Sparse Interactions: Identifying High-Dimensional Multilinear Systems via Compressed Sensing

Bobak Nazer and Robert D. Nowak
University of Wisconsin, Madison
ECE Department
Madison, WI, 53706, USA
Email: {bobak,nowak}@ece.wisc.edu

***Abstract*—This paper investigates the problem of identifying sparse multilinear systems. Such systems are characterized by multiplicative interactions between the input variables with sparsity meaning that relatively few of all conceivable interactions are present. This problem is motivated by the study of interactions among genes and proteins in living cells. The goal is to develop a sampling/sensing scheme to identify sparse multilinear systems using as few measurements as possible. We derive bounds on the number of measurements required for perfect reconstruction as a function of the sparsity level. Our results extend the notion of compressed sensing from the traditional notion of (linear) sparsity to more refined notions of sparsity encountered in nonlinear systems. In contrast to the linear sparsity models, in the multilinear case the pattern of sparsity may play a role in the sensing requirements.**

I. Introduction

The investigation in this paper is motivated by the following problem in systems biology. High-throughput experiments allow biologists to probe the effects of individual genes and their protein products. For many model organisms, such as yeast and the fruit fly, we now have so-called "single-deletion" cell libraries consisting of all possible variations of the normal cell with one gene removed or suppressed. By studying each single-deletion strain, biologists are able to deduce the relevance (or irrelevance) of a particular gene/protein to a specific function or process. For example, this kind of study was used to identify a small subset of the $13,071$ genes in the fruit fly that may be relevant to the replication of the influenza virus [1]. Similar studies have applied this methodology to identify genes involved in HIV virus replication. However, a meta-analysis of several independent studies has revealed a low degree of overlap between the sets of genes identified in the different studies [2].

One likely explanation is that there is redundancy in the genome. For example, two or more genes/proteins may have very similar functionalities. Removing a single gene in a situation like this may not produce a detectable effect, since a similar gene can perform the function of the deleted one. To detect the relevant genes it is necessary to remove them or modulate their expression levels simultaneously. A recent study tackled the problem directly for pairwise gene interactions in yeast [3]. This huge undertaking involved testing over 5 million gene pairs; yet only a small subset were identified as having potentially relevant interactions. Moving beyond pairwise interactions and/or to organisms with larger genomes is formidable to say the least. However, the results of this paper suggest that it may be possible to exploit the sparsity of the problem in order to drastically reduce the number of experiments required to identify the relatively small subset of significant interactions.

We propose and investigate a stylized version of the problem above and give bounds on the number of required measurements. A multilinear functional is used is used to model linear and nonlinear effects observable at the output of the underlying system. Through a change of variables, this identification problem can be expressed in a linear form and viewed as a compressed sensing problem with partially dependent measurement vectors. We use this formulation to leverage existing theory from compressed sensing. The novelty here is that the sensing matrices involved have a nonlinear dependency structure that requires a delicate analysis in order to establish a restricted isometry property (RIP) for our problem.

A. Related Work

Since the seminal papers of Candes, Romberg, Tao, and Donoho, there has been a great deal of interest in compressed sensing and its applications [4]–[6]. We do not attempt a full survey of the literature and only mention a few papers that are directly relevant to our considerations. The restricted isometry property and its connections to the Johnson-Lindenstrauss lemma were examined in [7] and we will use elements of that work in some of our own proofs. In [8], the authors showed how compressed sensing can be generalized to include sensing matrices with Toeplitz dependency structure. We will employ techniques from that paper to prove one of our bounds. Another bound we develop makes use of the framework recently developed in [9] for matrices with independent and isotropic rows or columns. We also mention that there have been proposals to use compressed sensing the design of genomics experiments [10], [11], but we are not aware of work that considers gene interactions.

This work was partially supported by the AFOSR grant FA9550-09-1-0140 and the NSF grant CCF-0728767. R. Nowak would also like to thank Trinity College and the Isaac Newton Institute at the University of Cambridge for support while this work was being completed.

II. Problem Statement

We now propose a mathematical model for interactions. There are M inputs $a_1, a_2, \ldots, a_M$ that take values in the reals. These inputs pass through a multilinear system with output

$$u = \sum_{1\le i_1<\cdots<i_D\le M} a_{i_1}\cdots a_{i_D}\; x_{i_1\cdots i_D}\ . \tag{1}$$

where each coefficient $x_{i_1 i_2\cdots i_D}$ takes values in the reals. Note that each combination of D inputs appears exactly once and that $N = \binom{M}{D}$ such combinations are possible. We will refer to N as the problem size. We are particularly interested in *sparse* multilinear functions in which only a small fraction of the coefficients are non-zero. Note that with $D = 1$ the multilinear function reduces to the standard linear model.

Example 1: For $D = 2$, the function can be written as the sum of all pairwise interactions:

$$u = \sum_{i_1=1}^{M}\sum_{i_2=i_1+1}^{M} a_{i_1}a_{i_2}\; x_{i_1 i_2}$$

Remark 1: We can generalize our framework to include all interactions of order D or less without affecting our results. In this case, we would have functions of the form

$$\sum_{d=1}^{D}\ \sum_{1\le i_1<\cdots<i_d\le M} a_{i_1}a_{i_2}\cdots a_{i_m}\; x^{(d)}_{i_1 i_2\cdots i_m}\ .$$

The restriction to interactions of order D simplifies the exposition and analysis.

The values of the coefficients $x_{i_1\cdots i_D}$ are unknown and our objective is to learn their values accurately and efficiently. To do so, we make measurements by choosing values for the M inputs and recording the resulting output. Let u_k denote the k^{th} measurement resulting from the input $\{a_{k1}, \ldots, a_{kM}\}$ and let K denote the total number of measurements taken. The goal is to determine bounds on the number measurements K needed to identify coefficients. In general, if $K < N$, then perfect reconstruction of the coefficients is impossible as the number of unknowns exceeds the number of equations.

However, if we assume that the number of non-zero coefficients is small compared to the problem size (i.e. the coefficients are sparse), then the coefficients can be identified with far less than N measurements. Formally, we say that the multilinear function in (1) is S-sparse if $x_{i_1 i_2\cdots i_D} \neq 0$ for at most S coefficients. Our bounds on the number of measurements needed for recovery rely on the use of random inputs. Also, although we focus on a noiseless measurement model in this paper, It is straightforward to also include the possibility of additive measurement noise using existing theory and methods (e.g., [12], [13]).

Throughout the paper, we will use c and C to denote generic positive numerical constants in our calculations, and they may represent different constants in different bounds.

III. Summary of Main Results

The relation between the measurements of the multilinear function in (1) and the parameters can be written as a linear system of equations. Using this representation, we apply results from compressed sensing to obtain upper bounds on the number of required measurements. Throughout the paper, we let $\log x$ denote the logarithm of x to the base 2 and $\ln x$ the natural logarithm.

Our first approach in Section VI combines the fact that our measurements preserve the input norm in expectation with a union bound over all sparsity patterns. In Section VII, we work with a framework that bounds all sparsity patterns simultaneously and yields optimal results to within polylogarithmic factors. Finally, in Section VIII, we work with tail bounds on our measurement vectors combined with a union bound over all patterns. This bound are based on a refined notion of sparsity called the *combinatorial dimension* of a multilinear functional which takes values in $1 \le \alpha \le D$ [14]. Overall, we arrive at the following theorem.

Theorem 1: Assume that the inputs are i.i.d. binary symmetric random variables. If the number of measurements K satisfies

$$K \ge c\min\left\{S\log^3(S)\log N, S^2\log\left(\frac{N}{S}\right), S^\alpha\log^\alpha\left(\frac{N}{S}\right)\right\}$$

then the measurements can be used to infer an arbitrary S-sparse multilinear function with overwhelming probability. This result is simply a combination of Theorems 3, 6, and 7, which we now set out to prove. Note that in our setting, α will tend to be close to be 1 as this corresponds to the case where each gene is involved in only one interaction.

IV. Compressed Sensing Formulation

We relate our situation to the canonical compressed sensing problem and use results from this area to derive an upper bound on the number of measurements. We begin by vectorizing the unknown coefficients according to some one-to-one index map:

$$\mathbf{x} = \{x_{i_1 i_2\cdots i_D}\}\ .$$

The same index map is used to create a measurement vector out of the products of the inputs:

$$\mathbf{a}_k = \{a_{i_1}a_{i_2}\cdots a_{i_D}\}\ .$$

This allows us to write each output in linear form:

$$u_k = \mathbf{a}_k^T\mathbf{x} \tag{2}$$

Next, we normalize each measurement by $\frac{1}{\sqrt{K}}$:

$$y_k = \frac{1}{\sqrt{K}}u_k.$$

Finally, we can write all K measurements in matrix form:

$$\mathbf{A} = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_K^T \end{bmatrix}, \qquad \mathbf{y} = [y_1\ y_2\ \cdots\ y_K]^T = \frac{1}{\sqrt{K}}\mathbf{A}\mathbf{x}.$$

Remark 2: Usually, the factor $\frac{1}{\sqrt{K}}$ is absorbed into $\mathbf{A}$ (the input). To account for the possibility of sums of interactions of different orders, here we normalize the measurements at the output.

We now review some standard definitions and results from compressed sensing that will be useful in our proofs. We begin with the notion of a restricted isometry property, first introduced in [5].

Definition 1: A matrix $\mathbf{A}$ satisfies the restricted isometry property (RIP) of order S with constant δ_S if

$$(1-\delta_S)\|\mathbf{x}\|_2^2 \leq \frac{1}{K}\|\mathbf{A}\mathbf{x}\|_2^2 \leq (1+\delta_S)\|\mathbf{x}\|_2^2$$

for all vectors $\mathbf{x}$ with support of size S or less, $\|\mathbf{x}\|_0 \leq S$.

If the matrix $\mathbf{A}$ satisfies the RIP with $\delta_{2S} < 1$, then it can be shown that solving the following ℓ_0 optimization problem

$$\min_{\hat{\mathbf{x}}\in\mathbb{R}^N} \|\hat{\mathbf{x}}\|_0 \text{ subject to } \frac{1}{\sqrt{K}}\mathbf{A}\hat{\mathbf{x}} = \mathbf{y} \tag{3}$$

yields an estimate $\hat{\mathbf{x}}$ that equals $\mathbf{x}$ if $\|\mathbf{x}\|_0 \leq S$. However, this approach quickly becomes computationally intractable as the problem size N grows. The landmark result of compressed sensing is that, under some technical conditions, the ℓ_0 criterion in the optimization in (3) can be relaxed to an ℓ_1 norm (which corresponds to solving a linear program) [4]–[6]. The theorem below is a good representative of the compressed sensing framework [15].

Theorem 2 (Candès): Assume the matrix $\mathbf{A}$ satisfies the RIP with $\delta_{2S} < \sqrt{2}-1$ and let $\mathbf{x}_S$ be the unique vector in $\mathbb{R}^N$ that is equal to $\mathbf{x} \in \mathbb{R}^N$ on its largest S values and zero elsewhere. Then, the solution $\mathbf{x}^*$ of the following ℓ_1 optimization problem:

$$\min_{\hat{\mathbf{x}}\in\mathbb{R}^N} \|\hat{\mathbf{x}}\|_1 \textit{ subject to } \frac{1}{\sqrt{K}}\mathbf{A}\hat{\mathbf{x}} = \mathbf{y} \tag{4}$$

satisfies

$$\|\mathbf{x}^*-\mathbf{x}\|_1 \leq c_1\|\mathbf{x}-\mathbf{x}_S\|_1$$
$$\|\mathbf{x}^*-\mathbf{x}\|_2 \leq c_2\frac{\|\mathbf{x}-\mathbf{x}_S\|_1}{\sqrt{S}}$$

for some positive constants c_1 and c_2.
For a full proof, see Theorem 1.2 in [15]. For extensions to cases where the measurements are contaminated with noise see, for example, [13].

Corollary 1: If the vector $\mathbf{x}$ has at most S non-zero entries and the matrix $\mathbf{A}$ satisfies the RIP with $\delta_{2S} < \sqrt{2}-1$, then the solution $\mathbf{x}^*$ of (4) is exactly equal to $\mathbf{x}$.

Thus, if we can demonstrate the matrix $\mathbf{A}$ satisfies the RIP, we can apply the tools of compressed sensing to efficiently infer the unknown coefficients.

V. Initial Observations

The RIP enforces that the norms all sparse vectors are approximately preserved by the measurement matrix. As a first step, we will show that our measurements preserve the norm in expectation.

Lemma 1: Assume the inputs are generated independently according to symmetric distributions with unit variance. Then, the expected value of the Gram matrix $\mathbf{G} = \frac{1}{K}\mathbf{A}^T\mathbf{A}$ of the measurement matrix $\frac{1}{\sqrt{K}}\mathbf{A}$ is an identity matrix:

$$\mathbb{E}\left[\mathbf{G}\right] = \mathbb{E}\left[\frac{1}{K}\mathbf{A}^T\mathbf{A}\right] = \mathbf{I}\ .$$

Proof: First, note that each diagonal element of $\mathbb{E}\left[\mathbf{G}\right]$ has the following form $\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[a_{ki_1}^2 a_{ki_2}^2 \cdots a_{ki_D}^2]$. Since the variables are independent and have variance 1 this takes the form $\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[a_{ki_1}^2]\mathbb{E}[a_{ki_2}^2]\cdots\mathbb{E}[a_{ki_D}^2]$ which is just equal to one. Next, each off-diagonal element of $\mathbb{E}\left[\mathbf{G}\right]$ can be written as a sum of products:

$$\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[a_{ki_1}a_{ki_2}\cdots a_{ki_D}a_{kj_1}a_{kj_2}\cdots a_{kj_D}]$$

and because each of the products in (1) has a unique form, it follows that two or more of the variables appears in just once in the expectation above. Without loss of generality, let i_1 and j_1 denote the indices of these variables. Then, the expectation can be written as:

$$\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[a_{ki_1}]\mathbb{E}[a_{kj_1}]\mathbb{E}[a_{ki_2}\cdots a_{ki_D}a_{kj_2}\cdots a_{kj_D}]$$

Since the a_{ki} are independent and symmetric random variables, $\mathbb{E}[a_{ki}] = 0$ so the summation is zero and each off-diagonal term has zero mean. ■

The lemma above has two key implications for our measurements. First, it implies that the expectation of the square of each measurement is equal to the norm of the vector:

$$\mathbb{E}[u_k^2] = \|\mathbf{x}\|_2^2\ .$$

Second, it shows that the sum of the squares of measurements is also equal to the norm in expectation:

$$\mathbb{E}\left[\|\mathbf{y}\|_2^2\right] = \frac{1}{K}\mathbf{x}^T\mathbb{E}\left[\mathbf{A}^T\mathbf{A}\right]\mathbf{x}\ =\ \|\mathbf{x}\|_2^2\ .$$

There are many input distributions that satisfy the assumptions of Lemma 1. We will assume from here forward that the inputs are independent and identically distributed (i.i.d.). Furthermore, for practical and theoretical reasons, we will assume that the input distribution is bounded. This is particularly natural in the systems biology setting motivating our investigation. Also, as we will see, nonlinear interactions lead to heavy-tailed distributions which are difficult to control for unbounded distributions. With these considerations in mind, it suffices to consider binary symmetric input distributions. Throughout the rest of the paper, we will consider i.i.d. inputs with distribution $\mathbb{P}(a_{ki} = 1) = \mathbb{P}(a_{ki} = -1) = 1/2$.

This distribution satisfies the requirements of Lemma 1 and so the norm of the output vector $\mathbf{y}$ will eventually converge to the norm of the input vector $\mathbf{x}$. To bound the number of measurements required for the RIP to hold with an appropriate constant, we need to characterize how quickly $\frac{1}{K}\sum u_k^2$ concentrates to its mean. Without loss of generality, let us assume that $\|\mathbf{x}\|_2 = 1$. Then we need to quantify how quickly $\frac{1}{K}\sum u_k^2$ concentrates about its mean value of 1. In the standard compressed sensing formulation, the elements of the measurement matrix are drawn i.i.d. according to a subgaussian distribution. In that case each measurement satisfies $\mathbb{P}(|u_k^2 - 1| > t) < \exp(-ct)$, for a constant $c > 0$, and from this tail bound it can be shown that the sum concentrates rapidly enough that only $S\log(N/S)$ measurements are required.

The binary symmetric distribution is subgaussian, but because of the nonlinear interactions the distribution of u^2, as defined in (1), can have much heavier tails. The tail behavior is intimately connected to the pattern of sparsity. Unlike the usual linear sparsity models, in the multilinear setting different sparsity patterns lead to different tail behaviors, depending on the amount of interaction and dependency in the terms involved. The following lemma characterizes the extremes of the tail behavior. A more refined analysis will be carried out later in Section VIII.

Lemma 2: Let u a multilinear function of the form (1) and let $a_1, \ldots, a_M$ be i.i.d. binary symmetric random variables. Assume $\|\mathbf{x}\|_2 = 1$ and let $\mathcal{T}$ denote the set of indices on which the coefficients $x_{i_1 i_2 \cdots i_D}$ are non-zero. Then there exists a constant $c > 0$ such that for sufficiently large positive t

$$\sup_{\mathcal{T}} \mathbb{P}\left(|u^2 - 1| > t\right) \geq \exp\left(-c\,t^{1/D}\right)$$
$$\inf_{\mathcal{T}} \mathbb{P}\left(|u^2 - 1| > t\right) \leq \exp\left(-ct\right) \ .$$

Proof: Let $|\mathcal{T}| = S \leq N$, the cardinality of $\mathcal{T}$. For the first bound, suppose $\mathcal{T}$ consists of all D-tuples in the set $1 \leq i_1 < \cdots < i_D \leq S^{1/D}$, where for convenience we assume that $S^{1/D}$ is an integer. In a sense, this is the most dependent configuration of an S-sparse D-linear form. Furthermore, assume that each non-zero coefficient takes the value $\frac{1}{\sqrt{S}}$ so that $\|\mathbf{x}\|_2 = 1$. The probability that $|u^2 - 1| \geq S - 1$ is lower bounded by

$$\mathbb{P}(|u^2 - 1| \geq S - 1) \geq 2^{-S^{1/D}} = \exp(-S^{1/D}\ln(2))$$

as this is the probability that $a_1 = \cdots = a_{S^{1/D}} = 1$. For the second bound, assume that each non-zero coefficient has a completely unique set of indices $i_1, \ldots, i_M$; i.e., no two non-zero coefficients have a single index value in common. In this case, the products $a_{i_1} \cdots a_{i_D}$ associated with the non-zero coefficients are i.i.d. Thus, u is equivalent to a weighted sum of i.i.d. binary symmetric random variables and is consequently subgaussian with tail $\mathbb{P}\left(|u_k^2 - 1| > t\right) \leq \exp\left(-ct\right)$, for some $c > 0$, as desired [16]. ■

Lemma 2 shows that the tails of a D-linear form can range from subgaussian to arbitrarily heavy-tailed, depending on D. Heavier tails generally translates into slower concentration about the mean, which presents challenges for the sparse recovery problem. Standard RIP bounds are not applicable to our situation due to the nonlinear dependencies. Therefore, we present three different attacks on our problem borrowing ideas from other approaches. The first and simplest approach is based on Geršgorin's Disk Theorem. The second attack applies a recent result for heavy-tailed restricted isometries [9], [17]. The third method uses results from the theory of Rademacher Chaos. None of these attacks yields the optimal RIP bounds in all situations, so our ultimate statement is a combination of the bounds derived from the three different approaches.

A natural question to ask is whether the pattern of sparsity, and hence the dependencies and tail behavior, has a real effect on the problem in practice. The simulation results depicted in Fig. 1 suggest that the dependencies have a significant impact. The simulations show that Gram matrices corresponding to higher order multilinear functions tend to have smaller minimum eigenvalues, which suggests that more measurements will be needed for higher order problems.

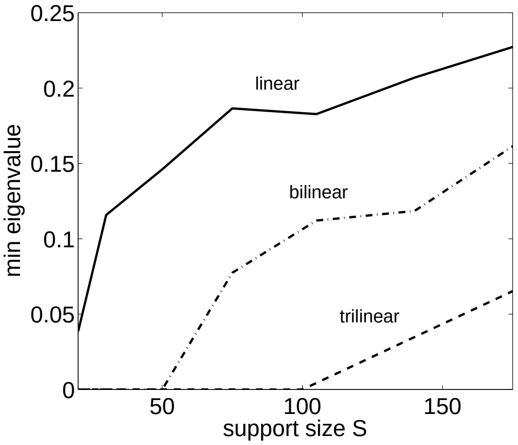

Fig. 1. Comparison of minimum eigenvalues of Gram matrices as a function support size S. Each curve depicts the smallest minimum eigenvalue observed in 10^5 independent draws of $S \times S$ Gram matrices generated by $5S$ independent vectors. Three different types of vectors are compared: vectors with S i.i.d. binary symmetric random entries (solid), with entries equal to all pairwise-products of $\sim \sqrt{S}$ i.i.d. binary symmetric variables (dash-dot), and with entries equal to all third order products of $\sim S^{1/3}$ i.i.d. binary symmetric variables (dashed).

VI. RIP from Geršgorin's Theorem

In this section, we will show how to get RIP constants arbitrarily close to 0 if the number of measurements K scales like $S^2 \log N$. We follow the proof strategy used by Haupt *et al.* to establish the RIP for Toeplitz matrices [8]. Let $\frac{1}{\sqrt{K}}\mathbf{A}_{\mathcal{R}}$ be the submatrix formed by taking the columns of $\frac{1}{\sqrt{K}}\mathbf{A}$ with indices in the set $\mathcal{R} \subset \{1, 2, \ldots, N\}$. The Gram matrix of $\frac{1}{\sqrt{K}}\mathbf{A}_{\mathcal{R}}$ is $\mathbf{G}_{\mathcal{R}} = \frac{1}{K}\mathbf{A}_{\mathcal{R}}^T\mathbf{A}_{\mathcal{R}}$. If we can show that the eigenvalues of $\mathbf{G}_{\mathcal{R}}$ lie in the range $[1 - \delta_S, 1 + \delta_S]$ for all subsets $\mathcal{R}$ of size S, $|\mathcal{R}| = S$, then the matrix must satisfy the RIP with constant δ_S. Haupt *et al.* bound the eigenvalues using Geršgorin's Disc Theorem which is reproduced below.

Theorem 3 (Geršgorin): Let $\mathbf{G} = \{g_{\ell m}\}$ be an $S \times S$ real-valued matrix. Then, each eigenvalue λ_ℓ lies in the following range

$$\lambda_\ell \in \Big[g_{\ell\ell} - \sum_{m \neq \ell} |g_{\ell m}|,\ g_{\ell\ell} + \sum_{m \neq \ell} |g_{\ell m}|\Big].$$

See, for instance, [18] for a proof. Thus, if we can show that the diagonal elements of each $\mathbf{G}_{\mathcal{R}}$ are close to one and the off-diagonal elements are close to zero, we can establish the RIP. These requirements can be checked via Hoeffding's concentration inequality.

Lemma 3 (Hoeffding): Let $v_1, v_2, \ldots, v_K$ be independent random variables satisfying $|v_i| \leq c_{\mathit{MAX}}$. Then, the probability that the sum $v_{\mathit{SUM}} = \sum_{i=1}^{K} v_i$ deviates from its mean $\mathbb{E}[v_{\mathit{SUM}}]$ is upper bounded as follows:

$$\mathbb{P}\left(\Big|v_{\mathit{SUM}} - \mathbb{E}[v_{\mathit{SUM}}]\Big| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2Kc_{\mathit{MAX}}^2}\right)$$

We can now prove our first main result.

Theorem 4: If the number of measurements satisfies

$$K \geq c_1 S^2 \log N$$

then the matrix $\mathbf{A}$ satisfies the RIP with probability at least $1 - \exp\left(-c_2 K/S^2\right)$ for some positive constants c_1 and c_2.

Proof: From Lemma 1, the expected value of the Gram matrix is the identity matrix. Note that since the inputs are binary and symmetric, the diagonal elements of the Gram matrix are exactly equal to 1:

$$g_{\ell\ell} = \frac{1}{K}\sum_{i=1}^{K} a_{i_1}^2 a_{i_2}^2 \cdots a_{i_D}^2 = \frac{1}{K}\sum_{i=1}^{K} 1 = 1 \ .$$

Since each off-diagonal element is the sum of binary symmetric random variables, we can use Lemma 3 to get $\mathbb{P}\left(|g_{\ell m}| \geq t\right) \leq 2\exp\left(-\frac{Kt^2}{2}\right)$. By the union bound,

$$\begin{aligned} \mathbb{P}\left(\bigcup_{\ell=1}^{M}\bigcup_{m\neq\ell}^{M}\{|g_{\ell m}| \geq t\}\right) &\leq 2N^2\exp\left(-\frac{Kt^2}{2}\right). \\ &\leq \exp\left(-\frac{Kt^2}{2} + \log N + 1\right). \end{aligned}$$

Therefore, using Theorem 3 and setting $t = \delta_S/S$, the eigenvalues of $\mathbf{G}_{\mathcal{R}}$ are in the range $\lambda_\ell \in [1-\delta_S, 1+\delta_S]$ with probability at least $1 - \exp\left(-\frac{K\delta_S^2}{2S^2} + \log N + 1\right)$. Therefore, if $K \geq \frac{2}{\delta_S^2}(S^2\log N + 1)$ we obtain the desired result. ■

Although this theorem does not establish the linear dependence on sparsity that we would like, the proof is quite simple and demonstrates that the number of required measurements does not depend on the interaction order D beyond the $\log N$ term which is approximately equal to $D\log M$.

VII. Heavy-Tailed Restricted Isometries

We now show how a recent result due to Vershynin [9], [17] can be applied to our problem. Vershynin's result is an extension of a framework pioneered in earlier work by Rudelson and Vershynin [17] that developed new bounds for random Fourier compressed sensing matrices as well as tighter constants for i.i.d. Gaussian matrices. The main result that is relevant to our discussion is the following theorem.

Theorem 5 (Vershynin): Let $\mathbf{A}$ be a $K \times N$ measurement matrix whose rows $\mathbf{a}_\ell^T$ are independent isotropic random vectors in $\mathbb{R}^N$. Assume the entries of $\mathbf{A}$ are bounded, $|a_{\ell m}| \leq 1$, almost surely. For every sparsity level $S \leq N$ and constant $0 < \epsilon < 1$, if the number of measurements satisfies

$$K \geq C\ \epsilon^{-2} S \log^3\left(\epsilon^{-2}S\right)\log N \tag{5}$$

then the RIP constant δ_S of the matrix $\mathbf{A}$ is upper bounded by ϵ in expectation: $\mathbb{E}[\delta_S] \leq \epsilon$.

This is Theorem 70 in [9]. The basic insight behind this result is that a union bound argument is not strong enough to establish the RIP with S polylog(N) heavy-tailed measurements. The arguments in the proof bound all possible sparsity patterns simultaneously.

The expectation bound on the RIP constant δ_S can easily be converted to a bound on the probability that δ_S exceeds some threshold using Markov's inequality. This leads us to the following bound on the number of required measurements to a RIP.

Theorem 6: If the number of measurements K is at least

$$K = C\frac{S}{\gamma^2\delta_S^2}\ \log^3\left(\frac{S}{\gamma^2\delta_S^2}\right)\ \log(N) \tag{6}$$

then the matrix $\mathbf{A}$ satisfies the RIP with constant δ_S with probability at least $1-\gamma$ for some positive constant C.

Proof: From Lemma 1, $\frac{1}{K}\mathbb{E}[\mathbf{A}^T\mathbf{A}] = \mathbf{I}$ so the rows of $\mathbf{A}$ are isotropic. Since we generate the inputs independently for each observation, the rows are also independent (i.e. all the dependencies introduced by the multilinear structure are across the columns). Applying Theorem 5, we get that the RIP constant satisfies $\mathbb{E}[\delta_S] \leq \epsilon$ if K satisfies (5). Now, by Markov's inequality, the probability that the RIP constant exceeds δ_S is upper bounded by ϵ/δ_S. Setting $\epsilon = \gamma\delta_S$ completes the proof. ■

Note that if the sparsity S scales linearly with the problem size N, this bound takes the simpler form $S\log^4(N)$.

VIII. RIP from Tail Bounds

To obtain the RIP based on tail bounds we give a straightforward generalization (to the heavy-tailed situation) of a well-known result for subgaussian tail bounds [7]. The proof is included in the Appendix. The basic idea is to start with a tail bound on the measurement vector and apply the union bound to bound the number of measurements needed to get a RIP.

Lemma 4: Assume that for any $\mathbf{x} \in \mathbb{R}^N$ that is S-sparse, $\|\mathbf{x}\|_0 \leq S$, the sensing matrix $\mathbf{A}$ satisfies the following concentration inequality for constants $\eta, \rho > 0$ and any $0 < \delta < 1$:

$$\mathbb{P}\left(\left|\frac{1}{K}\|\mathbf{A}\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2\right| > \epsilon\|\mathbf{x}\|_2^2\right) < \exp\Big(-c_0(\delta)K^\rho S^\eta\Big)\ .$$

If $K \geq c_1(\delta)S^{(1-\eta)/\rho}\left(\log(N/S)\right)^{1/\rho}$ (the number of rows in $\mathbf{A}$), then $\mathbf{A}$ satisfies the RIP with constant δ with probability at least $1 - \exp\left(-c_2(\delta)K^\rho S^\eta\right)$.

Vershynin's bound is within a poly-logarithmic factor of $S\log(N/S)$, the bound on K one would have for an i.i.d. subgaussian sensing matrix. As pointed out in Section V, certain sparsity patterns in the multilinear forms have no dependencies among the terms and thus are equivalent to linear forms. Vershynin's bound is not adaptive to the pattern of sparsity and consequently it is sometimes too conservative. In this section we use Rademacher chaos theory to obtain bounds that are adaptive in this sense. Before following this more sophisticated approach, we state a simple tail bound based on Bernstein's inequality.

Recall that $K^{-1}\|A\mathbf{x}\|_2^2 = K^{-1}\sum_{k=1}^K u_k^2$, where the u_k are i.i.d. Rademacher chaos variables (of the form of $u(M)$) as defined in (2). In the following sections we will assume (without loss of generality) that $\|\mathbf{x}\|_2 = 1$ and thus focus on bound $\mathbb{P}(|K^{-1}\sum_{k=1}^K u_k^2 - 1| > t)$.

Lemma 5: Assume that $u_1, \ldots, u_K$ are i.i.d. random variables of the form (2) with $a_1, \ldots, a_M$ i.i.d. binary symmetric and $\mathbf{x}$ is S-sparse. Then

$$\mathbb{P}\left(\Big|\frac{1}{K}\sum_{k=1}^K u_k^2 - 1\Big| > t\right) \leq \exp\left(\frac{-\min\left(3^{-2D}Kt^2, \frac{Kt}{3S}\right)}{4}\right)$$

Proof: Note that due to the sparsity and unit norm of $\mathbf{x}$, each measurement satisfies $u_k^2 \leq \|\mathbf{a}_k\|_\infty \|\mathbf{x}\|_1 \leq S$. Using Bonami's hypercontractive inequality (see equation (1.2) in [14]), it can be shown that $\left(\mathbb{E}[u_k^4]\right)^{1/4} \leq (4-1)^{D/2}\mathbb{E}[u_k^2] = 3^{D/2}$. By Bernstein's inequality, we get that

$$\mathbb{P}\left(\Big|\frac{1}{K}\sum_{k=1}^K u_k^2 - 1\Big| \geq t\right) \leq \quad \exp\left(-\frac{1}{4}\frac{Kt^2}{\max(3^{2D}, tS/3)}\right)$$

which completes the proof. ■

Using this tail bound in Lemma 4 produces a requirement of the form $K \geq C\max(3^{2D}S\log(N/S), S^2\log(N/S))$. This is similar to the bound obtained via Geršgorin's Theorem, but can be a bit tighter for larger values of S.

A. Rademacher Chaos Tail Bounds

A homogeneous Rademacher chaos of order $D \geq 1$ is a random variable of the same form as our observations, that is for any positive integer L

$$u(L) = \sum_{1\leq i_1<i_2<\cdots<i_D\leq L} a_{i_1}a_{i_2}\cdots a_{i_D}\; x_{i_1 i_2\cdots i_D}\ , \qquad (7)$$

where $x_{i_1,\ldots,i_D}$ are real-valued coefficients and $\{a_1, \ldots, a_L\}$ are i.i.d. binary symmetric (i.e., *Rademacher*) random variables, $\mathbb{P}(a_i = 1) = \mathbb{P}(a_i = -1) = 1/2$. Note that variable L is the number of independent variables involved in the chaos, and the multilinear function u defined in (1) corresponds to $u(M)$. As shown in Lemma 2, the tails of this random variable can be quite heavy or as light as subgaussian, depending on the pattern of non-zero $x_{i_1,\ldots,i_D}$. It turns out that the effect of this pattern is well-captured by the *combinatorial dimension* of the chaos [14].

First, we introduce the notion of a infinite series Rademacher chaos of the form

$$\sum_{1\leq i_1<\cdots<i_D<\infty} a_{i_1}a_{i_2}\cdots a_{i_D}\; x_{i_1 i_2\cdots i_D}\ ,$$

where $a_1, a_2, \ldots$ are i.i.d. Rademacher random variables. Let $\mathcal{T} \subseteq \mathbb{N}^D$ be the set of indices on which the coefficients of the Rademacher chaos are non-zero. For each $L = 1, 2, \ldots,$ let $\mathcal{T}_L$ be the restriction of $\mathcal{T}$ to inputs $1, 2, \ldots, L$:

$$\mathcal{T}_L = \mathcal{T} \cap \{1, 2, \ldots, L\}^D.$$

Note that truncating the infinite series to terms involving only $\{a_1, \ldots, a_L\}$ is equivalent to the multilinear form $u(L)$ in (7) and $\mathcal{T}_L$ is its coefficient support.

Different patterns of sparsity lead to varying degrees of dependency among the terms in the chaos. This dependency can be quantified in terms of the so-called *combinatorial dimension* which is defined below following the work of [14], [19].

Definition 2: We say that $\mathcal{T}$ has combinatorial dimension α if there exist positive constants c_1 and c_2 and a positive integer M_0 such that for each $L \geq L_0$,

$$\sup_{A_1,A_2,\ldots,A_D\subset\{1,\ldots,L\}^D} \frac{|\mathcal{T}_L \cap (A_1\times A_2\times\ldots\times A_D)|}{\left(\max_{1\leq j\leq D}|A_j|\right)^\alpha} \leq c_1$$

and $|\mathcal{T}_L| \geq c_2 L^\alpha$.

As shown in the following lemma (see Theorem 1.5 in [19] for a proof), the combinatorial dimension is intimately connected to the rate at which the chaos concentrates.

Lemma 6 (Blei-Janson): Let $u(L)$ be a sequence of Rademacher chaos of order D with combinatorial dimension α. For all $t \geq 2$, there exist positive constants c_1 and c_2 such that the upper tail is lower and upper bounded as follows:

$$\exp\left(-c_1 t^{2/\alpha}\right) \leq \sup_L \mathbb{P}\left(\big|u(L)\big| > t\right) \leq \exp\left(-c_2 t^{2/\alpha}\right)\ .$$

If we assume that $\|x\|_2 = 1$, then it follows that

$$\exp\left(-c_1 t^{1/\alpha}\right) \leq \sup_L \mathbb{P}\left(\big|u^2(L) - 1\big| > t\right) \leq \exp\left(-c_2 t^{1/\alpha}\right)$$

It can be shown that $1 \leq \alpha \leq D$ for order D Rademacher chaos. For example, in Lemma 2, the worst-case (heaviest) tail was generated by a chaos with combinatorial dimension of $\alpha = D$ and the best-case (lightest) tail corresponded to a situation where $\alpha = 1$. Those two tails corresponded to the largest lower bound and smallest upper bound, respectively, possible for a Rademacher chaos of order D.

The Blei-Janson result characterizes the tails of a single chaos. We are interested in the tails of $K^{-1}\|A\mathbf{x}\|_2^2 = K^{-1}\sum_{k=1}^K u_k^2$, where the u_k are i.i.d. and each is an order-D chaos in M variables. The tails of this sum are bounded in the following lemma.

Lemma 7: Assume that u_k, $k = 1, \ldots, K$, are i.i.d. Rademacher variables of order D with combinatorial dimension $1 \leq \alpha \leq D$. There exist constants $c, C > 0$ such that

$$\mathbb{P}\left(\Big|\frac{1}{K}\sum_{k=1}^K u_k^2 - 1\Big| > t\right) \leq C\exp(-c\min(Kt^2, K^{1/\alpha}\, t^{1/\alpha}))$$

The lemma is proved in the Appendix. A key element of the proof is a moment bound for sums of symmetric i.i.d. variables due to R. Latala [20]. Plugging the tail bounds from Lemma 5 and 7 into Lemma 4 yields the following theorem.

Theorem 7: If the number of measurements satisfies,

$$K \ge c_1 \min\left\{S^2 \log(N/S),\ S^\alpha \log^\alpha(N/S)\right\} \tag{8}$$

then the matrix $\mathbf{A}$ satisfies the RIP with probability at least $1 - \exp\left(-c_2 \max\left\{K/S,\ K^{1/\alpha}\right\}\right)$.
Note that in the case $\alpha = 1$, we have a bound that is strictly better than the bounds obtained above. The case $\alpha = 1$ may be common in many applications. For example, in the systems biology setting we may have pairs of genes interacting with each other but not with other genes.

IX. Extensions

It is possible in certain cases to obtain a tail bound that does not depend on the combinatorial dimension of the individual Rademacher chaos variables. Consider the (normalized) sum $K^{-1/2}\sum_{k=1}^K (u_k^2 - 1)$. This sum can be written as a sum of the form $K^{-1/2}\sum_{k=1}^K v_k$, where the v_k are i.i.d. and can be written as a sum of Rademacher chaos variables of order 2 through $2D$ in KM variables (with M variables contributed by each v_k in the sum). Observe that the constituent chaos variables include no interactions between the variables involved in v_i and v_j, for all $i \neq j$. This implies that the combinatorial dimension of $K^{-1/2}\sum_{k=1}^K v_k$ is less that that of u_k. In fact, as $K \to \infty$ the combinatorial dimension of $K^{-1/2}\sum_{k=1}^K v_k$ tends to 1. Suppose that each u_k is S-sparse, with $S < N$ a fixed constant. Take $K = cS\log(N/S)$ for some constant $c > 0$. Then it follows from Lemma 6 that $\lim_{N\to\infty} \mathbb{P}(|K^{-1/2}\sum_{k=1}^K v_k| > t) \le \exp(-c_2 t^2)$, and hence $\lim_{N\to\infty} \mathbb{P}(|\frac{1}{K}\sum_{k=1}^K v_k| > t) \le \exp(-c_2 K t^2)$. This tail bound suggests that if the sparsity level is fixed, then $K > cS\log(N/S)$ measurements will suffice for N sufficiently large.

It is possible to extend the results for the recovery of sparse multilinear functions to sparse polynomial functions. Polynomial functions include auto-interactions not present in the multilinear form. These interactions require the input distribution to be multi-level (binary will not suffice) and imply that the expectation of the Gram matrix $\mathbb{E}[\mathbf{A}^T\mathbf{A}]$ is not proportional to identity. Instead the expected Gram matrix can be arranged in a block diagonal structure. The techniques developed in this paper can be used to deal with this situation and the resulting RIP conditions will then depend on the eigenvalue spread of the expected Gram matrix.

Acknowledgement. The authors thank P. Alhquist for introducing them to the virus replication problem. R. Nowak would also like to thank R. Latala and R. Nickl for pointing out the moment estimates for the sums of i.i.d. variables. Both authors would like to thank B. Recht for pointing them to the work of M. Rudelson and R. Vershynin.

Appendix

A. Proof of Lemma 4

Without loss of generality, we can assume that $\|\mathbf{x}\| = 1$. Let $\mathcal{A}_\mathcal{T}$ denote the set of all vectors with ℓ_2-norm of 1 in $\mathbb{R}^N$ whose support is on pattern $\mathcal{T}$. Choose a cover $\mathcal{Q}_\mathcal{T} \subseteq \mathcal{A}_\mathcal{T}$ such that $\min_{\mathbf{q}\in\mathcal{Q}_\mathcal{T}} \|\mathbf{x}-\mathbf{q}\|_2 \le \frac{\delta}{4}$ for all $\mathbf{x} \in \mathcal{B}_\mathcal{T}$. It can be shown that there are choices of $\mathcal{Q}_\mathcal{T}$ with cardinality at most $\left(\frac{12}{\delta}\right)^S$. We now use the union bound to give an upper bound on the probability that $\mathbf{A}$ distorts the length of some $\mathbf{q}$ by more than $\frac{\delta}{2}$:

$$\begin{aligned}
&\mathbb{P}\left(\left\{\left|\|\mathbf{A}\mathbf{q}\|_2^2 - 1\right| \ge \delta/2 \ \text{ for some } \mathbf{q} \in \mathcal{Q}_\mathcal{T}\right\}\right) \\
&\quad \le \sum_{\mathbf{q}\in\mathcal{Q}_\mathcal{T}} \mathbb{P}\left(\left|\|\mathbf{A}\mathbf{q}\|_2^2 - 1\right| \ge \delta/2\right) \\
&\quad \le \sum_{\mathbf{q}\in\mathcal{Q}_\mathcal{T}} \exp(-c_0(\delta/2)K^\rho S^\eta) \\
&\quad \le \left(\frac{12}{\delta}\right)^S \exp(-c_0(\delta/2)K^\rho S^\eta).
\end{aligned}$$

This implies that

$$\left(1 - \frac{\delta}{2}\right) \le \|\mathbf{A}\mathbf{q}\|_2 \le \left(1 + \frac{\delta}{2}\right) \tag{9}$$

for all $\mathbf{q} \in \mathcal{Q}_\mathcal{T}$ with probability at least $1 - (12/\delta)^S \exp(-c_0(\delta/2)K^\rho S^\eta)$. To each vector $\mathbf{x} \in \mathcal{B}_\mathcal{T}$ assign a cover vector $\mathbf{q}_\mathbf{x} \in \mathcal{Q}_\mathcal{T}$ such that $\|\mathbf{q}_\mathbf{x} - \mathbf{x}\|_2 \le \frac{\delta}{4}$. Now, let $\gamma \ge 0$ be the smallest number such that $\|\mathbf{A}\mathbf{x}\|_2 \le (1+\gamma)\|\mathbf{x}\|_2$ for all $\mathbf{x} \in \mathcal{B}_\mathcal{T}$. Assume that (9) holds. By the triangle inequality,

$$\begin{aligned}
\|\mathbf{A}\mathbf{x}\|_2 &\le \|\mathbf{A}\mathbf{q}_\mathbf{x}\|_2 + \|\mathbf{A}(\mathbf{q}_\mathbf{x} - \mathbf{x})\|_2 && (10)\\
&\le 1 + \frac{\delta}{2} + (1+\gamma)\,\frac{\delta}{4}. && (11)
\end{aligned}$$

By assumption, $\gamma \le \delta/2 + (1+\gamma)\,\delta/4$, $\gamma\,(1-\delta/4) \le 3\delta/4$, and $\gamma \le \frac{3\delta/4}{1-\delta/4} \le \delta$. Plugging this back into (11) we get $\|\mathbf{A}\mathbf{x}\|_2 \le 1 + \delta/2 + (1+\delta)\delta/4 \le 1+\delta$ for all $\mathbf{x} \in \mathcal{B}_\mathcal{T}$. Similarly, by the reverse triangle inequality, we get that $\|\mathbf{A}\mathbf{x}\|_2 \ge \|\mathbf{A}\mathbf{q}_\mathbf{x}\|_2 - \|\mathbf{A}(\mathbf{x}-\mathbf{q}_\mathbf{x})\|_2 \ge 1 - \delta/2 - (1+\delta)\delta/4 \ge 1-\delta$.

Now that we have established that RIP holds for a single pattern $\mathcal{T}$, we can apply the union bound to show that it holds over all S-sparse vectors. There are $\binom{N}{S}$ patterns with support up to the sparsity level S. Recall that $\binom{N}{S} \le \left(\frac{eN}{S}\right)^S$. The probability that RIP does not hold for at least one pattern with support of size S (or less) is upper bounded as follows:

$$\begin{aligned}
&\left(\frac{eN}{S}\right)^S \left(\frac{12}{\delta}\right)^S \exp(-c_0(\delta/2)K^\rho S^\eta) \\
&\quad \le \exp\left(-c_0(\delta/2)K^\rho S^\eta + S\ln\left(\frac{eN}{S}\right) + S\ln\left(\frac{12}{\delta}\right)\right) \\
&\quad \le \exp(-c_2(\delta)K^\rho S^\eta) \quad \text{if } K^\rho \ge c_1(\delta)S^{1-\eta}\log\left(\frac{N}{S}\right)
\end{aligned}$$

B. Proof of Lemma 7

Each term in the sum u_k^2 can be written as the sum of the form

$$\begin{aligned} u_k^2 &= \left(\sum_{1 \leq i_1 < i_2 < \cdots < i_D \leq M} a_{i_1} a_{i_2} \cdots a_{i_D} \; x_{i_1 i_2 \cdots i_D} \right)^2 , \\ &= \sum_{1 \leq i_1 < i_2 < \cdots < i_D \leq M} a_{i_1}^2 a_{i_2}^2 \cdots a_{i_D}^2 \; x_{i_1 i_2 \cdots i_D}^2 \; + \; v_k \; , \\ &= \|x\|_2^2 + v_k \end{aligned}$$

where the v_k are the cross-terms of the sum and we have used the fact that $a_i^2 = 1$, $i = 1, \ldots, M$. Recall that we are assuming (wlog) that $\|x\|_2^2 = 1$. Therefore, $\mathbb{P}(|\frac{1}{K}\sum_{k=1}^K u_k^2 - 1| > t) = \mathbb{P}(|\frac{1}{K}\sum_{k=1}^K v_k| > t)$.

The v_k are i.i.d. and symmetrically distributed. Lemma 6 implies that $\mathbb{P}\left(|v_k| > t\right) \leq \exp\left(-c_2 t^{1/\alpha}\right)$, where $1 \leq \alpha \leq D$ is the combinatorial dimension of the chaos variables $\{u_k\}$. The moments of sums of i.i.d. symmetric random variables can be bounded using Corollary 2 in [20] which states that for $r > 2$

$$\begin{aligned} &\left(\mathbb{E}\left[\left(\sum_{k=1}^K v_k \right)^r \right] \right)^{1/r} \\ &\quad \leq \sup \left\{ \frac{r}{s} \left(\frac{K}{r} \right)^{1/s} \left(\mathbb{E}\,|v_k|^s \right)^{1/s} : \max\left(2, \frac{r}{K}\right) \leq s \leq r \right\} \\ &\quad \leq C \left(\sqrt{Kr} + K^{1/r} \left(\mathbb{E}\,|v_k|^r \right)^{1/r} \right) . \end{aligned}$$

Note that $\mathbb{E}|v_k|^r \leq \int_0^\infty e^{-ct^{1/(\alpha r)}} dt = c^{-\alpha r}\Gamma(1 + \alpha r)$, where Γ is the gamma function. It follows from Stirling's approximation that $\Gamma(1 + \alpha r)^{1/r} \leq C r^\alpha$, and so we have

$$\begin{aligned} \left(\mathbb{E}\left[\left(\sum_{k=1}^K v_k \right)^r \right] \right)^{1/r} &\leq C'(\sqrt{Kr} + K^{1/r} r^\alpha) \\ &\leq C(\sqrt{Kr} + r^\alpha) \; . \end{aligned}$$

The last inequality above follows by showing that for any $r \geq 2$, $1 \leq \alpha < \infty$, and $K \geq 1$, we have $\sqrt{Kr} + K^{1/r} r^\alpha \leq 3\sqrt{Kr} + e^{4\alpha} r^\alpha$. This is established by considering three cases:

1) If $2 \leq r \leq 4$, then $K^{1/r} r^\alpha \leq K^{1/2} r^\alpha \leq 4^{\alpha - 1/2}\sqrt{Kr}$.
2) If $K^{1/r} r^\alpha \leq \sqrt{Kr}$, then the claim is trivial.
3) If $r \geq 4$ and $K^{1/r} r^\alpha \geq \sqrt{Kr}$, then $K^{1/2 - 1/r} \leq r^{\alpha - 1/2} \leq r^\alpha$, so $K^{1/r} \leq r^{2\alpha/(r-2)} \leq r^{4\alpha/r} \leq e^{4\alpha}$.

So we have a bound of the form

$$\left(\mathbb{E}\left[\left(\sum_{k=1}^K v_k \right)^r \right] \right)^{1/r} \leq C \max\left\{ \sqrt{Kr}, r^\alpha \right\} \; . \qquad (12)$$

Markov's inequality yields

$$\mathbb{P}\left(\left| \sum_{k=1}^K v_k \right| \geq t \right) \leq t^{-r} \left(C \max\left\{ \sqrt{Kr}, r^\alpha \right\} \right)^r .$$

In both cases the upper bound has the form $C(t^{-1}K^\gamma r^\zeta)^r$, for appropriate γ and ζ. Taking $r = (tK^{-\gamma}e^{-1})^{1/\zeta}$ in each case we obtain the following bound for large t (so that $r > 2$):

$$\mathbb{P}\left(\left| \sum_{k=1}^K v_k \right| \geq t \right) \leq C' \exp\left(-c \min\left\{ t^2/K, t^{1/\alpha} \right\} \right) .$$

Substituting Kt for t, we get that

$$\mathbb{P}\left(\left| \frac{1}{K} \sum_{k=1}^K v_k \right| \geq t \right) \leq C' \exp\left(-c \min\left\{ Kt^2, K^{1/\alpha} t^{1/\alpha} \right\} \right) .$$

REFERENCES

[1] L. Hao, A. Sakurai, T. Watanabe, E. Sorenson, and C. A. Nidom *et al.*, "*Drosophila* RNAi screen identifies host genes important for influenza virus replication," *Nature*, vol. 454, pp. 890–893, August 2008. doi:10.1038/nature07151.

[2] F. D. Bushman, N. Malani, J. Fernandes, I. D'Orso, and G. Cagney *et al.*, "Host cell factors in HIV replication: Meta-analysis of genome-wide studies," *PLoS Pathogens*, vol. 5, p. e1000437, May 2009. doi:10.1371/journal.ppat.1000437.

[3] M. Costanzo, A. Baryshnikova, J. Bellay, Y. Kim, and E. D. Spear *et al.*, "The genetic landscape of a cell," *Science*, vol. 327, pp. 425–431, January 2010. doi:10.1126/science.1180823.

[4] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Info. Theory*, vol. 52, pp. 489–509, February 2006.

[5] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies," *IEEE Trans. Info. Theory*, vol. 52, pp. 5406–5425, December 2006.

[6] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, pp. 1289–1306, April 2006.

[7] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property of random matrices," *Constructive Approximation*, vol. 28, pp. 253–263, 2008.

[8] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Info. Theory*, Submitted August 2008.

[9] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," tech. rep., University of Michigan, August 2010.

[10] W. Dai, O. Milenkovic, M. Sheikh, and R. Baraniuk, "Probe design for compressive sensing DNA microarrays," *IEEE Intl. Conf. on Bioinformatics and Biomedicine*, pp. 163–169, 2008.

[11] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays," *IEEE Journal of STSP*, vol. 2, pp. 275–285, June 2008.

[12] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Info. Theory*, vol. 52, pp. 4036–4048, Sept. 2006.

[13] E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, pp. 2313–2351, Dec. 2007.

[14] R. Blei, *Analysis in Integer and Fractional Dimensions*. Cambridge, UK: Cambridge University Press, 2001.

[15] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. Paris, Ser. I*, vol. 346, pp. 589–592, May 2008.

[16] D. Achlioptas, "Database-friendly random projections: Johnson-Lindenstrauss with binary coins," *J. of Comp. and Sys. Sci.*, vol. 66, pp. 671–687, 2003.

[17] M. Rudelson and R. Vershynin, "On sparse reconstruction from fourier and gaussian measurements," *Comm. Pure Appl. Math.*, vol. 61, pp. 1025–1045, 2008.

[18] R. A. Brualdi and S. Mellendorf, "Regions in the complex plane containing the eigenvalues of a matrix," *The Am. Math. Monthly*, vol. 101, pp. 975–985, December 1994.

[19] R. Blei and S. Janson, "Rademacher chaos: tail estimates versus limit theorems," *Arkiv för Matematik*, vol. 42, pp. 13–29, April 2004.

[20] R. Latala, "Estimation of moments of sums of independent real random variables," *The Annals of Prob.*, vol. 25, no. 3, pp. 1502–1513, 1997.