Non-Asymptotic Analysis of Compressed Sensing Random Matrices: An U-Statistics Approach

Fabian Lim and Vladimir Marko Stojanovic
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139
{flim, vlada}@mit.edu

***Abstract*—We apply Heoffding's U-statistics to obtain non-asymptotic analysis for compressed sensing (CS) random matrices. These powerful (U-statistics) tools appear to apply naturally to CS theory, in particular here we focus on one particular large deviation result. We chose two applications to outline how U-statistics may apply to various CS recovery guarantees. Pros, cons, and further directions of the approach, are discussed.**

Restricted isometries of random matricies have well-regarded importance in CS. They guarantee i) uniqueness of sparse solutions, and ii) robust recovery. The fraction of size-k submatrices (out of all $\binom{n}{k}$ of them), that satisfy CS-type restricted isometeries, is an U-statistic. Concentration of U-statistics predict the "average-case" behavior of such isometries.

U-statistics related to Fuchs' conditions for ℓ_1-minimization support recovery, are derived. This leads to bounds on the fraction of recoverable k-supports. Empirically, we observe significant improvement over a recent large deviation (non-asymptotic) bound by Donoho & Tanner, for some practical system sizes with large undersampling.

The results apply regardless of column distribution, e.g. Gaussian, Bernoulli, etc. Similar concentration behavior has been empirically observed, when the sampling matrix is constructed using pseudorandom sequences (important in practice).

***Index Terms*—compressed sensing, pseudorandom sequences, random matrices, restricted isometries**

I. Introduction

Randomized sampling in CS is simple and efficient, applicable to a wide variety of sparse signals [1], [2]. Signals of block-length n are potentially compressed down to $m \ll n$ samples, where m is roughly of the order of k - the number of sparse signal components (or information rate). The theory began with asymptotics in the "large-systems limit" where sizes - m, k and n go to infinity, and ratios k/m and m/n are fixed [1], [2]. Current research tend towards practice, specifically *non-asymptotic* analysis [3], [4], *deterministic-type* sampling techniques [5], [6], [7], [8], and *hardware implementation* [9], [10], [11]. This work aims to contribute in the first respect - by studying *U-statistics* [12] of CS random matrices. Our interest is analysis for fixed (finite) sizes, important for practical system design - a recent signal acquisition system implementation [9], finds the number of samples m to significantly impact radio energy, CMOS circuit energy, and hardware area.

In practice, *pseudorandom binary sequences* (PRBS) are used to mimic statistical behavior. Empirical evidence is provided to support such an approach in the U-statistical setting. Tested in implementations [9], [11], [10], PRBS's are found to require low energy and negligible storage cost. In deterministic CS, it has come to our recent attention that matrices constructed using *sparse expander graphs* are recently implemented [13]. It would be interesting to compare both approaches from a hardware standpoint. Deterministic CS techniques are important, though its known random arguments make certain tasks somewhat simpler (*e.g.*, extending arguments to non-time sparse cases [2], *redundent* signal bases [14], even random sub-dictionary sampling [15]).

This work was supported by NSF grant ECCS-1128226.

U-statistics are meant for *combinatorial-type* settings, ala CS theory, and this approach seems overlooked. We initiate first efforts via two different techniques. In Section II-B, main concepts are illustrated using average-case behaviors of *restricted isometries* (quantities related to solution uniqueness, and robust recovery). In Section III-A, differences/similarities with past analysis is discussed: i) Bah & Tanners's asymptotic restricted isometry analysis, and ii) *statistical restricted isometries* (stRIP) of deterministic matrices [5]. In Section III-B we look at *Karush-Kuhn-Tucker* (KKT) analysis for ℓ_1-minimization support recovery; we focus on the simplest (noiseless) case studied by Fuchs [16]. U-statistics seem promising compared with past results [4], for measurement sizes $m = 50 \sim 150$ and $n = 1000 \sim 3000$ (high undersampling - similar sizes considered in implementation [9]).

Notation: Deterministic vectors and matrices are denoted using bold fonts (e.g. $\mathbf{a}$ and $\mathbf{A}$, respectively). Random variables (RV's) are denoted using upper-case italics (e.g., A). The (cumulative) distribution function (d.f.) of a RV A is denoted $F_A(a)$. The ℓ_2-norm is denoted $||\mathbf{a}||_{\ell_2}$, and the ℓ_0-pesudo and ℓ_1-norms denoted similarly. We write $\mathbb{1}\{A \leq a\}$ as the indicator of the event $\{A \leq a\}$. Real numbers are denoted $\mathbb{R}$. The matrix *pseudoinverse* of $\mathbf{A}$ is denoted $\mathbf{A}^\dagger$.

II. U-statistics and CS Parameters

A. Background: Restricted Isometries in CS Theory

Let $\mathbf{x} = [x_1, x_2, \cdots, x_n]^T \in \mathbb{R}^n$ denote a signal vector. For a positive constant $k \leq n$, let $\overline{\mathbf{x}}_k \in \mathbb{R}^n$ denote the *best k-term approximation* of $\mathbf{x}$, i.e. $\overline{\mathbf{x}}_k$ is the best approximation of $\mathbf{x}$ involving exactly k non-zero terms, see [2]. If the error $||\overline{\mathbf{x}}_k - \mathbf{x}||_{\ell_1}$ is small, the approx. is good. Vector $\mathbf{a}$ is k-sparse, if *at most* k coefficients are non-zero (i.e., $||\mathbf{a}||_{\ell_0} \leq k$).

In CS compression, we form a *measurement* vector $\mathbf{y} = [x_1, x_2, \cdots, x_m]^T \in \mathbb{R}^m$ for some $m \leq n$, by projecting $\mathbf{y} = \mathbf{\Phi x}$. Hence $\mathbf{\Phi}$ has a *non-trivial null-space*, so recovering $\mathbf{x}$ from $\mathbf{y}$ is challenging. Signal $\mathbf{x}$ is recoverable if i) $\mathbf{x}$ is well-approximated by some sufficiently sparse $\overline{\mathbf{x}}_k$, and if ii) matrix $\mathbf{\Phi}$ satisfies certain "conditioning" properties. Hence certain conditions guarantee that (NP) *ℓ_0-minimization*

$$\min_{\tilde{\mathbf{x}} \in \mathbb{R}^n} ||\tilde{\mathbf{x}}||_{\ell_0} \quad \text{s. t.} \quad \mathbf{y} = \mathbf{\Phi}\tilde{\mathbf{x}}, \tag{1}$$

has a unique solution, also that the *ℓ_1-minimization*

$$\min_{\tilde{\mathbf{x}} \in \mathbb{R}^n} ||\tilde{\mathbf{x}}||_{\ell_1} \quad \text{s. t.} \quad ||\tilde{\mathbf{y}} - \mathbf{\Phi}\tilde{\mathbf{x}}||_{\ell_2} \leq \epsilon \tag{2}$$

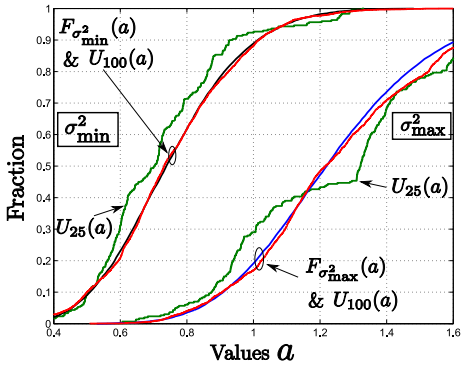

Fig. 1. Concentration of U-statistic U_n for kernels g corresp. to $\sigma^2_{\min}$ and $\sigma^2_{\max}$, see (5). Here $m = 25, k = 2$, and two values 25 to 100 for n.

recovers (in some sense) the solution to (1). Note $\tilde{\mathbf{y}}$ is allowed to be *noisy*, bounded as $\epsilon \geq ||\tilde{\mathbf{y}} - \mathbf{y}||_{\ell_2}$. Such conditions are usually framed via so-called *restricted isometry constants*.

Definition 1 (Restricted Isometry Constants [2])**.** *For each $k < n$, the k-th **restricted isometry constant** δ_k satisfies*

$$(1 - \delta_k)||\mathbf{a}||^2_{\ell_2} \leq ||\mathbf{\Phi a}||^2_{\ell_2} \leq (1 + \delta_k)||\mathbf{a}||^2_{\ell_2},$$

for all k-sparse $\mathbf{a} \in \mathbb{R}^n$ s.t. $||\mathbf{a}||_{\ell_0} \leq k$.

Theorem 1.1, c.f. [17] *Assume that $\delta_{2k} < \sqrt{2}-1$ and $||\tilde{\mathbf{y}} - \mathbf{y}||_{\ell_2} \leq \epsilon$. Then the ℓ_1-minimum solution $\mathbf{x}^*$ to (2) satisfies*

$$||\mathbf{x}^* - \mathbf{x}||_{\ell_2} \leq C_1 k^{-\frac{1}{2}} ||\mathbf{x} - \overline{\mathbf{x}}_k||_{\ell_1} + C_2 \epsilon,$$

for constants $C_1 = 4\sqrt{1+\delta_{2k}}/(1-\delta_{2k}(1+\sqrt{2}))$ and $C_2 = 2(\delta_{2k}(1-\sqrt{2})-1)/(\delta_{2k}(1+\sqrt{2})-1)$.

Theorem 1.1 bounds the recovery error $||\mathbf{x}^* - \mathbf{x}||_{\ell_2}$ in terms of the best k-term approximation error $||\mathbf{x} - \overline{\mathbf{x}}_k||_{\ell_1}$, and the noise ϵ. Also, $\delta_{2k} < 1$ guarantees that if any solution of (1) is k-sparse, that is the *unique* solution. Definition 1 w.r.t. δ_k implies combinatoric nature. Let $\mathcal{S}$ denote a size-k subset of indices $\{1, 2, \cdots, n\}$. Denote $\mathbf{\Phi}_{\mathcal{S}}$ to be the size $m \times k$ submatrix of $\mathbf{\Phi}$ indexed on (column indices) $\mathcal{S}$. Respectively denote $\sigma^2_{\min}(\mathbf{\Phi}_{\mathcal{S}})$ and $\sigma^2_{\max}(\mathbf{\Phi}_{\mathcal{S}})$ as the min. and max. *squared singular values* of $\mathbf{\Phi}_{\mathcal{S}}$. Assuming columns $\boldsymbol{\phi}_i$ of $\mathbf{\Phi}$ are properly normalized, the constant δ_k bounds

$$\delta_k \;\geq\; \max\{\sigma^2_{\max}(\mathbf{\Phi}_{\mathcal{S}}) - 1, 1 - \sigma^2_{\min}(\mathbf{\Phi}_{\mathcal{S}})\}. \qquad (3)$$

Computing δ_k is NP-hard, but can be estimated if $\mathbf{\Phi}$ is *randomly sampled*. Assume throughout that the n columns $\boldsymbol{\phi}_i$ are *independently* sampled with fixed column d.f. $F(\boldsymbol{\phi})$. Given $F(\boldsymbol{\phi})$ admits random concentration, a union bound over all $\binom{n}{k}$ subsets $\mathcal{S}$ shows (3) is satisfied with exponentially decreasing probability, when the number of measurements m grows at rate $k \log(n/k)$, see [2], [18], [19].

There are suspicions expressed on tightness of such union bounds [18]. Donoho suggested to alternatively derive results for fractions of subsets [20] - which we believe this to be worthwhile for non-asymptotic analysis. U-statistics coincide with Donoho's viewpoint, addressing "average-case" behavior of combinatorial-type quantities such as (3).

B. U-Statistics of Restricted Isometries

The concept behind U-statistics is most naturally introduced using the example of restricted isometries. Denote a function $g : \mathbb{R}^{m \times k} \times \mathbb{R} \mapsto [0, 1]$, with inputs an $m \times k$ submatrix $\mathbf{\Phi}_{\mathcal{S}}$, and $a \in \mathbb{R}$. Furthermore, let $g(\mathbf{\Phi}_{\mathcal{S}}, a)$ be

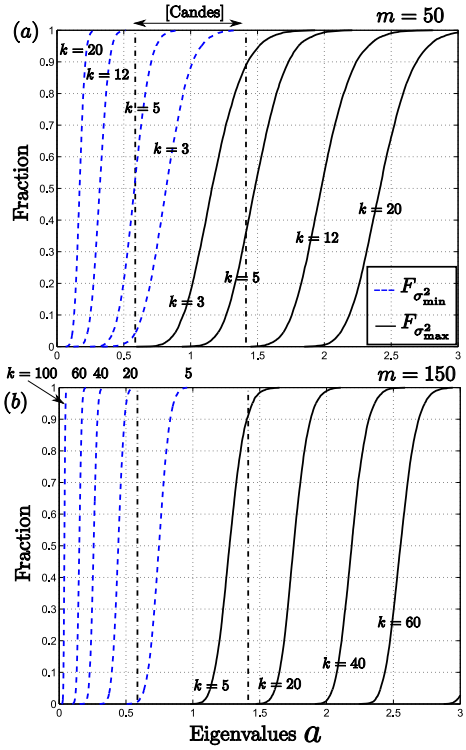

Fig. 2. Means $\mathbb{E}U_n = F_{\sigma^2_{\min}}$ and $F_{\sigma^2_{\max}}$ for (a) $m = 50$ and (b) $m = 150$ predict "average-case" concentration of U_n. Shown for Gaussian case.

invariant under permutation of columns of $\mathbf{\Phi}_{\mathcal{S}}$. An U-statistic w.r.t. matrix $\mathbf{\Phi}$ and function g is defined as, see [12],

$$U_n = U_n(a) \triangleq \frac{1}{\binom{n}{k}} \sum_{\mathcal{S}} g(\mathbf{\Phi}_{\mathcal{S}}, a). \qquad (4)$$

We say g is a *kernel* of degree k. We appropriately select g. In the context of restricted isometries (Thm. 1.1.), we select

$$\begin{aligned} g(\mathbf{\Phi}_{\mathcal{S}}, a) &= \mathbb{1}\left\{\sigma^2_{\max}(\mathbf{\Phi}_{\mathcal{S}}) \leq a\right\}, \text{ or} \\ g(\mathbf{\Phi}_{\mathcal{S}}, a) &= \mathbb{1}\left\{\sigma^2_{\min}(\mathbf{\Phi}_{\mathcal{S}}) \leq a\right\}. \end{aligned} \qquad (5)$$

The robust recovery guarantees require $\sigma^2_{\max}(\mathbf{\Phi}_{\mathcal{S}})$ and $\sigma^2_{\min}(\mathbf{\Phi}_{\mathcal{S}})$, to be respectively upper and lower bounded by $a = \sqrt{2}$ and $a = 2 - \sqrt{2}$. As such, the statistic $U_n(a)$ in (4) will count the fraction of subsets that satisfy the required isometeries. As (4) has many terms for large k, we therefore use concentration to predict the behavior of the U-statistic.

Vary $a \in \mathbb{R}$, and Figure $1(a)$ illustrates for $m = 25$ the empirical distribution of $U_n(a)$ for kernels g chosen as in (5) with degree $k = 2$. For both cases, the U-statistic $U_{100}(a)$ lies closer to the *mean* $\mathbb{E}U_n(a)$, than $U_{25}(a)$. Here column d.f. $F(\boldsymbol{\phi})$ is i.i.d. Gaussian, see [2]. If $F_{\sigma^2_{\max}}(a)$ and $F_{\sigma^2_{\min}}(a)$ respectively denote the marginal d.f.'s of $\sigma^2_{\min}$ and $\sigma^2_{\max}$, then $\mathbb{E}U_n(a) = F_{\sigma^2_{\min}}(a)$ and $\mathbb{E}U_n(a) = F_{\sigma^2_{\max}}(a)$, respectively. The following *large deviation* result can be shown by piecing together (5.5) in [21] and Lemma 2.1 in [22]. Note $\mathbb{E}U_n(a)$ does not depend on n, and for general kernels g write $\mathbb{E}U_n(a) = F_A(a)$. Bounds on $F_A(a)$ are not discussed (with exception Cor. 1) - $F_A(a)$ mostly obtained by simulation.

Theorem 1 (c.f. [21], [22])**.** *Let $U_n(a)$ be a U-statistic (4) of a matrix $\mathbf{\Phi}$, whose columns are i.i.d. sampled. Let g be a kernel with degree k. Let $F_A(a) = \mathbb{E}U_n(a)$. Almost surely, the deviation $|U_n(a) - \mathbb{E}U_n(a)| \leq \epsilon_n$ is bounded by error $\epsilon_n^2 = 2F_A(a)(1-F_A(a)) \cdot (n/k)^{-1} \log(n/k) = \mathcal{O}(n^{-1} \log n)$.*

See later Section IV for a sketch proof. Thm. 1 is a non-asymptotic result for fixed k, n, where F_A depends implicitly on m. The concentration is around the mean

$$\mathbb{E}U_n(a) = F_A(a) = \mathbb{E}g(\boldsymbol{A}, a),$$

where $\boldsymbol{A}$ is a random $m \times k$ matrix with column d.f. $F(\boldsymbol{\phi})$. Note $F_A(a)$ is a marginal quantity (no n dependence) and relatively (depending on g, m and k) easy to obtain. Thm. 1 is interesting, because of the natural "average-case" interpretation. One might try to investigate what happens if (3) is allowed to be satisfied *only* for a fraction of subsets S. In Subsection III-A we discuss differences with "worst-case" analyses of restricted isometries [2], [19], whereby the concentration in Thm. 1 gets stronger (i.e. error ϵ_n decreases) as n increases; this contrasts with "worst-case" where concentration worsens. Thm. 1 holds for general kernels g, and (in Subsection III-B) we discuss U-statistics related to noiseless ℓ_1-min. recovery. Thm. 1 holds for general column d.f.'s, see Subsection III-C for discussion on Bernoulli/PRBS.

Regarding issues not discussed, we point out relevant references on U-statistics. For a "worst-case" interpretation, see [23], Ch. 8. For weak-dependence amongst columns, see [24]. For normal approximation, see [12]. Finally for kernels[1] with degree n, e.g. if g is set to count subsets $\mathcal{S}$ of size-k, for which there is at least some disjoint size-k subset $\mathcal{R}$ such that $\sigma^2_{\min}(\boldsymbol{\Phi}_{\mathcal{S}\cup\mathcal{R}}) = 0$ (fraction of $\mathcal{S}$ which fail the condition for unique sparsity), see [25].

III. Further Discussion on U-Statistics

A. Comparing existing results - Bah & Tanner and stRIP

We illustrate differences between "average-" and "worst-case" analysis. Figures $2(a)$ and (b) shows the U-statistics means $\mathbb{E}U_n = F_{\sigma^2_{\min}}$ and $F_{\sigma^2_{\max}}$ for the two kernels in (5), and measurement sizes $m = 50$ and $m = 150$. To compare, Table I reproduces from [26], for "worst-case" squared sing. values (smallest/largest $\sigma^2_{\min}(\boldsymbol{\Phi}_{\mathcal{S}})$ and $\sigma^2_{\max}(\boldsymbol{\Phi}_{\mathcal{S}})$ over all $\mathcal{S}$, c.f. (3)).

Table I shows values for k/m versus m/n, "large-system limit" results for large m, k and n. These numbers seem different. For $k/m = 0.3$, we see for all undersamp. ratios m/n that $\sigma^2_{\min} \approx 0.01$. But for fixed $m = 50$ and $m = 150$, Figures $2(a)$ and (b) show for $k = 0.3 \cdot (150) = 15$ and $k = 45$, practically all $\sigma^2_{\min} > 0.1$. Again as n increases, the U-statistic concentrates with vanishing error ϵ_n (see Thm. 1)); contrast with Table I whereby these estimates do not concentrate (i.e. get larger) for fixed m and inc. n.

Similar observations are made for the $\sigma^2_{\max}$ case. For $k/m = 0.1$, we see $\sigma^2_{\max} > 3.45$ for various m/n. Comparing with Figures $2(a)$ and (b), we see that practically all $\sigma^2_{\max}$ seems smaller than 2, where at $m = 50$ and $k = 0.1 \cdot (50) = 5$ at least a fraction $\mathbb{E}U_n \approx 0.4$ of $\sigma^2_{\max}$ lie within the required (Candes') range for robust recovery. Do note Thm. 1.1 c.f. [2] strictly requires "worst-case" constants δ_k (see (3)). We move on to compare with "average-case" notions of restricted isometries in deterministic CS analysis.

Theorem 3, c.f. [8] *Let $\boldsymbol{\Phi}$ be a deterministic matrix with $||\boldsymbol{\phi}_i||_{\ell_2} = 1$, zero row sums (i.e. $\sum_{i=1}^n \boldsymbol{\phi}_i = \mathbf{0}$), and assume $|\boldsymbol{\phi}_i^T \boldsymbol{\phi}_j| \leq 1/\sqrt{m}$ for all $i \neq j$. Let $\boldsymbol{X}$ be a length-n random vector with equiprobable entries X_i, $||\boldsymbol{X}||_{\ell_0} = k$ and $\sum_{i=1}^n X_i = 0$. For $\delta > 1/(n-1)$, then $\Pr\{|||\boldsymbol{\Phi}\boldsymbol{X}||^2_{\ell_2} - ||\boldsymbol{X}||^2_{\ell_2}| < \delta||\boldsymbol{X}||^2_{\ell_2}\}$ is at least*

$$1 - 2\exp\left(-\frac{m(\delta - 1/(n-1))^2}{16k}\right).$$

[1]The weaker notion of unique sparsity in Thm. 13, [7], can be checked using a degree-$(k+1)$ U-statistic, similar to upcoming Subsect. III-B.

TABLE I
Asymptotic estimates [Bah & Tanner]

		Minimum: $\sigma^2_{\min}$			Maximum: $\sigma^2_{\max}$		
		m/n			m/n		
		0.1	*0.3*	*0.5*	*0.1*	*0.3*	*0.5*
k/m	*0.1*	0.095	0.118	0.130	3.952	3.610	3.459
	0.2	0.015	0.026	0.034	5.587	4.892	4.535
	0.3	0.003	0.006	0.010	6.939	5.806	5.361

Such results are termed *statistical* restricted isometry property (stRIP); the "statistical" notion comes from including a random $\boldsymbol{X}$. In context of Thm. 1, an analogue follows.

Corollary 1. *Let $F_{\sigma^2_{min}}(a)$ and $F_{\sigma^2_{max}}(a)$ belong to size-k Wishart matrices. For a suitably chosen constant $\delta = \delta(k/m)$, the fraction of size k-subsets $\mathcal{S}$ satisfying (3) is at least*

$$\begin{aligned} & 1 - F_{\sigma^2_{min}}(1 - \delta + \epsilon_n) - (1 - F_{\sigma^2_{max}}(1 + \delta - \epsilon_n)) \\ = \;& 1 - 2\exp\left(-\frac{mt^2}{2}\right), \end{aligned}$$

where constant $t = \sqrt{1 + \delta + \epsilon_n} - (1 + \sqrt{k/m}) > 0$.

Proof: Apply[2] Corollary 35 in [3] to Thm. 1, for the $1 - U_n$ and U_n resp. corresp. to the two kernels g in (5) ■

We point out that Cor. 1 makes a stronger claim (in the almost sure sense), and requires no random signal model.

B. An application of U-statistics for CS recovery

Fuchs' sufficient conditions prove equality between ℓ_1- and ℓ_0-minimum solutions (under noiseless conditions).

Proposition 1 (Fuchs [16])**.** *Let $\mathbf{x}$ be the unique solution to (1), let $\mathcal{S}$ denote its support. Assume $\boldsymbol{\Phi}_{\mathcal{S}}$ has full rank. Then for $\epsilon = 0$, the unique solution to (2) is also $\mathbf{x}$, if and only if every $i \notin \mathcal{S}$ satisfies*

$$\left|(\boldsymbol{\Phi}_{\mathcal{S}}^{\dagger}\boldsymbol{\phi}_i)^T \operatorname{sgn}(\tilde{\mathbf{x}})\right| < 1, \tag{6}$$

where $\tilde{\mathbf{x}}$ is the short vector of non-zero coefficients $\tilde{\mathbf{x}} = [x_{i_1}, x_{i_2}, \cdots, x_{i_k}]^T$, i.e., here $i_j \in \mathcal{S}$.

Let $\mathcal{R}$ denote a size-$(k+1)$ subset. and let $\mathcal{R} \setminus \{j\}$ is the size-k subset excluding j. Define a kernel $g : \mathbb{R}^{m\times(k+1)} \times \mathbb{R} \mapsto \mathbb{R}$ that satisfies

$$g(\boldsymbol{\Phi}_{\mathcal{R}}, a) = \frac{1}{2^k(k+1)} \sum_{\ell=1}^{2^k} \sum_{j\in\mathcal{R}} \mathbb{1}\left\{\left|(\boldsymbol{\Phi}_{\mathcal{R}\setminus\{j\}}^{\dagger}\boldsymbol{\phi}_j)^T \mathbf{b}_\ell\right| \geq a\right\} \tag{7}$$

where $\mathbf{b}_\ell$ enumerates the "sign" vectors in $\{-1, 1\}^k$.

Proposition 2. *Let g be the kernel (7). For any $\mathbf{x}$, condition (6) will be violated for at most a fraction $(n-k) \cdot U_n(1)$ of size k-subsets $\mathcal{S}$ and "sign" combinations $\mathbf{b} \in \{-1, 1\}^k$.*

Proof: Let indicator $\mathbb{1}_{\mathcal{S},j}$ be indexed on size-k subset $\mathcal{S}$ and index j. Using the identity $\binom{n}{k+1} \cdot (k+1) = \binom{n}{k} \cdot (n-k)$, by "pigeon-hole" argument $\sum_{\mathcal{R}} \sum_{j\in\mathcal{R}} \mathbb{1}_{\mathcal{R}\setminus\{j\},j} = \sum_{\mathcal{S}} \sum_{j\notin\mathcal{S}} \mathbb{1}_{\mathcal{S},j}$. Use expression (4) and the previous counting formula to get that $(n-k)U_n(1)$ equals

$$\frac{n-k}{2^k(k+1)\binom{n}{k+1}} \sum_{\ell=1}^{2^k} \sum_{\mathcal{S}} \sum_{j\notin\mathcal{S}} \mathbb{1}\left\{\left|(\boldsymbol{\Phi}_{\mathcal{S}}^{\dagger}\boldsymbol{\phi}_j)^T \mathbf{b}_\ell\right| > 1\right\}$$

[2]Corollary 35 in [3] omits some deviation constants.

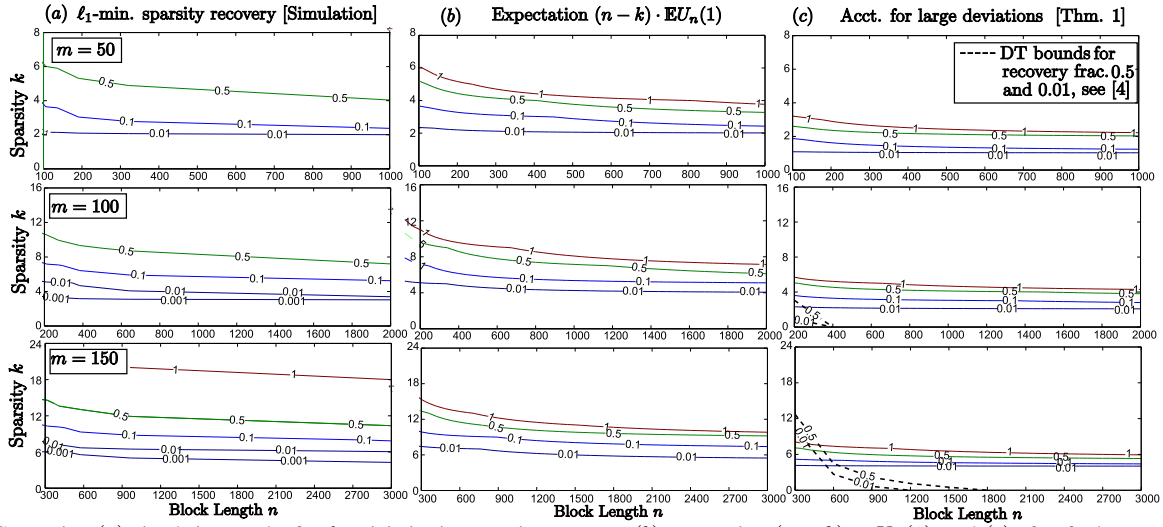

Fig. 3. Comparing (a) simulation results for ℓ_1-minimization sparsity recovery, (b) expectation $(n-k)\cdot \mathbb{E}U_n(1)$, and (c) after further accounting for large deviations (Thm. 2). We show cases $m = 50, 100$ and 150. We also compare with Donoho & Tanners' (DT) large deviation bounds [4].

lower bounded as (also apply previous binomial identity)

$$\frac{1}{2^k\binom{n}{k}}\sum_{\ell=1}^{2^k}\sum_{\mathcal{S}}\mathbb{1}\left\{\max_{j\notin\mathcal{S}}\left|(\mathbf{\Phi}_{\mathcal{S}}^{\dagger}\boldsymbol{\phi}_j)^T\mathbf{b}_\ell\right| > 1\right\}.$$

For $\mathbf{x}$ with "sign" $\mathbf{b} = \mathrm{sgn}(\tilde{\mathbf{x}})$ and some subset $\mathcal{S}$, the final indicator equals 1 when *at least* one $j \notin \mathcal{S}$ violates (6). ■

By Prop. 1, condition (6) needs to be enforced along with uniqueness solution conditions of problems (1) and (2); though by small random perturbation of $\mathbf{\Phi}$ uniqueness should hold a.s.. The counting argument is "rough", but Figure 3 reveals promising observations for some practical system sizes (inferred from [9]). Fig. $3(a)$ plots simulation results[3] for 3 measurement sizes $m = 50, 100$ and 150 and block sizes $n \geq 200$ and $n \leq 3000$. Fig. $3(b)$ plots the expectation $(n-k)\cdot\mathbb{E}U_n(1)$. Good match is observed between all cases where the fraction ≤ 0.5. Fig. $3(c)$ shows that even after accounting for deviation error ϵ_n (see Thm. 1), the bound is reasonably tight for fractions ≤ 0.5; stronger results could be hoped for $m = 150$. Compare with bounds obtained by Donoho & Tanner [4]. For system parameters $m = 50$ and $n \leq 1000$ chosen in hardware implementation [9], the results of [4] do not offer reasonable predictions. For $m = 100$, the bounds [4] work only for very small block lengths $n \leq 300$. Only $m = 150$ is reasonable, where [4] is better for $n \leq 400$.

The analysis [4] requires m to grow with n. Similarly here if m, k are left fixed, the "rough" counting will eventually fail for some large n; note $(n-k)\cdot\mathbb{E}U_n(1)$ grows at rate $\mathcal{O}(n)$ (the deviation $n\cdot\epsilon_n$ grows sublinearly, i.e. $\mathcal{O}(n^{\frac{1}{2}}\log n)$). Also the geometrical thresholds in [4] and Fuchs' recovery conditions (see Proposition 1) are not exactly the same[4]. Finally for brevity, as seen from Pro. 1, we omitted discussion on i) the soln. uniqueness of (1), and ii) the invertibility of $\mathbf{\Phi}_{\mathcal{S}}$. For Gaussian matrices i) and ii) occur with probability 0 if $2k < m$.

C. Empirical evidence supporting PRBS

Figure 4 is counterpart to Figure 2, when column distribution $F(\boldsymbol{\phi})$ is Bernoulli. Included is also an empirical

[3]In our experiments, the support $\mathcal{S}$ is deemed to be recovered, when the recovered signal coefficients $|x_i| \leq 10^{-16}$ for all indices $i \notin \mathcal{S}$.

[4]Fuchs' conditions are observed to be more conservative.

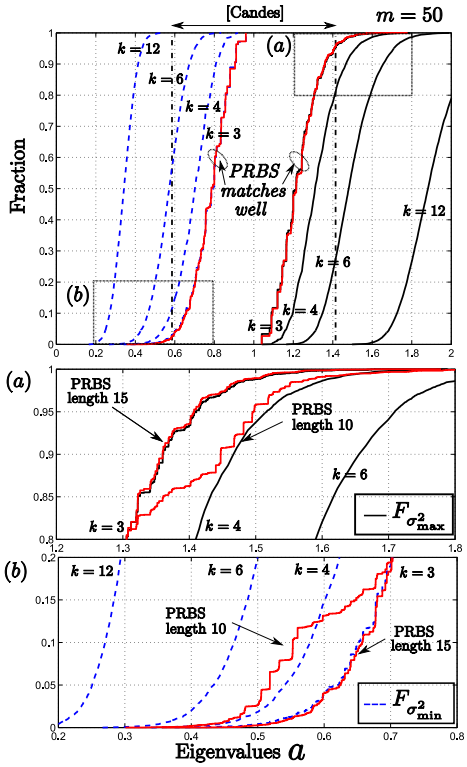

Fig. 4. Means $\mathbb{E}U_n(a) = F_{\sigma^2_{\min}}(a)$ and $F_{\sigma^2_{\max}}(a)$ for the Bernoulli case and $m = 50$. Empirical evidence (for $k = 3$) shows PRBS matches well with prediction. Also shown are zoom-in's on (a) upper and (b) lower tails.

curve generated from $\mathbf{\Phi}$ constructed using a PRBS of period $2^{15} - 1 = 32767$. For the case $k = 3$, we observe good empirical match (we used the same $n = 100$ as in Fig. 1). Figure $4(a)$ shows a closeup of the upper tail for the $\sigma^2_{\max}$ case. It seems that a smaller period $2^{10} - 1 = 1023$ is insufficient to model $m\cdot n = 5000$ Bernoulli matrix entries (here $n = 100$). Similar observations are made in $4(b)$ for the other $\sigma^2_{\min}$ case.

Bernoulli matrices $\mathbf{\Phi}$ are interesting, not only because their statistical properties can be approximated using PRBS, but also because they outperform Gaussian matrices. We see in Fig. 4 for $k = 3$, a larger fraction of $\sigma^2_{\max}$ lie within the range (for noisy recovery), as compared to Gaussian in Fig. 2

IV. Sketch Proof of Main Theorem

For brevity's sake, we shall not show the full proof of Thm. 1 here, but simply point out the critical ideas. Thm 1 is essentially a law of large numbers result for U-statistics, but here U_n is a sum of dependent variates, so techniques developed for i.i.d. sequences do not directly apply.

New notation: partition the index set $\{1, 2, \cdots, n\} = (\cup_{i=1}^{\nu_n} \mathcal{S}_i) \cup \mathcal{R}$, where there are $\nu_n = \lfloor n/k \rfloor$ subsets $\mathcal{S}_i$ of size $|\mathcal{S}_i| = k$, and a subset $\mathcal{R}$ of size $|\mathcal{R}| < k$. Let π denote some index *permutation* on $\{1, 2, \cdots, n\}$, and let $\pi(\mathcal{S})$ denote the permutation *image* of the set $\mathcal{S}$ under π.

Lemma 1 ([21]). *We have the following identity*

$$U_n(a) \;=\; \frac{1}{n!} \sum_{\pi} \left(\frac{1}{\nu_n} \sum_{i=1}^{\nu_n} g(\mathbf{\Phi}_{\pi(\mathcal{S}_i)}, a) \right) \tag{8}$$

Proof: This follows by observing that for any partition $\mathcal{S}_i$ and any arbitrary k-subset $\mathcal{S}$, there are altogether $k!(n-k)!$ permutations π that send $\mathcal{S}_i$ to $\mathcal{S}$, i.e. $\mathcal{S} = \pi(\mathcal{S}_i)$. ∎

Let $\mathcal{D}(b||b')$ denote the binary information divergence between two Bernoulli distributions b and b'.

Lemma 2 (Hoeffding, c.f. [21]). *Let the RV A be an average of ν_n i.i.d. RV's, each bounded between 0 and 1. Then for some small ϵ satisfying $0 < \epsilon < 1 - \mathbb{E}A$ and $h > 0$*

$$\mathbb{E}e^{h(A - \mathbb{E}A - \epsilon)} \leq e^{-\nu_n \mathcal{D}(\mathbb{E}A + \epsilon || \mathbb{E}A)}. \tag{9}$$

Proof of Lem. 2 is standard; we refer the reader to [21].

Sketch Proof of Thm. 1: We need a probability inequality on $\Pr\{|U_n - \mathbb{E}U_n| > \epsilon\}$, for some $\epsilon > 0$. In what follows we only discuss the one-sided inequality $\Pr\{U_n - \mathbb{E}U_n > \epsilon\}$ (the other side follows similarly). Use the Bernstein bound (for some $h > 0$) followed by Lemma 1

$$\begin{aligned} \Pr\{U_n - \mathbb{E}U_n > \epsilon\} \;&\leq\; \mathbb{E}e^{h(U_n - \mathbb{E}U_n - \epsilon)} \\ &=\; \mathbb{E}e^{\frac{1}{n!}\left(\sum_{\pi} h(A_\pi - \mathbb{E}A_\pi - \epsilon)\right)}. \end{aligned}$$

where $A_\pi(a)$ is shorthand for the inner sum of (8), i.e. $A_\pi(a) = \frac{1}{\nu_n} \sum_{i=1}^{\nu_n} g(\mathbf{\Phi}_{\pi(\mathcal{S}_i)}, a)$ denotes the sum corresponding to some perm. π. Use the convexity of $\exp(\cdot)$ to write

$$\Pr\{U_n - \mathbb{E}U_n > \epsilon\} \;\leq\; \frac{1}{n!} \sum_{\pi} \mathbb{E}e^{h(A_\pi - \mathbb{E}A_\pi - \epsilon)}.$$

For each term $\mathbb{E}e^{h(A_\pi - \mathbb{E}A_\pi - \epsilon)}$, observe that A_π is a sum of ν_n i.i.d. RV's; this is because the subsets $\mathcal{S}_i$ are disjoint, and the columns of $\mathbf{\Phi}$ are i.i.d. Note $\mathbb{E}A_\pi = \mathbb{E}U_n$ for any π. Then use Hoeffding's Lemma 2 to get that $e^{-\nu_n \mathcal{D}(\mathbb{E}U_n + \epsilon || \mathbb{E}U_n)}$ upper bounds $\Pr\{U_n - \mathbb{E}U_n > \epsilon\}$. Details are omitted here, see the exposition regarding Lemma 2.1 in [22] on how to manipulate the function $\mathcal{D}(\mathbb{E}U_n + \epsilon || \mathbb{E}U_n)$, to show that the error term $\epsilon = \epsilon_n$ detailed in the statement of Thm. 1, is sufficiently large to guarantee that

$$\sum_{n=1}^{\infty} \Pr\{|U_n - \mathbb{E}U_n| > \epsilon_n\} < \infty$$

holds. The sketch proof is completed. ∎

V. Conclusion

We proposed taking a U-statistic approach to analyzing CS random matrices. We discussed certain results which are naturally interpreted in the "average-case" sense. We showed two applications with respect to i) restricted isometries, ii) stRIP-results and fractions of supports not satisfying Fuchs' conditions (that guarantee signal recoverability). Possible future extensions of this work have been discussed, whereby one can draw from the wealth of U-statistics literature.

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. on Inform. Theory*, vol. 52, no. 4, pp. 1289 – 1306, Apr. 2006.

[2] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. on Inform. Theory*, vol. 51, pp. 4203–4215, 2005.

[3] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *Tech. Rep., University of Michigan*, Aug. 2010.

[4] D. L. Donoho and J. Tanner, "Exponential bounds implying construction of compressed sensing matrices, error-correcting codes, and neighborly polytopes by random sampling," *IEEE Trans. on Inform. Theory*, vol. 56, no. 4, pp. 2002–2016, Apr. 2010.

[5] R. Calderbank, S. Howard, and S. Jafarpour, *Deterministic compressed sensing matrices with groups of random variables*. Online: http://www.dsp.rice.edu/files/cs/strip-more.pdf, 2009.

[6] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank, "Efficient and robust compressed sensing using optimized expander graphs," *IEEE Trans. on Inform. Theory*, vol. 55, no. 9, pp. 4299–4308, Sep. 2009.

[7] J. A. Tropp, "On the conditioning of random subdictionaries," *App. Comp. Harm. Anal.*, vol. 25, no. 1, pp. 1–24, 2008.

[8] L. Gan, C. Long, T. T. Do, and T. D. Tran, *Analysis of the Statistical Restricted Isometry Property for Deterministic Sensing Matrices using Stien's Method*. Online: http://www.dsp.rice.edu/files/cs/Gan_StatRIP.pdf, 2009.

[9] F. Chen, A. P. Chandrakasan, and V. Stojanovic, "A signal-agnostic compressed sensing acquisition system for wireless and implantable systems," in *Proc. IEEE Custom Integrated Circuits Conference (CICC '10)*, San Jose, CA, Sep. 2010, pp. 1–4.

[10] M. Mishali and Y. C. Eldar, "Expected RIP: Conditioning of the modulated wideband converter," in *Proc. IEEE Information Theory Workshop (ITW '09)*, Taormina, Sicily, Oct. 2009, pp. 343–347.

[11] K. Kanoun, H. Mamaghanian, N. Khaled, and D. Atienza, "A real-time compressed sensing-based personal EKG monitoring system," in *Proc. IEEE/ACM Design, Automation and Test in Europe Conference (DATE '11)*, Grenoble, France, Mar. 2011, pp. 1–6.

[12] W. Hoeffding, "A class of statistics with asymptotically normal distributions," *Annals of Statistics*, vol. 19, pp. 293–325, 1948.

[13] J. Yoo, A. Khajehnejad, B. Hassibi, and A. Emami, "An analog sublinear time sparse signal acquisition framework based on structured matrices," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '12)*, Kyoto, Japan, Mar. 2012.

[14] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed Sensing with Coherent and Redundant Dictionaries," *App. Comp. Harm. Anal.*, vol. 31, no. 1, pp. 1–21, 2010.

[15] J. Tropp, "On the conditioning of random subdictionaries," *App. Comp. Harm. Anal.*, vol. 25, no. 1, pp. 1–24, Jul. 2008.

[16] J. J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. on Inform. Theory*, vol. 50, no. 6, pp. 1341–1344, 2004.

[17] E. Candes, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences, Paris, Serie I*, vol. 336, pp. 589–592, 2008.

[18] J. D. Blanchard, C. Cartis, and J. Tanner, "Phase transitions for restricted isometry properties," in *Proc. Signal Proc. with Adaptive Sparse Structured Reps. (SPARS '09)*, Saint-Malo, Fance, Apr. 2009.

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

[20] D. L. Donoho, "Neighorly polytopes and sparse solutions of underdetermined linear equations," *Tech. Rep., Stanford University*, Dec. 2004.

[21] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Am. Stat. Assoc.*, vol. 58, no. 301, pp. 13–30, Mar. 1963.

[22] P. K. Sen, "Asymptotic normality of sample quantiles for m-dependent processes," *Ann. Math. Statist.*, vol. 39, no. 5, pp. 1724–1730, 1968.

[23] A. D. Barbour, L. Holst, and S. Janson, *Poisson Approximation*. Oxford: Clarendon Press, 1992.

[24] M. Wendler, "Bahadur representation for U-quantiles of dependent data," *J. Multivariate Anal.*, vol. 102, no. 6, pp. 1–21, 2011.

[25] C. Heilig and D. Nolan, "Limit Theorems for the infinite-degree U-process," *Statistica Sinica*, vol. 11, pp. 289–302, 2001.

[26] B. Bah and J. Tanner, "Improved bounds on restricted isometry constants for gaussian matrices," *SIAM J. Matrix Analysis*, vol. 31, no. 5, pp. 2882–2898, 2010.