf{\Sigma}_{X_\epsilon Y_\epsilon}\mathbf{\Sigma}_{Y_\epsilon}^{-1/2}$ respectively, in a decreasing order. These requirements ensure that (X_ϵ, Y_ϵ) remain non-singular (and thus have finite common information), while the joint distribution of (X_ϵ, Y_ϵ) converges to that of X, Y as $\epsilon \downarrow 0$.

Remark 2. The conditions in (6) force each singular value of $\mathbf{\Sigma}_{X_\epsilon}^{-1/2}\mathbf{\Sigma}_{X_\epsilon Y_\epsilon}\mathbf{\Sigma}_{Y_\epsilon}^{-1/2}$ to go to the corresponding singular value of $\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}$ at an identical rate ϵ. This enables us to study how the common information increases as a function of ϵ. It is easy to show that, the same results we prove also hold when considering different convergence rates for each singular value, provided that these rates are of the same order, meaning they differ only by multiplicative constants.

Recall that from (3), the common information is infinite when $\sigma_i = 1$ for some i. Consider a sequence of covariance matrices that have singular values satisfying (6) when $\sigma_i = 1$ and share the same singular values with the target distribution for all other indices, i.e., $\rho_i = \sigma_i$ when $\sigma_i \neq 1$. It is easy to see that the same results we establish assuming the condition in (6) holds, also extend for the described sequence as well.

2) ϵ-approximation common information: This formulation looks at approximating a pair of Gaussian random variables X, Y that are jointly singular ($C(X,Y) = \infty$) with Gaussian random variables $\hat{X}, \hat{Y}$ that (i) are non-singular ($C(\hat{X}, \hat{Y})$ is finite) and (ii) have a distribution close to the distribution of X, Y. In other words, we ask, if we are restricted to using a finite number of bits as common information, how well can we (approximately) simulate X, Y.

We use the Frobenius-norm between covariance matrices to measure how close two Gaussian distributions are. For some $\epsilon > 0$, we define the *ϵ-approximation common information* as

$$C_\epsilon(X,Y) := \min_{\|\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}\|_F \leq \epsilon} C(\hat{X}, \hat{Y}), \tag{7}$$

where the optimization is over all pairs $(\hat{X}, \hat{Y})$ with covariance matrix $\hat{\mathbf{\Sigma}}$ and $\|\cdot\|_F$ is the Frobenius norm of a matrix.

Remark 3. The feasible solution set of (7) includes a special instance[3]: $\hat{X} = X + U_X, \hat{Y} = Y + U_Y$, where U_X and U_Y are independent Gaussian perturbations with small variance δ_i, such that $\sqrt{\sum_i^{d_X+d_Y} \delta_i^2} \leq \epsilon$.

Remark 4. The results on $C_\epsilon(X,Y)$ extend if we replace the Frobenius norm with any distribution distance $\text{dist}(XY, \hat{X}\hat{Y})$ that satisfies $a\|\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}\|_F \leq \text{dist}(XY, \hat{X}\hat{Y}) \leq b\|\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}\|_F$ for all Gaussian variables $XY, \hat{X}\hat{Y}$ and some constants a and b.

Remark 5. Note that formulation 1 in (6) studies a more restricted set of sequences than the sequences included in the feasible set of the optimization problem in (7). However, the result we show for formulation 1 is stronger as it holds for *all sequences* that satisfy the condition in (6). In contrast, the results in formulation 2 only hold for the sequence with the minimum common information (that achieves the optimal value of the minimization problem). It can be easily shown that there exist sequences in the feasible set of formulation 2 that have different asymptotics. For example, if some singular values of the approximation matrix take the value 1 or approach 1 at a rate different from $\Theta(\epsilon)$(e.g., ϵ^2 or 2^ϵ).

III. Main Results

In this section, we present our main results and proof outlines for the two formulations described in Section II-C. The detailed proofs are provided in Appendices D, E, and F in [23].

Before stating our main results, we present two properties of covariance matrices and the common information dimension, which are important to Theorems 1 and 2. As stated in (3) the common information is determined by the singular values of the normalized cross-covariance matrix $\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}$. Lemma 1 proves a bound on these singular values.

Lemma 1: Let $X \in \mathbb{R}^{d_X}$ and $Y \in \mathbb{R}^{d_Y}$ be jointly Gaussian variables with covariance matrix $\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_X & \mathbf{\Sigma}_{XY}^\top \\ \mathbf{\Sigma}_{XY} & \mathbf{\Sigma}_Y \end{bmatrix}$, and $d = \min\{d_X, d_Y\}$. Then the singular values of $\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}$, denoted as $\{\sigma_i\}_{i=1}^d$, satisfy

$$0 \leq \sigma_i \leq 1, \forall i \in [d] \tag{8}$$

The following lemma shows the relationship between the common information dimension and the singular values of $\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}$, which will enable us to connect the quantities $C(X_\epsilon, Y_\epsilon), C_\epsilon(X,Y)$ with the common information dimension $d(X,Y)$.

[3]Another way of approximation, that quantizes X and Y into discrete variables, is studied in our extended work [23].

Lemma 2: Assume $X \in \mathbb{R}^{d_X}$, $Y \in \mathbb{R}^{d_Y}$ are jointly Gaussian variables with covariance matrix $\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_X & \mathbf{\Sigma}_{XY}^\top \\ \mathbf{\Sigma}_{XY} & \mathbf{\Sigma}_Y \end{bmatrix}$, and $\{\sigma_i\}$ are the singular values of $\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}$. Then the common information dimension between X and Y, with respect to linear functions, satisfies

$$d(X,Y) = \sum_{i=1}^{\min\{d_X,d_Y\}} \mathbb{1}\{\sigma_i = 1\} \tag{9}$$

A. Common information of nearly singular sources

We consider a sequence of pairs of Gaussian random variables $\{(X_\epsilon, Y_\epsilon)\}_{\epsilon>0}$ satisfying (6). The following result shows that the growth rate of the common information $C(X_\epsilon, Y_\epsilon)$ is determined by the common information dimension $d(X,Y)$ with respect to linear functions.

Theorem 1: Let $X \in \mathbb{R}^{d_X}$ and $Y \in \mathbb{R}^{d_Y}$ be a pair of jointly Gaussian random variables, and $\{(X_\epsilon, Y_\epsilon)\}_{\epsilon>0}$ be a sequence as defined in (6). Then the common information $C(X_\epsilon, Y_\epsilon)$ satisfies

$$\lim_{\epsilon\downarrow 0} \frac{C(X_\epsilon, Y_\epsilon)}{\frac{1}{2}\log(\frac{1}{\epsilon})} = d(X,Y) \tag{10}$$

Proof Outline. The main technical challenge in proving Theorem 1 is the fact that there exist multiple sequences of random variables X_ϵ, Y_ϵ, with different values of $C(X_\epsilon, Y_\epsilon)$, that satisfy the constraints in (6). To address this issue, we prove the result by deriving an upper and a lower bound on $C(X_\epsilon, Y_\epsilon)$ that have the same asymptotic behavior.

The proof focuses on showing that $\lim_{\epsilon\downarrow 0} \frac{C(X_\epsilon, Y_\epsilon)}{\frac{1}{2}\log(\frac{1}{\epsilon})} = \sum_i^{\min\{d_X,d_Y\}} \mathbb{1}\{\sigma_i = 1\}$, where $\{\sigma_i\}$ are the singular values of the matrix $\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}$. We prove this by providing an upper and lower bound on $\frac{C(X_\epsilon, Y_\epsilon)}{\frac{1}{2}\log(\frac{1}{\epsilon})}$ that have the same limit when $\epsilon \downarrow 0$. Then we relate $\sum_i^{\min\{d_X,d_Y\}} \mathbb{1}\{\sigma_i = 1\}$ to the common information dimension $d(X,Y)$ using Lemma 2.

B. Approximate simulation

The following result shows that the ϵ-approximation common information $C_\epsilon(X,Y)$, defined in (7), for Gaussian variables grows at a rate determined by the common information dimension $d(X,Y)$ with respect to linear functions.

Theorem 2: Let $X \in \mathbb{R}^{d_X}$ and $Y \in \mathbb{R}^{d_Y}$ be a pair of jointly Gaussian random variables, then

$$\lim_{\epsilon\downarrow 0} \frac{C_\epsilon(X,Y)}{\frac{1}{2}\log(1/\epsilon)} = d(X,Y) \tag{11}$$

Proof Outline. The main technical challenge in proving Theorem 2 is the difficulty in finding a closed form solution of the optimization problem defining $C_\epsilon(X,Y)$. To address this issue, we follow a similar approach as in Theorem 1 by deriving an upper and a lower bound on C_ϵ that have the same asymptotics. However, it turns out that finding upper and lower bounds that have the same asymptotics is more involved than in the case of Theorem 1.

The proof uses the upper and lower bounds, derived as described next, to show that $\lim_{\epsilon\downarrow 0} \frac{C(X_\epsilon, Y_\epsilon)}{\frac{1}{2}\log(\frac{1}{\epsilon})} = \sum_i^{\min\{d_X,d_Y\}} \mathbb{1}\{\sigma_i = 1\}$, where $\{\sigma_i\}$ are the singular values of the matrix $\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}$. From Lemma 2, this concludes the proof of Theorem 2.

Upper Bound. As $C_\epsilon(X,Y)$ is the optimal value of a minimization problem, any feasible solution provides an upper bound. To find a feasible solution we use $\mathbf{\Sigma}_{\hat{X}} = \mathbf{\Sigma}_X, \mathbf{\Sigma}_{\hat{Y}} = \mathbf{\Sigma}_Y$. Then, we design the singular values of $\mathbf{\Sigma}_{\hat{X}}^{-1/2}\mathbf{\Sigma}_{\hat{X}\hat{Y}}\mathbf{\Sigma}_{\hat{Y}}^{-1/2}$, denoted as $\{\rho_i\}$, as follows. We set $\rho_i = \sigma_i$ when $\sigma_i \neq 1$. Recall that choosing a singular value to be 1 results in an infinite value for the common information. Hence, when $\sigma_i = 1$ we choose $\rho_i = 1-\delta$ where δ is the largest value that does not violate the constraint $\|\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}\|_F \leq \epsilon$.

Lower Bound. To find a lower bound, we relax the constraints set $\|\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}\|_F \leq \epsilon$, resulting in a smaller optimal value, to make it possible to find a closed form solution of the problem. The proof of the lower bound hinges on showing that $\|\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}\|_F \leq \epsilon$ implies

$$\|\Lambda - \hat{\Lambda}\|_F \leq c\epsilon, \tag{12}$$

where $\Lambda = \text{diag}(\sigma_i), \hat{\Lambda} = \text{diag}(\rho_i)$ are matrices containing the singular values of $\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}, \mathbf{\Sigma}_{\hat{X}}^{-1/2}\mathbf{\Sigma}_{\hat{X}\hat{Y}}\mathbf{\Sigma}_{\hat{Y}}^{-1/2}$ respectively, and c is a constant that may depend on $\mathbf{\Sigma}_X, \mathbf{\Sigma}_Y$. To further simplify the problem, we remove from the objective function the terms corresponding to $\sigma_i < 1$, and also remove the value $\log(1+\rho_i)$ from each term (recall the common information in (3)). We note that each term in the objective function is non-negative, and hence, removing terms will not increase the optimal solution value. Furthermore, we expect the asymptotics of the common information to be influenced by the singular values corresponding to $\sigma_i = 1$. This results in the following optimization problem

$$\begin{aligned} \min_{\rho} \quad & \frac{1}{2}\sum_{i:\sigma_i=1} \log\frac{1}{1-\rho_i} \\ \text{s.t.} \quad & \sum_{i:\sigma_i=1} (\sigma_i - \rho_i)^2 \leq \epsilon^2,\ 0 \leq \rho_i \leq 1, \end{aligned} \tag{13}$$

which can be solved in a closed form using symmetry and concavity of the log function.

Remark 6. We note that we can efficiently construct random variables for each ϵ with common information that has the asymptotic behavior in Theorem 2 (and thus can be used to approximate the target singular distribution with (nearly) the smallest common information). A possible choice is $\mathbf{\Sigma}_{\hat{X}} = \mathbf{\Sigma}_X, \mathbf{\Sigma}_{\hat{Y}} = \mathbf{\Sigma}_Y, \mathbf{\Sigma}_{\hat{X}}^{-1/2}\mathbf{\Sigma}_{\hat{X}\hat{Y}}\mathbf{\Sigma}_{\hat{Y}}^{-1/2} = U\hat{\Lambda}V$, where U, V are orthonormal matrices of the singular value decomposition of $\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}$, and $\hat{\Lambda}$ can be obtained using the solution of problem (13) (refer to Appendix F in [23] for details).

Remark 7. Why do these two theorems have the same bound? It may seem at first surprising that even though $C(X_\epsilon, Y_\epsilon)$ and $C_\epsilon(X,Y)$ have different definitions, they both grow (nearly) as $\frac{1}{2}d(X,Y)\log(1/\epsilon)$. Indeed, as we observed in Remark 5 the feasible set defining $C_\epsilon(X,Y)$ in (7) contains different sequences of random variables than those satisfying the conditions in (6). However, the proof of Theorem 2 shows that the random variables which minimize the common information satisfy a constraint similar to (6); namely, the singular values ρ_i corresponding to $\sigma_i = 1$ have the same distance to 1, where σ_i and ρ_i are the singular value of $\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}$ and $\mathbf{\Sigma}_{\hat{X}}^{-1/2}\mathbf{\Sigma}_{\hat{X}\hat{Y}}\mathbf{\Sigma}_{\hat{Y}}^{-1/2}$ respectively. Intuitively, to minimize the common information in (3), we need ρ_i to be as far as possible from the value 1, however, the distance constraint in (7) restricts us from choosing ρ_i too far from 1 whenever $\sigma_i = 1$. If one ρ_i is very close to 1, it will dominate the summation in (3) resulting in large common information. Hence, a good solution to (7) distributes the distance budget ϵ evenly across the ρ_i's corresponding to $\sigma_i = 1$.

IV. Numerical Evaluation

In this section, we numerically evaluate the growth rate of $C_\epsilon(X,Y)$ and $C(X_\epsilon, Y_\epsilon)$, as a function of the approximation error ϵ. Next, we present two examples.

Example 1. We let $X \in \mathbb{R}^4$ and $Y \in \mathbb{R}^4$ be jointly Gaussian vectors with zero means and covariance matrices

$$\mathbf{\Sigma}_X, \mathbf{\Sigma}_Y = \begin{bmatrix} 1 & 0.5 & 0 & 0 \\ 0.5 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{\Sigma}_{XY} = \begin{bmatrix} 1 & 0.5 & 0 & 0 \\ 0.5 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.3 \end{bmatrix}.$$

It is evident that $X_1 = Y_1, X_2 = Y_2$ almost surely, and $\text{rank}(\mathbf{\Sigma}) = 5 < \text{rank}(\mathbf{\Sigma}_X) + \text{rank}(\mathbf{\Sigma}_Y)$, thus, X and Y are jointly singular.

Figure 1 illustrates the normalized common information $\frac{C_\epsilon(X,Y)}{\frac{1}{2}\log(1/\epsilon)}$ and $\frac{C(X_\epsilon,Y_\epsilon)}{\frac{1}{2}\log(1/\epsilon)}$ for both formulations 1, 2 in Section II-C plotted against the approximation error ϵ. To calculate $C(X_\epsilon, Y_\epsilon)$, note that there exist multiple sequences[4] $\{(X_\epsilon, Y_\epsilon)\}_{\epsilon>0}$ that satisfy the requirements in (6). Here, we choose two representative sequences and plot the results for both: $\{(\underline{X}_\epsilon, \underline{Y}_\epsilon)\}$ which has the minimum common information among such sequences for all $\epsilon > 0$, and $\{(\overline{X}_\epsilon, \overline{Y}_\epsilon)\}$ which has the maximum common information. We calculate the $C((\underline{X}_\epsilon, \underline{Y}_\epsilon))$ and $C(\overline{X}_\epsilon, \overline{Y}_\epsilon)$ using the closed-form solution in (3) [17]. To calculate the ϵ-approximation common information $C_\epsilon(X,Y)$ we solve the optimization problem in (7) numerically using SciPy [24].

We observe in Figure 1 that both $\frac{C_\epsilon(X,Y)}{\frac{1}{2}\log(1/\epsilon)}$ and $\frac{C(X_\epsilon,Y_\epsilon)}{\frac{1}{2}\log(1/\epsilon)}$ converge to $d(X,Y) = \text{rank}(\mathbf{\Sigma}_X) + \text{rank}(\mathbf{\Sigma}_Y) - \text{rank}(\mathbf{\Sigma}) = 3$ as ϵ approaches 0. Moreover, they reach a value that is close to $d(X,Y)$ (e.g., a value < 4) quickly, even with relatively large values of ϵ. The trade-off between ϵ-approximate common information $C_\epsilon(X,Y)$ and error ϵ also indicates that the common information dimension $d(X,Y)$ provides a theoretical limit on the maximum achievable accuracy given a finite number of bits to represent the common randomness; or equivalently, the minimum number of bits required for the shared randomness to achieve a target simulation accuracy.

Example 2. In this example, we use $X \in \mathbb{R}^7$ and $Y \in \mathbb{R}^7$ with $\mathbf{\Sigma}_X = \mathbf{\Sigma}_Y = \boldsymbol{I}_7$, while we choose the cross-covariance matrices to be a diagonal matrix with $d(X,Y)$ diagonal elements set to be 1 and the rest to be 0.5. In Figure 2, we plot the minimum number of bits required to approximate X,Y versus the different common information dimensions $d(X,Y)$. We use two different levels of accuracy: $\epsilon = 2^{-5}$ and $\epsilon = 2^{-20}$. We observe from Figure 2, that the approximate common information grows linearly with the common information dimension $d(X,Y)$, where the slope is given by $\frac{1}{2}\log(1/\epsilon)$, as we also proved in Theorem 2. In addition, this plot provides a guide on the minimum number of bits that need to be shared to perform the distribution simulation within a given error. For instance, to simulate a target distribution with $d(X,Y) = 5$, we need to share 19 bits to achieve a relatively low 2^{-5} accuracy, or 53 bits to achieve a relatively high 2^{-20} accuracy.

[4] Note that there are at most $2^{\min\{d_X, d_Y\}}$ (X_ϵ, Y_ϵ) that satisfy (6), for each ϵ.

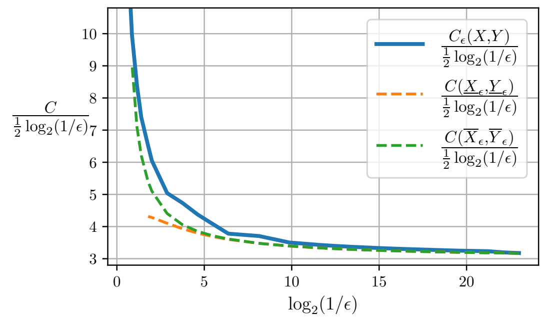

Fig. 1. Growth rate of $C_\epsilon(X,Y)$ and $C(X_\epsilon, Y_\epsilon)$.

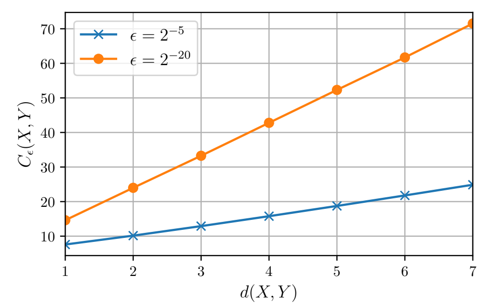

Fig. 2. The approximate common information $C_\epsilon(X,Y)$ vs common information dimension $d(X,Y)$.

V. Conclusion

In this paper, we studied how the common information between a pair of random variables increases as their distance ϵ to jointly singular Gaussian random variables (X,Y) decreases. We also studied the minimum amount of common information required to approximately simulate jointly singular Gaussian random variables X,Y with at most ϵ error. We proved that in both scenarios, the common information grows as $\frac{1}{2}d(X,Y)\log(1/\epsilon)$, where $d(X,Y)$ is the common information dimension of (X,Y) with respect to linear functions. Our results give an interpretation of the common information dimension in the context of approximate simulation and finite common information. It is interesting to note that the common information dimension restricted to linear functions determines the scaling behavior even when we do not impose any linearity constraints on the common information extraction. Future directions include generalizing the results to other distance measures and beyond Gaussian distributions.

REFERENCES

[1] O. A. Hanna, X. Li, S. Diggavi, and C. Fragouli, "Common information dimension," in *2023 IEEE International Symposium on Information Theory (ISIT)*, pp. 406–411, IEEE, 2023.

[2] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, no. 2, pp. 149–162, 1973.

[3] A. Wyner, "The common information of two dependent random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 163–179, 1975.

[4] G. R. Kumar, C. T. Li, and A. El Gamal, "Exact common information," in *2014 IEEE International Symposium on Information Theory*, pp. 161–165, IEEE, 2014.

[5] P. W. Cuff, H. H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4181–4206, 2010.

[6] L. Yu, H. Li, and C. W. Chen, "Generalized common informations: Measuring commonness by the conditional maximal correlation," 2017.

[7] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *The Bell System Technical Journal*, vol. 53, no. 9, pp. 1681–1721, 1974.

[8] G. Xu, W. Liu, and B. Chen, "A lossy source coding interpretation of wyner's common information," *IEEE Transactions on Information Theory*, vol. 62, no. 2, pp. 754–768, 2015.

[9] E. Sula and M. Gastpar, "The gray-wyner network and wyner's common information for gaussian sources," *IEEE Transactions on Information Theory*, vol. 68, no. 2, pp. 1369–1384, 2022.

[10] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. i. secret sharing," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1121–1132, 1993.

[11] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. ii. cr capacity," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 225–240, 1998.

[12] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE transactions on information theory*, vol. 39, no. 3, pp. 733–742, 1993.

[13] Y. Li, M. Yang, and Z. Zhang, "A survey of multi-view representation learning," *IEEE transactions on knowledge and data engineering*, vol. 31, no. 10, pp. 1863–1883, 2018.

[14] T. M. Sutter, I. Daunhawer, and J. E. Vogt, "Generalized multimodal ELBO," in *International Conference on Learning Representations*, 2021.

[15] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with clip latents," *arXiv preprint arXiv:2204.06125*, 2022.

[16] M. Kleinman, A. Achille, S. Soatto, and J. Kao, "Gacs-korner common information variational autoencoder," in *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[17] S. Satpathy and P. Cuff, "Gaussian secure source coding and wyner's common information," in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 116–120, 2015.

[18] K. B. Viswanatha, E. Akyol, and K. Rose, "The lossy common information of correlated sources," *IEEE Transactions on Information Theory*, vol. 60, no. 6, pp. 3238–3253, 2014.

[19] C. T. Li and A. El Gamal, "Distributed simulation of continuous random variables," *IEEE Transactions on Information Theory*, vol. 63, no. 10, pp. 6329–6343, 2017.

[20] L. Yu and V. Y. Tan, "On exact and ∞-rényi common informations," *IEEE Transactions on Information Theory*, vol. 66, no. 6, pp. 3366–3406, 2020.

[21] O. A. Hanna, X. Li, C. Fragouli, and S. Diggavi, "Can we break the dependency in distributed detection?," in *2022 IEEE International Symposium on Information Theory (ISIT)*, pp. 2720–2725, IEEE, 2022.

[22] Y. Wu and S. Verdú, "Rényi information dimension: Fundamental limits of almost lossless analog compression," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3721–3748, 2010.

[23] O. Hanna, X. Li, S. Diggavi, and C. Fragouli, "Common information dimension," 2024. https://arxiv.org/abs/2305.06469.

[24] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.