Differentially Private Distributed Matrix Multiplication: Fundamental Limits on the Accuracy-Privacy Trade-off

Ateet Devulapalli[†], Viveck R. Cadambe[†], Flavio P. Calmon[‡], Haewon Jeong[‡]
[†]Pennsylvania State University, [‡] Harvard University

***Abstract*—The classic BGW algorithm of Ben Or, Goldwasser and Wigderson for secure multiparty computing demonstrates that the secure distributed matrix multiplication over finite fields is possible over $2t+1$ computation nodes, while keeping the input matrices private from every t colluding computation nodes. In this paper, we develop and study a novel coding formulation to explore the trade-offs between computation accuracy and privacy in secure multiparty computing for real-valued data, even with fewer than $2t+1$ nodes, through a differential privacy perspective. For the case of $t = 1$, we develop achievable schemes and converse arguments that bound ϵ - the differential privacy parameter that measures the privacy loss - for a given accuracy level. Our achievable coding schemes are specializations of Shamir secret sharing applied to real-valued data, coupled with appropriate choice of evaluation points. We develop converse arguments that apply for general additive noise based schemes.**

***Index Terms*—Differential privacy, privacy-utility tradeoff, mean square error, secure multiparty computation, coded computing, distributed matrix multiplication.**

I. Introduction

The task of accurate and efficient distributed data processing while preserving data privacy is among the most important engineering problems in modern machine learning. The desire to keep data private inevitably requires the source adding some noise to the data before sharing it with the computation nodes. Secure multiparty computing (MPC) is a paradigm that ensures that data remains private from any t computing nodes, yet ensures computation of functions of the data[1]. The celebrated BGW algorithm [2], [3] provides a method to perform *information-theoretically* private computation of a wide class of functions building on Shamir secret sharing technique [4], which in turn builds on Reed Solomon codes. The goal of this paper is to study coding schemes for secure matrix multiplication for real-valued data.

Consider two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{F}^{L\times L}$, where $\mathbb{F}$ is a field, and a set of P computation nodes. Let $\mathbf{R}_i, \mathbf{S}_i, i = 1, 2 \dots, t$ are statistically independent random matrices. Suppose node i gets $\tilde{\mathbf{A}}_i = p_1(x_i), \tilde{\mathbf{B}}_i = p_2(x_i)$, where, $x_1, x_2, \dots, x_P$ are distinct non-zero scalars and $p_1(x), p_2(x)$ are matrix-valued polynomials:

$$p_1(x) = \mathbf{A} + \sum_{j=1}^{t} \mathbf{R}_j x^j, p_2(x) = \mathbf{B} + \sum_{i=1}^{t} \mathbf{S}_j x_j.$$

If the field $\mathbb{F}$ is finite, and the entries of $\mathbf{R}_i, \mathbf{S}_i, i = 1, 2, \dots, P$ are chosen randomly i.i.d. uniformly over the elements of the field, then the input to any subset $\mathcal{S}$ of t nodes is, in fact, independent of $\mathbf{A}, \mathbf{B}$. If node i computes $\tilde{\mathbf{A}}_i + \tilde{\mathbf{B}}_i$, the sum $\mathbf{A}+\mathbf{B}$ - which is the constant in the polynomial $p_1(x)+p_2(x)$ - can be recovered from the computation output of any $t+1$ of the P nodes by polynomial interpolation. Observe, similarly, that $\tilde{A}_i\tilde{B}_i$ can be interpreted as an evaluation, at $x = x_i$, of the degree $2t$ polynomial $p_1(x)p_2(x)$, whose constant term is $\mathbf{AB}$. Thus, the matrix product can be recovered from any $2t+1$ nodes via polynomial interpolation. The BGW algorithm uses the above coding scheme to perform secure MPC for the universal class of computations that can be expressed as sums and products, while maintaining (perfect) data privacy among every set of t nodes. Because of its universality, the BGW algorithm is a powerful technique and forms the basis of several secure MPC protocols. Notably, an overhead of $2t + 1$ computation nodes (i.e., $t + 1$ redundant nodes) are required to keep the data private from any t computation nodes and perform multiplications, as compared to mere data access as in Shamir secret sharing, or addition/aggregation wherein the computation output can be recovered from $t + 1$ computation nodes. In fact, for more complex functions, the overhead can be prohibitively large inevitably leading to multiple communication rounds [2], or more redundant nodes [**?**]. For instance, for polynomial computations (of which multiplication is a special case), the number of redundant nodes can scale linearly as the number of nodes. Consequently, the overheads of computation as well as communication can be quite large for complicated functions.

We are motivated by the question of reducing the the amount of redundant resources required to enable private distributed computation. We study the canonical and universal operation of multiplication, and aim to answer the following question: Can a set of fewer than $2t + 1$ nodes compute the matrix product by keeping the data (matrices) private from any t nodes? While impossibility results preclude this possibility if we aim for the dual goals of exact recovery of the matrix product, and perfect privacy, we study the question by allowing for approximations on both fronts.

For several machine learning applications that operate over real-valued data, approximate computation of the output typically suffices. Further, a prevalent[1] [5], [6]) paradigm for data privacy in machine learning applications in practice is *differential privacy (DP)* [7], which aims to keep small perturbations of the data private from the nodes. In particular, it requires that the extent of *privacy loss* (see Sec. III) to be bounded by a non-negative parameter ϵ; the smaller the ϵ, the lesser the privacy loss, and the case of matrices $\mathbf{A}, \mathbf{B}$ being independent of the nodes' input corresponds to the special case of $\epsilon = 0$[2]. While the DP framework allows us to tune the degree of privacy, to effectively use the framework in practice, it is important to understand how to set parameter ϵ based on the application at hand (See [8]); our paper aims to bring this understanding to the context of secure matrix multiplication.

Summary of Contributions

Our key technical contribution is an explict analytical characterization of the trade-off between computation accuracy and privacy for the case of $t = 1$, that is, where the data is kept private from every single computation node in the system. We consider the following problem. Assume that a computation node gets an input of the form $\mathbf{A}+\mathbf{R}, \mathbf{B}+\mathbf{S}$ and multiplies them, where $\mathbf{R}, \mathbf{S}$ are random noise matrices that independent of $(\mathbf{A}, \mathbf{B})$ designed to ensure data privacy. The goal of the decoder is to recover an estimate $\tilde{\mathbf{C}}$ of the product $\mathbf{AB}$ from N computation outputs at a certain accuracy level. The noise $\mathbf{R}, \mathbf{S}$ should ensure that the input should be kept ϵ-differentially private from every $t = 1$ node in the system.

We characterize, via an achievable coding scheme and a converse, the trade-off between mean square error $||\tilde{\mathbf{C}} - \mathbf{AB}||_2$ and the DP parameter ϵ. For both the achievable coding scheme and the converse, we follow a two step procedure. We first develop bounds on the mean square error in terms of the expected singular values of the noise matrices $\mathbf{R}, \mathbf{S}$. Then, we bound the DP parameter ϵ in terms of these expected singular values, applying a specific distribution for our achievable scheme, and bounding ϵ over all distributions for the converse. For the case where $\mathbf{A}, \mathbf{B}$ are scalars, our first step translates to a *tight* characterization between the mean square and the standard deviations of $\mathbf{R}, \mathbf{S}$. Our achievable scheme is a specialization of the Shamir secret sharing technique applied for real numbers with a careful choice of evaluation points, followed by a DP analysis. Our converse, notably, makes no assumptions on the structure of codes used, and applies to arbitrary schemes that work for additive noise.

A. Related Work

Differential Privacy and Secure MPC: Several prior works are aimed at connecting secure multiparty computation and differential privacy [9],[10],[11],[12],[13],[14]. Usually to ensure privacy secure MPC protocols often employ computationally expensive or communication expensive algorithms. Differential privacy framework could potentially be used to alleviate some of the complications while still maintaining a degree of privacy. Secure MPC protocols usually have the security requirement that given the output of computation the input must not be inferred, which does not address presence of auxiliary information. For example in [15] the authors were able to identify the medical records of governor of Massachusetts from a publicly released anonymized records, in which any personal identification was removed, by using additional information available to the public. Differential privacy provides a strong privacy guarantee against such linkage attacks [7]. Reference [9] tackles the problem of label private training and their protocols work by releasing differentially private intermediate computations while training, whereas standard MPC protocols use multiple communication rounds to ensure a computation circuit produces secure intermediate outputs. References [10],[11],[14] provides methods in a similar vein for sample aggregation algorithms, private record linkage, and private distributed median computation. In comparison we try to reduce overhead of t redundant nodes and use differential privacy framework to study the impact of such an action on privacy. References [12],[13] consider a problem where each party has an input and computes their desired function based on inputs of other parties. They show a strong and interesting result where the optimal protocol, in the sense that accuracy is maximized simultaneously at all parties, is the randomized response protocol [16], which is providing a random response where truth is revealed with some probability that depends on desired level of privacy, which fits perfectly with the differential privacy framework. They show the result for arbitrary function and arbitrary utilities and our work could be thought of as a specialization to distributed matrix multiplication case, while also providing additional theory specific to this case such as achievability and converse results.

VC: TODO: Secure MPC over reals. Cite papers and mention differences.

Coded Computing

Reducing overhead of redundant nodes has been of interest in our previous work [17] where we reduce the recovery threshold for straggler mitigation problem by allowing for approximate recovery and making evaluations arbitrarily close to 0. We arrive at similar conclusions in our current work explained in the following sections. Reference [18] provides an extension of BGW protocol to straggler mitigation using finite fields. Reference [19] is a further extension of [18] to real field, where they also study privacy accuracy tradeoff for $N \geq 2t+1$, where accuracy is measured in terms of precision lost due to floating point representation. A different straggler mitigation coding scheme along with security given by BGW inspired code is employed by [20] that also focuses on exact recovery. In [21] the authors aim to reduce communication overhead in information theoretically secure MPC matrix mulitplication schemes by using coding and give matching achievable and

[1] See, for example, Google's Tensorflow Privacy Framework.

[2] The Shamir secret sharing in fact can be applied for real-valued data as well. Specifically, by allowing evaluation points x_i to grow arbitrarily large, the DP parameter ϵ can be made arbitrarily small and still allow for perfectly accurate decoding of the matrix product from any $2t+1$ nodes.

converse rates for such a code. In constrast to all these works, we look at private matrix multiplication for $N \geq t+1$ instead of $2t+1$ and also introduce the differential privacy framework. Since at least $2t+1$ responses are needed in order to exactly recover the multiplication output, we measure accuracy in terms of recovery error

Privacy-Utility Trade-offs [22],[23]

II. Background on Differential Privacy

VC: This section can probably be removed, and kept only in the appendix/extended version.

Let $X \in \mathbb{R}$ be a random variable and $f : \mathbb{R} \to \mathbb{R}$ a mapping applied to X. If $f(X)$ is publicly released, an adversary may infer sensitive information about X. Such inferences can be mitigated by an *additive-noise privacy mechanism* which adds noise to $f(X)$ prior to disclosure. Denoting the additive noise by Z, the output of the privacy mechanisms is

$$Y = f(X) + Z.$$

The desideratum of *differential privacy* [7] is to ensure that small perturbations of X cannot be inferred from the disclosed random variable Y. Specifically, the mapping $\mathbb{P}_{Y|X}$ satisfies ϵ-differential privacy [7] for ℓ_1-neighboring inputs if

$$\frac{\Pr(Y \in \mathcal{A} | X = x_1)}{\Pr(Y \in \mathcal{A} | X = x_2)} \leq e^{\epsilon}$$

for any $\mathcal{A} \subseteq \mathbb{R}$ and $|x_1 - x_2| \leq 1$. The level of privacy achieved via noise addition will depend on the distribution of Z and the sensitivity of the function f, defined as

$$\Delta f = \max_{|x_1 - x_2| \leq 1} |f(x_1) - f(x_2)|.$$

Intuitively, ϵ-DP ensures that neighboring inputs x_1 and x_2 are nearly indistinguishable when any output Y is disclosed, thus preserving privacy. Note that ϵ is inversely proportional to variance of noise. Note that ϵ-DP is a property of the distribution of Z, thus saying Z satisfies ϵ-DP is equivalent to saying that $\mathbb{P}_{Y|X}$ satisfies ϵ-DP and we will use such language interchangeably.

III. System Model and Problem Statement

A. Notations

We define $[n] \triangleq \{1, 2, \cdots, n\}$. We use bold fonts for vectors and matrices. We define $(\mathbf{x})_i$ to be the i^{th} component of a vector $\mathbf{x}$ and $(\mathbf{X})_{k,l}$ be the $(k,l)^{\text{th}}$ element of a matrix $\mathbf{X}$. Denote $\kappa(\mathbf{X}), ||\mathbf{X}||_2$ and $||\mathbf{X}||_F$ to be the minimum singular value, ℓ_2 norm and Frobenius norm of a matrix $\mathbf{X}$ respectively. We use $X \sim \mathbb{Q}$ to say that the random variable X has the probability distribution $\mathbb{Q}$.

B. System Model

Although our We consider a computation system with P computation nodes. $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{L \times L}$ are random matrices, and node $i \in [P]$ receives:

$$\tilde{\mathbf{A}}_i = \mathbf{A} + \mathbf{R}_i$$

$$\tilde{\mathbf{B}}_i = \mathbf{B} + \mathbf{S}_i$$

where $\mathbf{R}_i, \mathbf{S}_i \in \mathbb{R}^{L \times L}$ are random matrices such that $(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_P, \mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_P)$ is statistically independent of $(\mathbf{A}, \mathbf{B})$. We denote by $\mathbf{R}, \mathbf{S} \in \mathbb{R}^{L \times PL}$, the following:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 & \ldots & \mathbf{R}_P \end{bmatrix}$$
$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_1 & \mathbf{S}_2 & \ldots & \mathbf{S}_P \end{bmatrix}.$$

In this paper we assume no shared randomness between $\mathbf{R}, \mathbf{S}$, i.e., they are statistically independent: $\mathbb{P}_{\mathbf{R},\mathbf{S}} = \mathbb{P}_{\mathbf{R}}\mathbb{P}_{\mathbf{S}}$. We denote by $\mathcal{P}_{\mathbf{R},\mathbf{S}}$ as the set of all possible joint distributions of $\mathbf{R}, \mathbf{S}$ where $\mathbf{R}, \mathbf{S}$ are independent.

For $i \in [P]$, computation node i outputs

$$\tilde{\mathbf{C}}_i = \tilde{\mathbf{A}}_i \tilde{\mathbf{B}}_i. \tag{3.1}$$

A decoder receives the computation output of an arbitrary set $\mathcal{S}$ of N nodes and performs a map: $d_{\mathcal{S}} : (\mathbb{R}^{L \times L})^{|\mathcal{S}|} \to \mathbb{R}^{L \times L}$ that is linear over $\mathbb{R}$. That is, the decoder outputs:

$$\widetilde{\mathbf{C}}_{\mathcal{S}} = d_{\mathcal{S}}(\mathbf{C}_i|_{i \in \mathcal{S}}) = \sum_{i \in [|\mathcal{S}|]} w_{i,\mathcal{S}} \tilde{\mathbf{C}}_i \tag{3.2}$$

where the co-efficients $w_{i,\mathcal{S}} \in \mathbb{R}, i \in \mathcal{S}$ specify the linear map $d_{\mathcal{S}}$. A (P, N) *coding scheme* for positive integers $P \geq N$ consists of the joint distribution $\mathbb{P}_{\mathbf{R},\mathbf{S}} \in \mathcal{P}_{\mathbf{R},\mathbf{S}}$, and the decoding maps $\prod_{\mathcal{S} \subseteq P : |\mathcal{S}| = N} \{d_{\mathcal{S}} : (\mathbb{R}^{L \times L})^{|\mathcal{S}|} \to \mathbb{R}^{L \times L}\}$.

The performance of a coding scheme is measured by two metrics: privacy and accuracy.

Remark 1. The standard secure multiparty computing set up assumes that $P = N$. We keep our system model general and allow P to be larger than N. When P is larger than N, the developed schemes have the benefit of tolerance to $P - N$ failures/stragglers, in addition to data privacy and accurate computations.

Privacy of a (P, N) coding scheme

Definition 3.1. (t-node ϵ-DP) The distribution $\mathbb{P}_{\mathbf{R},\mathbf{S}}$ satisfies t-node ϵ-DP if, for any $\epsilon \geq 0$, and for arbitrary matrices $\mathbf{A}_0, \mathbf{B}_0, \mathbf{A}_1, \mathbf{B}_1 \in \mathbb{R}^{L \times L}$ that satisfy $\left|\left| \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{B}_0 \end{bmatrix} - \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{B}_1 \end{bmatrix} \right|\right|_{\max} \leq 1$, and for every $i \in [P]$,

$$\frac{\mathbb{P}\left(\mathbf{Y}_{\mathcal{T}}^{(0)} \in \mathcal{A}\right)}{\mathbb{P}\left(\mathbf{Y}_{\mathcal{T}}^{(1)} \in \mathcal{A}\right)} \leq e^{\epsilon}, \tag{3.3}$$

for all subsets $\mathcal{T} \subseteq [P], |\mathcal{T}| = t$, for all subsets $\mathcal{A} \subset \mathbb{R}^{L \times L}$ in the Borel σ-field, where

$$\mathbf{Y}_{\mathcal{T}}^{(\ell)} \triangleq \begin{bmatrix} \mathbf{A}_\ell + \mathbf{R}_{i_1} & \mathbf{A}_\ell + \mathbf{R}_{i_2} & \ldots & \mathbf{A}_\ell + \mathbf{R}_{i_{|\mathcal{T}|}} \\ \mathbf{B}_\ell + \mathbf{S}_{i_1} & \mathbf{B}_\ell + \mathbf{S}_{i_2} & \ldots & \mathbf{B}_\ell + \mathbf{S}_{i_{|\mathcal{T}|}} \end{bmatrix}, \ell = 0, 1$$

where $\mathcal{T} = \{i_1, i_2, \ldots, i_{|\mathcal{T}|}\}$.

We denote by $\mathcal{P}_{\mathbf{R},\mathbf{S}}^{\epsilon,t}$, the set of all possible joint distributions $\mathbb{P}_{\mathbf{R},\mathbf{S}} \in \mathcal{P}_{\mathbf{R},\mathbf{S}}$ that satisfy t-node ϵ-DP. Note that (3.3) depends only on the joint distribution $\mathbb{P}_{\mathbf{R},\mathbf{S}}$ and does not depend on the distributions of $\mathbf{A}, \mathbf{B}$, since the definition applies for arbitrary vectors $\mathbf{A}_0, \mathbf{B}_0, \mathbf{A}_1, \mathbf{B}_1$ — that is, those that are not necessarily drawn from $\mathbb{P}_{\mathbf{A},\mathbf{B}}$.

Accuracy of a (P, N) coding scheme

The main goal of this paper is to characterize the trade-off between privacy and accuracy of estimation of the matrix-product $\mathbf{AB}$. In particular, we develop schemes that guarantee a certain level of differential privacy (i.e., a certain value of parameter ϵ), irrespective of the distribution of the inputs. It is, however, necessary (and standard, see VC: [refs]) to account for the data distribution and its parameters when evaluating the accuracy of coding schemes. The accuracy guarantees of the coding schemes developed in this paper rely on the following key assumptions:

Assumption 3.1. $\mathbf{A}$ and $\mathbf{B}$ are statistically independent random matrices.

Assumption 3.2. There is a parameter $\eta > 0$ such that:

$$\mathbb{E}\left[||\mathbf{A}||_F^2\right] = \mathbb{E}\left[||\mathbf{B}||_F^2\right] \leq \eta.$$

We measure the accuracy of a coding scheme via the mean square error. Specifically, we define:

Definition 3.2 (Linear Mean Square Error (LMSE))**.** For a (P, N) coding scheme Γ consisting of joint distribution $\mathbb{P}_{\mathbf{R},\mathbf{S}}$ decoding maps $\prod_{\mathcal{S}\subseteq P:|\mathcal{S}|=N}\{d_{\mathcal{S}} : (\mathbb{R}^{L\times L})^{|\mathcal{S}|} \to \mathbb{R}^{L\times L}\}$, the LMSE for subset $\mathcal{S} \subseteq [P], |\mathcal{S}| = N$ is defined as:

$$\mathsf{LMSE}_{\mathcal{S}}(\Gamma) = \mathbb{E}[||\mathbf{AB} - \widehat{\mathbf{C}}_{\mathcal{S}}||_F^2]. \tag{3.4}$$

where $\widehat{\mathbf{C}}_{\mathcal{S}}$ is defined in (3.2). The LMSE of the coding scheme Γ is defined to be:

$$\mathsf{LMSE}(\Gamma) = \max_{\mathcal{S}} \mathsf{LMSE}_{\mathcal{S}}(\Gamma).$$

It is worth noting that the expectation in the above definition is over the joint distributions of the random variables $\mathbf{A}, \mathbf{B}, \mathbf{R}, \mathbf{S}$. In particular, the accuracy of a coding scheme can depend on the parameters[3] of the joint distribution of $\mathbf{A}, \mathbf{B}$. Sometimes we will explicitly denote the distribution in the LMSE notation as: $\mathsf{LMSE}_{\mathcal{S}}^{\mathbb{P}_{\mathbf{A},\mathbf{B}}(\Gamma)}$ or $\mathsf{LMSE}^{\mathbb{P}_{\mathbf{A},\mathbf{B}}(\Gamma)}$.

Definition 3.3 (Optimal Mean Square Error ($\mathsf{LMSE}^*(P, N, \epsilon, t)$))**.**

$$\mathsf{LMSE}^*(P, N, \epsilon, t) = \inf_{\Gamma} \mathsf{LMSE}(\Gamma) \tag{3.5}$$

where the infimum is over the set of all possible (P, N) coding schemes Γ whose joint distribution $\mathbb{P}_{\mathbf{R},\mathbf{S}}$ satisfies t-node ϵ-DP.

The goal of this paper is to characterize $\mathsf{LMSE}^*(P, N, t, \epsilon)$. It is a simple exercise to verify that, if for $N \geq 2t + 1$, coding schemes used by the BGW algorithm achieve $\mathsf{LMSE}^*(P, N, t, \epsilon) = 0$ for all $\epsilon \geq 0$. That is, perfect privacy[4] and perfect accuracy are achievable for $N \geq 2t + 1$ nodes. Thus, our goal is to characterize $\mathsf{LMSE}^*(P, N, t, \epsilon)$ for $N \leq 2t$.

[3]It can be readily verified from the LMSE definition that the accuracy simply depends on the means, variances and pairwise correlations of all the random variables involved in $\mathbf{A}, \mathbf{B}, \mathbf{R}_i|_{i=1}^P, \mathbf{S}_i|_{i=1}^P$.

IV. Summary of Results

The main contribution of this paper is the characterization of an explicit tradeoff between accuracy (LMSE) and privacy (ϵ) for distributed matrix multiplication for the case[5] where $t = 1, N = 2$,. We present our results for the case where $\eta = 1$. All our results can be readily extended for general values of η (see extended version []). The key to our approach is to utilize the variance of the noise as proxy metric for DP, and develop a tight relation between privacy and accuracy under this metric. Then, the obtained results are translated to bounds on the privacy-accuracy trade-off for ϵ-DP.

We present two technical results. The first is an achievability result that shows that there exists a (P, N) coding scheme with random variables $(\mathbf{R}, \mathbf{S})$ with $\mathbb{E}[||\mathbf{R}_i||_F^2], \mathbb{E}[||\mathbf{S}_i||_F^2] \geq \sigma_{\mathsf{Ach}}^2, \forall i \in [P]$, such that

$$\mathsf{LMSE}(\Gamma) \leq \frac{1}{\left(1 + \frac{1}{\sigma_{\mathsf{Ach}}^2}\right)^2} + \Delta$$

for every $\Delta > 0$. The second is a converse that states that, for *any* $\mathcal{S} \subseteq [P], |\mathcal{S}| = 2$:

$$\mathsf{LMSE}_{\mathcal{S}}(\Gamma) \geq \frac{1}{\left(1 + \frac{1}{\sigma_{\mathsf{Con}}^2}\right)^2}.$$

so long as the expected all singular values of $\mathbf{R}_i, \mathbf{S}_i$ are lower bounded by σ_{Con}^2 in expectation, that is: $\mathbb{E}[\kappa(\mathbf{R}_i)^2], \mathbb{E}[\kappa(\mathbf{S}_i)^2] \geq \sigma_{\mathsf{Con}}^2, \forall i \in \mathcal{S}$.

The parameters σ_{Ach}^2 and σ_{Con}^2 in the above intuitively determine the (minimum) variance of the noise added to the inputs at the nodes, and therefore they indirectly control the degree of privacy of the inputs at the nodes. In particular, larger values of σ_{Ach} and σ_{Con} corresponds to greater amount of privacy, and correspondingly poorer LMSE. We next discuss details behind the achievability and converse stated above. Our discussions also include bounds implied on the LMSE in terms of the DP parameter ϵ. We only sketch the main achievability argument here, all missing theorem proofs and details can be found in the extended version [].

A. Achievability

Theorem 4.1. *Let $\mathbf{\Lambda}, \mathbf{\Theta}$ be $L \times L$ independent zero-mean random matrices with i.i.d. entries each with a variance of $1/L^2$. For any $\Delta, \sigma_{\mathsf{Ach}} > 0$ there exist scalars $u_i, i \in [P]$ with $|u_i| \geq \sigma_{\mathsf{Ach}}$ such that, if $\mathbf{R}_i = u_i\mathbf{\Lambda}, \mathbf{S}_i = u_i\mathbf{\Theta}, i \in [P]$, then*

[4]More precisely, ϵ can be made arbitrarily small by adding Laplacian noise of correspondingly large variances, and yet the LMSE can be kept 0 [].

[5]The case of $t = 1, N = 1$ is simple to analyze along the arguments of this paper, and is presented in the extended version of this paper [].

there is a $(P, N=2)$ coding scheme Γ with distribution $\mathbb{P}_{\mathbf{R},\mathbf{S}}$ and

$$\mathsf{LMSE}(\Gamma) \leq \frac{1}{\left(1+\frac{1}{\sigma_{\mathsf{Ach}}^2}\right)^2} + \Delta. \tag{4.1}$$

Proof Sketch (Missing details in []). For simplicity of notation, we sketch the argument for the subset $\mathcal{S} = \{1, 2\}$; the same argument readily extends to an arbitrary two-element subset of $[P]$. As stated in theorem statement, we show an achievable scheme for the subset of distributions $(\mathbf{R}, \mathbf{S})$, where node $i \in [P]$ gets evaluations as,

$$\tilde{\mathbf{A}}_i = \mathbf{A} + \mathbf{\Lambda} u_i, \tilde{\mathbf{B}}_i = \mathbf{B} + \mathbf{\Theta} u_i,$$

where $|u_i| \geq \sigma_{\mathsf{Ach}}, \forall i \in [P]$. For convenience of illustration we drop the dependence on $\mathcal{S}$ for decoding weights, i.e., $w_{1,\mathcal{S}}, w_{2,\mathcal{S}}$ will be written as w_1, w_2 respectively. Thus,

$$\widehat{\mathbf{C}}_{\mathcal{S}} = w_1 \tilde{\mathbf{A}}_1 \tilde{\mathbf{B}}_1 + w_2 \tilde{\mathbf{A}}_2 \tilde{\mathbf{B}}_2.$$

Then, it can be shown that from the independence of $\mathbf{A}, \mathbf{B}, \mathbf{\Lambda}, \mathbf{\Theta}$ and from the theorem hypothesis that $\mathbb{E}[\mathbf{\Lambda}], \mathbb{E}[\mathbf{\Theta}] = \mathbf{0}, E[||\mathbf{\Lambda}||_F^2], E[||\mathbf{\Theta}||_F^2] = 1$ that:

$$\mathsf{LMSE}_{\mathcal{S}}(\Gamma) = \mathbb{E}[||\mathbf{A}\mathbf{B} - \widehat{\mathbf{C}}_{\mathcal{S}}||_F^2]$$

$$= (w_1 + w_2 - 1)^2 + 2(w_1 u_1 + w_2 u_2)^2 + (w_1 u_1^2 + w_2 u_2^2)^2.$$

Then minimizing the above expression over w_1, w_2 and then substituting back, we get the following expression,

$$\mathsf{LMSE}_{\mathcal{S}}(\Gamma) = \frac{2u_1^2 u_2^2}{2u_1^2 u_2^2 + (u_1 + u_2)^2 + 2}$$

so long as u_1, u_2 are distinct. By choosing *distinct* $u_i, i \in [P]$ arbitrarily close to σ_{Ach}, we obtain, for any $\Delta > 0$,

$$\mathsf{LMSE}_{\mathcal{S}}(\Gamma) \leq \frac{1}{\left(1+\frac{1}{\sigma_{\mathsf{Ach}}^2}\right)^2} + \Delta$$

□

It is worth noting that the achievable coding scheme is indeed Shamir secret sharing over real field with an appropriate choice of evaluation points. An intriguing aspect of the theorem is that the choice of evaluation points u_i is arbitrarily close to σ_{Ach}. Consider the case where $u_1 = \sigma_{\mathsf{Ach}}$, and examine the LMSE with respect to u_2. When u_2 is *equal* to σ_{Ach}, the computation of both nodes $1, 2$ are identical, and this would lead to a poor LMSE. But even a small deviation translates to a near optimal choice of u_2. Interestingly, the LMSE is, in fact, a discontinuous function of u_2 at $u_2 = \sigma_{\mathsf{Ach}}$.

We translate the result of Theorem 4.1 to ϵ-DP by restricting $\mathbf{\Lambda}, \mathbf{\Theta}$ to independent Laplace distributions.

Theorem 4.2. *Let $\mathbf{\Lambda}, \mathbf{\Theta}$ be independent zero-mean random matrices with i.i.d. Laplacian distributed entries each with a variance of $1/L^2$. For any $\Delta > 0, \epsilon \geq 0$ there exist scalars $u_i, i \in [P]$ such that, if $\mathbf{R}_i = u_i \mathbf{\Lambda}, \mathbf{S}_i = u_i \mathbf{\Theta}, i \in [P]$, then there exists a $(P, N = 2)$ coding scheme with distribution $\mathbb{P}_{\mathbf{R},\mathbf{S}} \in \mathcal{P}_{\mathbf{R},\mathbf{S}}^{\epsilon,1}$ and*

$$\mathsf{LMSE}(\Gamma) \leq \frac{1}{\left(1+\frac{\epsilon^2}{8L^6}\right)^2} + \Delta.$$

The full proof is given in [], but the idea is essentially similar to the proof of Laplacian mechanism satisfying ϵ-DP in differential privacy literature [7].

B. Converse

We also derive converse results that lower bound the LMSE for a fixed level of privacy. Similar to our approach to achievability, we first derive a lower bound in Theorem 4.3 in terms of the variances of the noise distributions that are added.

Theorem 4.3. *For any (P, N) code Γ whose distribution $\mathbb{P}_{\mathbf{R},\mathbf{S}}$ satisfies $E\left[\kappa(\mathbf{R}_i)^2\right], E\left[\kappa(\mathbf{S}_i)^2\right] \geq \sigma_{\mathsf{Con}}^2, \forall i \in [P]$, there exists a distribution $\mathbb{P}_{\mathbf{A},\mathbf{B}}$ for the inputs such that*

$$\mathsf{LMSE}_{\mathcal{S}}(\Gamma) \geq \frac{1}{\left(1+\frac{1}{\sigma_{\mathsf{Con}}^2}\right)^2}.$$

The proof has similar tones to the proof of Theorem 4.1, so omit a proof sketch here and urge the reader to look at our extended version [].

Next, the above converse is translated to a bound in terms of ϵ-the DP parameter. It is worth noting that the converse does not necessarily assume Laplace distributions, and is applicable to *any* distribution $\mathbb{P}_{\mathbf{R},\mathbf{S}}$ that satisfies ϵ-DP.

Theorem 4.4.

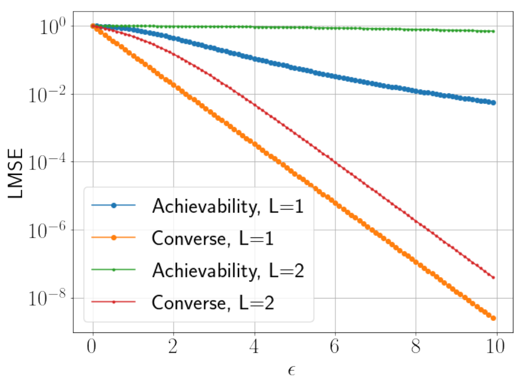
$$\mathsf{LMSE}^*(P, 2, \epsilon, 1) \geq \frac{1}{\left(\frac{e^\epsilon - 1}{L^2} + 1\right)^2} \tag{4.2}$$

Proof Sketch (Missing details in []). We observe that lower bounding σ_{Con}^2 in Theorem 4.3 should give us the desired relation. We show a proof sketch for the scalar case $L = 1$, the general proof is given in []. Without loss of generality assume k^{th} node attains the minimum variance $\sigma_{\mathsf{Con}}^2 = \mathbb{E}[R_k^2]$. Denote the pdf of R_k to be p_{R_k}. Then we can write from the differential privacy Definition 3.1,

$$\frac{p_{R_k}(r-1)}{p_{R_k}(r)} \leq e^\epsilon.$$

$$\sigma_{\mathsf{Con}}^2 = \int_{-\infty}^{\infty} r^2 p_{R_k}(r) dr \tag{4.3}$$
$$\geq \int_{-\infty}^{\infty} r^2 p_{R_k}(r-1) dr \tag{4.4}$$
$$= e^{-\epsilon} \int_{-\infty}^{\infty} (r+1)^2 p_{R_k}(r) dr \tag{4.5}$$
$$= e^{-\epsilon}(\sigma_{\mathsf{Con}}^2 + 1) \tag{4.6}$$
$$\implies \sigma_{\mathsf{Con}}^2 \geq \frac{1}{e^{\epsilon}-1} \tag{4.7}$$

Substituting the above in the result of Theorem 4.3 gives us the desired relation. □



(a) LMSE vs ϵ

Figure 1: Figure shows in log scale, achievable LMSE when using Laplace noise, and converse on LMSE given by theorem 4.4, for various ϵ.

C. Tight characterization for $t = 1, N = 2, L = 1$

It is readily seen that for $L = 1$, $\sigma_{\mathsf{Ach}} = \sigma_{\mathsf{Con}} \triangleq \sigma$. To highlight that this is the scalar case, we remove the boldface from notations.

$$\tilde{A}_i = A + R_i, \tilde{S}_i = B + S_i,$$
$$\tilde{C}_i = \tilde{A}_i \tilde{B}_i.$$

Then from the achievability Theorem 4.1 and converse Theorem 4.3 and from Lemma B.1 we provide the following result.

Theorem 4.5. *For any $\sigma > 0$,*

$$\inf_{\Gamma} \mathsf{LMSE}(\Gamma) = \frac{1}{\left(1+\frac{1}{\sigma^2}\right)^2}. \tag{4.8}$$

where the infimum is over all coding schemes Γ whose probability distribution $\mathbb{P}_{\mathbf{R},\mathbf{S}}$ satisfies

$$\mathbb{E}\left[R_i^2\right], \mathbb{E}\left[S_i^2\right] \geq \sigma^2, \forall i \in [P].$$

V. Discussion and Future Work

This work opens a new direction via the search of codes that optimize privacy-utility trade-off for secure multiparty computing. There are several open questions motivated by our work.

First, the study of optimal code design for $t \geq 1$ is a natural open question. While an achievable scheme can be developed along the same lines as our paper by assessing the Shamir secret sharing scheme with arbitrary close evaluation points, development of a tight characterization (even for the scalar case of $L = 1$, with the standard deviation measure on the privacy loss) is an open problem. Second, our coding schemes do not assume shared randomness, that is, they assume $\mathbf{R}, \mathbf{S}$ are statistically independent. The question of whether shared randomness can improve the accuracy-privacy trade-off is an interesting open question. Finally, because our schemes require evaluation points that are arbitrarily close to each other, the computation nodes need to perform computations at a high level of precision. This can involve hidden computation and storage costs (see, a similar phenomenon in [17], [24]). An explicit characterization of these hidden costs is an interesting area of future work.

REFERENCES

[1] D. Evans, V. Kolesnikov, and M. Rosulek, "A pragmatic introduction to secure multi-party computation," *Foundations and Trends® in Privacy and Security*, vol. 2, no. 2-3, 2017.

[2] M. Ben-Or, S. Goldwasser, and A. Wigderson, "Completeness theorems for non-cryptographic fault-tolerant distributed computation," in *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, ser. STOC '88. New York, NY, USA: Association for Computing Machinery, 1988, p. 1–10. [Online]. Available: https://doi.org/10.1145/62212.62213

[3] G. Asharov and Y. Lindell, "A full proof of the bgw protocol for perfectly secure multiparty computation," *J. Cryptol.*, vol. 30, no. 1, p. 58–151, jan 2017. [Online]. Available: https://doi.org/10.1007/s00145-015-9214-4

[4] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, no. 11, p. 612–613, nov 1979. [Online]. Available: https://doi.org/10.1145/359168.359176

[5] H. B. McMahan, G. Andrew, U. Erlingsson, S. Chien, I. Mironov, N. Papernot, and P. Kairouz, "A general approach to adding differential privacy to iterative training procedures," *arXiv preprint arXiv:1812.06210*, 2018.

[6] "Tensorflow privacy." [Online]. Available: https://www.tensorflow.org/responsible_ai/privacy/guide

[7] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014. [Online]. Available: http://dx.doi.org/10.1561/0400000042

[8] C. Dwork, N. Kohli, and D. Mulligan, "Differential privacy in practice: Expose your epsilons!" *Journal of Privacy and Confidentiality*, vol. 9, no. 2, 2019.

[9] S. Yuan, M. Shen, I. Mironov, and A. C. A. Nascimento, "Practical, label private deep learning training based on secure multiparty computation and differential privacy," Cryptology ePrint Archive, Report 2021/835, 2021. [Online]. Available: https://ia.cr/2021/835

[10] M. Pettai and P. Laud, "Combining differential privacy and secure multiparty computation," in *Proceedings of the 31st Annual Computer Security Applications Conference*, ser. ACSAC 2015. New York, NY, USA: Association for Computing Machinery, 2015, p. 421–430. [Online]. Available: https://doi.org/10.1145/2818000.2818027

[11] X. He, A. Machanavajjhala, C. Flynn, and D. Srivastava, "Composing differential privacy and secure computation: A case study on scaling private record linkage," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1389–1406. [Online]. Available: https://doi.org/10.1145/3133956.3134030

[12] P. Kairouz, S. Oh, and P. Viswanath, "Secure multi-party differential privacy," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28. Curran Associates, Inc., 2015. [Online]. Available: https://proceedings.neurips.cc/paper/2015/file/a01610228fe998f515a72dd730294d87-Paper.pdf

[13] ——, "Differentially private multi-party computation," in *2016 Annual Conference on Information Science and Systems (CISS)*, 2016, pp. 128–132. [Online]. Available: https://doi.org/10.1109/CISS.2016.7460489

[14] J. Böhler and F. Kerschbaum, "Secure multi-party computation of differentially private median," in *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, Aug. 2020, pp. 2147–2164. [Online]. Available: https://www.usenix.org/conference/usenixsecurity20/presentation/boehler

[15] L. Sweeney, "Simple demographics often identify people uniquely," 2000. [Online]. Available: http://dataprivacylab.org/projects/identifiability/

[16] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965, pMID: 12261830. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01621459.1965.10480775

[17] H. Jeong, A. Devulapalli, V. R. Cadambe, and F. P. Calmon, "ϵ-approximate coded matrix multiplication is nearly twice as efficient as exact multiplication," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 3, pp. 845–854, 2021. [Online]. Available: https://doi.org/10.1109/JSAIT.2021.3099811

[18] Q. Yu, S. Li, N. Raviv, S. M. M. Kalan, M. Soltanolkotabi, and S. A. Avestimehr, "Lagrange coded computing: Optimal design for resiliency, security, and privacy," in *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and M. Sugiyama, Eds., vol. 89. PMLR, 16–18 Apr 2019, pp. 1215–1225. [Online]. Available: https://proceedings.mlr.press/v89/yu19b.html

[19] M. Soleymani, H. Mahdavifar, and A. S. Avestimehr, "Analog lagrange coded computing," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 283–295, 2021. [Online]. Available: https://doi.org/10.1109/JSAIT.2021.3056377

[20] H. Akbari-Nodehi and M. A. Maddah-Ali, "Secure coded multi-party computation for massive matrix operations," *IEEE Transactions on Information Theory*, vol. 67, no. 4, pp. 2379–2398, 2021.

[21] W.-T. Chang and R. Tandon, "On the capacity of secure distributed matrix multiplication," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–6.

[22] Q. Geng, W. Ding, R. Guo, and S. Kumar, "Tight analysis of privacy and utility tradeoff in approximate differential privacy," in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, S. Chiappa and R. Calandra, Eds., vol. 108. PMLR, 26–28 Aug 2020, pp. 89–99. [Online]. Available: https://proceedings.mlr.press/v108/geng20a.html

[23] Q. Geng and P. Viswanath, "The optimal noise-adding mechanism in differential privacy," *IEEE Transactions on Information Theory*, vol. 62, no. 2, pp. 925–951, 2016. [Online]. Available: https://doi.org/10.1109/TIT.2015.2504967

[24] J. Wang, Z. Jia, and S. A. Jafar, "Price of precision in coded distributed matrix multiplication: A dimensional analysis," in *2021 IEEE Information Theory Workshop (ITW)*. IEEE, 2021, pp. 1–6.

[25] Wikipedia contributors, "Weyl's inequality — Wikipedia, the free encyclopedia," https://en.wikipedia.org/w/index.php?title=Weyl%27s_inequality&oldid=1065059511, 2022, [Online; accessed 3-February-2022].

[26] J. F. Grcar, "A matrix lower bound," *Linear Algebra and its Applications*, vol. 433, no. 1, pp. 203–220, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0024379510000923

Appendix A
Proofs

In this section we provide proofs for the achievability and converse theorems stated in the previous section.

A. Achievability

We do the analysis for some $N = 2$ nodes i and j, which can be applied to any two nodes. Thus let the subset $\mathcal{S} = \{i, j\}$. As stated in Theorem 4.1, we show an achievable scheme for a subset of distributions $(\mathbf{R}, \mathbf{S})$, where node i gets evaluations as,

$$\tilde{\mathbf{A}}_i = \mathbf{A} + \mathbf{\Lambda} u_i, \tilde{\mathbf{B}}_i = \mathbf{B} + \mathbf{\Theta} u_i.$$

We pick $\mathbb{P}_{\mathbf{\Lambda},\mathbf{\Theta}}$ such that $\mathbb{E}[\mathbf{\Lambda}] = \mathbb{E}[\mathbf{\Theta}] = 0$ and $|u_i| \geq \sigma_{\mathsf{Ach}}$. For convenience of illustration we drop the dependence on $\mathcal{S}$ for decoding weights, i.e., $w_{1,\mathcal{S}}, w_{2,\mathcal{S}}$ will be written as w_1, w_2 respectively. Thus,

$$\widehat{\mathbf{C}}_{\mathcal{S}} = w_1 \tilde{\mathbf{A}}_i \tilde{\mathbf{B}}_i + w_2 \tilde{\mathbf{A}}_j \tilde{\mathbf{B}}_j.$$

Let

$$\begin{aligned}\bar{E}(w_1, w_2, u_i, u_j) &\triangleq (w_1 + w_2 - 1)^2 + 2(w_1 u_i + w_2 u_j)^2 \\ &+ (w_1 u_i^2 + w_2 u_j^2)^2.\end{aligned}$$

We will now prove an upper bound on $\bar{E}(w_1, w_2, u_i, u_j)$ which will be used to prove Theorem 4.1.

Lemma A.1. *For any $\Delta > 0$ there exist some $w_1^*, w_2^* \in \mathbb{R}$ and $u_i, u_j \in \mathbb{R}$ satisfying $|u_i|, |u_j| \geq \sigma_{\mathsf{Ach}}$ such that*

$$\bar{E}(w_1^*, w_2^*, u_i, u_j) \leq \frac{1}{(1 + \frac{1}{\sigma_{\mathsf{Ach}}^2})^2} + \Delta.$$

Proof. We find w_1^*, w_2^* by minimizing $\bar{E}(w_1, w_2, u_i, u_j)$ over w_1, w_2. Equating the Jacobian of $\bar{E}(w_1, w_2, u_i, u_j)$ with respect to w_1, w_2 to 0 and solving for w_1, w_2 gives,

$$w_1^* = \frac{-u_i u_j^2 - u_j^3 - 2u_j}{(u_i - u_j)\left(2u_i^2 u_j^2 + 2u_i u_j + u_i^2 + u_j^2 + 2\right)} \tag{A.1}$$

$$w_2^* = \frac{u_i^2 u_j + u_i^3 + 2u_i}{(u_i - u_j)\left(2u_i^2 u_j^2 + 2u_i u_j + u_i^2 + u_j^2 + 2\right)}. \tag{A.2}$$

Observe that $\bar{E}(w_1, w_2, u_i, u_j)$ is a convex quadratic in w_1, w_2. Substituting w_1^*, w_2^* in $\bar{E}(w_1, w_2, u_i, u_j)$.

$$\begin{aligned}\min_{w_1, w_2 \in \mathbb{R}} \bar{E}(w_1, w_2, u_i, u_j) &= \bar{E}(w_1^*, w_2^*, u_i, u_j) \\ &= \frac{2u_i^2 u_j^2}{2u_i^2 u_j^2 + (u_i + u_j)^2 + 2}\end{aligned} \tag{A.3}$$

Now, let $\epsilon_1 = \frac{1}{\sigma_{\mathsf{Ach}}^2} - \frac{1}{u_i^2}$, $\epsilon_2 = \frac{1}{\sigma_{\mathsf{Ach}}^2} - \frac{1}{u_j^2}$ and $\epsilon_3 = (u_i - u_j)^2$.

To bring about the desired relation, we do the following,

$$\begin{aligned}\left(1 + \frac{1}{u_i^2}\right)\left(1 + \frac{1}{u_j^2}\right) - \frac{1}{\bar{E}(w_1^*, w_2^*, u_i, u_j)} \\ = \frac{1}{2} \frac{(u_i - u_j)^2}{u_i^2 u_j^2} \\ \leq \frac{1}{2\sigma_{\mathsf{Ach}}^4} \epsilon_3\end{aligned} \tag{A.4}$$

$$\begin{aligned}\left(1 + \frac{1}{\sigma_{\mathsf{Ach}}^2} - \epsilon_1\right)\left(1 + \frac{1}{\sigma_{\mathsf{Ach}}^2} - \epsilon_2\right) - \frac{1}{\bar{E}(w_1^*, w_2^*, u_i, u_j)} \\ \leq \frac{1}{2\sigma_{\mathsf{Ach}}^4} \epsilon_3\end{aligned} \tag{A.5}$$

$$\begin{aligned}\left(1 + \frac{1}{\sigma_{\mathsf{Ach}}^2}\right)^2 - \frac{1}{\bar{E}(w_1^*, w_2^*, u_i, u_j)} \\ \leq \left(1 + \frac{1}{\sigma_{\mathsf{Ach}}^2}\right)(\epsilon_1 + \epsilon_2) - \epsilon_1 \epsilon_2 + \frac{1}{2\sigma_{\mathsf{Ach}}^4} \epsilon_3 \triangleq h.\end{aligned} \tag{A.6}$$

$$\implies \bar{E}(w_1^*, w_2^*, u_i, u_j) \leq \frac{1}{(1 + \frac{1}{\sigma_{\mathsf{Ach}}^2})^2 - h}. \tag{A.7}$$

Remark 2. We ideally want h to be close to 0. Observe that picking $|u_i|$ and $|u_j|$ close to σ_{Ach} makes $\epsilon_1, \epsilon_2, \epsilon_3$ close to 0 which gives h close to 0.

Taylor series expansion about $h = 0$ gives

$$\bar{E}(w_1^*, w_2^*, u_i, u_j) \leq \frac{1}{(1 + \frac{1}{\sigma_{\mathsf{Ach}}^2})^2} + O(h). \tag{A.8}$$

Taking $\Delta = O(h)$ gives,

$$\implies \bar{E}(w_1^*, w_2^*, u_i, u_j) \leq \frac{1}{(1 + \frac{1}{\sigma_{\mathsf{Ach}}^2})^2} + \Delta. \tag{A.9}$$

$\square$

We now use the above result to prove Theorem 4.1

Proof of Theorem 4.1.

$$\begin{aligned}\mathsf{LMSE}_{\mathcal{S}}(\Gamma) &= \mathbb{E}[||\mathbf{AB} - \widehat{\mathbf{C}}_{\mathcal{S}}||_F^2] \\ &= \mathbb{E}[||w_1(\mathbf{AB} + (\mathbf{\Lambda B} + \mathbf{A\Theta})u_i + \mathbf{\Lambda\Theta} u_i^2) \\ &\quad + w_2(\mathbf{AB} + (\mathbf{\Lambda B} + \mathbf{A\Theta})u_j + \mathbf{\Lambda\Theta} u_j^2) - \mathbf{AB}||_F^2]\end{aligned} \tag{A.10}$$

$$\begin{aligned}&= \mathbb{E}[||(w_1 + w_2 - 1)\mathbf{AB} + (w_1 u_i + w_2 u_j)(\mathbf{\Lambda B} + \mathbf{A\Theta}) \\ &\quad + (w_1 u_i^2 + w_2 u_j^2)\mathbf{\Lambda\Theta}||_F^2]\end{aligned} \tag{A.11}$$

Since $\mathbf{A}, \mathbf{B}, \mathbf{\Lambda}, \mathbf{\Theta}$ are independent and $\mathbf{\Lambda}, \mathbf{\Theta}$ are zero mean, the cross products vanish,

$$\begin{aligned} &= (w_1 + w_2 - 1)^2 \mathbb{E}[||\mathbf{AB}||_F^2] \\ &\quad + (w_1 u_i + w_2 u_j)^2 (\mathbb{E}[||\mathbf{\Lambda B}||_F^2] \\ &\quad + \mathbb{E}[||\mathbf{A\Theta}||_F^2]) + (w_1 u_i^2 + w_2 u_j^2)^2 \mathbb{E}[||\mathbf{\Lambda\Theta}||_F^2] \end{aligned} \tag{A.12}$$

From system model and from theorem statement we know $\mathbb{E}||\mathbf{A}||_F^2], \mathbb{E}[||\mathbf{B}||_F^2] \leq 1, \mathbb{E}[||\mathbf{\Lambda}||_F^2] = \mathbb{E}[||\mathbf{\Theta}||_F^2] = 1$ and using sub-multiplicative property of Frobenius norm and independence of $\mathbf{A}, \mathbf{B}, \mathbf{\Lambda}, \mathbf{\Theta}$ gives,

$$\leq (w_1 + w_2 - 1)^2 + 2(w_1 u_i + w_2 u_j)^2 + (w_1 u_i^2 + w_2 u_j^2)^2 \tag{A.13}$$

From Corollary A.1

$$\leq \frac{1}{(1 + \frac{1}{\sigma_{\mathsf{Ach}}^2})^2} + \Delta \tag{A.14}$$

□

Making the distributional assumptions given in Theorem 4.2, we provide a relation between σ_{Ach} and ϵ.

Proof of Theorem 4.2. Observe that for the i^{th} node if $(\mathbf{\Lambda}_i)_{m,n} \sim \text{Laplace}(0, \frac{1}{\sqrt{2}L})$, then the distribution of $(\mathbf{R}_i)_{m,n}$ is,

$$f_{(\mathbf{R}_i)_{m,n}}(z) = \frac{L}{\sqrt{2}|u_i|} \exp\left(-\sqrt{2}L\frac{|z|}{|u_i|}\right) \tag{A.15}$$

Similarly,

$$f_{(\mathbf{S}_i)_{m,n}}(z) = \frac{L}{\sqrt{2}|u_i|} \exp\left(-\sqrt{2}L\frac{|z|}{|u_i|}\right) \tag{A.16}$$

Let's evaluate the ratio in Definition 3.1 at a point $\bar{\mathbf{Y}} = \begin{bmatrix} \bar{\mathbf{Y}}_0 \in \mathbb{R}^{L \times L} \\ \bar{\mathbf{Y}}_1 \in \mathbb{R}^{L \times L} \end{bmatrix}$ i.e., let $\mathcal{A} = \{\bar{\mathbf{Y}}\}$. Let $\mathbf{X}_l = \begin{bmatrix} \mathbf{A}_l \\ \mathbf{B}_l \end{bmatrix}$.

$$\frac{\mathbb{P}(\mathbf{Y}^{(0)} \in \mathcal{A})}{\mathbb{P}(\mathbf{Y}^{(1)} \in \mathcal{A})} = \frac{\mathbb{P}_{\mathbf{R}_i}(\bar{\mathbf{Y}} - \mathbf{X}_0)}{\mathbb{P}_{\mathbf{R}_i}(\bar{\mathbf{Y}} - \mathbf{X}_1)} \tag{A.17}$$

$$\begin{aligned} &= \prod_{k,l \in [L]} \frac{\exp\left(-\sqrt{2}L\frac{|(\bar{\mathbf{Y}}_0)_{m,n} - (\mathbf{A}_0)_{m,n}|}{|u_i|}\right)}{\exp\left(-\sqrt{2}L\frac{|(\bar{\mathbf{Y}}_0)_{m,n} - (\mathbf{A}_1)_{m,n}|}{|u_i|}\right)} \\ &\qquad \frac{\exp\left(-\sqrt{2}L\frac{|(\bar{\mathbf{Y}}_1)_{m,n} - (\mathbf{B}_0)_{m,n}|}{|u_i|}\right)}{\exp\left(-\sqrt{2}L\frac{|(\bar{\mathbf{Y}}_1)_{m,n} - (\mathbf{B}_1)_{m,n}|}{|u_i|}\right)} \end{aligned} \tag{A.18}$$

$$\begin{aligned} &\leq \prod_{k,l \in [L]} \exp\left(\sqrt{2}L\frac{|(\mathbf{A}_0)_{m,n} - (\mathbf{A}_1)_{m,n}|}{|u_i|}\right) \\ &\qquad \exp\left(\sqrt{2}L\frac{|(\mathbf{B}_0)_{m,n} - (\mathbf{B}_1)_{m,n}|}{|u_i|}\right) \end{aligned} \tag{A.19}$$

Since $|u_i| \geq \sigma_{\mathsf{Ach}}$ and letting $\beta_{m,n} = |(\mathbf{A}_0)_{m,n} - (\mathbf{A}_1)_{m,n}| + |(\mathbf{B}_0)_{m,n} - (\mathbf{B}_1)_{m,n}|$,

$$\leq \prod_{k,l \in [L]} \exp\left(\frac{\sqrt{2}L}{\sigma_{\mathsf{Ach}}}\beta_{m,n}\right) \tag{A.20}$$

$$= \exp\left(\frac{\sqrt{2}L}{\sigma_{\mathsf{Ach}}}\sum_{k,l \in [L]} \beta_{m,n}\right) \tag{A.21}$$

$$\leq \exp\left(\frac{2\sqrt{2}L^3}{\sigma_{\mathsf{Ach}}}||\mathbf{X}_0 - \mathbf{X}_1||_{\max}\right) \tag{A.22}$$

$$\leq \exp\left(\frac{2\sqrt{2}L^3}{\sigma_{\mathsf{Ach}}}\right) = \exp(\epsilon) \tag{A.23}$$

Thus, we take $\sigma_{\mathsf{Ach}}^2 = \frac{8L^6}{\epsilon^2}$. We obtain the last equation by letting $\sigma_{\mathsf{Ach}} = \frac{2\sqrt{2}L^3}{\epsilon}$. Finally, plugging this σ_{Ach} in equation (4.1) completes the proof. □

B. Converse

Proof of Theorem 4.3.
Consider a (P, N) coding scheme Γ which satisfies $\mathbb{E}(\kappa(\mathbf{R}_i)^2) \geq \sigma_{\mathsf{Con}}^2$ and $\mathbb{E}(\kappa(\mathbf{S}_i)^2) \geq \sigma_{\mathsf{Con}}^2$. Similar to achievability section we do the analysis for some two nodes i and j, which can be applied to any two nodes. Thus let the subset $\mathcal{S} = \{i, j\}$. AD: Ideally if the appendix matrix case lemma is proven before deadline, here we can state that we are choosing the correlated noises with zero mean. Otherwise we have state this as an assumption both here and in theorem statmemt. Take,

$$\tilde{\mathbf{A}}_i = \mathbf{A} + \mathbf{\Lambda}u_i, \tilde{\mathbf{B}}_i = \mathbf{B} + \mathbf{\Theta}v_i,$$

with $\mathbb{E}[\kappa(\mathbf{\Lambda})^2], \mathbb{E}[\kappa(\mathbf{\Theta})^2] \geq 1$ and $|u_i|, |v_i| \geq \sigma_{\mathsf{Con}}$ for all $i \in [P]$.

Pick a distribution $\mathbb{P}_{\mathbf{A},\mathbf{B}}$ such that $\mathbf{A}, \mathbf{B}$ i.i.d. diagonal matrices, with i.i.d. entries having Bernoulli distribution, i.e.,

$$\Pr((\mathbf{A})_{m,n} = 0) = 1 \; \forall \; m \neq n, \; m, n \in [L],$$

$$\Pr\left((\mathbf{A})_{m,m} = \frac{1}{\sqrt{L}}\right) = \Pr\left((\mathbf{A})_{m,m} = -\frac{1}{\sqrt{L}}\right) = \frac{1}{2} \; \forall \; m \in [L].$$

Observe that $\mathbb{E}[||\mathbf{A}||_F^2], \mathbb{E}[||\mathbf{B}||_F^2] = 1$ and $\mathbb{E}[\mathbf{A}] = \mathbb{E}[\mathbf{B}] = 0$.

Again for convenience of illustration we drop the dependence on $\mathcal{S}$ for decoding weights, i.e., $w_{1,\mathcal{S}}, w_{2,\mathcal{S}}$ will be written as w_1, w_2 respectively. Thus,

$$\widehat{\mathbf{C}}_{\mathcal{S}} = w_1 \tilde{\mathbf{A}}_i \tilde{\mathbf{B}}_i + w_2 \tilde{\mathbf{A}}_j \tilde{\mathbf{B}}_j.$$

$\square$

$$\mathsf{LMSE}_{\mathcal{S}}(\Gamma) = \mathbb{E}[||\mathbf{AB} - \widehat{\mathbf{C}}_{\mathcal{S}}||_F^2] \tag{A.24}$$

$$= \mathbb{E}[||w_1(\mathbf{AB} + \mathbf{\Lambda B}u_i + \mathbf{A\Theta}v_i + \mathbf{\Lambda\Theta}u_iv_i) + w_2(\mathbf{AB} + \mathbf{\Lambda B}u_j + \mathbf{A\Theta}v_j + \mathbf{\Lambda\Theta}u_jv_j) - \mathbf{AB}||_F^2] \tag{A.25}$$

$$= \mathbb{E}[||(w_1 + w_2 - 1)\mathbf{AB} + (w_1u_i + w_2u_j)\mathbf{\Lambda B} + (w_1v_i + w_2v_j)\mathbf{A\Theta} + (w_1u_iv_i + w_2u_jv_j)\mathbf{\Lambda\Theta}||_F^2] \tag{A.26}$$

Since $\mathbf{A}, \mathbf{B}, \mathbf{\Lambda}, \mathbf{\Theta}$ are independent and $\mathbf{A}, \mathbf{B}$ are zero mean, the cross products vanish,

$$= (w_1 + w_2 - 1)^2\mathbb{E}[||\mathbf{AB}||_F^2] + (w_1u_i + w_2u_j)^2\mathbb{E}[||\mathbf{\Lambda B}||_F^2] + (w_1v_i + w_2v_j)^2\mathbb{E}[||\mathbf{A\Theta}||_F^2] + (w_1u_iv_i + w_2u_jv_j)^2\mathbb{E}[||\mathbf{\Lambda\Theta}||_F^2] \tag{A.27}$$

$$\geq (w_1 + w_2 - 1)^2\mathbb{E}[||\mathbf{A}||_F^2]\mathbb{E}[||\mathbf{B}||_F^2] + (w_1u_i + w_2u_j)^2\mathbb{E}[\kappa(\mathbf{\Lambda})^2]\mathbb{E}[||\mathbf{B}||_F^2] + (w_1v_i + w_2v_j)^2\mathbb{E}[||\mathbf{A}||_F^2]\mathbb{E}[\kappa(\mathbf{\Theta})^2] + (w_1u_iv_i + w_2u_jv_j)^2\mathbb{E}[\kappa(\mathbf{\Lambda})^2]\mathbb{E}[\kappa(\mathbf{\Theta})^2] \tag{A.28}$$

$$\geq (w_1 + w_2 - 1)^2 + (w_1u_i + w_2u_j)^2 + (w_1v_i + w_2v_j)^2 + (w_1u_iv_i + w_2u_jv_j)^2 \tag{A.29}$$

$$= ||\mathbf{M}||_F^2 \tag{A.30}$$

where $\mathbf{M}$ is defined to be:

$$\mathbf{M} = \left(w_1 \begin{bmatrix} 1 \\ u_i \end{bmatrix} \begin{bmatrix} 1 & v_i \end{bmatrix} + w_2 \begin{bmatrix} 1 \\ u_j \end{bmatrix} \begin{bmatrix} 1 & v_j \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right). \tag{A.31}$$

Multiplying by $\begin{bmatrix} -v_j \\ 1 \end{bmatrix}$ on both sides and taking ℓ_2-norm,

$$\left\| \mathbf{M} \begin{bmatrix} -v_j \\ 1 \end{bmatrix} \right\|_2^2 = (w_1(v_i - v_j) - v_j)^2 + w_1^2(v_i - v_j)^2u_i^2 \tag{A.32}$$

The above is a convex quadratic equation in w_1 and is minimized at,

$$w_1^* = \frac{v_j}{(1 + u_i^2)(v_i - v_j)}.$$

Substituting w_1^* in equation (A.32) and simplifying we write,

$$\frac{v_j^2u_i^2}{1 + u_i^2} \leq \left\| \mathbf{M} \begin{bmatrix} -v_j \\ 1 \end{bmatrix} \right\|_2^2 \tag{A.33}$$

$$\leq ||\mathbf{M}||_2^2(1 + v_j^2) \tag{A.34}$$

$$\implies ||\mathbf{M}||_F^2 \geq ||\mathbf{M}||_2^2 \geq \frac{u_i^2}{1 + u_i^2}\frac{v_j^2}{1 + v_j^2} \tag{A.35}$$

$$\geq \frac{1}{(1 + \frac{1}{\sigma_{\mathsf{Con}}^2})^2} \tag{A.36}$$

Therefore,

$$\mathsf{LMSE}_{\mathcal{S}}(\Gamma) \geq \frac{1}{(1 + \frac{1}{\sigma_{\mathsf{Con}}^2})^2} \tag{A.37}$$

Similar as in achievability, we now prove a lower bound on σ_{Con}^2 and use it the above result.

Proof of Theorem 4.4. Consider a (P, N) coding scheme Γ which satisfies $\mathbb{E}(\kappa(\mathbf{R}_i)^2) \geq \sigma_{\mathsf{Con}}^2$ and $\mathbb{E}(\kappa(\mathbf{S}_i)^2) \geq \sigma_{\mathsf{Con}}^2$. We aim to lower bound σ_{Con}^2 for a given ϵ, i.e., we say that given some ϵ we cannot use σ_{Con}^2 less than some quantity and still satisfy ϵ-DP. Thus, we assume worst case inputs and derive σ_{Con}^2 achievable for that worst case scenario. We further use this to show that LMSE below some quantity is not achievable given an ϵ. Without loss of generality assume that $\sigma_{\mathsf{Con}}^2 = \mathbb{E}[\kappa(\mathbf{R}_k)^2]$, it suffices to concentrate on k^{th} node. Let $\mathbf{X}_l = \begin{bmatrix} \mathbf{A}_l \\ \mathbf{B}_l \end{bmatrix}$. And let,

$$\mathbf{Y}^{(0)} = \mathbf{X}_0 + \begin{bmatrix} \mathbf{R}_k \\ \mathbf{S}_k \end{bmatrix},$$

$$\mathbf{Y}^{(1)} = \mathbf{X}_1 + \begin{bmatrix} \mathbf{R}_k \\ \mathbf{S}_k \end{bmatrix},$$

From differential privacy Definition 3.1, for any subset $\mathcal{A} \subset \mathbb{R}^{2L \times L}$ for any $||\mathbf{X}_0 - \mathbf{X}_1||_{\max} \leq 1$ it is true that

$$\frac{P(\mathbf{Y}^{(1)} \in \mathcal{A})}{P(\mathbf{Y}^{(0)} \in \mathcal{A})} \leq e^\epsilon \text{ and } \frac{P(\mathbf{Y}^{(0)} \in \mathcal{A})}{P(\mathbf{Y}^{(1)} \in \mathcal{A})} \leq e^\epsilon.$$

The worst case inputs have $||\mathbf{X}_0 - \mathbf{X}_1||_{\max} = 1$. Without loss of generality we pick $\mathbf{X}_0 = \mathbf{0}_{2L \times L}$ and $\mathbf{X}_1 = \mathbf{1}_{2L \times L}$, where $\mathbf{0}_{2L \times L}$ and $\mathbf{1}_{2L \times L}$ are $2L \times L$ matrices with all elements 0's and 1's respectively. For some $\bar{\mathbf{R}} \in \mathbb{R}^{L \times L}$, pick $\mathcal{A} = \left\{ \begin{bmatrix} \bar{\mathbf{R}} \\ \mathbb{R}^{L \times L} \end{bmatrix} \right\}$. Denote the pdf of $\mathbf{R}_k$ to be $p_{\mathbf{R}_k}$ and $\mathbf{1}$ an $L \times L$ matrix with all elements to be 1. Then by independence of $\mathbf{R}_k$ and $\mathbf{S}_k$ and from our choice of $\mathcal{A}$, we can effectively ignore the role of $\mathbf{S}_k$ and rewrite the above as,

$$\frac{p_{\mathbf{R}_k}(\bar{\mathbf{R}} - \mathbf{1})}{p_{\mathbf{R}_k}(\bar{\mathbf{R}})} \leq e^\epsilon \text{ and } \frac{p_{\mathbf{R}_k}(\bar{\mathbf{R}})}{p_{\mathbf{R}_k}(\bar{\mathbf{R}} - \mathbf{1})} \leq e^\epsilon.$$

Using post processing property [7] of differential privacy we can use $\kappa(\cdot)$ map without violating ϵ-DP. Denote the pdf of $\kappa(\mathbf{R}_k)$ to be $p_{\kappa(\mathbf{R}_k)}$. After $\kappa(\cdot)$ mapping we have two cases,

1) $\kappa(\bar{\mathbf{R}} - \mathbf{1}) \leq \kappa(\bar{\mathbf{R}})$: Take

$$\frac{p_{\kappa(\mathbf{R}_k)}(\kappa(\bar{\mathbf{R}} - \mathbf{1}))}{p_{\kappa(\mathbf{R}_k)}(\kappa(\bar{\mathbf{R}}))} \leq e^\epsilon.$$

$$\implies P(\kappa(\mathbf{R}_k) \leq \kappa(\bar{\mathbf{R}} - \mathbf{1})) \leq e^\epsilon P(\kappa(\mathbf{R}_k) \leq \kappa(\bar{\mathbf{R}}))$$

Using Weyl's inequality for singular values [25], [26] $\kappa(\bar{\mathbf{R}} - \mathbf{1}) \geq \kappa(\bar{\mathbf{R}}) - ||\mathbf{1}||_2$, we write,

$$P(\kappa(\mathbf{R}_k) \leq \kappa(\bar{\mathbf{R}}) - ||\mathbf{1}||_2) \leq e^\epsilon P(\kappa(\mathbf{R}_k) \leq \kappa(\bar{\mathbf{R}}))$$

$$\implies \frac{p_{\kappa(\mathbf{R}_k)}(\kappa(\bar{\mathbf{R}}) - ||\mathbf{1}||_2)}{p_{\kappa(\mathbf{R}_k)}(\kappa(\bar{\mathbf{R}}))} \leq e^\epsilon.$$

Let $\kappa(\bar{\mathbf{R}}) = r$ and we know that $||\mathbf{1}||_2 = L$, we write,

$$\frac{p_{\kappa(\mathbf{R}_k)}(r-L)}{p_{\kappa(\mathbf{R}_k)}(r)} \leq e^{\epsilon}.$$

2) $\kappa(\bar{\mathbf{R}} - \mathbf{1}) \geq \kappa(\bar{\mathbf{R}})$: Take

$$\frac{p_{\kappa(\mathbf{R}_k)}(\kappa(\bar{\mathbf{R}}))}{p_{\kappa(\mathbf{R}_k)}(\kappa(\bar{\mathbf{R}} - \mathbf{1}))} \leq e^{\epsilon}.$$

$$\implies P(\kappa(\mathbf{R}_k) \leq \kappa(\bar{\mathbf{R}})) \leq e^{\epsilon} P(\kappa(\mathbf{R}_k) \leq \kappa(\bar{\mathbf{R}} - \mathbf{1}))$$

Using Weyl's inequality for singular values [25], [26], $\kappa(\bar{\mathbf{R}} - \mathbf{1}) \leq \kappa(\bar{\mathbf{R}}) + ||\mathbf{1}||_2$, we write,

$$P(\kappa(\mathbf{R}_k) \leq \kappa(\bar{\mathbf{R}} - \mathbf{1}) - ||\mathbf{1}||_2) \leq e^{\epsilon} P(\kappa(\mathbf{R}_k) \leq \kappa(\bar{\mathbf{R}} - \mathbf{1}))$$

$$\implies \frac{p_{\kappa(\mathbf{R}_k)}(\kappa(\bar{\mathbf{R}} - \mathbf{1}) - ||\mathbf{1}||_2)}{p_{\kappa(\mathbf{R}_k)}(\kappa(\bar{\mathbf{R}} - \mathbf{1}))} \leq e^{\epsilon}.$$

Let $\kappa(\bar{\mathbf{R}} - \mathbf{1}) = r$ and we know that $||\mathbf{1}||_2 = L$, we write,

$$\frac{p_{\kappa(\mathbf{R}_k)}(r-L)}{p_{\kappa(\mathbf{R}_k)}(r)} \leq e^{\epsilon}.$$

Thus, we have shown that the above relation is true for any r, in turn true for arbitrary choice of $\bar{\mathbf{R}}$. Now,

$$\sigma^2_{\mathsf{Con}} = \mathbb{E}[\kappa(\mathbf{R}_k)^2] = \int_{-\infty}^{\infty} r^2 p_{\kappa(\mathbf{R}_k)}(r) dr \tag{A.38}$$

$$\geq e^{-\epsilon} \int_{-\infty}^{\infty} r^2 p_{\kappa(\mathbf{R}_k)}(r-L) dr \tag{A.39}$$

$$= e^{-\epsilon} \int_{-\infty}^{\infty} (r+L)^2 p_{\kappa(\mathbf{R}_k)}(r) dr \tag{A.40}$$

$$= e^{-\epsilon} (\sigma^2_{\mathsf{Con}} + L^2 + 2L\mathbb{E}[\kappa(\mathbf{R}_k)]) \tag{A.41}$$

$$\geq e^{-\epsilon} (\sigma^2_{\mathsf{Con}} + L^2) \tag{A.42}$$

$$\implies \sigma^2_{\mathsf{Con}} \geq \frac{L^2}{e^{\epsilon} - 1} \tag{A.43}$$

Substituting the above in (A.37) gives us the desired relation. □

APPENDIX B

========================WIP==============

In this appendix, we will prove a lemma, which pertains to minimization of LMSE. We aim to show that for our system model, considering $\mathbb{E}[R_i] = \mathbb{E}[S_i] = 0$ for all $i \in [P]$, and $R_i = \bar{k}_j R_j$ for all $i \neq j$ and some constants $\bar{k}_j \in \mathbb{R}$, provides minimum LMSE. We write these conditions formally. We represent the linear relation statement in terms of the Pearson's correlation coefficient, which for any two random variables Λ and Θ is defined as,

$$\rho_{\Lambda,\Theta} = \frac{\mathsf{Cov}(\Lambda, \Theta)}{\sqrt{\mathsf{Var}(\Lambda)\mathsf{Var}(\Theta)}}.$$

Recall that $|\rho_{\Lambda,\Theta}| \leq 1$. Let the decoding weights be $w_{1,\mathcal{S}}, w_{2,\mathcal{S}}$ and let $\mathcal{S} = \{i, j\}$. Expanding the LMSE given in (3.2),

$$\begin{aligned}\mathsf{LMSE}_{\mathcal{S}}(\Gamma) = \mathbb{E}[&-w_{1,\mathcal{S}} (AB + AS_i + BR_i + R_i S_i) \\ &- w_{2,\mathcal{S}} (AB + AS_j + BR_j + R_j S_j) + AB]^2.\end{aligned} \tag{B.1}$$

Let

$$k_{1,\mathcal{S}} = \frac{1}{\sqrt{\mathbb{E}[R_i^2]\mathbb{E}[R_j^2]}} \text{ and } k_{2,\mathcal{S}} = \frac{1}{\sqrt{\mathbb{E}[S_i^2]\mathbb{E}[S_j^2]}}.$$

Condition B.1. Random vectors $\mathbf{R}$ and $\mathbf{S}$ have zero mean.

Condition B.2. For all i, j and for some $w_{1,\mathcal{S}}, w_{2,\mathcal{S}}$ random vectors $\mathbf{R}$ and $\mathbf{S}$ satisfy,

$$(\rho_{R_i,R_j}, \rho_{S_i,S_j}) = \begin{cases} (1,1) & \text{if } w_{1,\mathcal{S}} w_{2,\mathcal{S}} < 0, \\ (-1,1) & \text{if } w_{1,\mathcal{S}} w_{2,\mathcal{S}} > 0 \text{ and } k_{1,\mathcal{S}} \leq k_{2,\mathcal{S}}, \\ (1,-1) & \text{if } w_{1,\mathcal{S}} w_{2,\mathcal{S}} > 0 \text{ and } k_{1,\mathcal{S}} > k_{2,\mathcal{S}}. \end{cases}$$

To prove that using these conditions is optimal for LMSE we first construct a multivariate normal distribution with the desired moments, since normal distribution is completely parametrized by its mean and covariance matrix. Then we show that any normal distribution not having these moments has a worse LMSE than that is obtained by using the specified normal distribution. After which we use the fact that LMSE is only dependent on second order statistics to argue that this in fact applies to any distribution with the desired properties.

Let $\Gamma^{\mathbb{Q}_{\mathbf{R},\mathbf{S}}}$ be a $(P, N = 2)$ coding scheme consisting of some joint distribution $\mathbb{Q}_{\mathbf{R},\mathbf{S}}$ and decoding maps $\prod_{\mathcal{S} \subseteq P: |\mathcal{S}| = N} \{d_{\mathcal{S}} : (\mathbb{R}^{L \times L})^{|\mathcal{S}|} \to \mathbb{R}^{L \times L}\}$.

Lemma B.1. *For any distribution $\mathbb{H}_{\mathbf{R},\mathbf{S}} \in \mathcal{P}_{\mathbf{R},\mathbf{S}}$, for any set $\mathcal{S} \subseteq [P], |\mathcal{S}| = 2$, there exists a $\mathbb{G}_{\mathbf{R},\mathbf{S}}$ such that*

$$\mathsf{LMSE}_{\mathcal{S}}(\Gamma^{\mathbb{H}_{\mathbf{R},\mathbf{S}}}) \geq \mathsf{LMSE}_{\mathcal{S}}(\Gamma^{\mathbb{G}_{\mathbf{R},\mathbf{S}}}),$$

where $\mathbb{G}_{\mathbf{R},\mathbf{S}} \in \mathcal{P}_{\mathbf{R},\mathbf{S}}$ is a multivariate normal distribution with $\mathbb{E}_{R_i \sim \mathbb{G}}[R_i^2] = \mathbb{E}_{R_i \sim \mathbb{H}}[R_i^2]$ and $\mathbb{E}_{S_i \sim \mathbb{G}}[S_i^2] = \mathbb{E}_{S_i \sim \mathbb{H}}[S_i^2]$ and satisfies conditions B.1 and B.2.

Proof. For simplicity we drop the dependence on $\mathcal{S}$ for the decoding weights. Let the decoding weights be w_1 and w_2 and let the variances of R_i, S_i, R_j, S_j be fixed, i.e., only means and covariances can be varied. Note that this is possible since we assumed normal distribution.

$$\begin{aligned}\mathsf{LMSE}_{\mathcal{S}}(d_{\mathcal{S}}, G'_{\mathbf{R},\mathbf{S}}) = \mathbb{E}[(&-w_1 (AB + AS_i + BR_i + R_i S_i) \\ &- w_2 (AB + AS_j + BR_j + R_j S_j) + AB)^2]\end{aligned} \tag{B.2}$$

$$\begin{aligned}&= \mathbb{E}[(R_i^2 S_i^2 + R_i^2 + S_i^2) w_1^2 + (R_j^2 S_j^2 + R_j^2 + S_j^2) w_2^2 \\ &+ (w_1 + w_2 - 1)^2] + 2w_1 w_2 \mathbb{E}[R_i R_j S_i S_j + R_i R_j + S_i S_j]\end{aligned} \tag{B.3}$$

Using the fact that $\mathbb{E}[R_i^2] = \mathsf{Var}(R_i) + \mathbb{E}(R_i)^2, \mathbb{E}[R_iR_j] = \mathsf{Cov}(R_i, R_j) + \mathbb{E}(R_i)\mathbb{E}(R_j)$ and expanding,

$$
\begin{aligned}
&= (\mathsf{Var}(R_i)\mathsf{Var}(S_i) + \mathsf{Var}(R_i) + \mathsf{Var}(S_i))w_1^2 \\
&+ (\mathsf{Var}(R_j)\mathsf{Var}(S_j) + \mathsf{Var}(R_j) + \mathsf{Var}(S_j))w_2^2 + (w_1 + w_2 - 1)^2 \\
&+ 2w_1w_2(\mathsf{Cov}(R_i, R_j)\mathsf{Cov}(S_i, S_j) + \mathsf{Cov}(R_i, R_j) + \mathsf{Cov}(S_i, S_j)) \\
&+ (w_1\mathbb{E}(R_i) + w_2\mathbb{E}(R_j))^2 + (w_1\mathbb{E}(S_i) + w_2\mathbb{E}(S_j))^2 \\
&+ (w_1\mathbb{E}(R_i)\mathbb{E}(S_i) + w_2\mathbb{E}(R_j)\mathbb{E}(S_j))^2 \\
&+ (w_1^2\mathbb{E}(R_i)^2\mathsf{Var}(S_i) + w_2^2\mathbb{E}(R_j)^2\mathsf{Var}(S_j) \\
&+ 2w_1w_2\mathsf{Cov}(S_i, S_j)\mathbb{E}[R_i]\mathbb{E}[R_j]) \\
&+ (w_1^2\mathbb{E}(S_i)^2\mathsf{Var}(R_i) + w_2^2\mathbb{E}(S_j)^2\mathsf{Var}(R_j) \\
&+ 2w_1w_2\mathsf{Cov}(R_i, R_j)\mathbb{E}[S_i]\mathbb{E}[S_j]) \qquad \text{(B.4)}
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(1)}{\geq} (\mathsf{Var}(R_i)\mathsf{Var}(S_i) + \mathsf{Var}(R_i) + \mathsf{Var}(S_i))w_1^2 \\
&+ (\mathsf{Var}(R_j)\mathsf{Var}(S_j) + \mathsf{Var}(R_j) + \mathsf{Var}(S_j))w_2^2 + (w_1 + w_2 - 1)^2 \\
&+ 2w_1w_2(\mathsf{Cov}(R_i, R_j)\mathsf{Cov}(S_i, S_j) + \mathsf{Cov}(R_i, R_j) + \mathsf{Cov}(S_i, S_j)) \\
&+ (w_1\mathbb{E}(R_i) + w_2\mathbb{E}(R_j))^2 + (w_1\mathbb{E}(S_i) + w_2\mathbb{E}(S_j))^2 \\
&+ (w_1\mathbb{E}(R_i)\mathbb{E}(S_i) + w_2\mathbb{E}(R_j)\mathbb{E}(S_j))^2 \\
&+ \left(w_1\mathbb{E}(R_i)\sqrt{\mathsf{Var}(S_i)} - \mathsf{sgn}(w_1w_2)w_2\mathbb{E}(R_j)\sqrt{\mathsf{Var}(S_j)}\right)^2 \\
&+ \left(w_1\mathbb{E}(S_i)\sqrt{\mathsf{Var}(R_i)} - \mathsf{sgn}(w_1w_2)w_2\mathbb{E}(S_j)\sqrt{\mathsf{Var}(R_j)}\right)^2 \qquad \text{(B.5)}
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(2)}{\geq} (\mathsf{Var}(R_i)\mathsf{Var}(S_i) + \mathsf{Var}(R_i) + \mathsf{Var}(S_i))w_1^2 \\
&+ (\mathsf{Var}(R_j)\mathsf{Var}(S_j) + \mathsf{Var}(R_j) + \mathsf{Var}(S_j))w_2^2 + (w_1 + w_2 - 1)^2 \\
&+ 2w_1w_2(\mathsf{Cov}(R_i, R_j)\mathsf{Cov}(S_i, S_j) + \mathsf{Cov}(R_i, R_j) + \mathsf{Cov}(S_i, S_j)) \qquad \text{(B.6)}
\end{aligned}
$$

where,

$$
\mathsf{sgn}(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0. \end{cases}
$$

Inequality (1) is obtained by using the fact that $|\mathsf{Cov}(\Lambda, \Theta)| \leq \sqrt{\mathsf{Var}(\Lambda)\mathsf{Var}(\Theta)}$. A simple way to obtain inequality (2) from inequality (1) is to choose

$$\mathbb{E}(R_i) = \mathbb{E}(R_j) = \mathbb{E}(S_i) = \mathbb{E}(S_j) = 0.$$

This shows condition B.1.

Denote correlation coefficients between R_i and R_j to be ρ_{R_i,R_j}, and S_i and S_j to be ρ_{S_i,S_j}. We start from Equation (B.6) and account the zero mean condition.

$$
\begin{aligned}
&\mathsf{LMSE}_{\mathcal{S}}(d_{\mathcal{S}}, G'_{\mathbf{R},\mathbf{S}}) \geq \mathbb{E}[(R_i^2S_i^2 + R_i^2 + S_i^2)w_1^2 \\
&+ (R_j^2S_j^2 + R_j^2 + S_j^2)w_2^2 + (w_1 + w_2 - 1)^2] \\
&+ 2w_1w_2(\mathbb{E}[R_iR_j] + \mathbb{E}[S_iS_j] + \mathbb{E}[R_iR_j]\mathbb{E}[S_iS_j]) \qquad \text{(B.7)} \\
&= \bar{k} + \sqrt{\mathbb{E}[R_i^2]\mathbb{E}[R_j^2]\mathbb{E}[S_i^2]\mathbb{E}[S_j^2]} \\
&\quad 2w_1w_2\left(k_{2,\mathcal{S}}\rho_{R_i,R_j} + k_{1,\mathcal{S}}\rho_{S_i,S_j} + \rho_{R_i,R_j}\rho_{S_i,S_j}\right) \qquad \text{(B.8)}
\end{aligned}
$$

where $\bar{k} = \mathbb{E}[(R_i^2S_i^2 + R_i^2 + S_i^2)w_1^2 + (R_j^2S_j^2 + R_j^2 + S_j^2)w_2^2 + (w_1 + w_2 - 1)^2]$.

Now we minimize LMSE over various correlations ρ_{R_i,R_j} and ρ_{S_i,S_j}. Again, this is possible since we assumed normal distribution. We re-write the final equation as two optimization problems over the correlations, for some $c, d > 0$,

$$
\begin{aligned}
(\mathrm{O}_1)\ \min_{x,y}\quad & cx + dy + xy \\
\text{s.t.}\quad & 1 - x^2 \geq 0, \\
& 1 - y^2 \geq 0.
\end{aligned}
$$

$$
\begin{aligned}
(\mathrm{O}_2)\ \max_{x,y}\quad & cx + dy + xy \\
\text{s.t.}\quad & 1 - x^2 \geq 0, \\
& 1 - y^2 \geq 0.
\end{aligned}
$$

where $x = \rho_{R_i,R_j}, y = \rho_{S_i,S_j}$ and $c = k_{2,\mathcal{S}}, d = k_{1,\mathcal{S}}$. O_1 is solved if w_1w_2 is positive and O_2 is solved if w_1w_2 is negative. Let the slack variables be s and t.

$$
\begin{aligned}
(\mathrm{O}_1)\ \min_{x,y}\quad & cx + dy + xy \\
\text{s.t.}\quad & 1 - x^2 - s^2 = 0, \\
& 1 - y^2 - t^2 = 0.
\end{aligned}
$$

$$
\begin{aligned}
(\mathrm{O}_2)\ \max_{x,y}\quad & cx + dy + xy \\
\text{s.t.}\quad & 1 - x^2 - s^2 = 0, \\
& 1 - y^2 - t^2 = 0.
\end{aligned}
$$

Both problems have same Langrangian formulation. Let $\lambda_1, \lambda_2 \in \mathbb{R}$ be the lagrange multipliers.

$$
\begin{aligned}
\mathcal{L}(x, y, \lambda_1, \lambda_2, s, t) &= cx + dy + xy \\
&\quad - \lambda_1(1 - x^2 - s^2) - \lambda_2(1 - y^2 - t^2). \qquad \text{(B.9)}
\end{aligned}
$$

Taking gradient of $\mathcal{L}$ w.r.t. x, y, s, t and equating to 0, combined with constraints gives following equations,

$$c + y = -2\lambda_1 x \qquad \text{(B.10)}$$
$$d + x = -2\lambda_2 y \qquad \text{(B.11)}$$
$$0 = -2\lambda_1 s \qquad \text{(B.12)}$$
$$0 = -2\lambda_2 t \qquad \text{(B.13)}$$
$$1 - x^2 - s^2 = 0 \qquad \text{(B.14)}$$
$$1 - y^2 - t^2 = 0 \qquad \text{(B.15)}$$

Let $f(x, y) = cx + dy + xy$. The complementary slackness conditions gives the 4 cases:

1) $\lambda_1 = 0, \lambda_2 = 0, s \neq 0, t \neq 0$:

$$\implies x = -d, y = -c \implies 0 \leq c \leq 1, 0 \leq d \leq 1.$$

$$f(-d, -c) = -cd,\ 0 \leq c \leq 1, 0 \leq d \leq 1.$$

2) $\lambda_1 = 0, \lambda_2 \neq 0, s \neq 0, t = 0$:

$$\implies y = -c = -1, x = 2\lambda_2 - d.$$

$$f(2\lambda_2 - d, -1) = -d$$

Observe that in this case, for any $d > 0$, there's multiple x that satisfy the equations, and we choose λ_2 such that $x = 1$.

$$f(1, -1) = -d,\ c = 1, d > 0$$

3) $\lambda_1 \neq 0, \lambda_2 = 0, s = 0, t \neq 0$:

$$\implies x = -d = -1, y = 2\lambda_1 - c.$$

$$f(-1, 2\lambda_1 - c) = -c$$

Observe that in this case, for any $c > 0$, there's multiple y that satisfy the equations, and we choose λ_1 such that $y = 1$.

$$f(-1, 1) = -c,\ c > 0, d = 1$$

4) $\lambda_1 \neq 0, \lambda_2 \neq 0, s = 0, t = 0$:

$$\implies x = \pm 1, y = \pm 1.$$

We choose appropriate values for x and y based on c and d such that $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$. Therefore for any $c > 0$ and $d > 0$,

a) $f(1, 1) = 1 + c + d$.
b) $f(1, -1) = -1 + c - d$.
c) $f(-1, 1) = -1 - c + d$.
d) $f(-1, -1) = 1 - c - d$.

Observe that Case 2 is identical to Case 4b and Case 3 is identical to Case 4c. Also observe that for $0 \leq c \leq 1$ and $0 \leq d \leq 1$, $\min(-1 + c - d, -1 - c + d) < -cd$. To solve O_1 we have take the minimum of the function evaluations at these stationary points, and for any $c > 0$ and $d > 0$ that value is $\min(-1 + c - d, -1 - c + d)$ attained at $(x, y) = (1, -1)$ or $(-1, 1)$. To solve O_2 we have take the maximum of the function evaluations at these stationary points, and for any $c > 0$ and $d > 0$ that value is $1 + c + d$ attained at $(x, y) = (1, 1)$.

Therefore, we obtain the condition B.2 and that

$$\mathsf{LMSE}_{\mathcal{S}}(d_S, G'_{\mathbf{R},\mathbf{S}}) \geq \mathsf{LMSE}_{\mathcal{S}}(d_S, G_{\mathbf{R},\mathbf{S}}).$$

□

Since LMSE only depends on second order statistics, we say the following.

Corollary B.1.1. *For any distribution $H'_{\mathbf{R},\mathbf{S}} \in \mathcal{P}_{\mathbf{R},\mathbf{S}}$, for any set $\mathcal{S} \subseteq [P], |\mathcal{S}| = 2$ and any $d_{\mathcal{S}}$, there exists a $H_{\mathbf{R},\mathbf{S}}$ such that*

$$\mathsf{LMSE}_{\mathcal{S}}(d_{\mathcal{S}}, H'_{\mathbf{R},\mathbf{S}}) \geq \mathsf{LMSE}_{\mathcal{S}}(d_{\mathcal{S}}, H_{\mathbf{R},\mathbf{S}}),$$

where $H_{\mathbf{R},\mathbf{S}} \in \mathcal{P}_{\mathbf{R},\mathbf{S}}$ is some distribution with $\mathbb{E}_{R_i \sim H}[R_i^2] = \mathbb{E}_{R_i \sim H'}[R_i^2]$ and $\mathbb{E}_{S_i \sim H}[S_i^2] = \mathbb{E}_{S_i \sim H'}[S_i^2]$ and satisfies conditions B.1 and B.2.

Proof. Given $H'_{\mathbf{R},\mathbf{S}}$, we compute its mean and covariance matrix and use them construct a multivariate normal distribution $G'_{\mathbf{R},\mathbf{S}}$. And since LMSE only depends on the second order statistics and not on the underlying distribution, we write

$$\mathsf{LMSE}_{\mathcal{S}}(d_S, H'_{\mathbf{R},\mathbf{S}}) = \mathsf{LMSE}_{\mathcal{S}}(d_S, G'_{\mathbf{R},\mathbf{S}}) \tag{B.16}$$

$$\overset{\text{Lemma } B.1}{\geq} \mathsf{LMSE}_{\mathcal{S}}(d_S, G_{\mathbf{R},\mathbf{S}}) \tag{B.17}$$

$$= \mathsf{LMSE}_{\mathcal{S}}(d_S, H_{\mathbf{R},\mathbf{S}}). \tag{B.18}$$

□