Rank Minimization over Finite Fields

Vincent Y. F. Tan*†, Laura Balzano* and Stark C. Draper*

* Dept. of ECE, University of Wisconsin-Madison, Email: {vtan@,sunbeam@ece.,sdraper@ece.}wisc.edu
† LIDS, Massachusetts Institute of Technology, Email: vtan@mit.edu

***Abstract*—This paper establishes information-theoretic limits in estimating a finite field low-rank matrix given random linear measurements of it. Necessary and sufficient conditions on the number of measurements required are provided. It is shown that these conditions are sharp. The reliability function associated to the minimum-rank decoder is also derived. Our bounds hold even in the case where the sensing matrices are sparse. Connections to rank-metric codes are discussed.**

***Index Terms*—Rank minimization, Finite fields, Reliability function, Sparse measurement matrices, Rank-metric codes**

I. Introduction

The problem of matrix completion [1]–[3] has been well-studied in recent years. Essentially, one is given a (small) subset of entries of a low-rank matrix and one is required to estimate all the the remaining entries. This problem has a variety of applications from collaborative filtering to obtaining the minimal realization of a linear system. Algorithms based on the nuclear-norm (sum of singular values) convex relaxation of the rank function have enjoyed tremendous successes.

A generalization of the matrix completion problem is the rank minimization problem [4] where, instead of being given entries of the low-rank matrix, one is given linear measurements of it. The nuclear-norm heuristic has also been proven to be extremely effective in finding the unknown matrix. Results from the literature are typically of the following flavor: If the number of measurements exceeds a small multiple of the product of the dimension of the matrix and its rank, then the nuclear-norm heuristic is guaranteed to yield the same (optimal) solution as the exact rank-minimization problem.

The bulk (if not all) of the rank minimization literature deals with the case where the entries of the unknown matrix are from the reals $\mathbb{R}$. In this paper, we focus on the case where the elements of the matrix belong to a *finite field* $\mathbb{F}_q$. Arithmetic is performed in the field and the rank is also computed with respect to $\mathbb{F}_q$. Such a problem has applications in coding theory as well as distributed storage and interference alignment [5]. We derive information-theoretic limits on the number of measurements needed for estimating the matrix given linear measurements. In this paper, we are not as concerned with the computational complexity of recovering the unknown low-rank matrix as compared to the fundamental limits of doing so. However, we will comment on possible practical rank minimization techniques over finite fields by drawing analogies between rank minimization and the class of codes known as rank-metric codes [6]–[8].

V. Tan is funded by A*STAR, Singapore.

A. Summary of Contributions

There are four key contributions in this paper. Firstly, by using Fano's inequality and [9], we derive a necessary condition on the number of measurements required for estimating a matrix from linear measurements. We have a weak and strong converse; the latter providing the rate of convergence of the error probability to unity (over all estimators). Secondly, we demonstrate that the so-called *min-rank decoder* (similar to Csiszár's α-decoder [10]) achieves the information-theoretic lower bound on the number of measurements under the condition that the elements of the sensing (or measurement) matrices are drawn independently and uniformly from $\mathbb{F}_q$. Thirdly, we derive the reliability function (error exponent) of the min-rank decoder by using de Caen's lower bound on the probability of a union [11]. Finally, we show that even if the fraction of non-zero entries of the sensing matrices scales as $\Theta(\frac{\log n}{n})$, the min-rank decoder achieves the information-theoretic lower bound. This result opens the possibility for the development of decoding algorithms based on sparse parity-check matrices.

B. Related Work

Our work is partially inspired by [12] where fundamental limits for compressed sensing over finite fields were derived. To the best of the authors' knowledge, the work by Vishwanath in [3] is the only one that deals with matrix completion (or rank minimization) over finite alphabets in an information-theoretic setting. It was shown using typicality arguments that the number of measurements required is within a logarithmic factor of the lower bound. Our setting is different because we assume that we have linear measurements instead of randomly sampled entries. We are able to show that the achievability and converse match for a family of sensing matrices.

The family of codes known as rank-metric codes [6]–[8] bears a striking similarity to the rank minimization problem over finite fields. We comment on connections in Section VI.

II. Notation and System Model

In this paper we adopt the following set of notation: San-serif font (e.g., y), bold-face upper-case (e.g., $\mathbf{X}$), bold-face lower-case (e.g., $\mathbf{w}$) denote random quantities, matrices and vectors respectively. Sets (and events) are denoted with calligraphic font. For a prime power q, we denote the Galois (finite) field with q elements as $\mathbb{F}_q$. The set of $m \times n$ matrices with entries in $\mathbb{F}_q$ is denoted as $\mathbb{F}_q^{m\times n}$. For simplicity, we let $[k] := \{1, \ldots, k\}$ and $\mathsf{y}^k := (\mathsf{y}_1, \ldots, \mathsf{y}_k)$. For a matrix $\mathbf{M}$, $\|\mathbf{M}\|_0$ and $\mathrm{rank}(\mathbf{M})$ denote the number of non-zero elements in $\mathbf{M}$ and the rank of $\mathbf{M}$ in $\mathbb{F}_q$ respectively.

We are interested in the following model: Let $\mathbf{X}$ be an unknown square[1] matrix in $\mathbb{F}_q^{n\times n}$ whose rank is less than or equal to r, i.e., $\mathrm{rank}(\mathbf{X}) \leq r$. We assume that the *rank-dimension ratio* $\gamma := r/n$ is constant as n grows. We would like to recover $\mathbf{X}$ from *linear measurements*

$$\mathsf{y}_a = \langle \mathbf{H}_a, \mathbf{X}\rangle := \sum_{(i,j)\in[n]^2} [\mathbf{H}_a]_{i,j}[\mathbf{X}]_{i,j} \qquad a \in [k]. \tag{1}$$

In (1), the *sensing matrices* $\mathbf{H}_a \in \mathbb{F}_q^{n\times n}, a \in [k]$ are chosen according to some probability mass function (pmf). The k scalar *measurements* $\mathsf{y}_a \in \mathbb{F}_q, a \in [k]$ are available to estimate $\mathbf{X}$. We will allow k to depend on n, i.e., $k = k(n)$. Multiplication and addition in (1) are performed in $\mathbb{F}_q$.

Our model is somewhat different from the matrix completion problem [1]–[3]. In the matrix completion setup, a subset of entries $\Omega \subset [n]^2$ in the matrix $\mathbf{X}$ is observed and one would like to "fill in" the rest assuming the matrix is low-rank. This corresponds to the case where the sensing matrix $\mathbf{H}_a$ is only non-zero in a single position and assuming $\mathbf{H}_a \neq \mathbf{H}_{a'}$ for all $a \neq a'$, the number of measurements is $k = |\Omega|$. In our measurement model in (1), we do not assume that $\|\mathbf{H}_a\|_0 = 1$.

We are interested in estimating the matrix $\mathbf{X}$ given y^k. Our goal in this paper is to characterize necessary and sufficient conditions on the number of measurements k as n becomes large assuming a particular pmf governing the sensing matrices $\mathbf{H}_a$. In the sequel, we will leverage the following lemma:

Lemma 1 (Bounds on number of low-rank matrices [6])**.** *Let $\Psi_q(n,r)$ be the number of matrices in $\mathbb{F}_q^{n\times n}$ of rank less than or equal to r. Then the following bounds hold:*

$$q^{(2n-2)r-r^2} \leq \Psi_q(n,r) \leq q^{(2n+1)r-r^2+1}. \tag{2}$$

In other words, $|\log_q \Psi_q(n,r) - 2\gamma(1-\gamma/2)n^2| = o(n^2)$.

III. Necessary Conditions for Recovery

This section presents necessary conditions on the scaling of k with n for the matrix $\mathbf{X}$ to be recovered *reliably*, i.e., for the error probability in estimating $\mathbf{X}$ to tend to zero as n grows. As with all other converse statements in information theory, it is necessary to assume a statistical model on the unknown object, in this case $\mathbf{X}$. Hence, in this section, we denote the unknown low-rank matrix as $\mathbf{X}$. We also assume that $\mathbf{X}$ is drawn *uniformly at random* from the set of matrices in $\mathbb{F}_q^{n\times n}$ of rank less than or equal to $r = \gamma n$. For an *estimator* $\hat{\mathbf{X}}(\mathsf{y}^k, \mathbf{H}^k)$ whose range is the set of all $\mathbb{F}_q^{n\times n}$-matrices whose rank is less than or equal to r, we define the event $\mathcal{E}_n := \{\hat{\mathbf{X}} \neq \mathbf{X}\}$.

Proposition 2 (Weak converse)**.** *Fix $\varepsilon > 0$ and assume that $\mathbf{X}$ is drawn uniformly at random from all matrices of rank less than or equal to r. Also, assume $\mathbf{X}$ is independent of $\mathbf{H}^k$. If,*

$$k < (2-\varepsilon)\gamma\,(1-\gamma/2)\,n^2 \tag{3}$$

then for any estimator $\hat{\mathbf{X}}$ whose range is the set of $\mathbb{F}_q^{n\times n}$-matrices whose rank is less than or equal to r, $\mathbb{P}(\mathcal{E}_n) \geq \varepsilon/3$ for all n sufficiently large.

[1]Our results are not restricted to the case where $\mathbf{X}$ is square but in this paper, we focus on the case when $\mathbf{X}$ is square for ease of exposition.

Proposition 2 states that the number of measurements k must exceed $2nr - r^2 = 2\gamma(1-\gamma/2)n^2$ for recovery of $\mathbf{X}$ to be reliable. From a linear algebraic perspective, this means we need at least as many measurements as degrees of freedom. The proof involves an elementary application of Fano's inequality and is included for completeness.

Proof: Consider the following lower bounds:

$$\begin{aligned}\mathbb{P}(\hat{\mathbf{X}} \neq \mathbf{X}) &\overset{(a)}{\geq} \frac{H(\mathbf{X}|\mathsf{y}^k,\mathbf{H}^k)-1}{\log_q \Psi_q(n,r)} = \frac{H(\mathbf{X})-I(\mathbf{X};\mathsf{y}^k,\mathbf{H}^k)-1}{\log_q \Psi_q(n,r)} \\ &= \frac{H(\mathbf{X})-I(\mathbf{X};\mathsf{y}^k|\mathbf{H}^k)-1}{\log_q \Psi_q(n,r)} = \frac{H(\mathbf{X})-H(\mathsf{y}^k|\mathbf{H}^k)-1}{\log_q \Psi_q(n,r)} \\ &\overset{(b)}{\geq} \frac{H(\mathbf{X})-k-1}{\log_q \Psi_q(n,r)} \overset{(c)}{=} 1 - \frac{k}{\log_q \Psi_q(n,r)} - o(1),\end{aligned} \tag{4}$$

where (a) is by Fano's inequality, (b) because $\mathsf{y}_a \in \mathbb{F}_q$ so $H(\mathsf{y}^k|\mathbf{H}^k) \leq H(\mathsf{y}^k) \leq kH(\mathsf{y}_1) \leq 1$ and finally, (c) is due to the uniformity of $\mathbf{X}$. Hence, if k satisfies (3) for some $\varepsilon > 0$, then $k/\log_q \Psi_q(n,r) \leq 1-\varepsilon/2$ for all n sufficiently large by (2). Hence, (4) is larger than $\varepsilon/3$ for all n sufficiently large. ∎

It is tempting to wonder whether one can show that the error probability $\mathbb{P}(\mathcal{E}_n)$ tends to one for *any* estimator $\hat{\mathbf{X}}$. By using ideas from [9], we can indeed demonstrate a strong converse.

Proposition 3 (Strong converse)**.** *If all the hypotheses in Proposition 2 are met,*

$$\limsup_{n\to\infty} \frac{1}{n^2}\log_q[1-\mathbb{P}(\mathcal{E}_n)] \leq -2\gamma(1-\gamma/2) + \limsup_{n\to\infty}\frac{k}{n^2}. \tag{5}$$

It can be seen that if (3) holds, the term on the RHS of (5) is negative and hence $\mathbb{P}(\mathcal{E}_n) \to 1$ as $n \to \infty$. In addition, the rate of convergence of $\mathbb{P}(\mathcal{E}_n)$ to unity is also provided in terms of k and γ. The proof of Proposition 3 is omitted[2] as it is basically follows along the lines of Theorem 1 in [9].

IV. Uniformly Random Sensing Matrices

In this section, we provide sufficient conditions for the recovery of $\mathbf{X}$ (a deterministic low-rank matrix) given y^k. To do so consider the following optimization problem:

$$\begin{aligned}&\text{minimize} \quad \mathrm{rank}(\mathbf{X}) \\ &\text{subject to} \quad \langle \mathbf{H}_a, \mathbf{X}\rangle = \mathsf{y}_a, \quad a \in [k]\end{aligned} \tag{6}$$

That is, among all matrices that satisfy the linear constraints in (1), we find the one whose rank is the minimum. We call (6) the *min-rank decoder*. We denote the set of minimizers to (6) as $\mathcal{S} \subset \mathbb{F}_q^{n\times n}$. If $\mathcal{S}$ is a singleton set, we denote the unique optimizer to (6) as $\mathbf{X}^*$. The optimization problem in (6) is, in general, intractable unless there is additional structure on the sensing matrices $\mathbf{H}_a$ (See Section VI). The form of the minimization problem is reminiscent of Csiszár's so-called α-decoder for linear codes [10].

In this section, we will also provide the functional form of the reliability function (error exponent) for this recovery problem. We will then generalize the model in (1) to the case where the measurements y^k may be noisy.

[2]See http://homepages.cae.wisc.edu/~vtan/isit11 for all the proofs.

A. The Noiseless Case

We now assume that each element in each sensing matrix is drawn independently and uniformly at random from $\mathbb{F}_q$, i.e., $\mathbb{P}([\mathbf{H}_a]_{i,j} = h) = 1/q$ for all $h \in \mathbb{F}_q$. We call this the *uniform measurement model*. We also define the error event to be

$$\mathcal{E}_n := \{|\mathcal{S}| > 1\} \cup (\{|\mathcal{S}| = 1\} \cap \{\mathbf{X}^* \neq \mathbf{X}\}). \tag{7}$$

Note that as we consider $\{|\mathcal{S}| > 1\}$ to be an error, we demand the solution to (6) to be unique. We can now exploit ideas from [12] to demonstrate the following result:

Proposition 4 (Achievability)**.** *Fix $\varepsilon > 0$. Under the uniform measurement model, if*

$$k > (2+\varepsilon)\gamma\,(1-\gamma/2)\,n^2 \tag{8}$$

then $\mathbb{P}(\mathcal{E}_n) \to 0$ as $n \to \infty$.

Note that the number of measurements stipulated by Proposition 4 matches the information-theoretic lower bound in (3). In this sense, the min-rank decoder prescribed by the optimization problem in (6) is optimal. We remark that the packing-like achievability proof [12] is much simpler than in the one presented in [3] (albeit in a slightly different setting).

Proof: To each $\mathbf{Z} \in \mathbb{F}_q^{n\times n}$ that is not equal to $\mathbf{X}$ and whose rank is less than or equal to $\operatorname{rank}(\mathbf{X})$, define the event

$$\mathcal{A}_{\mathbf{Z}} := \{\langle \mathbf{Z}, \mathbf{H}_a\rangle = \langle \mathbf{X}, \mathbf{H}_a\rangle, \forall\, a \in [k]\}. \tag{9}$$

Then we note that $\mathbb{P}(\mathcal{E}_n) = \mathbb{P}(\cup_{\mathbf{Z}\neq\mathbf{X}:\operatorname{rank}(\mathbf{Z})\leq\operatorname{rank}(\mathbf{X})}\mathcal{A}_{\mathbf{Z}})$ since an error occurs if and only if there exists a $\mathbf{Z} \neq \mathbf{X}$ such that (i) $\mathbf{Z}$ satisfies the linear constraints (ii) its rank is less than the rank of $\mathbf{X}$. Furthermore, we claim that $\mathbb{P}(\mathcal{A}_{\mathbf{Z}}) = q^{-k}$ for every $\mathbf{Z} \neq \mathbf{X}$. This follows because

$$\begin{aligned}\mathbb{P}(\mathcal{A}_{\mathbf{Z}}) &= \mathbb{P}(\langle \mathbf{Z}-\mathbf{X}, \mathbf{H}_a\rangle = 0, a \in [k]) \\ &\overset{(a)}{=} \mathbb{P}(\langle \mathbf{Z}-\mathbf{X}, \mathbf{H}_1\rangle = 0)^k \overset{(b)}{=} q^{-k}\end{aligned} \tag{10}$$

where (a) follows from the fact that the $\mathbf{H}_a$ are i.i.d. matrices and (b) from the fact $\mathbf{Z}-\mathbf{X} \neq 0$ and every non-zero element in a finite field has a (unique) multiplicative inverse so $\mathbb{P}(\langle \mathbf{Z}-\mathbf{X}, \mathbf{H}_1\rangle = 0) = q^{-1}$ for every $\mathbf{Z} \neq \mathbf{X}$. Now by combining (10) with the use of the union of events bound, we have

$$\begin{aligned}\mathbb{P}(\mathcal{E}_n) &\leq \sum_{\mathbf{Z}\neq\mathbf{X}:\operatorname{rank}(\mathbf{Z})\leq\operatorname{rank}(\mathbf{X})} q^{-k} \overset{(a)}{\leq} q^{2(n+1)r-r^2+1-k} \\ &= q^{n^2[2\gamma(1-\gamma/2)+o(1)-k/n^2]},\end{aligned} \tag{11}$$

where (a) follows from the upper bound in (2). Thus, we see that if k satisfies (8), the exponent in (11) is less than $-\varepsilon\gamma(1-\gamma/2) + o(1)$ and hence $\mathbb{P}(\mathcal{E}_n) \to 0$. ∎

B. The Reliability Function

We have shown in the previous section that the min-rank decoder is optimal in the sense that the number of measurements required for it to decode $\mathbf{X}$ reliably matches the lower bound on k. It is also interesting to analyze the *rate* at which $\mathbb{P}(\mathcal{E}_n)$ decays to zero for the min-rank decoder.

To do so, we define[3] $R := 1 - \lim_{n\to\infty} k/n^2$, assuming the limit exists. Also define the *reliability function* or *error exponent* $E : [0,1] \to [0,1]$ as

$$E(R) := \lim_{n\to\infty} -\frac{1}{n^2}\log_q \mathbb{P}(\mathcal{E}_n). \tag{12}$$

Unlike the usual definition of the reliability function, the normalization in (12) is $1/n^2$ since $\mathbf{X}$ is $n \times n$.[4] The following proposition provides an upper bound for the reliability function assuming the min-rank decoder is used.

Proposition 5 (Upper bound on $E(R)$)**.** *Let $\tilde{\gamma} := \operatorname{rank}(\mathbf{X})/n$ and assume that $\tilde{\gamma}$ is constant. Under the uniform measurement model and assuming the min-rank decoder is used,*

$$E(R) \leq \max\{(1-R) - 2\tilde{\gamma}\,(1-\tilde{\gamma}/2)\,, 0\}. \tag{13}$$

The proof of this result hinges on the *pairwise independence* of the events $\mathcal{A}_{\mathbf{Z}}$ and de Caen's inequality [11].

Proof: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The inequality by de Caen states that for events $\mathcal{B}_1, \ldots, \mathcal{B}_M \in \mathcal{F}$, the probability of the union can be lower bounded as

$$\mathbb{P}\left(\bigcup_{m=1}^{M}\mathcal{B}_m\right) \geq \sum_{m=1}^{M}\frac{\mathbb{P}(\mathcal{B}_m)^2}{\sum_{m'=1}^{M}\mathbb{P}(\mathcal{B}_m \cap \mathcal{B}_{m'})}. \tag{14}$$

In order to apply (14) to our setup, we need to compute the probabilities $\mathbb{P}(\mathcal{A}_{\mathbf{Z}})$ and $\mathbb{P}(\mathcal{A}_{\mathbf{Z}} \cap \mathcal{A}_{\mathbf{Z}'})$. The former is q^{-k} as argued in (10). Note that since the the uniform measurement model is assumed, $\mathbb{P}(\mathcal{A}_{\mathbf{Z}}\cap\mathcal{A}_{\mathbf{Z}'}) = q^{-2k}$ if $\mathbf{Z} \neq \mathbf{Z}'$ and $\mathbb{P}(\mathcal{A}_{\mathbf{Z}}\cap\mathcal{A}_{\mathbf{Z}'}) = \mathbb{P}(\mathcal{A}_{\mathbf{Z}}) = q^{-k}$ if $\mathbf{Z} = \mathbf{Z}'$. Now, we apply (14) to $\mathbb{P}(\mathcal{E}_n)$ noting that $\mathcal{E}_n$ is the union of all $\mathcal{A}_{\mathbf{Z}}$ such that $\mathbf{Z} \neq \mathbf{X}$ and $\operatorname{rank}(\mathbf{Z}) \leq \operatorname{rank}(\mathbf{X}) =: \tilde{r}$. Then,

$$\begin{aligned}\mathbb{P}(\mathcal{E}_n) &\geq \sum_{\substack{\mathbf{Z}\neq\mathbf{X}\\ \operatorname{rank}(\mathbf{Z})\leq\operatorname{rank}(\mathbf{X})}} \frac{q^{-2k}}{q^{-k}\left(1 + \sum_{\substack{\mathbf{Z}'\neq\mathbf{X},\mathbf{Z}\\ \operatorname{rank}(\mathbf{Z}')\leq\operatorname{rank}(\mathbf{X})}} q^{-k}\right)} \\ &\overset{(a)}{\geq} \frac{q^{2n\tilde{r}-\tilde{r}^2-o(n^2)-k}}{1+q^{2n\tilde{r}-\tilde{r}^2+o(n^2)-k}} = \frac{q^{n^2[2\tilde{\gamma}(1-\tilde{\gamma}/2)-o(1)-k/n^2]}}{1+q^{n^2[2\tilde{\gamma}(1-\tilde{\gamma}/2)+o(1)-k/n^2]}},\end{aligned}$$

where (a) is from the bounds in (2). Assuming $1 - R \geq 2\tilde{\gamma}\,(1-\tilde{\gamma}/2)$, the normalized logarithm of the error probability can now be simplified as

$$\limsup_{n\to\infty} -\frac{1}{n^2}\log_q \mathbb{P}(\mathcal{E}_n) \leq -2\tilde{\gamma}\,(1-\tilde{\gamma}/2) + \lim_{n\to\infty}\frac{k}{n^2}, \tag{15}$$

where we used the fact that $q^{n^2[2\tilde{\gamma}(1-\tilde{\gamma}/2)+o(1)-k/n^2]} \to 0$. This shows that $E(R)$ is upper bounded by the RHS of (15). The case where $1 - R < 2\tilde{\gamma}\,(1-\tilde{\gamma}/2)$ results in $E(R) = 0$ because $\mathbb{P}(\mathcal{E}_n)$ fails to converge to zero as $n \to \infty$. ∎

Corollary 6. *Under the assumptions of Proposition 5,*

$$E(R) = \max\{(1-R) - 2\tilde{\gamma}\,(1-\tilde{\gamma}/2)\,, 0\} \tag{16}$$

Proof: The lower bound on $E(R)$ follows directly from achievability in (11). The upper bound is given in (15). ∎

[3] The quantity R can be interpreted as a lower bound of the rate of the code $\mathfrak{C} := \{\mathbf{C} \in \mathbb{F}_q^{n\times n} : \langle \mathbf{C}, \mathbf{H}_a\rangle = 0, a \in [k]\}$.

[4] The "block-length" of the code $\mathfrak{C}$ is n^2.

We observe that pairwise independence of the events $\mathcal{A}_{\mathbf{Z}}$ is crucial. This is a consequence of the linear measurement model in (1). Note that the events $\mathcal{A}_{\mathbf{Z}}$ are *not* jointly independent. But the beauty of de Caen's bound allows us to exploit the pairwise independence to lower bound $\mathbb{P}(\mathcal{E}_n)$ and thus to obtain a tight upper bound on $E(R)$. To draw an analogy, just as linear codes achieve capacity in symmetric DMCs as only pairwise independence is required, de Caen's inequality allows us to move the exploitation of pairwise independence into the error exponent domain and make statements about the error exponent behavior of ensembles of linear codes.

We have a precise characterization of the reliability function $E(R)$ defined in (12) for the min-rank decoder. Note that the form of $E(R)$ in (16) also coincides with the exponential rate at which $\mathbb{P}(\mathcal{E}_n)$ converges to one as given in Proposition 3.

C. The Noisy Case

We conclude of this section by commenting how the min-rank decoder can be modified if the measurements y^k are corrupted by noise from the finite field. Now, instead of having access to noiseless measurements, we have

$$\mathsf{y}_a = \langle \mathbf{H}_a, \mathbf{X} \rangle + w_a \qquad a \in [k]. \tag{17}$$

where $\mathbf{w} = (w_1, \ldots, w_k) \in \mathbb{F}_q^k$ is a deterministic but unknown noise vector. We assume that $\|\mathbf{w}\|_0 = \lfloor \sigma n^2 \rfloor$ for some *noise level* $\sigma \in (0, k/n^2]$. Consider the following generalization of the min-rank decoder:

$$\begin{aligned} &\text{minimize} \quad \operatorname{rank}(\mathbf{X}) + \lambda \|\mathbf{w}\|_0 \\ &\text{subject to} \quad \langle \mathbf{H}_a, \mathbf{X} \rangle + w_a = \mathsf{y}_a, \quad a \in [k] \end{aligned} \tag{18}$$

The optimization variables are $\mathbf{X} \in \mathbb{F}_q^{n \times n}$ and $\mathbf{w} \in \mathbb{F}_q^k$. The parameter $\lambda > 0$ (which is allowed to depend on n) governs the tradeoff between the rank of $\mathbf{X}$ and the sparsity of $\mathbf{w}$. Let $H_{\mathrm{b}}(\,\cdot\,)$ denote the binary entropy function.

Proposition 7 (Achievability under noisy measurement model)**.** *Fix $\varepsilon > 0$ and choose $\lambda = \frac{1}{n}$. Assuming the uniform measurement model and that $\|\mathbf{w}\|_0 = \lfloor \sigma n^2 \rfloor$, if*

$$k > \frac{(3+\varepsilon)(\gamma+\sigma)[1-(\gamma+\sigma)/3]}{1 - H_{\mathrm{b}}[1/(3-(\gamma+\sigma))] \log_q 2}\, n^2, \tag{19}$$

then $\mathbb{P}(\mathcal{E}_n) \to 0$ as $n \to \infty$.

Since the prefactor in (19) is a monotonically increasing function in the noise level σ, the number of measurements degrades as σ increases, agreeing with intuition. Note that the regularization parameter λ is independent of σ. The factor of 3 (instead of 2) in (19) arises due in part to the uncertainty in the locations of the non-zero elements of $\mathbf{w}$. The proof of Proposition 7 is deferred to a longer version of this paper.

V. Sparse Random Sensing Matrices

In the previous section, we focused exclusively on the case where the elements of the sensing matrices $\mathbf{H}_a, a \in [k]$ are drawn independently and uniformly at random from $\mathbb{F}_q$. However, there is substantial motivation to consider different ensembles of sensing matrices. For example, in low-density parity check (LDPC) codes, the parity check matrix (analogous to the set of $\mathbf{H}_a$ matrices) is sparse. The sparsity aids in decoding via the sum-product (belief propagation) algorithm as the resulting Tanner (factor) graph is sparse.

In this section, we analyze the scenario where the sensing matrices are sparse. More precisely, each element of each matrix $\mathbf{H}_a$ is assumed to be an independently and identically distributed random variable with associated pmf

$$P(h; \delta, q) := \begin{cases} 1-\delta & h = 0 \\ \delta/(q-1) & h \in \mathbb{F}_q \setminus \{0\} \end{cases}. \tag{20}$$

Observe that if δ is small, then the probability that a randomly selected entry in $\mathbf{H}_a$ is zero is close to unity. In the rest of this section, we allow δ to depend on n but we do not make the dependence of δ on n explicit for ease of exposition. The question we would like to answer is: *How fast can δ decay with n such that the min-rank decoder is still reliable?*

Proposition 8 (Achievability under sparse measurement model)**.** *Fix $\varepsilon > 0$ and let δ be any sequence in $\Theta(\frac{\log n}{n})$. If there exists a positive integer N_ε such that* (8) *holds for all $n > N_\varepsilon$, then $\mathbb{P}(\mathcal{E}_n) \to 0$ as $n \to \infty$.*

Note that the sparsity-factor δ of the sensing matrices is allowed to tend to zero albeit at a controlled rate of $\Theta(\frac{\log n}{n})$. Thus, each $\mathbf{H}_a$ is allowed to have, on average, $\Theta(n \log n)$ non-zero entries (out of n^2 entries). The scaling rate is reminiscent of the number of trials required for success in the so-called *coupon collector's problem*. Indeed, we need at least an entry in a row and an entry in a column of $\mathbf{X}$ to be sensed (by a sensing matrix $\mathbf{H}_a$) for the min-rank decoder to succeed. The number of measurements required in the sparse sensing case is exactly the same as in the case where the elements of $\mathbf{H}_a$ are drawn *uniformly* at random from $\mathbb{F}_q$ in Proposition 4. In fact it also matches the information-theoretic lower bound. The following lemma is used in the proof of Proposition 8.

Lemma 9. *Let $d_{\mathbf{Z}} := \|\mathbf{X} - \mathbf{Z}\|_0$. The probability of $\mathcal{A}_{\mathbf{Z}}$, denoted as $\theta(d_{\mathbf{Z}}; \delta, q, k)$, is only a function of $d_{\mathbf{Z}}$ and is given as*

$$\theta(d; \delta, q, k) = \left[q^{-1} + (1 - q^{-1}) \left(1 - \frac{\delta}{1 - q^{-1}} \right)^d \right]^k. \tag{21}$$

Lemma 9 can be proved by induction on d and by observing that the pmf of $[\mathbf{H}_a]_{1,1} + [\mathbf{H}_a]_{1,2}$ is the circular convolution of $P(h; \delta, q)$ with itself. The function $\theta(d; \delta, q, k)$ is monotonically decreasing in d and is also upper bounded by $(1-\delta)^k$ for all d. We now provide a sketch of the proof of Proposition 8. The basic idea is to partition all possibly "misleading" matrices $\mathbf{Z}$ into subsets based on their Hamming distance from $\mathbf{X}$.

Proof: We can upper bound the probability $\mathbb{P}(\mathcal{E}_n)$ as:

$$\begin{aligned} \mathbb{P}(\mathcal{E}_n) &\overset{(a)}{\leq} \sum_{d=1}^{n^2} \sum_{\substack{\mathbf{Z} \neq \mathbf{X}, \operatorname{rank}(\mathbf{Z}) \leq \operatorname{rank}(\mathbf{X}) \\ \|\mathbf{X}-\mathbf{Z}\|_0 = d}} \theta(d; \delta, q, k) \\ &\overset{(b)}{\leq} \sum_{d=1}^{\lfloor \beta n^2 \rfloor} \sum_{\substack{\mathbf{Z} \neq \mathbf{X}, \operatorname{rank}(\mathbf{Z}) \leq \operatorname{rank}(\mathbf{X}) \\ \|\mathbf{X}-\mathbf{Z}\|_0 = d}} (1-\delta)^k + \ldots \end{aligned}$$

$$+ \sum_{d=\lceil \beta n^2 \rceil}^{n^2} \sum_{\substack{\mathbf{Z} \neq \mathbf{X}, \mathrm{rank}(\mathbf{Z}) \leq \mathrm{rank}(\mathbf{X}) \\ \|\mathbf{X}-\mathbf{Z}\|_0 = d}} \theta(\lceil \beta n^2 \rceil; \delta, q, k), \quad (22)$$

where in (a) we partition the sum over $\mathbf{Z}$ into classes of matrices which differ from $\mathbf{X}$ by d entries and in (b), we partition the resulting sum into two parts in accordance to the fractional parameter β (which is allowed to depend on n). We denote the two sums in (22) as A_n and B_n respectively. Now,

$$\begin{aligned} A_n &\leq n^2 |\{\mathbf{Z} : \|\mathbf{Z}-\mathbf{X}\|_0 = \lfloor \beta n^2 \rfloor\}| (1-\delta)^k \\ &\overset{(a)}{\leq} n^2 2^{n^2 H_{\mathrm{b}}(\beta)} (q-1)^{\beta n^2} (1-\delta)^k. \end{aligned} \quad (23)$$

Similarly, the second term in (22) can be upper bounded as

$$\begin{aligned} B_n &\leq n^2 |\{\mathbf{Z} : \mathrm{rank}(\mathbf{Z}) \leq \mathrm{rank}(\mathbf{X})\}| \theta(\lceil \beta n^2 \rceil; \delta, q, k) \\ &\overset{(a)}{\leq} n^2 q^{2\gamma(1-\gamma/2)n^2 + o(n^2)} \theta(\lceil \beta n^2 \rceil; \delta, q, k) \quad (24) \\ &\overset{(b)}{=} n^2 q^{n^2 \left[2\gamma(1-\gamma/2) + o(1) + \frac{k}{n^2} \log_q (q^{-1} + (1-q^{-1})(1 - \frac{\delta}{1-q^{-1}})^{\lceil \beta n^2 \rceil}) \right]}, \end{aligned}$$

where in (a) we used Lemma 1 and in (b) we used Lemma 9. Consider the choice of parameters: $\beta = \Theta(\frac{1}{n})$ and $\delta = \Theta(\frac{\log n}{n})$. Then, both (23) and (24) tend to zero as $n \to \infty$ if k satisfies (8) for all n sufficiently large. ∎

The natural question at this juncture is whether the "reliability function" [cf. (12)] can be computed for this sparse measurement model. The pairwise independence among the events $\mathcal{A}_{\mathbf{Z}}$ no longer holds and thus, de Caen's lower bound may not be tight as in the case where the entries of the sensing matrices are drawn uniformly at random from $\mathbb{F}_q$. It is not straightforward to compute $\mathbb{P}(\mathcal{A}_{\mathbf{Z}} \cap \mathcal{A}_{\mathbf{Z}'})$ as in the proof of Proposition 5. By the choice of (β, δ) our bounding technique for Proposition 8 only ensures that

$$\limsup_{n \to \infty} \frac{1}{n \log n} \log_q \mathbb{P}(\mathcal{E}_n) \leq -C \quad (25)$$

for some $C \in (0, \infty)$. Thus, instead of having a speed[5] of n^2 in the large-deviations upper bound, we have a speed of $n \log n$. This is because δ is allowed to decay to zero. Whether the speed $n \log n$ is optimal is open.

VI. Discussion and Conclusions

There is a natural correspondence between the rank minimization problem and rank-metric decoding [6], [7]. In the former, we solve a problem of the form (6). In the latter, the *code* typically consists of all matrices that lie in the kernel of some linear operator $\mathbf{H}$, i.e., the code $\mathfrak{C} = \ker(\mathbf{H}) \subset \mathbb{F}_q^{n \times n}$. A particular *codeword* $\mathbf{C} \in \mathfrak{C}$ is transmitted. The received word is given as $\mathbf{R} = \mathbf{C} + \mathbf{X}$, where $\mathbf{X}$ is assumed to be a low-rank "error" matrix. The optimization problem is then

$$\text{minimize} \quad \mathrm{rank}(\mathbf{R} - \mathbf{C}) \quad \text{subject to} \quad \mathbf{C} \in \mathfrak{C} \quad (26)$$

which is identical to the min-rank problem in (6) with the identification of the error matrix $\mathbf{X} \equiv \mathbf{R} - \mathbf{C}$. In general, solving (26) is intractable (NP-hard) but it is known that if the linear operator $\mathbf{H}$ admits a favorable algebraic structure, then learning a sufficiently low-rank $\mathbf{X}$ (and thus $\mathbf{C}$) from $\mathbf{R}$ can be done in polynomial time. More precisely, the class of *Gabidulin codes* [13], which are rank-metric analogs of Reed-Soloman codes, not only achieves the Singleton bound (and thus has maximum rank distance), but decoding can be achieved using a modified form of the Berlekamp-Massey algorithm. However, the structured nature of codes (and in particular the mutual dependence between the equivalent H_a matrices) does not permit the line of analysis we adopted.

Another promising direction was adopted in [8] where the authors assumed that the error matrix $\mathbf{X}$ is drawn uniformly at random from all matrices of *known* rank r. The authors constructed a code in which they first learned the rowspace of $\mathbf{X}$ before adopting a message-passing strategy to complete the reconstruction. However, the structured codebook violates the assumptions for our preceding analyses to hold. Nonetheless, by writing $\mathbf{X} = \mathbf{U}\mathbf{V}^T$ for matrices $\mathbf{U}, \mathbf{V} \in \mathbb{F}_q^{n \times r}$, we see that if the rowspace is known, all that remains unknown in $\mathbf{X}$ is the $\mathbf{U}$ matrix. Thus, in principle, we can solve the system of linear equations $\langle \mathsf{H}_a, \mathbf{U}\mathbf{V}^T \rangle = \mathsf{y}_a, a \in [k]$ for $\mathbf{U}$ to complete the recovery of $\mathbf{X}$. This is a subject of current investigation. The ideas in [8] were extended in [14] where the authors computed the capacity of various matrix-valued channels over finite fields as well as devised "error trapping" codes to achieve capacity.

In this paper, we derived information-theoretic limits for recovering a low-rank matrix over a finite field given linear measurements. We are currently working on developing tractable algorithms to solve the min-rank optimization (6) for particular classes of sensing matrices.

[5] The term *speed* is in direct analogy to the theory of large-deviations where $\mathbb{P}_n$ is said to satisfy a large-deviations upper bound with *speed* a_n and *rate function* $I(\cdot)$ if $\limsup_{n \to \infty} a_n^{-1} \log \mathbb{P}_n(\mathcal{E}) \leq -\inf_{x \in \mathrm{cl}(\mathcal{E})} I(x)$.

References

[1] E. J. Candès and T. Tao, "The power of convex relaxation: near-optimal matrix completion," *IEEE Trans. on Inf. Th.*, vol. 56, no. 5, 2010.

[2] B. Recht, "A simpler approach to matrix completion," *To appear in J. Mach. Learn. Research*, 2009, arXiv:0910.0651.

[3] S. Vishwanath, "Information theoretic bounds for low-rank matrix completion," in *Intl. Symp. Inf. Th.*, Austin, TX, July 2010.

[4] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 2, no. 52, pp. 471–501, 2009.

[5] D. S. Papailiopoulos and A. G. Dimakis, "Distributed Storage Codes Meet Multiple-Access Wiretap Channels," in *Proc. of Allerton*, 2010.

[6] P. Loidreau, "Properties of codes in rank metric," 2006, arXiv:0610057.

[7] D. Silva, F. R. Kschischang, and R. Kötter, "A rank-metric approach to error control in random network coding," *IEEE Trans. on Inf. Th.*, vol. 54, no. 9, 2008.

[8] A. Montanari and R. Urbanke, "Coding for network coding," 2007, arXiv:0711.3935.

[9] G. Bresler, E. Mossel, and A. Sly, "Reconstruction of Markov random fields from samples: Some observations and algorithms," in *11th International workshop APPROX*, 2008, pp. 343–356.

[10] I. Csiszár, "Linear codes for sources and source networks: Error exponents, universal coding," *IEEE Trans. on Inf. Th.*, vol. 28, no. 4, 1982.

[11] D. de Caen, "A lower bound on the probability of a union," *Discrete Math.*, vol. 69, pp. 217–220, May 1997.

[12] S. Draper and S. Malekpour, "Compressed Sensing over Finite Fields," in *Intl. Symp. Inf. Th.*, Seoul, Korea, July 2009.

[13] E. M. Gabidulin, "Theory of codes with maximum rank distance," *Probl. Inform. Transm.*, vol. 21, no. 1, pp. 1–12, 1985.

[14] D. Silva, F. R. Kschischang, and R. Kötter, "Communication over finite-field matrix channels," *IEEE Trans. on Inf. Th.*, vol. 56, no. 3, 2010.