Learning Trigonometric Polynomials from Random Samples and Exponential Inequalities for Eigenvalues of Random Matrices

Karlheinz Gröchenig[*], Benedikt M. Pötscher[†], Holger Rauhut[*]

January 26, 2007

Abstract

Motivated by problems arising in random sampling of trigonometric polynomials, we derive exponential inequalities for the operator norm of the difference between the sample second moment matrix $n^{-1}U^*U$ and its expectation where U is a complex random $n \times D$ matrix with independent rows. These results immediately imply deviation inequalities for the largest (smallest) eigenvalues of the sample second moment matrix, which in turn lead to results on the condition number of the sample second moment matrix. We also show that trigonometric polynomials in several variables can be learned from $const \cdot D \ln D$ random samples.

Keywords: eigenvalues; exponential inequality; learning theory; random matrix; random sampling; trigonometric polynomial.

1 Introduction

Let U be a complex random $n \times D$ matrix with independent rows. The matrix of (non-centered) sample second moments is then given by $n^{-1}U^*U$. We provide exponential probability inequalities for the operator norm of the difference between the sample second moment matrix and its expectation. These results immediately imply deviation inequalities for the largest (smallest) eigenvalues of the sample second moment matrix. As a consequence we obtain probability inequalities for the condition number of the sample second moment matrix. Sample second moment matrices arise

---

as central objects of interest in many areas, such as multivariate analysis, stochastic linear regression, time series analysis, and learning theory.

Our motivation comes from learning theory and, in particular, from random sampling of trigonometric polynomials. Random sampling is a strategy of choice for learning an unknown function in a given class of functions. This idea is predominant in the version of learning theory and sampling theory by Cucker, Smale, and Zhou [7, 15]. In Bass and Gröchenig [1] the randomization of the samples was used for the justification of numerical algorithms. Random sampling and random measurements are central in the emerging field of sparse reconstruction, also referred to as compressed sensing [3, 4, 6, 8, 10, 11, 13].

In this paper, we revisit the random sampling of trigonometric polynomials with a given degree or support, which was first studied in [1]. We first review and supplement the probability inequalities for the condition number of the associated Fourier sample second moment matrix in [1] (Section 2). In Section 3 we replace Fourier matrices by general random matrices with independent rows and derive probability estimates for sample second moment matrix obtained from general random matrices U. Our main result is an exponential probability inequality for the condition number of the sample second moment matrix for a vast class of random matrices. These include random matrices with independent identically distributed (i.i.d.) rows and bounded entries. The boundedness assumption on the entries can be relaxed to the existence of finite moment generating functions. Our proof is much simpler than the one in [1] and allows us to incorporate the case of random matrices with independent, but not necessarily identically distributed rows. This rather technical extension is treated in Section 3.2.

A further feature of our results is that all constants are given explicitly as a function of the parameters that describe the distribution of the random matrices. The explicit form of the constants is important to determine the sample size for which the condition number of the sample second moment matrix is small with high ("overwhelming") probability.

Mendelson and Pajor [9] have recently found a related, beautiful and deep inequality for the sample second moment matrix of random matrices. While in the same spirit, their assumptions, methods, and conclusions are different, and so their and our results are not directly comparable in that neither result implies the other. The inequality of [9] involves an unspecified absolute constant so that the main contribution is an asymptotic bound for the condition number as the sample size n tends to infinity. A detailed comparison between [9] and our result will be given in Section 3.1.1.

In learning theory one is often interested in the efficiency of the sampling procedure, i.e., in Cucker and Smale's words [7], "how many random samples do we need to assert, with confidence $1 - \delta$, that the condition number does not exceed a given threshold." For random sampling of trigonometric polynomials in several variables inspection shows that the probability inequalities in [1], as well as the ones in Section 3 of the present paper, lead to lower bounds for the required sample size

that are typically of the order $D^2 \ln D$. We show in Section 3.1.1 how the result in [9] can be used to improve this order to $D \ln D$. In Section 4 this result is further improved by using the method developed in [10] (after inspiration from [3]). To put it more casually, these results show that we need $const \cdot D \ln D$ random samples to learn a trigonometric polynomial taken from a D-dimensional space. This seems to be the optimal order that can be expected in a probabilistic setting.

Notation. By $\| \cdot \|_2$ we denote the usual Euclidean norm on $\mathbb{C}^D$. For a (hermitian) matrix A we denote by $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ the maximal and minimal eigenvalues of A. The condition number of A is then given by $\lambda_{\max}(A)/\lambda_{\min}(A)$. For a matrix A its transpose is denote by A' and its conjugate-transpose by A^*. The operator norm of a matrix is $\|A\| = \lambda_{\max}(A^*A)^{1/2}$. By $\mathbb{P}$ we denote the probability measure on the probability space supporting all the random variables used subsequently, and $\mathbb{E}$ denotes the corresponding expectation operator.

2 Random Sampling of Trigonometric Polynomials

Let Γ be a (non-empty) finite subset of $\mathbb{Z}^d$. By $\mathcal{P}_\Gamma$ we denote the space of all trigonometric polynomials in dimension d with coefficients supported on Γ. Such a polynomial has the form

$$f(x) = \sum_{k \in \Gamma} a_k e^{2\pi i k \cdot x}, \quad x \in [0,1]^d$$

with coefficients $a_k \in \mathbb{C}$. If $\Gamma = \{-m, -m+1, \ldots, m-1, m\}^d$, then $\mathcal{P}_\Gamma$ is the space of all trigonometric polynomials of degree at most m. We let $D = |\Gamma|$ be the dimension of $\mathcal{P}_\Gamma$.

Let $x_1, \ldots, x_n \in [0,1]^d$. We are interested in the reconstruction of a trigonometric polynomial f from its sample values $f(x_1), \ldots, f(x_n)$. Let $y = (f(x_1), \ldots, f(x_n))'$ be the vector of sampled values of f and let U be the $n \times D$ matrix with entries

$$u_{tk} = e^{2\pi i k \cdot x_t}, \quad k \in \Gamma, \ t = 1, \ldots, n. \tag{2.1}$$

The reconstruction of f amounts to solving the linear system

$$Ua = y$$

for the coefficient vector $a = (a_k)_{k \in \Gamma}$. Alternatively, one may try to solve the normal equation

$$U^*Ua = U^*y.$$

We note that the invertibility of U^*U is equivalent to the sampling inequalities

$$A\|f\|_2^2 \leq \sum_{t=1}^n |f(x_t)|^2 = a^*U^*Ua \leq B\|f\|_2^2 \qquad \text{for all } f \in \mathcal{P}_\Gamma, \tag{2.2}$$

3

for some positive real numbers A and B, and that the condition number of U^*U is bounded by B/A.

In the spirit of learning theory, one assumes that the sampling points are taken at random. Then the matrix U^*U is a random matrix depending on the sampling points (x_t). Several questions arise:

1. Determine the probability that U^*U is invertible.

2. Determine the probability that the condition number of U^*U does not exceed a given threshold.

3. Determine the number of random samples required to achieve such estimates. This is the effectivity problem for random sampling.

For trigonometric polynomials Question 1 has been answered in [1, Thm. 1.1]: If the x_t are i.i.d. with a distribution that is absolutely continuous with respect to the Lebesgue-measure on $[0,1]^d$, then U^*U is invertible almost surely provided $n \geq D$.

Furthermore, some answers to Questions 2 and 3 are also provided in [1]. It is shown that U^*U is well-conditioned whenever the number of samples is large enough [1, Thms. 5.1, 6.2]:

Theorem 2.1. *Assume that $x_1, \ldots, x_n$ are i.i.d. random variables uniformly distributed on $[0,1]^d$. Let U be the associated random Fourier matrix defined in (2.1). Let $\varepsilon \in (0,1)$. There exist positive constants A, B depending only on $D = |\Gamma|$ such that the event*

$$1 - \varepsilon \leq \lambda_{\min}(n^{-1}U^*U) \leq \lambda_{\max}(n^{-1}U^*U) \leq 1 + \varepsilon$$

has probability at least

$$1 - Ae^{-Bn\varepsilon^2/(1+\varepsilon)}. \tag{2.3}$$

In particular, with probability not less than (2.3), the condition number of U^*U is bounded by $(1+\varepsilon)/(1-\varepsilon)$.

A careful analysis of the constants A and B in (2.3) reveals that the number of samples required in (2.3) to guarantee a probability $\geq 1 - \delta$ is

$$n \geq CD^2 \ln D, \tag{2.4}$$

where C depends on δ and ε. If $\Gamma = \{-m, \ldots, m\}^d$ (trigonometric polynomials of degree m in d variables), then $D = (2m+1)^d$, and this bound on the number of samples is unfortunately too large to be useful at more realistic sample sizes like $n \sim D$ or $n \sim D \ln D$.

For the case $\Gamma = \{-m, \ldots, m\}^d$ a better estimate for the condition number can be extracted from [1]. We work, however, with a slightly different matrix. Given the sampling points $x_1, \ldots, x_n$, we define the Voronoi regions

$$V_t := \{y \in [0,1]^d : \|y - x_t\|_2 \leq \|y - x_s\|_2, s \neq t, 1 \leq s \leq n\}, \quad t = 1, \ldots, n$$

4

and let $w_t = |V_t|$ be the Lebesgue measure of V_t. We consider the weighted matrix T^w

$$T^w := U^*WU$$

where W is the diagonal matrix with the weights w_t, $t = 1, \ldots, n$, on the diagonal. Note that a is also the solution of $T^w a = U^*Wy$. The following result is implicit in [1].

Theorem 2.2. *Let $\Gamma = \{-m, \ldots, m\}^d$, i.e., we consider trigonometric polynomials of d variables of degree m. Suppose that $x_1, \ldots, x_n$ are i.i.d random variables which are uniformly distributed on $[0,1]^d$. Choose $\gamma \in (0,1)$. If*

$$n \geq \left(\frac{2\pi d}{\gamma \ln 2}\right)^d m^d \ln \left(\left(\frac{2\pi d}{\gamma \ln 2}\right)^d \frac{m^d}{\delta}\right), \qquad (2.5)$$

then with probability at least $1-\delta$ the condition number of T^w is bounded by $(1 - 2^{\gamma-1})^{-2}$.

<u>Proof:</u> By combining a deterministic estimate with a probabilistic covering result, the following estimate was derived in [1, Thm. 4.2]: Let $N \in \mathbb{N}$ be arbitrary; then with probability at least $1 - N^d e^{-n/N^d}$ we have

$$(2 - e^{2\pi md/N})^2 \leq \lambda_{\min}(T^w) \leq \lambda_{\max}(T^w) \leq 4.$$

For the condition number to be bounded by $4 \left(2 - 2^\gamma\right)^{-2}$ with probability at least $1 - \delta$, we need that

$$2\pi md/N \leq \gamma \ln 2 \qquad \text{and} \qquad N^d e^{-n/N^d} \leq \delta.$$

By solving for n, we find that n must satisfy the inequality (2.5). ∎

Since $D = |\Gamma| = |\{-m, -m+1, \ldots, m-1, m\}^d| = (2m+1)^d$, Theorem 2.2 becomes effective for

$$n \approx (\pi d)^d D \ln \left((\pi d)^d D\right). \qquad (2.6)$$

Thus Theorem 2.2 is a genuine improvement over (2.4) for fixed value of d. The dependence $n \sim D \ln D$ on the dimension of the function space seems to be of the correct order. However, the constant $(\pi d/\gamma)^d$ depends strongly on the number of variables d, and so Theorem 2.2 does not escape the curse of dimensionality.

In Theorem 4.1 we will prove a much better result for the condition number of U^*U where the constants do neither depend on d nor on the special form of the spectrum Γ. See also Corollary 3.5.

3 Exponential Inequalities For Sample Second Moment Matrices

In this section we abstract from the concrete form of U as given in (2.1) and consider arbitrary complex random matrices with independent rows satisfying some regularity

conditions. Apart from being of interest in its own, this more general setting allows one to study random sampling not only for trigonometric polynomials but also for more general types of finite-dimensional function spaces, such as random sampling of algebraic polynomials on domains, or of spaces of spherical harmonics on the sphere (see [1, Sect. 6] for a list of examples).

3.1 The I.I.D. Case

We assume first that the random matrix $U \in \mathbb{C}^{n \times D}$ has independent identically distributed rows and delay the discussion of the case of independent, but not identically distributed rows to Section 3.2. Furthermore, we assume that the rows $u_{t\cdot} = (u_{t1}, \ldots, u_{tD})$ of U satisfy the following condition: The moment generating functions of the random variables $\mathrm{Re}(\overline{u_{1k}}u_{1j})$ and $\mathrm{Im}(\overline{u_{1k}}u_{1j})$ exist for all $1 \leq k, j \leq D$; i.e., there exists $x_0 > 0$ such that for all $1 \leq k, j \leq D$

$$\mathbb{E}\left[\exp(x\,\mathrm{Re}(\overline{u_{1k}}u_{1j}))\right] < \infty, \qquad \mathbb{E}\left[\exp(x\,\mathrm{Im}(\overline{u_{1k}}u_{1j}))\right] < \infty \qquad (3.1)$$

hold for all $x < x_0$. Note that a sufficient condition for (3.1) is that the moment generating function of $|u_{1k}|^2 + |u_{1j}|^2$ exists for all k, j. Further, we let

$$Q := \mathbb{E}(u_{1\cdot}^* u_{1\cdot}) \in \mathbb{C}^{D \times D}$$

with entries q_{kj}. We note that by the strong law of large numbers $n^{-1}U^*U$ converges to $Q = \mathbb{E}[n^{-1}U^*U]$ almost surely.

Assumption (3.1) is easily seen to be equivalent to the existence of finite constants $M \geq 0$ and $v_{kj} \geq 0$ such that for all $\ell \geq 2$

$$\mathbb{E}\left[|\,\mathrm{Re}(\overline{u_{1k}}u_{1j} - q_{kj})|^\ell\right] \leq 2^{-1}\ell!\,M^{\ell-2}v_{kj}, \qquad (3.2)$$

$$\mathbb{E}\left[|\,\mathrm{Im}(\overline{u_{1k}}u_{1j} - q_{kj})|^\ell\right] \leq 2^{-1}\ell!\,M^{\ell-2}v_{kj} \qquad (3.3)$$

hold for all $1 \leq k, j \leq D$. For a generalization leading to a slightly better, but more complex bound see Section 3.2.

Remark 3.1. If the random variables u_{1k} are bounded, i.e.,

$$|\,\mathrm{Re}(\overline{u_{1k}}u_{1j} - q_{kj})| \leq C \quad \text{and} \quad |\,\mathrm{Im}(\overline{u_{1k}}u_{1j} - q_{kj})| \leq C$$

holds with probability 1 for all $1 \leq k, j \leq D$, then (3.2) and (3.3) hold with $M = C/3$ and

$$v_{kj} = \max\{\mathbb{E}\left[(\mathrm{Re}(\overline{u_{1k}}u_{1j} - q_{kj}))^2\right], \mathbb{E}\left[(\mathrm{Im}(\overline{u_{1k}}u_{1j} - q_{kj}))^2\right]\}. \qquad (3.4)$$

This claim is obvious for $\ell = 2$. For $\ell \geq 3$ it follows from a general inequality for arbitrary real-valued bounded random variables X:

$$\mathbb{E}\left[|X|^\ell\right] = \mathbb{E}\left[X^2|X|^{\ell-2}\right] \leq C^{\ell-2}\mathbb{E}[X^2] = \ell!\,\frac{C^{\ell-2}}{\ell!}\sigma^2 \leq \ell!\,\frac{C^{\ell-2}}{2 \cdot 3^{\ell-2}}\sigma^2,$$

where $\sigma^2 = \mathbb{E}[X^2]$ and $|X| \leq C$ holds with probability 1. In particular, this shows that the random Fourier matrix given in (2.1) satisfies (3.2) and (3.3). $\qquad \square$

The proof of the main result in this section will make use of the following Bernstein-type inequality for unbounded random variables given in Bennett [2, eq. (7)], see also [16, Lemma 2.2.11]:

Let $X_1, \ldots, X_n$ be independent real-valued random variables with zero mean such that $E|X_t|^\ell \leq \ell! M^{\ell-2} v_t / 2$ holds for every $\ell \geq 2$ and $t = 1, \ldots, n$ for some finite constants $M \geq 0$ and $v_t \geq 0$. Then for every $x > 0$

$$\mathbb{P}\left(\left| \sum_{t=1}^{n} X_t \right| \geq x \right) \leq 2 e^{-\frac{x^2}{2} \left(\sum_{t=1}^{n} v_t + Mx \right)^{-1}}, \tag{3.5}$$

with the convention that the right-hand side in (3.5) is zero if $M = 0$ and $\sum_{t=1}^{n} v_t = 0$.

Note that Bennett [2] assumes $\sum_{t=1}^{n} v_t > 0$ but the inequality (3.5) trivially also holds for $\sum_{t=1}^{n} v_t = 0$ in which case the probability on the left-hand side is zero. Inequality (3.5), and hence the subsequent results, can be somewhat improved, see Bennett [2, eq. (7a)]. Since this does not result in any significant gain, we do not give the details.

Set

$$v := \max_{1 \leq k, j \leq D} v_{kj}.$$

Note that neither v nor M depend on n because the rows are identically distributed. However, they depend on the distribution of the random vector u_1. and hence may depend on D. Our main result now reads as follows.

Theorem 3.1. *Assume that the rows $u_1, \ldots, u_n$. of U are i.i.d. random vectors in $\mathbb{C}^D$ whose entries satisfy the moment bounds (3.2) and (3.3). Then, for every $\varepsilon > 0$, the operator norm satisfies*

$$\left\| n^{-1} U^* U - Q \right\| < \varepsilon$$

with probability at least

$$1 - 4D^2 \exp\left(-\frac{n\varepsilon^2}{D^2 \left(4v + 2\sqrt{2} D^{-1} M \varepsilon \right)} \right). \tag{3.6}$$

In particular, with probability not less than (3.6) the extremal eigenvalues of $n^{-1} U^* U$ satisfy

$$\lambda_{\min}(Q) - \varepsilon < \lambda_{\min}(n^{-1} U^* U) \leq \lambda_{\max}(n^{-1} U^* U) < \lambda_{\max}(Q) + \varepsilon. \tag{3.7}$$

Consequently, if Q is non-singular and $\varepsilon \in (0, \lambda_{\min}(Q))$, then the condition number of $U^* U$ is bounded by $\frac{\lambda_{\max}(Q) + \varepsilon}{\lambda_{\min}(Q) - \varepsilon}$ with probability not less than (3.6).

In connection with (3.7) we note that $\lambda_{\min}(n^{-1} U^* U) \geq 0$ holds trivially, since the matrix $n^{-1} U^* U$ is nonnegative definite.

Proof: We first note that inequality (3.7) for the extremal eigenvalues of $n^{-1} U^* U$ follows from the inequality $\| n^{-1} U^* U - Q \| < \varepsilon$ for the operator norm. Hence, it

suffices to concentrate on the operator norm, which we majorize with Schur's test by using that $\|A\| \leq \max_k \sum_j |a_{kj}|$ for self-adjoint A. In this way we obtain that

$$\mathbb{P}(\|n^{-1}U^*U - Q\| \geq \varepsilon) \leq \mathbb{P}\left(\max_{k=1,\ldots,D} \sum_{j=1}^{D} \left| n^{-1}\sum_{t=1}^{n}(\overline{u_{tk}}u_{tj} - q_{kj}) \right| \geq \varepsilon \right)$$

$$\leq \sum_{k=1}^{D} \mathbb{P}\left(\sum_{j=1}^{D} \left| n^{-1}\sum_{t=1}^{n}(\overline{u_{tk}}u_{tj} - q_{kj}) \right| \geq \varepsilon \right) \leq \sum_{k,j=1}^{D} \mathbb{P}\left(\left| n^{-1}\sum_{t=1}^{n}(\overline{u_{tk}}u_{tj} - q_{kj}) \right| \geq \varepsilon/D \right)$$

$$= \sum_{k,j=1}^{D} \mathbb{P}\left(\left| n^{-1}\sum_{t=1}^{n}(\overline{u_{tk}}u_{tj} - q_{kj}) \right|^2 \geq (\varepsilon/D)^2 \right) \tag{3.8}$$

$$= \sum_{k,j=1}^{D} \mathbb{P}\left(\left| n^{-1}\sum_{t=1}^{n}\mathrm{Re}(\overline{u_{tk}}u_{tj} - q_{kj}) \right|^2 + \left| n^{-1}\sum_{t=1}^{n}\mathrm{Im}(\overline{u_{tk}}u_{tj} - q_{kj}) \right|^2 \geq (\varepsilon/D)^2 \right)$$

$$\leq \sum_{k,j=1}^{D} \mathbb{P}\left(\left| \sum_{t=1}^{n}\mathrm{Re}(\overline{u_{tk}}u_{tj} - q_{kj}) \right| \geq \frac{n\varepsilon}{\sqrt{2}D} \right) + \sum_{k,j=1}^{D} \mathbb{P}\left(\left| \sum_{t=1}^{n}\mathrm{Im}(\overline{u_{tk}}u_{tj} - q_{kj}) \right| \geq \frac{n\varepsilon}{\sqrt{2}D} \right).$$

For each index k, j the inequality (3.5) gives

$$\mathbb{P}\left(\left| \sum_{t=1}^{n}\mathrm{Re}(\overline{u_{tk}}u_{tj} - q_{kj}) \right| \geq \frac{n\varepsilon}{\sqrt{2}D} \right) \leq 2\exp\left(-\frac{n\varepsilon^2}{D^2\left(4v_{kj} + 2\sqrt{2}D^{-1}M\varepsilon\right)} \right) \tag{3.9}$$

and similarly for the imaginary part. Hence, we finally obtain

$$\mathbb{P}\left(\|n^{-1}U^*U - Q\| \geq \varepsilon\right) \leq 4\sum_{k,j=1}^{D}\exp\left(-\frac{n\varepsilon^2}{D^2\left(4v_{kj} + 2\sqrt{2}D^{-1}M\varepsilon\right)} \right)$$

$$\leq 4D^2\exp\left(-\frac{n\varepsilon^2}{D^2\left(4v + 2\sqrt{2}D^{-1}M\varepsilon\right)} \right) \tag{3.10}$$

with v as defined above. Thus (3.6) follows. ∎

Remark 3.2. If $u_{1.}$ possesses an absolutely continuous distribution then Q is automatically non-singular. More generally, this holds as long as the distribution of $u_{1.}$ is not concentrated on a $(D-1)$-dimensional linear subspace of $\mathbb{C}^D$. To see this, consider the quadratic forms z^*Qz for $z \in \mathbb{C}^D$ and note that

$$z^*Qz = \mathbb{E}[z^*u_{1.}^*u_{1.}z] = \mathbb{E}[|u_{1.}z|^2] \geq 0.$$

Hence, if $z^*Qz = 0$, then $|u_{1.}z|^2 = 0$ with probability 1; thus the distribution of $u_{1.}$ would have to reside in the orthogonal complement of the one-dimensional subspace spanned by z^*. □

Remark 3.3. For real-valued random matrices U we can improve the probability bound (3.6) to

$$1 - 2D^2 \exp\left(-\frac{n\varepsilon^2}{2D^2\left(v + \frac{M\varepsilon}{D}\right)}\right).$$

A similar improvement for real-valued U applies to the subsequent corollary and remark as well as to the results in Section 3.2. □

Corollary 3.2. *Assume that the rows* $u_{1.}, \ldots, u_{n.}$ *of* U *are i.i.d. random vectors in* $\mathbb{C}^D$ *that are bounded, i.e.,*

$$|\operatorname{Re}(\overline{u_{1k}}u_{1j} - q_{kj})| \leq C \quad and \quad |\operatorname{Im}(\overline{u_{1k}}u_{1j} - q_{kj})| \leq C$$

holds with probability 1 for or all $1 \leq k, j \leq D$. *Let*

$$b := \max_{k,j=1,\ldots,D} \left\{\mathbb{E}\left[(\operatorname{Re}(\overline{u_{1k}}u_{1j} - q_{kj}))^2\right], \mathbb{E}\left[(\operatorname{Im}(\overline{u_{1k}}u_{1j} - q_{kj}))^2\right]\right\}.$$

Then the conclusions of Theorem 3.1 hold and (3.6) becomes

$$1 - 4D^2 \exp\left(-\frac{n\varepsilon^2}{D^2\left(4b + 2\sqrt{2}D^{-1}C\varepsilon/3\right)}\right). \tag{3.11}$$

Proof: By Remark 3.1 conditions (3.2) and (3.3) hold with $M = C/3$ and v_{kj} as in (3.4). Then the statement follows from Theorem 3.1. ∎

Remark 3.4. Corollary 3.2 can also be derived by using the classical Bernstein inequality instead of inequality (3.5) in the proof of Theorem 3.1. Furthermore, the bound in (3.11) can be somewhat improved by using an improved form of Bernstein's inequality [2, eq. (8)] (see also [5, Corollary A.2]) for bounded random variables instead of (3.5) in that step: If we use that inequality in the estimate (3.9), we arrive at the following improved bound (provided $C > 0$, $b > 0$):

$$1 - 4D^2 \exp\left(-C^{-2}nb\left(\left(1 + \frac{C\varepsilon}{\sqrt{2}Db}\right)\ln\left(1 + \frac{C\varepsilon}{\sqrt{2}Db}\right) - \frac{C\varepsilon}{\sqrt{2}Db}\right)\right).$$

□

Let us now apply our findings to random sampling of trigonometric polynomials.

Corollary 3.3. *Let* $x_1, \ldots, x_n$ *be independent random variables uniformly distributed on* $[0,1]^d$. *Let* U *be the associated* $n \times D$ *random Fourier matrix (2.1). Let* $\varepsilon > 0$. *Then with probability at least*

$$1 - 4D(D - 1) \exp\left(-\frac{n\varepsilon^2}{2\left((D-1)^2 + \sqrt{2}(D-1)\varepsilon/3\right)}\right) \tag{3.12}$$

we have

$$\left\|n^{-1}U^*U - Q\right\| < \varepsilon, \tag{3.13}$$

9

and hence

$$1 - \varepsilon < \lambda_{\min}(n^{-1}U^*U) \le \lambda_{\max}(n^{-1}U^*U) < 1 + \varepsilon. \qquad (3.14)$$

Consequently, for $0 < \varepsilon < 1$, the condition number of U^*U is bounded by $(1+\varepsilon)/(1-\varepsilon)$ with probability not less than (3.12).

Proof: In this case $(n^{-1}U^*U)_{kj} = n^{-1}\sum_{t=1}^{n} e^{2\pi i(j-k)\cdot x_t}$ and consequently $Q = I$, so $\lambda_{\min}(Q) = \lambda_{\max}(Q) = 1$. [By abuse of notation, k denotes both an element of Γ and a column index.] Furthermore, $\sum_{t=1}^{n}(\overline{u_{tk}}u_{tj} - q_{kj}) = 0$ for $k = j$. Hence, the double sum in the second line of (3.8) only extends over $j \ne k$ and consequently ε/D can be replaced by $\varepsilon/(D-1)$ in the subsequent steps in (3.8). Furthermore, when deducing (3.10) from the union bound in (3.8) we only have to take into account $D(D-1)$ instead of D^2 summands; cf. also Remark 3.8 below. Moreover, $|\operatorname{Re}(\overline{u_{tk}}u_{tj} - q_{kj})| \le 1$ for all k, j. For $k \ne j$ we have

$$\mathbb{E}\left[\operatorname{Re}(\overline{u_{tk}}u_{tj} - q_{kj})^2\right] = \int_{[0,1]^d}(\operatorname{Re}(\exp(2\pi i(j-k)\cdot x))^2 dx = \frac{1}{2},$$

hence $v_{kj} = 1/2$. The same holds for the imaginary part. In view of Remark 3.1 the result follows. $\blacksquare$

From the previous result it is easy to determine the minimal number of sampling points sufficient to provide a small condition number with high probability.

Corollary 3.4. *Let $x_1, \ldots, x_n$ be independent random variables uniformly distributed on $[0,1]^d$. Let U be the associated $n \times D$ random Fourier matrix (2.1). Let $0 < \varepsilon < 1, 0 < \delta < 1$ and suppose*

$$n \ge \frac{2}{\varepsilon^2}\left((D-1)^2 + \frac{\sqrt{2}(D-1)\varepsilon}{3}\right)\ln\left(\frac{4D(D-1)}{\delta}\right). \qquad (3.15)$$

Then (3.13) and (3.14) hold with probability at least $1 - \delta$.

We note that (3.15) is implied by the more compact inequality

$$n \ge \frac{CD^2 \ln(D/\delta)}{\varepsilon^2} \qquad (3.16)$$

for an appropriate constant C. We will improve on this result in Corollary 3.5 and in Section 4, see in particular (4.2).

3.1.1 Comparison With Other Results

Recently Mendelson and Pajor [9] provide a related exponential inequality for random matrices with i.i.d. real-valued rows. They assume the following properties:

(a) There exists $\rho > 0$ such that for every $\theta \in \mathbb{R}^D, \|\theta\|_2 = 1, (\mathbb{E}|\langle u_1, \theta\rangle|^4)^{1/4} \le \rho < \infty$.

(b) Set $Z = \|u_1.\|_2$, then $\|Z\|_{\psi_\alpha} < \infty$ for some $\alpha \geq 1$.

Here the Orlicz norm $\|\cdot\|_{\psi_\alpha}$ of a real-valued random variable Y with respect to $\psi_\alpha(x) = \exp(x^\alpha) - 1$ is defined as $\|Y\|_{\psi_\alpha} = \inf\{C > 0 : \mathbb{E}\psi_\alpha(|Y|/C) \leq 1\}$. Note that if $\alpha \geq 2$, condition (b) is stronger than our assumption (3.1), hence condition (b) implies (3.2)-(3.3).

Under conditions (a) and (b), Mendelson and Pajor [9, Theorem 2.1] show that there exists an absolute constant $c > 0$ such that for every $\varepsilon > 0$ the operator norm satisfies

$$\mathbb{P}(\|n^{-1}U^*U - Q\| < \varepsilon) \geq 1 - 2\exp\left(-\left(\frac{c\varepsilon}{\max\{B_n, A_n^2\}}\right)^{\alpha/(\alpha+2)}\right), \qquad (3.17)$$

where

$$A_n = \|Z\|_{\psi_\alpha} \frac{\sqrt{\ln(\min(D,n))}(\ln n)^{1/\alpha}}{\sqrt{n}}, \qquad B_n = \frac{\rho^2}{\sqrt{n}} + \lambda_{\max}^{1/2}(Q)A_n.$$

We have added a factor 2 on the right-hand side of (3.17) to correct a missing constant in [9, Theorem 2.1]. Since the constant c is not specified, the value of (3.17) is mainly for asymptotics as $n \to \infty$, whereas the results of Section 3 yield estimates with explicit constants for given n. Moreover, the probability estimate (3.17) is only subexponential. For fixed dimension D, the right-hand side of (3.17) is of the order

$$1 - 2\exp\left(-c_1 n^{\alpha/(2\alpha+4)}(\ln n)^{-1/(\alpha+2)}\right),$$

which is only subexponential, whereas the bound in (3.6) is exponential of the form

$$1 - c_2 \exp(-c_3 n).$$

Here c_1, c_2, and c_3 are constants that depend on D. We also note that the proof of Theorem 3.1 is quite simple and easily extends to the case of independent but not necessarily identically distributed rows as shown in the next section.

Nevertheless, the result of Mendelson and Pajor [9, Theorem 2.1] can be used to improve upon (3.15) in the special case where the set Γ is symmetric in the sense that $k \in \Gamma$ implies $-k \in \Gamma$.

Corollary 3.5. *Let $x_1, \ldots, x_n$ be independent random variables uniformly distributed on $[0,1]^d$. Let U be the associated $n \times D$ random Fourier matrix (2.1) and assume that Γ is symmetric. Let $0 < \varepsilon < 1, 0 < \delta < 1$ and suppose*

$$n \geq \max\left\{D, c^{-1}\varepsilon^{-1}\ln\left(\frac{2}{\delta}\right)D\ln D, \left[c^{-1}\varepsilon^{-1}\ln\left(\frac{2}{\delta}\right)\right]^2(\sqrt{D} + \sqrt{D\ln D})^2\right\} \quad (3.18)$$

where c is the absolute constant in (3.17). Then (3.13) and (3.14) hold with probability at least $1 - \delta$.

Proof: Although it is possible to adapt [9] to complex-valued random matrices, we will use the result as stated.

Since Γ is symmetric by assumption, we may write it as $\Delta \cup (-\Delta)$ with $\Delta \cap (-\Delta) \subseteq \{0\}$. Define the real $n \times D$ matrix W by $w_{t(-k)} = \sqrt{2}\cos(2\pi k \cdot x_t)$ and $w_{tk} = \sqrt{2}\sin(2\pi k \cdot x_t)$ for $k \in \Delta \backslash \{0\}$ and set $w_{t0} = 1$ if $0 \in \Delta$. Then clearly $U = WS$ where S is a unitary $D \times D$ matrix and consequently $\|n^{-1}U^*U - I\| = \|n^{-1}W^*W - I\|$. To apply (3.17) we note that $\|w_{1\cdot}\|_2 = D^{1/2}$ and hence $\|Z\|_{\psi_\alpha} = \|\|w_{1\cdot}\|_2\|_{\psi_\alpha} = D^{1/2}(\ln 2)^{-1/\alpha}$ for every $\alpha \geq 1$. Furthermore,

$$\sup_{\|\theta\|_2=1,\theta\in\mathbb{R}^D} (\mathbb{E}|\langle w_{1\cdot},\theta\rangle|^4)^{1/4} = \sup_{\|\theta\|_2=1,\theta\in\mathbb{R}^D} (\mathbb{E}|\langle u_{1\cdot}S^*,\theta\rangle|^4)^{1/4}$$

$$= \sup_{\|\theta\|_2=1,\theta\in\mathbb{R}^D} (\mathbb{E}|\langle u_{1\cdot},S\theta\rangle|^4)^{1/4}$$

$$\leq \sup_{\|\theta\|_2=1,\theta\in\mathbb{C}^D} (\mathbb{E}|\langle u_{1\cdot},\theta\rangle|^4)^{1/4}.$$

Now, since $\left|\sum_{k\in\Gamma}\bar{\theta}_k\exp(2\pi ik\cdot x_1)\right|^2 \leq \left(\sum_{k\in\Gamma}|\theta_k|\right)^2$ we obtain

$$\mathbb{E}|\langle u_{1\cdot},\theta\rangle|^4 = \mathbb{E}\left[\left|\sum_{k\in\Gamma}\bar{\theta}_k\exp(2\pi ik\cdot x_1)\right|^4\right]$$

$$\leq \left(\sum_{k\in\Gamma}|\theta_k|\right)^2\mathbb{E}\left[\left|\sum_{k\in\Gamma}\bar{\theta}_k\exp(2\pi ik\cdot x_1)\right|^2\right]$$

$$\leq D\|\theta\|_2^2 = D.$$

This shows that the rows $w_{t\cdot}$ satisfy condition (a) in [9, Theorem 2.1] with $\rho = D^{1/4}$. As a consequence, (3.17) applies to W, and hence to the Fourier matrix U, for every $\alpha \geq 1$. Since the left-hand side of (3.17) does not depend on α, we may let $\alpha \to \infty$ and obtain for $n \geq D$ the bound

$$\mathbb{P}(\|n^{-1}U^*U - Q\| < \varepsilon) \geq 1 - 2\exp\left(-c\varepsilon\min\{n^{1/2}/(\sqrt{D} + \sqrt{D\ln D}), n/(D\ln D)\}\right).$$

The probability is not less than $1 - \delta$ whenever condition (3.18) holds. ∎

Comparing (3.18) with (3.15), we have gained on the exponent of D. However, the quantity $\ln(\delta^{-1})$ now enters quadratically instead of linearly, and an unspecified constant appears in the lower bound for n.

3.2   The Non-I.I.D. Case and Other Generalizations

In this section we generalize the results to the case where the random matrix $U \in \mathbb{C}^{n\times D}$ has independent rows which, however, need not be identically distributed. In the course of this generalization we also obtain some slight improvements in the case

of i.i.d. rows discussed above. Apart from the assumption of independent rows, we assume that the matrix U satisfies the following condition: The moment generating functions of the random variables $\mathrm{Re}(\overline{u_{tk}}u_{tj})$ and $\mathrm{Im}(\overline{u_{tk}}u_{tj})$ exist for all $1 \le t \le n$ and $1 \le k, j \le D$; i.e., there exists $x_0 > 0$ such that for all $1 \le t \le n$ and $1 \le k, j \le D$

$$\mathbb{E}\left[\exp(x\,\mathrm{Re}(\overline{u_{tk}}u_{tj}))\right] < \infty, \qquad \mathbb{E}\left[\exp(x\,\mathrm{Im}(\overline{u_{tk}}u_{tj}))\right] < \infty \qquad (3.19)$$

holds for all $x < x_0$. Note that x_0 will depend on the distribution of U and thus may depend on n and D. Furthermore, we set

$$Q^{(t)} := \mathbb{E}(u_{t\cdot}^* u_{t\cdot}) \in \mathbb{C}^{D \times D}$$

with entries $q_{kj}^{(t)}$ and

$$Q_n := n^{-1} \sum_{t=1}^{n} Q^{(t)} = \mathbb{E}[n^{-1}U^*U] \in \mathbb{C}^{D \times D}. \qquad (3.20)$$

As in Section 3.1, assumption (3.19) is seen to be equivalent to the existence of finite constants $M_{kj1}^{(t)} \ge 0$, $M_{kj2}^{(t)} \ge 0$, $v_{kj1}^{(t)} \ge 0$, $v_{kj2}^{(t)} \ge 0$, such that for all $\ell \ge 2$

$$\mathbb{E}\left[|\,\mathrm{Re}(\overline{u_{tk}}u_{tj} - q_{kj}^{(t)})|^\ell\right] \le 2^{-1}\ell!\,(M_{kj1}^{(t)})^{\ell-2}v_{kj1}^{(t)}, \qquad (3.21)$$

$$\mathbb{E}\left[|\,\mathrm{Im}(\overline{u_{tk}}u_{tj} - q_{kj}^{(t)})|^\ell\right] \le 2^{-1}\ell!\,(M_{kj2}^{(t)})^{\ell-2}v_{kj2}^{(t)} \qquad (3.22)$$

hold for all $1 \le t \le n$ and $1 \le k, j \le D$. If $M_{kji}^{(t)}v_{kji}^{(t)} = 0$ then we may assume without loss of generality that $M_{kji}^{(t)} = v_{kji}^{(t)} = 0$.

For fixed n it is always possible to choose the constants on the right-hand side of (3.21) and (3.22) independent of t. However, for $n \to \infty$ the resulting conditions in Theorem 3.6 below would become unnecessarily restrictive in the non-identically distributed case. Furthermore, allowing the constants to depend on k, j and to be different in (3.21) and (3.22), provides some extra flexibility which results in an improved, albeit more complex bound even in the case of i.i.d. rows.

Remark 3.5. Condition (3.21) necessarily implies $v_{kj1}^{(t)} \ge \sigma_{kj1}^{(t)2}$, where $\sigma_{kj1}^{(t)2}$ denotes the variance of $\mathrm{Re}(\overline{u_{tk}}u_{tj} - q_{kj}^{(t)})$. Furthermore, observe that given condition (3.21) is satisfied, it is also always satisfied with $v_{kj1}^{(t)} = \sigma_{kj1}^{(t)2}$. [This is obvious if $\sigma_{kj1}^{(t)2} = 0$, and otherwise follows by replacing $M_{kj1}^{(t)}$ with $M_{kj1}^{(t)}v_{kj1}^{(t)}/\sigma_{kj1}^{(t)2}$, observing that $v_{kj1}^{(t)}/\sigma_{kj1}^{(t)2} \ge 1$ as noted before.] Similar comments apply to condition (3.22). $\qquad\square$

Remark 3.6. If the random variables u_{tk} are bounded, i.e.,

$$|\,\mathrm{Re}(\overline{u_{tk}}u_{tj} - q_{kj}^{(t)})| \le C_{kj1}^{(t)} \quad \text{and} \quad |\,\mathrm{Im}(\overline{u_{tk}}u_{tj} - q_{kj}^{(t)})| \le C_{kj2}^{(t)}$$

holds with probability 1 for all $1 \le t \le n$, $1 \le k, j \le D$, then (3.21) and (3.22) hold with $M_{kj1}^{(t)} = C_{kj1}^{(t)}/3$, $M_{kj2}^{(t)} = C_{kj2}^{(t)}/3$, and

$$v_{kj1}^{(t)} = \mathbb{E}\left[(\mathrm{Re}(\overline{u_{tk}}u_{tj} - q_{kj}^{(t)}))^2\right] \text{ and } v_{kj2}^{(t)} = \mathbb{E}\left[(\mathrm{Im}(\overline{u_{tk}}u_{tj} - q_{kj}^{(t)}))^2\right]. \qquad (3.23)$$

This follows exactly as in Remark 3.1. $\qquad\square$

In order to present the generalization of Theorem 3.1 we introduce

$$v_{kj1n} := \sum_{t=1}^{n} v_{kj1}^{(t)}, \quad v_{kj2n} := \sum_{t=1}^{n} v_{kj2}^{(t)}$$

and

$$M_{kj1n} = \max\{M_{kj1}^{(t)} : 1 \leq t \leq n\}, \quad M_{kj2n} = \max\{M_{kj2}^{(t)} : 1 \leq t \leq n\}.$$

Furthermore, set $v_n = \max\{v_{kj1n}, v_{kj2n} : 1 \leq k, j \leq D\}$ and $M_n = \max\{M_{kj1n}, M_{kj2n} : 1 \leq k, j \leq D\}$. Note that v_n and M_n depend on the distribution of the random matrix U and hence may depend on D. The expression on the right-hand side of (3.24) below is the direct generalization of (3.6) to the non-identically distributed case, whereas the bound $1 - \Psi$ given in (3.27) below is an improvement (even in the case of i.i.d. rows).

Theorem 3.6. *Assume that the rows $u_{1.}, \ldots, u_{n.}$ of U are independent random vectors in $\mathbb{C}^D$ whose entries satisfy the moment bounds (3.21) and (3.22). Then, for every $\varepsilon > 0$, the operator norm satisfies*

$$\left\| n^{-1} U^* U - Q_n \right\| < \varepsilon$$

with probability at least $1 - \Psi$ where Ψ is defined in (3.27) below. Furthermore,

$$1 - \Psi \geq 1 - 4D^2 \exp\left(-\frac{n\varepsilon^2}{D^2 \left(4n^{-1} v_n + 2\sqrt{2} D^{-1} M_n \varepsilon \right)} \right). \tag{3.24}$$

In particular, with probability not less than $1-\Psi$, the extremal eigenvalues of $n^{-1} U^* U$ satisfy

$$\lambda_{\min}(Q_n) - \varepsilon < \lambda_{\min}(n^{-1} U^* U) \leq \lambda_{\max}(n^{-1} U^* U) < \lambda_{\max}(Q_n) + \varepsilon. \tag{3.25}$$

Consequently the condition number of $U^* U$ is bounded by $\frac{\lambda_{\max}(Q_n) + \varepsilon}{\lambda_{\min}(Q_n) - \varepsilon}$ with probability not less than $1 - \Psi$, provided that Q_n defined in (3.20) is non-singular and $\varepsilon \in (0, \lambda_{\min}(Q_n))$.

<u>Proof:</u> Exactly as in the proof of Theorem 3.1 we arrive at

$$\mathbb{P}(\left\| n^{-1} U^* U - Q_n \right\| \geq \varepsilon) \tag{3.26}$$

$$\leq \sum_{k,j=1}^{D} \mathbb{P}\left(\left| \sum_{t=1}^{n} \mathrm{Re}(\overline{u_{tk}} u_{tj} - q_{kj}^{(t)}) \right| \geq \frac{n\varepsilon}{\sqrt{2}D} \right) + \sum_{k,j=1}^{D} \mathbb{P}\left(\left| \sum_{t=1}^{n} \mathrm{Im}(\overline{u_{tk}} u_{tj} - q_{kj}^{(t)}) \right| \geq \frac{n\varepsilon}{\sqrt{2}D} \right).$$

Again using inequality (3.5) for each k, j gives

$$\mathbb{P}\left(\left| \sum_{t=1}^{n} \mathrm{Re}(\overline{u_{tk}} u_{tj} - q_{kj}^{(t)}) \right| \geq \frac{n\varepsilon}{\sqrt{2}D} \right) \leq 2 \exp\left(-\frac{n\varepsilon^2}{D^2 \left(4n^{-1} v_{kj1n} + 2\sqrt{2} D^{-1} M_{kj1n} \varepsilon \right)} \right)$$

14

and similarly for the imaginary part. Hence, we finally obtain $\mathbb{P}\left(\|n^{-1}U^*U - Q\| \geq \varepsilon\right) \leq \Psi$ where

$$
\begin{aligned}
\Psi &= 2\sum_{i=1}^{2}\sum_{k,j=1}^{D} \exp\left(-\frac{n\varepsilon^2}{D^2\left(4n^{-1}v_{kjin} + 2\sqrt{2}D^{-1}M_{kjin}\varepsilon\right)}\right) \\
&\leq 4D^2 \exp\left(-\frac{n\varepsilon^2}{D^2\left(4n^{-1}v_n + 2\sqrt{2}D^{-1}M_n\varepsilon\right)}\right).
\end{aligned}
\tag{3.27}
$$

∎

Remark 3.7. A sufficient condition for Q_n to be non-singular is that at least one of the matrices $Q^{(t)}$ has this property. The argument in Remark 3.2 shows that the latter is the case if the distribution of $u_{t\cdot}$ is not concentrated on a $(D-1)$-dimensional linear subspace of $\mathbb{C}^D$. However, note the possibility that nevertheless $\lambda_{\min}(Q_n) \to 0$ as $n \to \infty$. □

Remark 3.8. (i) In case the (k,j)-element of $n^{-1}U^*U - Q_n$ is zero with probability 1, the corresponding terms on the right-hand side of (3.26) are zero and do not contribute to the bound in (3.26). Due to the independence assumption, the (k,j)-element is zero if and only if $\overline{u_{tk}}u_{tj} - q_{kj}^{(t)} = 0$ with probability 1 for every t. Hence, we may set $v_{kj1}^{(t)} = v_{kj2}^{(t)} = M_{kj1}^{(t)} = M_{kj2}^{(t)} = 0$ which shows that the corresponding terms in the bound Ψ are also automatically zero. However, in this case the bound (3.26) and the subsequent bounds can be improved in that in the (k,j)-th term in both sums on the right-hand side of (3.26) the constant D can be replaced by D_k, where D_k denotes the number of non-zero elements in the k-th row of $n^{-1}U^*U - Q_n$.

(ii) A similar remark applies in the case that some or all elements of $n^{-1}U^*U - Q_n$ are real (or imaginary). Cf. Remark 3.3.

4 Random Sampling of Trigonometric Polynomials Revisited

We now return to the special case of sampling trigonometric polynomials on uniformly distributed random points and show how the results in the previous sections can be improved. The analysis is based on techniques developed in [10] for the recovery of sparse trigonometric polynomials from random samples by basis pursuit (ℓ_1-minimization) and orthogonal matching pursuit. Some of the ideas are inspired by the pioneering work of Candès, Romberg and Tao in [3].

Theorem 4.1. *Let* $\Gamma \subset \mathbb{Z}^d$ *of size* $|\Gamma| = D$ *and let* $x_1, \ldots, x_n$ *be i.i.d. random variables that are uniformly distributed on* $[0,1]^d$. *Let* U *be the associated random Fourier matrix given by (2.1). Choose* $0 < \varepsilon < 1$, $0 < \alpha < \varepsilon^2$, *and* $\delta > 0$. *If*

$$
\left\lfloor \frac{\alpha n}{3D} \right\rfloor \geq \left[\ln\left(\frac{\varepsilon^2}{\alpha}\right)\right]^{-1} \ln\left(\frac{D}{\delta(1-\alpha)}\right),
\tag{4.1}
$$

15

then, with probability at least $1 - \delta$, we have

$$\left\| n^{-1} U^* U - I \right\| < \varepsilon$$

and hence

$$1 - \varepsilon < \lambda_{\min}(n^{-1} U^* U) \leq \lambda_{\max}(n^{-1} U^* U) < 1 + \varepsilon.$$

Consequently, the condition number of $U^* U$ is bounded by $\frac{1+\varepsilon}{1-\varepsilon}$ with probability $\geq 1 - \delta$.

For instance, the choice $\alpha = \varepsilon^2/e$ gives

$$n \geq \frac{3De}{\varepsilon^2} \left[\ln\left(\frac{D}{\delta}\right) + 2 - \ln(e-1) \right] \tag{4.2}$$

as a simple sufficient condition.

Compared (4.2) with (3.15) or (3.16), we have gained on the exponent in D; compared with Theorem 2.2 and (2.6), the constants are now independent of the dimension d of the state space (= the number of variables); compared with (3.18), the term $\ln(\delta^{-1})$ only enters linearly in (4.1) and (4.2) instead of quadratically and there is now no restriction on Γ. Moreover, the constants are explicit and small.

4.1 Proof of Theorem 4.1

We introduce the polynomials

$$F_m(z) = \sum_{k=1}^{\lfloor m/2 \rfloor} S_2(m,k) z^k, \qquad m \in \mathbb{N}, \tag{4.3}$$

where $S_2(m,k)$ are the associated Stirling numbers of the second kind. These are connected to the combinatorics of certain set partitions, and they can be computed by means of their exponential generating function, see [12, formula (27), p.77] or Sloane's A008299 in [14],

$$\sum_{m=1}^{\infty} F_m(z) \frac{x^m}{m!} = \exp\left(z(e^x - x - 1) \right). \tag{4.4}$$

Further, we define

$$G_m(z) := z^{-m} F_m(z). \tag{4.5}$$

Using the G_m's, we first establish a more general result from which Theorem 4.1 will follow.

Theorem 4.2. *Let $\Gamma \subset \mathbb{Z}^d$ of size $|\Gamma| = D$ and let $x_1, \ldots, x_n$ be i.i.d. random variables that are uniformly distributed on $[0,1]^d$. Let U be the associated random Fourier matrix given by (2.1), and let $\varepsilon > 0$. Then, for every $m \in \mathbb{N}$, we have*

$$\left\| n^{-1} U^* U - I \right\| < \varepsilon,$$

16

and hence

$$1 - \varepsilon < \lambda_{\min}(n^{-1}U^*U) \leq \lambda_{\max}(n^{-1}U^*U) < 1 + \varepsilon\,,$$

with probability at least

$$1 - \varepsilon^{-2m}DG_{2m}(n/D).$$

Proof: Again, the estimates for the eigenvalues follow from the inequality $\|n^{-1}U^*U - I\| < \varepsilon$. Furthermore, since $n^{-1}U^*U - I$ is self-adjoint, we have for every $m \in \mathbb{N}$

$$\|n^{-1}U^*U - I\| = \|(n^{-1}U^*U - I)^m\|^{1/m} \leq \|(n^{-1}U^*U - I)^m\|_F^{1/m},$$

where $\|\cdot\|_F$ denotes the Frobenius norm, $\|A\|_F = \sqrt{\mathrm{Tr}(AA^*)}$. Consequently,

$$\mathbb{P}(\|n^{-1}U^*U - I\| \geq \varepsilon) \leq \mathbb{P}(\|(n^{-1}U^*U - I)^m\|_F \geq \varepsilon^m).$$

We now apply Markov's inequality and obtain that

$$\mathbb{P}(\|(n^{-1}U^*U - I)^m\|_F \geq \varepsilon^m) \leq \varepsilon^{-2m}\mathbb{E}\left[\|(n^{-1}U^*U - I)^m\|_F^2\right].$$

The latter expectation was studied in [10, Section 3.3], see also Lemma 3.3 in [10]: It was shown that

$$\mathbb{E}\left[\|(n^{-1}U^*U - I)^m\|_F^2\right] \leq D\,G_{2m}(n/D), \tag{4.6}$$

which concludes the proof. ∎

We now show how Theorem 4.1 follows from Theorem 4.2. This is done by estimating G_m and by a diligent choice of the free parameter m.

We set the oversampling rate to be $\theta = n/D$. In [10, Section 3.5] it was shown that

$$G_{2m}(\theta) \leq (3m/\theta)^m \frac{1 - (3m/\theta)^m}{1 - (3m/\theta)}.$$

For given $0 < \alpha < 1$ and θ, we choose $m = m(\theta) \in \mathbb{N}$ such that $(3m(\theta)/\theta) \leq \alpha < 1$. Note that this is possible since $\lfloor \alpha n/3D \rfloor \geq 1$ follows from the assumptions of the theorem. In the following we will take the value

$$m(\theta) = \lfloor \alpha\theta/3 \rfloor\,,$$

and obtain that

$$G_{2m(\theta)}(\theta) \leq \frac{\alpha^{m(\theta)}}{1 - \alpha}. \tag{4.7}$$

In view of Theorem 4.2 we want to achieve $\varepsilon^{-2m}D\,G_{2m}(\theta) \leq \delta$. By (4.7) this inequality is satisfied if

$$D\varepsilon^{-2m(\theta)}\frac{\alpha^{m(\theta)}}{1 - \alpha} \leq \delta,$$

which is equivalent to

$$\ln\left(\frac{\varepsilon^2}{\alpha}\right)m(\theta) \geq \ln\left(\frac{D}{(1-\alpha)\delta}\right).$$

Since $\alpha < \varepsilon^2$ by assumption, Theorem 4.1 follows.

Acknowledgments

We would like to thank Richard Nickl for helpful discussions.

References

[1] R. Bass and K. Gröchenig. Random sampling of multivariate trigonometric polynomials. *SIAM J. Math. Anal.*, 36(3):773–795, 2004.

[2] G. Bennett. Probability inequalities for the sum of independent random variables. *J. Am. Stat. Assoc.*, 57:33–45, 1962.

[3] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, 2006.

[4] E. Candès and T. Tao. Near optimal signal recovery from random projections and universal encoding strategies. *IEEE Trans. Inf. Theory*, 52(12):5406–5425, 2006.

[5] N. Cesa Bianchi and G. Lugosi. *Prediction, Learning and Games.* Cambridge University Press, 2006.

[6] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[7] F. Cucker and S. Smale. On the mathematical foundations of learning. *Bull. Amer. Math. Soc. (N.S.)*, 39(1):1–49 (electronic), 2002.

[8] S. Kunis and H. Rauhut. Random sampling of sparse trigonometric polynomials II - Orthogonal Matching Pursuit versus Basis Pursuit. *preprint*, 2006.

[9] S. Mendelson and A. Pajor. On singular values of matrices with independent rows. *Bernoulli*, 12:761–773, 2006.

[10] H. Rauhut. Random sampling of sparse trigonometric polynomials. *Appl. Comput. Harm. Anal.*, 22:16–42, 2007.

[11] H. Rauhut. Stability results for random sampling of sparse trigonometric polynomials. *preprint*, 2006.

[12] J. Riordan. *An Introduction to Combinatorial Analysis.* Wiley, New York, London, 1958.

[13] M. Rudelson and R. Vershynin. Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements. In *Proc. CISS 2006 (40th Annual Conference on Information Sciences and Systems)*, 2006.

[14] N. Sloane. On-Line Encyclopedia of Integer Sequences. Available: `http://www.research.att.com/~njas/sequences`.

[15] S. Smale and D.-X. Zhou. Shannon sampling and function reconstruction from point values. *Bull. Amer. Math. Soc. (N.S.)*, 41(3):279–305 (electronic), 2004.

[16] A. Van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes.* Springer-Verlag, 1996.