



HAL
open science

Multi-User Distributed Computing Via Compressed Sensing

Ali Khalesi, Sajad Daei, Marios Kountouris, Petros Elia

► **To cite this version:**

Ali Khalesi, Sajad Daei, Marios Kountouris, Petros Elia. Multi-User Distributed Computing Via Compressed Sensing. ITW 2023, IEEE Information Theory Workshop, 23-28 April 2023, Saint-Malo, France, IEEE, Apr 2023, Saint-Malo, France. pp.509-514, 10.1109/ITW55543.2023.10160241 . hal-04093028

HAL Id: hal-04093028

<https://hal.science/hal-04093028>

Submitted on 11 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multi-User Distributed Computing Via Compressed Sensing

Ali Khalesi, Sajad Daei, Marios Kountouris, Petros Elia

Communication Systems Department, EURECOM

06410 Sophia Antipolis, France

Email: {ali.khalesi;sajad.daei;marios.kountouris;petros.elia}@eurecom.fr

Abstract—The multi-user linearly-separable distributed computing problem is considered here, in which N servers help to compute the real-valued functions requested by K users, where each function can be written as a linear combination of up to L (generally non-linear) subfunctions. Each server computes a fraction γ of the subfunctions, then communicates a function of its computed outputs to some of the users, and then each user collects its received data to recover its desired function. Our goal is to bound the ratio between the computation workload done by all servers over the number of datasets.

To this end, we here reformulate the real-valued distributed computing problem into a matrix factorization problem and then into a basic sparse recovery problem, where sparsity implies computational savings. Building on this, we first give a simple probabilistic scheme for subfunction assignment, which allows us to upper bound the optimal normalized computation cost as $\gamma \leq \frac{K}{N}$ that a generally intractable ℓ_0 -minimization would give. To bypass the intractability of such optimal scheme, we show that if these optimal schemes enjoy $\gamma \leq -r \frac{K}{N} W_{-1}^{-1}(-\frac{2K}{eNr})$ (where $W_{-1}(\cdot)$ is the Lambert function and r calibrates the communication between servers and users), then they can actually be derived using a tractable Basis Pursuit ℓ_1 -minimization. This newly-revealed connection opens up the possibility of designing practical distributed computing algorithms by employing tools and methods from compressed sensing.

Index Terms—Distributed computing, Linearly-separable functions, Compressed sensing, Sparse representation.

I. INTRODUCTION

Distributed computing plays an important role in speeding up non-linear and computationally hard computing tasks. As the complexity of these tasks increases, there is an ever rising need for novel parallel processing techniques that efficiently offload computations to groups of distributed servers. This same complexity increase also brings about many challenges, including, to name a few, computing accuracy [1], [2], scalability [3]–[5], privacy and security [6]–[10], and latency and straggler mitigation [11]–[15]. There is also substantial research work on various explorations of communication vs. computation trade-off [14], [16], [17], [18]. Moreover, motivated by the applicability of deriving schemes that work on real numbers, other interesting related settings, such as the Lagrange-coded secret sharing over real numbers [19], coded distributed polynomial evaluation over complex matrices [20],

[21], as well as the setting of secure distributed multiplication of real or complex matrices [22], have emerged. For a detailed survey of related research works, the interested reader is referred to [23], [24].

This same aforementioned complexity increase, has also brought about various related frameworks such as MapReduce [25] and Spark [26] that apply to broad classes of functions. Focusing on functions over finite fields, the recent work in [27], [28]¹ proposed the so-called *Multi-User Linearly-Separable Distributed-Computing framework*, which allows for distributed computation of functions that adhere to the very broad linearly separable format, which in turn captures various classes of linear and non-linear functions of practical interest². Such functions have the form

$$F(D_1, D_2, \dots, D_L) = \sum_{l=1}^L f_l g_l(D_l)$$

where D_1, D_2, \dots, D_L are the L input datasets, $W_l = g_l(D_l)$ are the computed outputs of basis subfunctions $g_l(D_l)$, and f_l are scalar coefficients. In the multi-user (K users and N servers) setting where each user asks for its own function, the work in [27] transformed the distributed computing problem into a simple (preferably sparse) matrix factorization problem over finite fields, and then proceeded to make the direct connection between distributed computing, matrix factorization, and a new coding theoretic approach. In particular, for a $K \times L$ demand matrix \mathbf{F} where each row describes the coefficients that define the function requested by a user, the problem was transformed into the factorization problem $\mathbf{F} = \mathbf{D}\mathbf{E}$, where \mathbf{D} and \mathbf{E} are the decoding and encoding matrices respectively. \mathbf{E} dictated which server should compute which subfunction and then how each server should combine the computed outputs before transmitting, while the $K \times N$ decoding matrix \mathbf{D} dictated which user should each server communicate to and how each user should combine the various received signals. Then, a solution was proposed that derived from the powerful class of covering codes, and from a new class of so-called *partial covering codes*. In particular, the parity-check matrix from such codes played the role of \mathbf{D} , while then, after

⁰This work was supported by the European Research Council (ERC) through the EU Horizon 2020 Research and Innovation Programme under Grant 725929 (Project DUALITY) and Grant 101003431 (Project SONATA)

¹Multi-user variant of [14] which generalizes the many distributed computing problems such as gradient coding, matrix-matrix multiplication and etc.

²For more information on this, please see [29], [30].

considering the columns of \mathbf{F} as syndromes, the columns of \mathbf{E} were identified as the coset leaders with minimum weight, thus guaranteeing the sparsest \mathbf{E} and thus the least computational cost. For example, when \mathbf{D} is derived from the basic class of covering codes over a q -ary field, then the corresponding normalized computational cost $\gamma \in [0, 1]$ — describing the fraction of all subfunctions each server had to compute — took the form $\gamma = H_q^{-1}(K/N)$ where $H_q^{-1}(\cdot)$ is the functional inverse of the entropy function.

We are here though interested in computing functions directly over the reals, which will indeed constitute a substantial deviation from the finite field case. This emphasis on the real (or complex) domain is necessitated by the fact that computing a real-valued problem over a finite field (after discretization) may not be as practical as computing it directly over the reals, mainly because discretization may entail large precision costs and accuracy losses, as well as because finite field computations are notoriously slower than floating point operations. For that, we will consider real-valued functions over L real-valued datasets (or equivalently, with L component/basis subfunctions), and N computing servers and K users each demanding their own function. As we are now working in the field of real numbers, the coding theoretic approach in [27] does not directly apply, and thus a new approach is required.

Here, our approach is based on establishing, for the first time, a connection between distributed computing and compressed sensing. As a first step, we show (Proposition 1) that there exists an achievable scheme whose normalized computational cost is bounded above as $\gamma \leq \frac{K}{N}$. This is a probabilistic scheme, where \mathbf{D} is chosen from the Gaussian ensemble, and where the corresponding sparsity of \mathbf{E} is the outcome of a randomized process. Then we propose ℓ_0 -minimization, which takes as input \mathbf{D} and \mathbf{F} to yield a sparse \mathbf{E} . This minimization though is generally intractable, and for this reason, we draw from the rich literature of compressed sensing to suggest a more practical approach where we show (Theorem 1) that as long as there exists a scheme whose computational cost is bounded by $\gamma \leq -r \frac{K}{N} W_{-1}^{-1}(-\frac{2K}{eNr})$ (where $W_{-1}(\cdot)$ is the Lambert function and r is a parameter that calibrates the communication between servers and users) we can in fact employ a tractable basis pursuit ℓ_1 -minimization to derive such scheme.

Notations: For matrices \mathbf{A} and \mathbf{B} , then $[\mathbf{A}, \mathbf{B}]$ indicates the horizontal concatenation of the two matrices. We define $[n] \triangleq \{1, 2, \dots, n\}$. For any matrix $\mathbf{X} \in \mathbb{F}^{m \times n}$, then $\mathbf{X}(i, j)$, $i \in [m]$, $j \in [n]$, represents the entry in the i -th row and j -th column, while $\mathbf{X}(i, :)$, $i \in [m]$, represents the i -th row, and $\mathbf{X}(:, j)$, $j \in [n]$ represents the j -th column of \mathbf{X} . For two index sets $\mathcal{I} \subset [m]$, $\mathcal{J} \subset [n]$, then $\mathbf{X}(\mathcal{I}, \mathcal{J})$ represents the sub-matrix comprised of the rows in \mathcal{I} and columns in \mathcal{J} . We will use $\|\mathbf{X}\|_0$ to represent the number of nonzero elements of some matrix (or vector) \mathbf{x} . Also, \otimes is the Kronecker product and $\text{vec}(\mathbf{X})$ is the vectorization of \mathbf{X} . \mathbf{A}^* represents a conjugate of \mathbf{A} . $\mathbf{A}_{\mathcal{S}}$ represents a matrix that consists of only the columns of \mathbf{A} indexed in $\mathcal{S} \subset [N]$. $\mathbf{I}_{m \times m}$ represents an $m \times m$ identity matrix.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider the multi-user linearly-separable distributed computation setting (cf. Fig. 1), which consists of K users/clients, N active servers, and a master node that coordinates servers and users. The tasks performed on each server may entail substantial computational complexity as well as time constraints. We consider a setting where each server n can communicate in a single shot (a single time-slot) to some arbitrary user-set $\mathcal{T}_n \subset [K]$, via a dedicated broadcast channel.

In our setting, each user asks for a (generally non-linear) function from a space of linearly-separable functions, where each such function takes several datasets D_1, \dots, D_L as input, and is of the form of a linear combination of individual subfunctions $g_l(D_l) \in \mathbb{R}$, each taking a single dataset D_l as input. Thus, the function $F_k(D_1, \dots, D_L) \in \mathbb{R}$, demanded by user $k \in [K]$, is a real-valued function of the form

$$\begin{aligned} F_k(D_1, D_2, \dots, D_L) &\triangleq f_{k,1}g_1(D_1) \dots + f_{k,L}g_L(D_L) \quad (1) \\ &= f_{k,1}W_1 + \dots + f_{k,L}W_L \quad (2) \end{aligned}$$

where $W_l = g_l(D_l) \in \mathbb{R}$, $l \in [L]$ is a so-called ‘file’ output, and $f_{k,l} \in \mathbb{R}$, $k \in [K]$, $l \in [L]$ are the linear combination coefficients that define each desired function.

A. Phases of the Process

The model involves three phases, with the first being the *demand phase*, the second being the *assignment and computation phase*, and the final one the *transmission and decoding phase*. In the demand phase, each user $k \in [K]$ requests $F_k(\cdot)$ from the master node, who then deduces the decomposition as in (2). Then, based on these K desired functions, during the assignment and computation phase, the master assigns some of the subfunctions to each server n , which then proceeds to compute these and produce the corresponding files $W_l = f_l(D_l)$ for all the subfunctions $f_l(D_l)$, $l \in \mathcal{W}_n$ it is responsible for.

During the transmission phase, each server $n \in [N]$ broadcasts

$$z_n \triangleq \sum_{l \in [L]} e_{n,l} W_l, \quad n \in [N] \quad (3)$$

in a single shot its own linear combination of the locally computed output files, and does so to its own particular subset of users \mathcal{T}_n . The above is defined by the so-called *encoding coefficients* $e_{n,l} \in \mathbb{R}$ which are determined by the master.

Finally, during the decoding phase, each user k linearly combines the received signals as follows

$$F'_k \triangleq \sum_{n \in [N]} d_{k,n} z_n \quad (4)$$

for some decoding coefficients $d_{k,n} \in \mathbb{R}$, $n \in [N]$, determined again by the master node. Naturally $d_{k,n} = 0, \forall k \notin \mathcal{T}_n$. In the end, we say the exact decoding is successful when $F'_k = F_k$ for all $k \in [K]$.

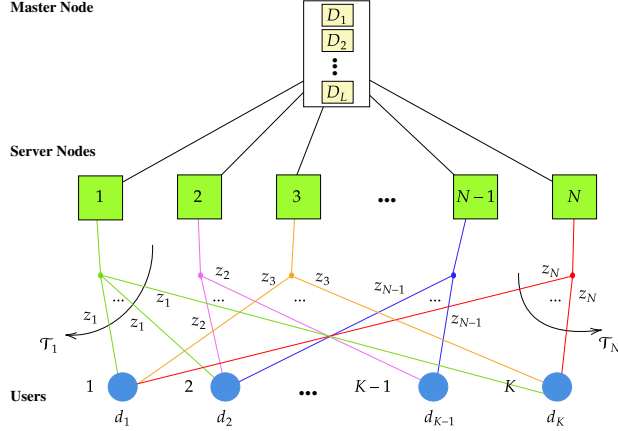


Fig. 1. The K -user, N -server, L -Dataset linearly-separable computation setting. Once each user informs the master of its desired function $F_k(\cdot)$, each server $n \in [N]$ computes a subfunction $W_l = f_l(D_l) \in \mathbb{R}$ in $\mathcal{W}_n \subseteq [L]$. Afterwards, server n broadcasts a linear combination z_n (of the locally available computed files) to all users in \mathcal{T}_n . This combination is defined by the coefficients $e_{n,l}$. Finally, for decoding, each user $k \in [K]$ linearly combines (based on decoding vectors \mathbf{d}_k) all the received signals from all of the servers it has received from. Decoding should produce for each user its desired function $F_k(D_1, \dots, D_L)$.

B. Problem Formulation

Similarly to the finite field case, also here, to formulate the problem we use $\mathbf{f} \triangleq [F_1, F_2, \dots, F_K]^\top$, $\mathbf{f}_k \triangleq [f_{k,1}, f_{k,2}, \dots, f_{k,L}]^\top$, $k \in [K]$, $\mathbf{w} \triangleq [W_1, W_2, \dots, W_L]^\top$ where \mathbf{f} represents the vector of the demanded functions outputs (cf. (2)), \mathbf{f}_k the vector of function coefficients for user k (cf. (2)), and \mathbf{w} the vector of output files combined over all subfunctions. We also have $\mathbf{e}_n \triangleq [e_{n,1}, e_{n,2}, \dots, e_{n,L}]^\top$, $n \in [N]$ $\mathbf{z} \triangleq [z_1, z_2, \dots, z_N]^\top$ respectively representing the encoding vector at server n , and the overall transmitted vector across all the servers (cf. (3)). Furthermore, we have $\mathbf{d}_k \triangleq [d_{k,1}, d_{k,2}, \dots, d_{k,N}]^\top$, $k \in [K]$ $\mathbf{f}'_k \triangleq [F'_1, F'_2, \dots, F'_K]^\top$

respectively representing the decoding vector at user k , and the vector of the decoded functions across all the users. In addition, we have $\mathbf{F} \triangleq [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_K]^\top \in \mathbb{R}^{K \times L}$, $\mathbf{E} \triangleq [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N]^\top \in \mathbb{R}^{N \times L}$, $\mathbf{D} \triangleq [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K]^\top \in \mathbb{R}^{K \times N}$ where \mathbf{F} represents the $K \times L$ so-called *jobs matrix* of all function coefficients across all the users, where \mathbf{E} represents the $N \times L$ *computing and encoding matrix* across all servers, and where \mathbf{D} represents the $K \times N$ *decoding matrix* across all the users.

Directly from (2), we have that

$$\mathbf{f} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_K]^\top \mathbf{w} \quad (5)$$

and from (3) we have the overall transmitted vector taking the form

$$\mathbf{z} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N]^\top \mathbf{w} = \mathbf{E}\mathbf{w}. \quad (6)$$

Furthermore, directly from (4) we have that

$$F'_k = \mathbf{d}_k^\top \mathbf{z} \quad (7)$$

and thus we have

$$\mathbf{f}' = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K]^\top \mathbf{z} = \mathbf{D}\mathbf{z}. \quad (8)$$

Recall that we must guarantee

$$\mathbf{f}' = \mathbf{f}. \quad (9)$$

After substituting (5), (6) and (8) into (9), we see that the above feasibility condition in (9) is satisfied if and only if

$$\mathbf{D}\mathbf{E}\mathbf{w} = \mathbf{F}\mathbf{w}. \quad (10)$$

For this to hold for any \mathbf{w} , we must thus have

$$\mathbf{D}\mathbf{E} = \mathbf{F}. \quad (11)$$

Note that throughout the whole paper, we assume that $K \leq N$.

C. Computational Cost

Recalling quickly that each server n computes the subfunctions whose index are in \mathcal{W}_n , and since $\mathcal{W}_n = \text{supp}(\mathbf{E}(n, :))$, then the *normalized computation cost* in our case naturally takes the form

$$\gamma \triangleq \frac{\sum_{n=1}^N |\mathcal{W}_n|}{NL} = \frac{\|\mathbf{E}\|_0}{NL}. \quad (12)$$

As one can see, γ simply describes the average fraction of subfunctions that must be computed by each server, which is also the fraction of non-zero elements in \mathbf{E} . It is now clear that decomposing \mathbf{F} into the product of two matrices \mathbf{D} and sparse \mathbf{E} , implies reduced computation cost which results in reduced delay. In particular, the fewer number of nonzero elements in a row of \mathbf{E} means less delay in finishing up a task for a server.

III. RESULTS

In this section, we first give a basic probabilistic scheme for subtask assignment, based on employing a Gaussian³ random matrix \mathbf{D} , where the scheme employs a simple zero-forcing approach that solves a determined linear system. Albeit basic, this will allow us to upper bound the optimal normalized computation cost — which a generally intractable ℓ_0 -minimization would give — as $\gamma \leq \frac{K}{N}$.

Proposition 1. *For the multi-user linearly-separable distributed computing problem, with K users, N servers and L datasets, employing a random Gaussian \mathbf{D} , guarantees that with probability 1, there exists a scheme with bounded normalized computation cost $\gamma \leq K/N$, which serves as an upper bound the ℓ_0 -minimal cost.*

Proof. From (11), we have that $\mathbf{F}(:, l) = \mathbf{D}\mathbf{E}(:, l)$, $\mathbf{F}(:, l) \in \mathbb{R}^{K \times 1}$, $\forall l \in [L]$ where for each l , in the context of the ℓ_0 -minimization in (14), we have $\mathbf{y} = \mathbf{F}(:, l)$, $\mathbf{D} = \mathbf{A}$ and $\mathbf{E}(:, l) = \mathbf{z}$, where again we have an underdetermined system of equations.

Consider choosing L arbitrary random subsets $\mathcal{S}_l \subset [N]$, $l \in [L]$ where $|\mathcal{S}_l| = K$. Now for each $l \in [L]$, we focus on the l -th column $\mathbf{E}([N], l)$ of \mathbf{E} and set

³This implies that each entry of \mathbf{D} is independently and identically picked from a Gaussian distribution.

the elements $\mathbf{E}([N] \setminus \mathcal{S}_l, l) = 0$, i.e., from column l , only the elements indexed by \mathcal{S}_l remain non-zero. Now since $\mathbf{F}(:, l) = \mathbf{D}([K], [N] \setminus \mathcal{S}_l) \mathbf{E}([N] \setminus \mathcal{S}_l, l) + \mathbf{D}([K], \mathcal{S}_l) \mathbf{E}(\mathcal{S}_l, l)$, we get $\mathbf{F}(:, l) = \mathbf{D}([K], \mathcal{S}_l) \mathbf{E}(\mathcal{S}_l, l)$, which is a determined system of equations that allows us to determine $\mathbf{E}(\mathcal{S}_l, l)$. In the above, $\mathbf{D}([K], \mathcal{S}_l)$ is the corresponding $K \times K$ submatrix of \mathbf{D} . Given the above, each such $\mathbf{D}([K], \mathcal{S}_l)$ is a $K \times K$ Gaussian sub-matrix, which is naturally nonsingular with probability one. Thus, the determined system of equations always has a unique solution for $\mathbf{E}(\mathcal{S}_l, l), \forall l \in [L]$, therefore the scheme works for all $F(:, l), l \in [L]$. This scheme guarantees that $\|\mathbf{E}(:, l)\|_0 \leq K$, then $\|\mathbf{E}\|_0 \leq KL$, and thus guarantees that $\gamma \leq \frac{K}{N}$. Better performance can be achieved by employing ℓ_0 -minimization as in (14). \square

As we will discuss in Section IV, ℓ_0 -minimization is known to be NP-hard [31], hence intractable. To offer a practical solution, the following theorem utilizes results from compressed sensing to describe a range of practical solutions, which now use ℓ_1 -minimization in order to find — as we will clarify later on — sparse (and unique) encoding matrices \mathbf{E} .

Theorem 1. *For the multi-user linearly-separable distributed computing problem, with K users, N servers and L datasets, if a scheme exists with a (κ, β) sub-Gaussian random matrix \mathbf{D} (cf. Lemma 2) for which ℓ_0 -minimisation would yield*

$$\gamma \leq -\frac{1}{r} \frac{K}{N} W_{-1}^{-1} \left(-\frac{2K}{erN} \right), \quad 0 \leq K/N \leq 12(2\beta + \kappa)/\kappa^2,$$

then the corresponding (and unique) \mathbf{E} can be found via basis pursuit ℓ_1 -minimization with probability at least $1 - 2e^{-\frac{KL}{r}}$, where $r = 12(4\beta + 2\kappa)/\kappa^2$.

Proof. The proof is provided in the following section, which starts with a brief primer on compressed sensing. \square

IV. PROOF OF THEOREM 1

Before proceeding with the proof, we quickly describe some basic properties of compressed sensing.

A. Brief Primer on Compressed sensing

We provide here a brief introduction of the compressed sensing results [31]–[41], which will be employed in our distributed computing problem. We will utilize notation common to the compressed sensing literature, and the link to the computing parameters will be clarified in the next subsection.

As described in [35], compressed sensing seeks to recover a sparse vector $\mathbf{x} \in \mathbb{R}^p$ from a few underdetermined linear measurements of the form:

$$\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{R}^m \quad (13)$$

where $\mathbf{A} \in \mathbb{R}^{m \times p}$, $m, p \in \mathbb{N}$ is the so-called measurement matrix, and $\mathbf{y} = [y_1, \dots, y_m]^T$ is the measurement vector. In our case, as we will see later on, \mathbf{y} will be associated to our computing and encoding matrix \mathbf{E} , then \mathbf{A} to the communication and decoding matrix \mathbf{D} , and \mathbf{x} will be associated to the jobs matrix \mathbf{F} . The general approach is to recover the

sparsest solution via a basic but computationally intractable ℓ_0 -minimization that takes the form

$$\min_{\mathbf{z} \in \mathbb{R}^p} \|\mathbf{z}\|_0 := \sum_{i=1}^p 1_{|z_i| \neq 0} \quad \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{z} \quad (14)$$

where $1_{|z_i| \neq 0}$ denotes the indicator function. This same optimization will lead to the sparsest solution for \mathbf{E} and thus will yield the smallest possible γ . To the best of our knowledge, there are no results that enables us to bound the weight of this sparsest solution, and for that we will use a basic constructive approach to bound γ .

The NP-hard nature of the optimization problem in (14) has led to the consideration of an ℓ_1 -norm minimization approach, also known as *basis pursuit*, which is considered as the closest convex tractable alternative for (14), and which is given by

$$\min_{\mathbf{z} \in \mathbb{R}^p} \|\mathbf{z}\|_1 := \sum_{i=1}^p |z_i| \quad (15)$$

$$\text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{z}. \quad (16)$$

It is well established in compressed sensing that the estimate $\hat{\mathbf{x}} \in \mathbb{R}^p$ obtained by solving (15), achieves the desired unique solution \mathbf{x} as long as some conditions are satisfied⁴. These conditions are closely related to certain properties of the measurement matrix, with one such condition being the well known restricted isometry property (RIP) [42], which dictates how well ℓ_1 -norm optimization algorithms, such as basis pursuit [43], can perform. This is captured in the following result, found in [31, Theorem 6.2], which is reproduced here.

Lemma 1. For a matrix $\mathbf{A} \in \mathbb{R}^{m \times p}$ and for

$$\delta_s(\mathbf{A}) \triangleq \max_{S \subset [p], |S| \leq s} \|\mathbf{A}_S^* \mathbf{A}_S - \mathbf{I}_{m \times m}\|_{2 \rightarrow 2}^2 \quad (17)$$

being the s th restricted isometry constant, and if $\delta_{2s}(\mathbf{A}) < \frac{1}{3}$, then every s -sparse vector $\mathbf{x} \in \mathbb{R}^p$ is the unique solution of

$$\minimize_{\mathbf{z} \in \mathbb{R}^p} \|\mathbf{z}\|_1 \quad \text{subject to } \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}. \quad (18)$$

In particular, the above Lemma shows that having a measurement vector $\mathbf{y} \in \mathbb{R}^m$, where we know a priori that it is a result of a linear system $\mathbf{A}\mathbf{x}$, $\mathbf{x} \in \mathbb{R}^p, m \leq p$, induced by a unique and s -sparse vector \mathbf{x} if $\delta_{2s}(\mathbf{A}) < \frac{1}{3}$, then via having a ℓ_1 -minimizer, we can find \mathbf{z} as a solution to $\mathbf{A}\mathbf{z}$, which it has minimum $\|\mathbf{z}\|_1$, then the above Lemma guarantees that the solution of this minimization is equal to \mathbf{x} . The above lemma shows that having $\delta_{2s}(\mathbf{A}) < 1/3$ is sufficient to guarantee the exact recovery of all unique s -sparse vectors via ℓ_1 -minimization. It basically states that if \mathbf{A} behaves relatively similar to orthonormal matrices when operating on sparse vectors, then ℓ_1 -minimization will act as an ℓ_0 -minimization.

In our problem here, it is important that we pick \mathbf{D} such that \mathbf{A} abides by the above property. The following two results tell us directly how to do that. The first, below, is directly adapted from [31, Theorem 9.2].

⁴For our computing problem. these will be conditions in the form of an upper bound on γ .

Lemma 2. Let $\mathbf{A} \in \mathbb{R}^{m \times p}$ be an i.i.d. (zero mean, unit variance) sub-Gaussian random matrix with parameters β and κ such that

$$\mathbb{P}(|A_{i,j}| \geq t) \leq \beta e^{-\kappa t^2}, \forall t > 0. \quad (19)$$

Then, $\delta_{2s}(\frac{\mathbf{A}}{\sqrt{m}}) \leq \delta$ is satisfied with probability at least $1 - 2e^{-\frac{\delta^2 m}{2c}}$ for any δ such that

$$m \geq 2c\delta^{-2}s \ln\left(\frac{ep}{2s}\right) \quad (20)$$

where

$$c = \frac{2(4\beta + 2\kappa)}{3k^2}. \quad (21)$$

Combining Lemma 2 and Lemma 1 after setting $\delta = 1/3$, implies that the uniform recovery of all s -sparse vectors is possible with high probability via the ℓ_1 -minimization in (15) as long as the number of measurements satisfies $m \geq (12(4\beta + 2\kappa)/\kappa^2)s \ln(\frac{ep}{2s})$.

B. Proof of Theorem 1

Directly from (11), we have

$$\text{vec}(\mathbf{F}) = (\mathbf{D} \otimes \mathbf{I}_{L \times L}) \times \text{vec}(\mathbf{E}), \quad (22)$$

which matches the compressed sensing setting $\mathbf{y} = \mathbf{A}\mathbf{x}$ in (13) when considering $\mathbf{y} = \text{vec}(\mathbf{F})$, $\mathbf{A} = \mathbf{D} \otimes \mathbf{I}_{L \times L}$, and $\mathbf{x} = \text{vec}(\mathbf{E})$, where now $m = KL$ and $p = NL$.

Furthermore, let us also note that directly from [44], we have

$$\delta_s(\mathbf{D} \otimes \mathbf{I}_{L \times L}) \leq \delta_s(\mathbf{D}). \quad (23)$$

With the elements of \mathbf{D} being chosen independently from a zero-mean, unit-variance sub-Gaussian distribution with parameters β , κ (cf. (19)), we can now employ Lemma 1 and (23), together with Lemma 2 after setting $\delta_{2s} < \delta = 1/3$, to conclude that the exact recovery threshold for a unique \mathbf{E} matrix via ℓ_1 -minimization driven by basis pursuit, takes the form

$$KL \geq r \|\mathbf{E}\|_0 \ln\left(\frac{eNL}{2r \|\mathbf{E}\|_0}\right) \quad (24)$$

where $r = 12(4\beta + 2\kappa)/\kappa^2$. Then after normalizing both sides by NL and applying (12), we get

$$\frac{K}{N} \geq r\gamma \ln\left(\frac{e}{2r\gamma}\right). \quad (25)$$

Let us now define $f(x) \triangleq -r^{-1}xW_{-1}^{-1}\left(-\frac{2x}{e}\right)$ and evaluate it on the both sides of (25), at $x_1 = K/N$ and $x_2 = r\gamma \ln(\frac{e}{2r\gamma})$ since $f(x)$ is a monotonically increasing function on $0 \leq x \leq r/2$ and $r/2 \geq x_1 \geq x_2 \geq 0$, we have $f(x_1) \geq f(x_2)$, in the view of the fact that its inverse⁵ function is $f^{-1}(x) = rx \ln(\frac{e}{2rx})$, we can retrieve the claim.

⁵To see this, for $g(x) \triangleq rx \ln(\frac{e}{2rx})$ and $f(x) \triangleq -r^{-1}xW_{-1}^{-1}(-2x/e)$, we see that $g^{-1}(x) = f(x)$ simply because $f(g(x)) = -r^{-1}g(x)W_{-1}^{-1}\left(-\frac{2g(x)}{e}\right) = -x \ln(\frac{e}{2rx})W_{-1}^{-1} = -x \ln(e/2rx)W_{-1}^{-1}\left(\frac{2rx}{e} \ln(2rx/e)\right) = -x \ln(e/2rx) / \ln(2rx/e) = x$.

V. DISCUSSION AND CONCLUSION

In the context of our distributed computing problem, it is interesting to observe some of the similarities that exist between the real case (which employed compressed sensing techniques) and the finite-field case in [27], which employed the structure of covering codes whose covering radius was a measure of the sparsity of the solution for \mathbf{E} . For example, in the extreme case of $L = q^K$, the work in [27] revealed⁶ the optimal normalized computational cost to be of the form of $\gamma \simeq H_q^{-1}(K/N)$, which almost matches $\gamma \simeq K/N$ in the limit of large q , derived in this paper.

It is also worth noting that our result in Theorem 1 automatically accepts an additional uniqueness property — on the sparsest solution \mathbf{x} in (13) — which is in fact not needed in our distributed computing problem. It would be interesting to explore further improvements in the computational costs, upon the removal of this uniqueness condition. Finally, one can imagine that further improvements in the distributed computing problem could also benefit from the deep connections revealed in [35] between compressed sensing and error correction.

REFERENCES

- [1] T. Jahani-Nezhad and M. A. Maddah-Ali, “Codedsketch: A coding scheme for distributed computation of approximated matrix multiplication,” *IEEE Transactions on Information Theory*, vol. 67, no. 6, pp. 4185–4196, 2021.
- [2] 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)*, pp. 1–6, IEEE, 2021.
- [3] S. Li, Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, “A scalable framework for wireless distributed computing,” *IEEE/ACM Transactions on Networking*, vol. 25, no. 5, pp. 2643–2654, 2017.
- [4] F. Haddadpour, M. M. Kamani, M. Mahdavi, and V. Cadambe, “Trading redundancy for communication: Speeding up distributed sgd for non-convex optimization,” in *International Conference on Machine Learning*, pp. 2545–2554, PMLR, 2019.
- [5] M. Soleymani, H. Mahdavi, and A. S. Avestimehr, “Analog Lagrange coded computing,” *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 283–295, 2021.
- [6] H. Sun and S. A. Jafar, “The capacity of private computation,” *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3880–3897, 2018.
- [7] M. Soleymani and H. Mahdavi, “Distributed multi-user secret sharing,” *IEEE Transactions on Information Theory*, vol. 67, no. 1, pp. 164–178, 2020.
- [8] A. Khalesi, M. Mirmohseni, and M. A. Maddah-Ali, “The capacity region of distributed multi-user secret sharing,” *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 3, pp. 1057–1071, 2021.
- [9] M. Soleymani, H. Mahdavi, and A. S. Avestimehr, “Privacy-preserving distributed learning in the analog domain,” *arXiv preprint arXiv:2007.08803*, 2020.
- [10] M. Soleymani, R. E. Ali, H. Mahdavi, and A. S. Avestimehr, “List-decodable coded computing: Breaking the adversarial toleration barrier,” *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 3, pp. 867–878, 2021.
- [11] N. Raviv, I. Tamo, R. Tandon, and A. G. Dimakis, “Gradient coding from cyclic mds codes and expander graphs,” *IEEE Transactions on Information Theory*, vol. 66, no. 12, pp. 7475–7489, 2020.
- [12] K. Lee, M. Lam, R. Pedarsani, D. Papailiopoulos, and K. Ramchandran, “Speeding up distributed machine learning using codes,” *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1514–1529, 2017.

⁶Over a q -ary alphabet, $H_q(x)$ denotes the entropy function, which takes the form $H_q(x) \triangleq x \log_q(q-1) - x \log_q(x) - (1-x) \log_q(1-x)$ for all $0 < x < 1 - 1/q$.

- [13] M. Egger, R. Bitar, A. Wachter-Zeh, and D. Gündüz, "Efficient distributed machine learning via combinatorial multi-armed bandits," *arXiv preprint arXiv:2202.08302*, 2022.
- [14] K. Wan, H. Sun, M. Ji, and G. Caire, "Distributed linearly separable computation," *IEEE Transactions on Information Theory*, vol. 68, no. 2, pp. 1259–1278, 2022.
- [15] A. Behrouzi-Far and E. Soljanin, "Efficient replication for straggler mitigation in distributed computing," *arXiv preprint arXiv:2006.02318*, 2020.
- [16] K. Wan, H. Sun, M. Ji, and G. Caire, "On the tradeoff between computation and communication costs for distributed linearly separable computation," *IEEE Transactions on Communications*, vol. 69, no. 11, pp. 7390–7405, 2021.
- [17] E. Parrinello, E. Lampiris, and P. Elia, "Coded distributed computing with node cooperation substantially increases speedup factors," in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 1291–1295, 2018.
- [18] S. Li, M. A. Maddah-Ali, Q. Yu, and A. S. Avestimehr, "A fundamental tradeoff between computation and communication in distributed computing," *IEEE Transactions on Information Theory*, vol. 64, no. 1, pp. 109–128, 2018.
- [19] K. Tjell and R. Wisniewski, "Privacy in distributed computations based on real number secret sharing," *arXiv preprint arXiv:2107.00911*, 2021.
- [20] 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.
- [21] M. Soleymani, H. MahdaviFar, and A. S. Avestimehr, "Analog secret sharing with applications to private distributed learning," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 1893–1904, 2022.
- [22] O. Makkonen and C. Hollanti, "Analog secure distributed matrix multiplication over complex numbers," in *2022 IEEE International Symposium on Information Theory (ISIT)*, pp. 1211–1216, 2022.
- [23] J. S. Ng, W. Y. B. Lim, N. C. Luong, Z. Xiong, A. Asheralieva, D. Niyato, C. Leung, and C. Miao, "A survey of coded distributed computing," *arXiv preprint arXiv:2008.09048*, 2020.
- [24] S. Li, S. Avestimehr, *et al.*, "Coded computing: Mitigating fundamental bottlenecks in large-scale distributed computing and machine learning," *Foundations and Trends® in Communications and Information Theory*, vol. 17, no. 1, pp. 1–148, 2020.
- [25] J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.
- [26] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: Cluster computing with working sets," in *2nd USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 10)*, 2010.
- [27] A. Khalesi and P. Elia, "Multi-user linearly-separable distributed computing," *arXiv preprint arXiv:2206.11119*, 2022.
- [28] A. Khalesi and P. Elia, "Multi-user linearly separable computation: A coding theoretic approach," in *2022 IEEE Information Theory Workshop (ITW)*, pp. 428–433, 2022.
- [29] A. Mallick, S. Smith, and G. Joshi, "Rateless codes for distributed non-linear computations," in *2021 11th International Symposium on Topics in Coding (ISTC)*, pp. 1–5, IEEE, 2021.
- [30] A. Mallick and G. Joshi, "Rateless sum-recovery codes for distributed non-linear computations," in *2022 IEEE Information Theory Workshop (ITW)*, pp. 160–165, IEEE, 2022.
- [31] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Birkhäuser Basel, 2013.
- [32] E. J. Candes and Y. Plan, "A probabilistic and riplless theory of compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7235–7254, 2011.
- [33] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [34] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes rendus mathématique*, vol. 346, no. 9–10, pp. 589–592, 2008.
- [35] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE transactions on information theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [36] T. T. Cai, L. Wang, and G. Xu, "New bounds for restricted isometry constants," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4388–4394, 2010.
- [37] T. T. Cai, L. Wang, and G. Xu, "Shifting inequality and recovery of sparse signals," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1300–1308, 2010.
- [38] S. Foucart, "A note on guaranteed sparse recovery via ℓ_1 -minimization," *Applied and Computational Harmonic Analysis*, vol. 29, no. 1, pp. 97–103, 2010.
- [39] S. Foucart and M.-J. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q \leq 1$," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395–407, 2009.
- [40] Q. Mo and S. Li, "New bounds on the restricted isometry constant δ_{2k} ," *Applied and Computational Harmonic Analysis*, vol. 31, no. 3, pp. 460–468, 2011.
- [41] S. Daei, F. Haddadi, and A. Amini, "Living near the edge: A lower-bound on the phase transition of total variation minimization," *IEEE Transactions on Information Theory*, vol. 66, no. 5, pp. 3261–3267, 2019.
- [42] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [43] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.
- [44] S. Jökar and V. Mehrmann, "Sparse solutions to underdetermined kronecker product systems," *Linear Algebra and its Applications*, vol. 431, no. 12, pp. 2437–2447, 2009.