frontiers in Applied Mathematics and Statistics

ORIGINAL RESEARCH
published: 06 June 2019
doi: 10.3389/fams.2019.00026

Derandomizing Compressed Sensing With Combinatorial Design

Peter Jung[1]*, *Richard Kueng*[2]* and *Dustin G. Mixon*[3]

[1] Communications and Information Theory Group, Technische Universität Berlin, Berlin, Germany, [2] Department of Computing and Mathematical Sciences, Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, CA, United States, [3] Department of Mathematics, Ohio State University, Columbus, OH, United States

Compressed sensing is the art of effectively reconstructing structured n-dimensional vectors from substantially fewer measurements than naively anticipated. A plethora of analytical reconstruction guarantees support this credo. The strongest among them are based on deep results from large-dimensional probability theory and require a considerable amount of randomness in the measurement design. Here, we demonstrate that derandomization techniques allow for a considerable reduction in the randomness required for such proof strategies. More precisely, we establish uniform s-sparse reconstruction guarantees for $Cs\log(n)$ measurements that are chosen independently from strength-4 orthogonal arrays and maximal sets of mutually unbiased bases, respectively. These are highly structured families of $\tilde{C}n^2$ vectors that imitate signed Bernoulli and standard Gaussian vectors in a (partially) derandomized fashion.

Keywords: compressed sensing, *k*-wise independence, orthogonal arrays, spherical design, derandomization

OPEN ACCESS

Edited by:
Jean-Luc Bouchot,
Beijing Institute of Technology, China

Reviewed by:
Nguyen Quang Tran,
Aalto University, Finland
Shao-Bo Lin,
Wenzhou University, China
Martin Lotz,
University of Warwick,
United Kingdom

***Correspondence:**
Peter Jung
peter.jung@tu-berlin.de
Richard Kueng
rkueng@caltech.edu

Specialty section:
This article was submitted to Mathematics of Computation and Data Science, a section of the journal Frontiers in Applied Mathematics and Statistics

Received: *23 September 2018*
Accepted: *01 May 2019*
Published: *06 June 2019*

Citation:
Jung P, Kueng R and Mixon DG (2019) Derandomizing Compressed Sensing With Combinatorial Design. Front. Appl. Math. Stat. 5:26. doi: 10.3389/fams.2019.00026

1. INTRODUCTION AND MAIN RESULTS

1.1. Motivation

Compressed sensing is the art of effectively reconstructing structured signals from substantially fewer measurements than would naively be required for standard techniques like least squares. Although not entirely novel, rigorous treatments of this observation [1, 2] spurred considerable scientific attention from 2006 on (see e.g., [3, 4]) and references therein. While deterministic results do exist, the strongest theoretic reconstruction guarantees still rely on randomness. Each measurement corresponds to the observed inner product of the unknown vector, with a vector chosen randomly from a fixed measurement ensemble. Broadly, these ensembles can be grouped into two families:

(i) *Generic measurements* such as independent Gaussian, or Bernoulli vectors. Such an abundance of randomness allows establishing very strong results by following comparatively simple and instructive proof techniques. The downside is that concrete implementations do require a lot of randomness. In fact, they might be too random to be useful for certain applications.

(ii) *Structured measurements* such as random rows of a Fourier, or Hadamard matrix. In contrast to generic measurements, these feature a lot of structure that is geared toward applications. Moreover, selecting random rows from a fixed matrix does require very little randomness. For example, $\log_2(n)$ random bits are sufficient to select a row of the DFT uniformly at random. In contrast, generating an i.i.d. Bernoulli vector would consume n bits of randomness. Structure and comparatively little randomness have a downside, though. Theoretic convergence guarantees tend to be weaker than their generic counterparts. It should also not come as a surprise that the necessary proof techniques become considerably more involved.

Typically, results for type (i) precede the results for type (ii). Phase retrieval via PhaseLift is a concrete example for such a development. Generic convergence guarantees [5, 6] preceded (partially) de-randomized results [7, 8]. Compressed sensing is special in this regard. The two seminal works [1, 2] from 2006 provided both results almost simultaneously. This had an interesting consequence. Despite considerable effort, to date there still seems to be a gap between both proof techniques.

Here, we try to close this gap by applying a method that is very well established in theoretical computer science: *partial derandomization*. We start with a proof technique of type (i) and considerably reduce the amount of randomness required for it to work. While doing so, we keep careful track of the amount of randomness that is still necessary. Finally, we replace the original (generic) random measurements with pseudo-random ones that mimic them in a sufficiently accurate fashion. Our results highlight that this technique *almost* allows bridging the gap between existing proof techniques for generic and structured measurements: the results are still strong but require slightly more randomness than choosing vectors uniformly from a bounded orthogonal system, such as Fourier or Hadamard vectors.

There is also a didactic angle to this work: within the realm of signal processing, partial-derandomization techniques have been successfully applied to matrix reconstruction [8, 9] and phase retrieval via PhaseLift [7, 10, 11]. Although similar in spirit, the more involved nature of these problems may obscure the key ideas, intuition and tricks behind such an approach. However, the same techniques have not yet been applied to the original problem of compressed sensing. Here, we fill this gap and, in doing so, provide an introduction to partial derandomization techniques by example. To preserve this didactic angle, we try to keep the presentation as simple and self-contained as possible.

Finally, one may argue that compressed sensing has not fully lived up to the high expectations of the community yet (see e.g., Tropp [12]). Arguably, one of the most glaring problems for applications is the requirement of choosing individual measurements at random[1]. While we are not able to fully overcome this drawback here, the methods described in this work do limit the amount of randomness required to generate individual structured measurements. We believe that this may help to reduce the discrepancy between "what can be proved" and "what can be done" in a variety of concrete applications.

1.2. Preliminaries on Compressed Sensing

Compressed sensing aims at reconstructing s-sparse vectors $\mathbf{x} \in \mathbb{C}^n$ from $m \ll n$ linear measurements:

$$\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{C}^m.$$

Since $m \ll n$, the matrix $\mathbf{A}$ is singular and there are infinitely many solutions to this equation. A convex penalizing function is used to promote sparsity among these solutions. Typically, this penalizing function is the ℓ_1-norm $\|\mathbf{z}\|_{\ell_1} = \sum_{i=1}^n |z_i|$:

$$\begin{aligned} \underset{\mathbf{z}\in\mathbb{C}^n}{\text{minimize}} \quad & \|\mathbf{z}\|_{\ell_1} \\ \text{subject to} \quad & \mathbf{A}\mathbf{z} = \mathbf{y} \end{aligned} \tag{1}$$

Strong mathematical proofs for correct recovery have been established. By and large, these statements require randomness in the sense that each row $\mathbf{a}_i \in \mathbb{C}^n$ of $\mathbf{A}$ is an *independent* copy of a random vector $\mathbf{a} \in \mathbb{R}^n$. Prominent examples include

(1) $m = Cs\log(n/s)$ standard complex Gaussian measurements: $\mathbf{a}_g \sim \mathcal{N}(\mathbf{0}, \mathbb{I}/\sqrt{2}) + i\mathcal{N}(\mathbf{0}, \mathbb{I}/\sqrt{2})$,
(2) $m = Cs\log(n/s)$ signed Bernoulli (Rademacher) measurements: $\mathbf{a}_{sb} \sim \{\pm 1\}^n$,
(3) $m = Cs\log^4(n)$ random rows of a DFT matrix: $\mathbf{a}_f \sim \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$,
(4) for $n = 2^d$: $m = Cs\log^4(n)$ random rows of a Hadamard matrix: $\mathbf{a}_h \sim \{\mathbf{h}_1, \dots, \mathbf{h}_n\}$.

A rigorous treatment of all these cases can be found in Foucart and Rauhut [3]. Here, and throughout this work, $C > 0$ denotes an absolute constant whose exact value depends on the context but is always independent of the problem parameters n, s, and m. It is instructive to compare the amount of randomness that is required to generate one instance of the random vectors in question. A random signed Bernoulli vector $\mathbf{a}_{sb} \in \mathbb{R}^n$ requires n random bits (one for each coordinate), while a total of $\log_2(n)$ random bits suffice to select a random row $\mathbf{a}_h \in \mathbb{R}^n$ of a Hadamard matrix. A comparison between complex standard Gaussian vectors $\mathbf{a}_g \in \mathbb{C}^n$ and random Fourier vectors $\mathbf{a}_f \in \mathbb{C}^n$ indicates a similar discrepancy. In summary: highly structured random vectors, like $\mathbf{a}_f, \mathbf{a}_h$ require exponentially fewer random bits to generate than generic random vectors, like $\mathbf{a}_g, \mathbf{a}_{sb}$. Importantly, this transition from generic measurements to highly structured ones comes at a price. The number of measurements required in cases (3) and (4) scales poly-logarithmically in n. More sophisticated approaches allow for converting this offset into a poly-logarithmic scaling in s rather than n [14, 15]. Another, arguably even higher price, is hidden in the proof techniques behind these results. They are considerably more involved.

The following two subsections are devoted to introducing formalisms that allow for partially de-randomizing signed Bernoulli vectors and complex standard Gaussian vectors, respectively.

1.3. Partially De-randomizing Signed Bernoulli Vectors

Throughout this work, we endow $\mathbb{C}^n$ with the standard inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n \bar{x}_i y_i$. We denote the associated (Euclidean) norm by $\|\mathbf{z}\|_{\ell_2}^2 = \langle \mathbf{z}, \mathbf{z} \rangle$. Let $\mathbf{a}_{sb} = (\epsilon_1, \dots, \epsilon_n)^T$ be a signed Bernoulli vector with coefficients $\epsilon_i \sim \{\pm 1\}$ chosen independently at random (Rademacher random variables). Then,

$$\mathbb{E}\left[\epsilon_i \bar{\epsilon_j}\right] = \mathbb{E}\left[\epsilon_i \epsilon_j\right] = \delta_{ij}. \tag{2}$$

[1] Existing deterministic constructions (see e.g., Bandeira et al. [13]), do not (yet) yield comparable statements.

This feature is equivalent to *isotropy* of signed Bernoulli vectors:

$$\mathbb{E}\left[|\langle \mathbf{z}, \mathbf{a}_{sb}\rangle|^2\right] = \sum_{i,j=1}^{n} \mathbb{E}\left[\epsilon_i \bar{\epsilon}_j\right] \bar{z}_i z_j = \langle \mathbf{z}, \mathbf{z}\rangle = \|\mathbf{z}\|_{\ell_2}^2. \quad (3)$$

for all $\mathbf{z} \in \mathbb{C}^n$.
Independent sign entries are sufficient, but not necessary for this feature. Indeed, suppose that $n = 2^d$ is a power of two. Then, the rows of a Sylvester Hadamard matrix correspond to a particular subset of n sign vectors that are also proportional to an orthonormal basis. The orthonormal basis property ensures that a randomly selected Hadamard row $\mathbf{a}_h \in \mathbb{R}^n$ is also isotropic. In turn, the coordinates $a_i \in \{\pm 1\}$ of $\mathbf{a}_h$ also obey (2), despite not being independent instances of random signs. This feature is called *pairwise independence* and naturally generalizes to $k \geq 2$:

Definition 1. (k-wise independence.) *Fix $k \geq 2$ and let ϵ_i denote independent instances of a signed Bernoulli random variable. We call a random sign vector $\mathbf{a} \in \{\pm 1\}^n$ k-wise independent, if its components $a_1, \ldots, a_n$ obey*

$$\mathbb{E}\left[\prod_{i=1}^{k} a_{i_k}\right] = \mathbb{E}\left[\prod_{i=1}^{k} \epsilon_{i_k}\right] \quad (4)$$

for all k-tuples of indices $1 \leq i_1, \ldots, i_k \leq n$.

For $k = 2$, this defining property reduces to Equation (2) and is equivalent to demanding isotropy of the distribution. Random rows of Hadamard matrices form a concrete example for pairwise independence.

What is more, explicit constructions for k-wise independent vectors in $\mathbb{R}^n$ are known for any k and n. In this work we focus on particular constructions that rely on generalizing the following instructive example. Fix $n = 3$ and consider the $M = 4$ rows of the following matrix:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

The first two columns summarize all possible tuples ($k = 2$ combinations) of ± 1. The coefficients of the third column correspond to their entry-wise product. Hence, the third column is completely characterized by the first two and therefore three components of a selected row are *not* mutually independent. Nonetheless, each subset of two coordinates does mimic independent behavior: all possible length-two combinations of ± 1 occur exactly once in these two columns. This ensures that a randomly selected row obeys Equation (4) for $k = 2$ (pairwise independence). The concept of *orthogonal arrays* [16] generalizes this simple example.

Definition 2. (Orthogonal array) *A (binary) orthogonal array of* strength *k with M rows and n columns is a $M \times n$ sign matrix such that every selection of k columns contains all elements of $\{\pm 1\}^k$ an equal number of times.*

The example from above is an 4×3 orthogonal array of strength 2. Strength-k orthogonal arrays are closely related to the concept of k-wise independence. The following implication is straightforward.

Lemma 1. *Selecting a row of an $M \times n$ orthogonal array of strength k uniformly at random produces k-wise independent random vectors in n dimensions.*

This correspondence identifies orthogonal arrays as general-purpose seeds for pseudo-random behavior. What is more, explicit constructions of orthogonal arrays are known for any k and any n. In contrast to the "full" array that lists all 2^n possible elements of $\{\pm 1\}^n$ as its rows, these constructions typically only require $M \geq \mathcal{O}\left(n^{k/2}\right)$ rows. This fundamental restriction is called *Rao's bound* and only scales polynomially in the dimension n. Choosing a random row from such an array only requires $\log_2(M) = \mathcal{O}(k \log_2(n))$ random bits and produces a random vector that is k-wise independent.

1.4. Partially Derandomizing Complex Standard Gaussian Vectors

Let us now discuss another general-purpose tool for (partial) de-randomization. Concentration of measure implies that n-dimensional standard complex Gaussian vectors concentrate sharply around the complex sphere of radius $\sqrt{n}$. Hence, they behave very similarly to vectors $\mathbf{a}_s \in \mathbb{C}^n$ chosen uniformly from this sphere. For any $k \in \mathbb{N}$, such spherical random vectors obey

$$\mathbb{E}\left[|\langle \mathbf{z}, \mathbf{a}_s\rangle|^{2k}\right] = n^k \int_{\mathbf{w} \in \mathbb{S}^{n-1}} |\langle \mathbf{z}, \mathbf{w}\rangle|^{2k} \mathrm{d}\mathbf{w} = n^k \binom{n+k-1}{k}^{-1} \|\mathbf{z}\|_{\ell_2}^{2k}.$$

for all $\mathbf{z} \in \mathbb{C}^n$.
Here, d$\mathbf{w}$ denotes the uniform measure on the complex unit sphere $\mathbb{S}^{n-1} = \left\{\mathbf{x} \in \mathbb{C}^n : \|\mathbf{x}\|_{\ell_2} = 1\right\}$. This formula characterizes even moments of the uniform distribution[2] on $\mathbb{S}^{n-1}$. The concept of k-designs [17] uses this moment formula as a starting point for partial de-randomization. Roughly speaking, a k-design is a finite subset of $\sqrt{n}$-length vectors such that the uniform distribution over these vectors reproduces the uniform measure on $\sqrt{n}\mathbb{S}^{n-1}$ up to k-th moments. More precisely:

Definition 3. *A set of N vectors $\{\mathbf{w}_i\}_{i=1}^N \subset \sqrt{n}\mathbb{S}^{n-1}$ is called a* (complex projective) k-design *if, for any $l \in \left[k\right] = \left\{1, \ldots, k\right\}$, a randomly selected element $\mathbf{a}_{(k)}$ obeys*

$$\mathbb{E}\left[|\langle \mathbf{z}, \mathbf{a}_{(k)}\rangle|^{2l}\right] = n^l \binom{n+l-1}{l}^{-1} \|\mathbf{z}\|_{\ell_2}^{2l} \quad \textit{for all} \quad \mathbf{z} \in \mathbb{C}^n.$$

(Spherical) k-designs were originally developed as cubature formulas for the real-valued unit sphere [17]. The concept has since been extended to other sets. A generalization to the complex projective space $\mathbb{C}P^{n-1}$ gives rise to Definition 3. Complex projective k-designs are known to exist for any k and any

[2]For comparison, a complex standard Gaussian vector obeys $\mathbb{E}\left[|\langle \mathbf{z}, \mathbf{a}_g\rangle|^{2k}\right] = k!\|\mathbf{z}\|_{\ell_2}^{2k}$ instead.

dimension n (see e.g., [18–20]). However, explicit constructions for $k \geq 3$ are notoriously difficult to find. In contrast, several explicit families of 2-designs have been identified. Here, we will focus on one such family. Two orthonormal bases $\{\mathbf{b}_i\}_{i=1}^n$ and $\{\mathbf{c}_i\}_{i=1}^n$ of $\mathbb{C}^n$ are called *mutually unbiased* if

$$\left|\langle \mathbf{b}_i, \mathbf{c}_j \rangle\right|^2 = \frac{1}{n} \quad \text{for all} \quad i,j \in [n]\,. \tag{5}$$

A prominent example for such a basis pair is the standard basis and the Fourier, or Hadamard, basis, respectively. One can show that at most $n+1$ different orthonormal bases may exist that have this property in a pairwise fashion [21, Theorem 3.5]. Such a set of $n+1$ bases is called a *maximal set of mutually unbiased bases* (MMUB). For instance, in $n = 2$ the standard basis together with

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}1\\i\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-i\end{pmatrix}$$

forms a MMUB. Importantly, MMUBs are always (proportional to) 2-designs [22]. Explicit constructions exist for any prime power dimension n and one can ensure that the standard basis is always included. Here we point out one construction that is particularly simple if the dimension $n \geq 5$ is prime [23]: The standard basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_n \in \mathbb{C}^n$ together with all vectors whose entry-wise coefficients correspond to

$$\left[\mathbf{b}_{\alpha,\lambda}\right]_k = \frac{1}{\sqrt{n}}\omega_n^{(k+\alpha)^3+\lambda(k+\alpha)} \tag{6}$$

form a MMUB. Here $\omega_n = \exp\left(\frac{2\pi i}{n}\right)$ is a n-th root of unity. The parameter $\alpha \in [n]$ singles out one of the n different bases, while $\lambda \in [n]$ labels the n corresponding basis vectors. Excluding the standard basis, this set of n^2 vectors corresponds to all time-frequency shifts of a discrete Alltop sequence $[\mathbf{f}]_k = \omega_n^{k^3}$ [24].

1.5. Main Results

In the following $\tilde{c} > 0$, like C, denotes an absolute constant whose precise value may depend on the context.

Theorem 1. (CS from orthogonal array measurements.) *Suppose that a matrix* $\mathbf{A}$ *contains* $m \geq Cs\log(2n)$ *rows that are chosen independently from strength-4 orthogonal array. Then, with probability at least* $1 - 2\mathrm{e}^{-\tilde{c}m}$, *any s-sparse* $\mathbf{x} \in \mathbb{C}^n$ *can be recovered from* $\mathbf{y} = \mathbf{A}\mathbf{x}$ *by solving the convex optimization problem (1).*

Theorem 2. (CS from time-frequency shifted Alltop sequences.) *Let* $n \geq 5$ *be prime and suppose that* $\mathbf{A}$ *contains* $m \geq Cs\log(2n)$ *rows that correspond to random time-frequency shifts of the n-dimensional Alltop sequence (6). Then, with probability at least* $1 - 2\mathrm{e}^{-\tilde{c}m}$, *any s-sparse* $\mathbf{x} \in \mathbb{R}^n$ *can be recovered from* $\mathbf{y} = \mathbf{A}\mathbf{x}$ *by solving (1).*

This result readily generalizes to measurements that are sampled from a maximal set of mutually unbiased bases (excluding the standard basis). Time-frequency shifts of the Alltop sequence are one concrete construction that applies to prime dimensions only.

Note that the cardinality of all Alltop shifts is n^2. Hence, $2\log_2(n)$ random bits suffice to select a random time-frequency shift. In turn, a total of

$$2\log_2(n)m \simeq 2Cs\log^2(n) \tag{7}$$

random bits are required for sampling a complete measurement matrix $\mathbf{A}$. This number is exponentially smaller than the number of random bits required to generate a matrix with independent complex Gaussian entries. A similar comparison holds true for random signed Bernoulli matrices and columns sampled from a strength-4 orthogonal array.

Highly structured families of vectors – such as rows of a Fourier, or Hadamard matrix – require even less randomness to sample from: only $\log_2(n)$ bits are required to select such a row uniformly at random. However, existing recovery guarantees are weaker than the main results presented here. They require an order of $Cs\mathrm{polylog}(s)\log(n)$ random measurements to establish comparable results. Thus, the total number of random bits required for such a procedure scales like $Cs\mathrm{polylog}(s)\log^2(n)$. Equation (7) still establishes a logarithmic improvement in terms of sparsity.

The recovery guarantees in Theorem 1 and Theorem 2 can be readily extended to ensure stability with respect to noise corruption in the measurements and robustness with respect to violations of the model assumption of sparsity. We refer to section 3 for details.

We also emphasize that there are results in the literature that establish compressed sensing guarantees with comparable, or even less, randomness. Obviously, deterministic constructions are the extreme case in this regard. Early deterministic results suffer from a "quadratic bottleneck." The number of measurements must scale quadratically in the sparsity: $m \simeq s^2$. Although this obstacle was overcome, existing progress is still comparatively mild. Bourgain et al. [25], Mixon [26], Bandeira et al. [27] establish deterministic convergence guarantees for $m \simeq s^{2-\epsilon}$, where $\epsilon > 0$ is a (very) small constant.

Closer in spirit to this work is Bandeira et al. [28]. There, the authors employ the Legendre symbol – which is well known for its pseudorandom behavior – to partially derandomize a signed Bernoulli matrix. In doing so, they establish uniform s-sparse recovery from $m \geq Cs\log^2(s)\log(n)$ measurements that require an order of $s\log(s)\log(n)$ random bits to generate. Compared to the main results presented here, this result gets by with less randomness, but requires more measurements. The proof technique is also very different.

To date, the strongest de-randomized reconstruction guarantees hail from a close connection between s-sparse recovery and Johnson-Lindenstrauss embeddings [29, 30]. These have a wide range of applications in modern data science. Kane and Nelson [31] established a very strong partial de-randomization for such embeddings. This result may be used to establish uniform s-sparse recovery for $m = Cs\log(n/s)$ measurements that require an order of $s\log\left(s\log(n/s)\log(n/s)\right)$ random bits. This result surpasses

the main results presented here in both sampling rate and randomness required.

However, this strong result follows from "reducing" the problem of s-sparse recovery to a (seemingly) very different problem: find Johnson-Lindenstrauss embeddings. Such a reduction typically does not preserve problem-specific structure. In contrast, the approach presented here addresses the problem of sparse recovery directly and relies on tools from signal processing. In doing so, we maintain structural properties that are common in several applications of s-sparse recovery. Orthogonal array measurements, for instance, have ± 1-entries. This is well-suited for the single pixel camera [32]. Alltop sequence constructions, on the other hand, have successfully been applied to stylized radar problems [33]. Both types of measurements also have the property that every entry has unit modulus. This is an important feature for communication applications like CDMA [34]. Having pointed out these high-level connections, we want to emphasize that careful, problem specific adaptations may be required to rigorously exploit them. The framework developed here may serve as a guideline on how to achieve this goal in concrete scenarios.

2. PROOFS

2.1. Textbook-Worthy Proof for Real-Valued Compressed Sensing With Gaussian Measurements

This section is devoted to summarizing an elegant argument, originally by Rudelson and Vershynin [14], see also [35–37] for arguments that are similar in spirit. This argument only applies to s-sparse recovery of real-valued signals. We will generalize a similar idea to the complex case later on.

In this work we are concerned with *uniform* reconstruction guarantees: with high probability, a single realization of the measurement matrix $\mathbf{A}$ allows for reconstructing *any* s-sparse vector $\mathbf{x}$ by means of ℓ_1-regularization (1). A necessary pre-requisite for uniform recovery is the demand that the kernel, or *nullspace*, of $\mathbf{A}$ must not contain any s-sparse vectors. This condition is captured by the *nullspace property* (NSP). Define,

$$T_s = \left\{\mathbf{z} \in \mathbb{S}^{n-1} : \ \|\mathbf{z}\|_{\ell_1} \geq 2\sigma_s(\mathbf{z})\right\}, \tag{8}$$

where $\sigma_s(\mathbf{x}) = \inf_{\|\mathbf{z}\|_0 \leq s} \|\mathbf{x} - \mathbf{z}\|_{\ell_1}$ is the approximation error (measured in ℓ_1-norm) one incurs when approximating $\mathbf{x}$ by s-sparse vectors.

A matrix $\mathbf{A}$ obeys the NSP of order s (s-NSP) if,

$$\inf_{\mathbf{z}\in T_s} \|\mathbf{A}\mathbf{z}\|_{\ell_2} > 0. \tag{9}$$

The set T_s is a subset of the unit sphere that contains all normalized s-sparse vectors. This justifies the informal definition of the NSP: no s-sparse vector is an element of the nullspace of $\mathbf{A}$. Importantly, the NSP is not only necessary, but also sufficient for uniform recovery (see e.g., Foucart and Rauhut [3, Theorem 4.5]). Hence, universal recovery of s-sparse signals readily follows from establishing Rel. (9). The nullspace property and its relation to s-sparse recovery has long been somewhat folklore. We refer to Foucart and Rauhut [3, Notes on section 4] for a discussion of its origin.

The following powerful statement allows for exploiting generic randomness in order to establish nullspace properties. It is originally due to Gordon [38], but we utilize a more modern reformulation, see Foucart and Rauhut [3, Theorem 9.21].

Theorem 3. (Gordon's escape through a mesh.) *Let* $\mathbf{A}$ *be a real-valued* $m \times n$ *standard Gaussian matrix*[3] *and let* $E \subseteq \mathbb{S}^{n-1} \subset \mathbb{R}^n$ *be a subset of the real-valued unit sphere. Define the* Gaussian width $\ell(E) = \mathbb{E}\sup_{\mathbf{z}\in E}\langle \mathbf{a}_g, \mathbf{z}\rangle$, *where the expectation is over realizations* $\mathbf{a}_g \sim \mathcal{N}(0, \mathbb{I})$ *of a standard Gaussian random vector. Then, for* $t \geq 0$ *the bound*

$$\inf_{\mathbf{z}\in E} \|\mathbf{A}\mathbf{z}\|_{\ell_2} \geq \sqrt{m-1} - \ell(E) - t$$

is true with probability at least $1 - \mathrm{e}^{-t^2/2}$.

This is a deep statement that connects random matrix theory to geometry: the Gaussian width is a rough measure of the size of the set $E \subseteq \mathbb{S}^{n-1}$. Setting $E = T_s$ allows us to conclude that a matrix $\mathbf{A}$ encompassing m independent Gaussian measurements is very likely to obey the s-NSP (9), provided that $(m-1)$ exceeds $\ell(T_s)^2$. In order to derive an upper bound on $\ell(T_s)$, we may use the following inclusion

$$T_s \subset 2\mathrm{conv}\left(\Sigma_s\right),$$

see e.g., Kabanava and Rauhut [35, Lemma 3] and Rudelson and Vershynin [14, Lemma 4.5]. Here, $\Sigma_s \subseteq \mathbb{S}^{n-1}$ denotes the set of all s-sparse vectors with unit length. In turn,

$$\ell(T_s) \leq 2\mathbb{E} \sup_{\mathbf{z}\in\mathrm{conv}(\Sigma_s)} \langle \mathbf{a}_g, \mathbf{z}\rangle = 2\mathbb{E} \sup_{\mathbf{z}\in\Sigma_s} \langle \mathbf{a}_g, \mathbf{z}\rangle, \tag{10}$$

because the linear function $\mathbf{z} \mapsto \langle \mathbf{a}_g, \mathbf{z}\rangle$ achieves its maximum value at the boundary Σ_s of the convex set $\mathrm{conv}\left(\Sigma_s\right)$. The right-hand side of (10) is the expected supremum of a Gaussian process indexed by $\mathbf{z} \in \Sigma_s$. *Dudley's inequality* [39], see also [3, Theorem 8.23], states

$$\mathbb{E}\sup_{\mathbf{z}\in\Sigma_s} \langle \mathbf{a}_g, \mathbf{z}\rangle \leq 4\sqrt{2}\int_0^1 \sqrt{\ln\left(\mathcal{C}\left(\Sigma_s, \|\cdot\|_{\ell_2}, u\right)\right)}\mathrm{d}u,$$

where $\mathcal{C}(\Sigma_s, \|\cdot\|_{\ell_2}, u)$ are covering numbers associated with the set Σ_s. They are defined as the smallest cardinality of an u-covering net with respect to the Euclidean distance. A volumetric counting argument yields $\mathcal{C}(\Sigma_s, \|\cdot\|_{\ell_2}, u) \leq \left(\frac{en}{s}\right)^s \left(1 + \frac{2}{u}\right)^s$ and Dudley's inequality therefore implies

$$\ell(T_s) \leq c\sqrt{s\log\left(en/s\right)},$$

[3] This is a $m \times n$ matrix where each entry is sampled independently from the standard normal distribution $\mathcal{N}(0, 1)$.

where c is an absolute constant. This readily yields the following assertion (choose $t = \sqrt{2\tilde{c}m}$ for $\tilde{c}$ sufficiently small).

Theorem 4. (NSP for Gaussian measurements.) *A number of* $m \geq c^2 s \log(en/s)$ *independent real-valued Gaussian measurements obeys the (real-valued) s-NSP with probability at least* $1 - \mathrm{e}^{-\tilde{c}m}$.

This argument is exemplary for generic proof techniques: strong results from probability theory allow for establishing close-to-optimal recovery guarantees in a relatively succinct fashion.

2.2. Extending the Scope to Subgaussian Measurements

The extended arguments presented here are largely due to Dirksen, Lecue and Rauhut [36]. Again, we will focus on the real-valued case.

Gordon's escape through a mesh is only valid for Gaussian random matrices $\mathbf{A}$. Novel methods are required to extend this proof technique beyond this idealized case. Comparatively recently, Mendelson provided such a generalization [40, 41].

Theorem 5. (Mendelson's small ball method, Tropp's formulation [37]). *Suppose that* $\mathbf{A}$ *is a random* $m \times n$ *matrix whose rows correspond to m independent realizations of a random vector* $\mathbf{a} \in \mathbb{R}^n$. *Fix a set* $E \subseteq \mathbb{R}^n$, *and define,*

$$Q_\xi(\mathbf{a}, E) = \inf_{\mathbf{z}\in E} \Pr\left[|\langle \mathbf{z}, \mathbf{a}\rangle| \geq \xi\right] \quad \text{for} \quad \xi > 0,$$

$$W_m(\mathbf{a}, E) = \mathbb{E} \sup_{\mathbf{z}\in E} \langle \mathbf{z}, \mathbf{h}\rangle \quad \text{where} \quad \mathbf{h} = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \epsilon_i \mathbf{a}_i \in \mathbb{R}^n,$$

is the empirical average over m independent copies of $\mathbf{a}$ *weighted by uniformly random signs* $\epsilon_i \sim \{\pm 1\}$. *Then, for any* $t, \xi > 0$

$$\inf_{\mathbf{z}\in E} \|\mathbf{A}\mathbf{z}\|_{\ell_2} \geq \xi\sqrt{m} Q_{2\xi}(\mathbf{a}, E) - 2W_m(\mathbf{a}, E) - \xi t$$

with probability at least $1 - 2\mathrm{e}^{-t^2/2}$.

It is worthwhile to point out that for real-valued Gaussian vectors this result recovers Theorem 3 up to constants. Fix $\xi > 0$ of appropriate size. Then, $E \subseteq \mathbb{S}^{n-1}$ ensures that $\xi Q_{2\xi}(\mathbf{a}_g, E)$ is constant. Moreover, $W_m(\mathbf{a}_g, E)$ reduces to the usual Gaussian width $\ell(E)$.

Mendelson's small ball method can be used to establish the nullspace property for independent random measurements $\mathbf{a} \in \mathbb{R}^n$ that exhibit *subgaussian* behavior:

$$\mathbb{E} \exp\left(\theta \langle \mathbf{z}, \mathbf{a}\rangle\right) \leq \exp\left(\frac{\theta^2}{2} \|\mathbf{z}\|_{\ell_2}^2\right) \quad \text{for all} \quad \mathbf{z} \in \mathbb{R}^n,\ \theta > 0. \tag{11}$$

Signed Bernoulli vectors are a concrete example. Such random vectors are isotropic (3) and direct computation also reveals

$$\begin{aligned}\mathbb{E}\left[\langle \mathbf{z}, \mathbf{a}_{\mathrm{sb}}\rangle^4\right] &= \sum_{i,j,k,l=1}^{n} \mathbb{E}\left[\epsilon_i\epsilon_j\epsilon_k\epsilon_l\right] z_i z_j z_k z_l = \sum_{i=1}^{n} \mathbb{E}\left[\epsilon_i^4\right] z_i^4 \\ &\quad + 3\sum_{i\neq j} \mathbb{E}\left[\epsilon_i^2\right] \mathbb{E}\left[\epsilon_j^2\right] z_i^2 z_j^2 \\ &= \sum_{i=1}^{n} z_i^4 + 3\sum_{i\neq j} z_i^2 z_j^2 = 3\|\mathbf{z}\|_{\ell_2}^4 - 2\|\mathbf{z}\|_{\ell_4}^4 \leq 3\|\mathbf{z}\|_{\ell_2}^4,\end{aligned} \tag{12}$$

because there are 3 possible pairings of four indices. Next, set $E = T_s \subset \mathbb{S}^{n-1}$. An application of the Paley-Zygmund inequality then allows for bounding the parameter $Q_{2\xi}(\mathbf{a}_{\mathrm{sb}}, T_s)$ in Mendelson's small ball method from below:

$$\begin{aligned}Q_{2\xi}(\mathbf{a}_{\mathrm{sb}}, T_s) &\geq \inf_{\mathbf{z}\in\mathbb{S}^{n-1}} \Pr\left[|\langle \mathbf{z}, \mathbf{a}_{sb}\rangle| \geq 2\xi\right] = \inf_{\mathbf{z}\in\mathbb{S}^{n-1}} \\ &\qquad \Pr\left[\langle \mathbf{z}, \mathbf{a}_{sb}\rangle^2 \geq 4\xi^2 \mathbb{E}\left[\langle \mathbf{z}, \mathbf{a}_{sb}\rangle^2\right]\right] \\ &\geq \inf_{\mathbf{z}\in\mathbb{S}^{n-1}} \left(1 - 4\xi^2\right)^2 \frac{\mathbb{E}\left[\langle \mathbf{z}, \mathbf{a}_{sb}\rangle^2\right]^2}{\mathbb{E}\left[\langle \mathbf{z}, \mathbf{a}_{sb}\rangle^4\right]} \geq \frac{(1-4\xi^2)^2}{3}.\end{aligned}$$

This lower bound is constant for any $\xi \in (0, 1/2)$.

Next, note that $X_{\mathbf{z}} = \langle \mathbf{z}, \mathbf{h}\rangle$ is a stochastic process that is indexed by $\mathbf{z} \in \mathbb{R}^n$. This process is centered ($\mathbb{E}X_{\mathbf{z}} = 0$) and Equation (11) implies that it is also subgaussian. Moreover, $\mathbb{E}\left[|X_{\mathbf{z}} - X_{\mathbf{y}}|^2\right] = \|\mathbf{z} - \mathbf{y}\|_{\ell_2}^2$ readily follows from isotropy (3). Unlike Gordon's escape through a mesh, Dudley's inequality does remain valid for such stochastic processes with subgaussian marginals. We can now repeat the width analysis from the previous section to obtain

$$W_m(\mathbf{a}_{sb}, T_s) \leq 2\mathbb{E} \sup_{\mathbf{z}\in\Sigma_s} \langle \mathbf{z}, \mathbf{h}\rangle \leq c\sqrt{s\log(en/s)}.$$

Fixing $\xi > 0$ sufficiently small, setting $t = \sqrt{2\tilde{c}m}$ and inserting these bounds into Equation (5) yields the following result.

Theorem 6. (NSP for signed Bernoulli measurements.) *A matrix* $\mathbf{A}$ *encompassing* $m \geq Cs\log(en/s)$ *random signed Bernoulli measurements obeys the (real-valued) s-NSP with probability at least* $1 - \mathrm{e}^{-\tilde{c}m}$.

A similar result remains valid for other classes of independent measurements with subgaussian marginals (11).

2.3. Generalization to Complex-Valued Signals and Partial De-Randomization

The nullspace property, as well as its connection to uniform s-sparse recovery readily generalizes to complex-valued s-sparse vectors (see e.g., Foucart and Rauhut [3], section 4). In turn, Mendelson's small ball method may also be generalized to the complex-valued case:

Theorem 7. (Mendelson's small ball method for complex vector spaces.) *Suppose that the rows of* $\mathbf{A}$ *correspond to m independent copies of a random vector* $\mathbf{a} \in \mathbb{C}^n$. *Fix a set* $E \subseteq \mathbb{C}^n$ *and define*

$$Q_\xi(\mathbf{a}, E) = \inf_{\mathbf{z}\in E} \Pr\left[|\langle \mathbf{z}, \mathbf{a}\rangle| \geq \xi\right] \quad \text{for} \quad \xi > 0,$$

$$W_m(\mathbf{a}, E) = \mathbb{E} \sup_{\mathbf{z}\in E} |\langle \mathbf{z}, \mathbf{h}\rangle| \quad \textit{where} \quad \mathbf{h} = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \epsilon_i \mathbf{a}_i.$$

Then, for any $t, \xi > 0$

$$\inf_{\mathbf{z}\in E} \|\mathbf{A}\mathbf{z}\|_{\ell_2} \geq \sqrt{2} \left(\xi \sqrt{m} Q_{2^{3/2}\xi}/2 - 2W_m(E, \mathbf{a}) - \xi t \right)$$

with probability at least $1 - 2\mathrm{e}^{-t^2/2}$.

Such a generalization was conjectured by Tropp [37], but we are not aware of any rigorous proof in the literature. We provide one in subsection 5.1 and believe that this generalization may be of independent interest. In particular, this extension allows for generalizing the arguments from the previous subsection to the complex-valued case.

Let us now turn to the main scope of this work: partial de-randomization. Effectively, Mendelson's small ball method reduces the task of establishing nullspace properties to bounding the two parameters $Q_{2^{3/2}\xi}(\mathbf{a}, T_s)$ and $W_m(\mathbf{a}, T_s)$ in an appropriate fashion. A lower bound on the former readily follows from the Paley-Zygmund inequality, provided that the following relations hold for any $\mathbf{z} \in \mathbb{C}^n$:

$$\begin{aligned} \mathbb{E}\left[|\langle \mathbf{a}, \mathbf{z}\rangle|^2\right] &= \|\mathbf{z}\|_{\ell_2}^2 \qquad \text{(isotropy)}, \\ \mathbb{E}\left[|\langle \mathbf{a}, \mathbf{z}\rangle|^4\right] &\leq C_4 \|\mathbf{z}\|_{\ell_2}^4 \text{ (4h moment bound)}. \end{aligned}$$

Here, $C_4 > 0$ is constant.
Indeed, inserting these bounds into the Paley-Zygmund inequality yields

$$Q_{2^{3/2}\xi}(\mathbf{a}, T_s) \geq C_4^{-1} \left(1 - 8\xi^2\right)^2 \quad \text{for any} \quad \xi \in \left(0, 2^{-3/2}\right). \tag{13}$$

In contrast, establishing an upper bound on $W_m(\mathbf{a}, T_s)$ via Dudley's inequality requires subgaussian marginals (11) (that must not depend on the ambient dimension). This implicitly imposes stringent constraints on *all* moments of $\mathbf{a}$ simultaneously. An additional assumption allows to considerably weaken these demands:

$$\max_{1\leq k\leq n} |\langle \mathbf{e}_k, \mathbf{a}\rangle|^2 = 1 \quad \text{almost surely} \qquad \text{(incoherence)}. \tag{14}$$

Here, $\mathbf{e}_1, \ldots, \mathbf{e}_n$ denotes the standard basis of $\mathbb{C}^n$. Incoherence has long been identified as a key ingredient for developing s-sparse recovery guarantees. Here, we utilize it to establish an upper bound on $W_m(\mathbf{A}, T_s)$ that does not rely on subgaussian marginals.

Lemma 2. *Let* $\mathbf{a} \in \mathbb{C}^n$ *be a random vector that is isotropic and incoherent. Let* $T_s \subset \mathbb{C}^n$ *be the complex-valued generalization of the set defined in Equation (8) and assume* $m \geq \log(2n)$. *Then,*

$$W_m(\mathbf{a}, T_s) \leq 4\sqrt{2s \log(2n)}. \tag{15}$$

This bound only requires an appropriate scaling of the first two moments (isotropy) but comes at a price. The bound scales logarithmically in n rather than n/s. We defer a proof of this statement to subsection 5.2 below. Inserting the bounds (13) and (15) into the assertion of Theorem 7 readily yields the main technical result of this work:

Theorem 8. *Suppose that* $\mathbf{a} \in \mathbb{C}^n$ *is a random vector that obeys incoherence, isotropy and the 4th moment bound. Then, choosing*

$$m \geq Cs \log(n)$$

instances of $\mathbf{a}$ *uniformly at random results in a measurement matrix* $\mathbf{A}$ *that obeys the complex-valued nullspace property of order* s *with probability at least* $1 - 2\mathrm{e}^{-\tilde{c}m}$.

In complete analogy to the real-valued case, the complex nullspace property ensures uniform recovery of s-sparse vectors $\mathbf{x} \in \mathbb{C}^n$ from $\mathbf{y} = \mathbf{A}\mathbf{x}$ via solving the convex optimization problem (1).

2.4. Recovery Guarantee for Strength-4 Orthogonal Arrays

Suppose that $\mathbf{a}_{oa} \in \{\pm 1\}^n$ is chosen uniformly from a strength-4 orthogonal array. By definition, each component a_i of $\mathbf{a}$ has unit modulus. This ensures *incoherence*. Moreover, $\mathbb{E}\left[a_i a_j\right] = \mathbb{E}\left[\epsilon_i \epsilon_j\right] = \delta_{ij}$, because 4-wise independence necessarily implies 2-wise independence. Isotropy then readily follows from (3). Finally, 4-wise independence suffices to establish the 4th moment bound. By assumption $\mathbb{E}\left[a_i a_j \bar{a}_k \bar{a}_l\right] = \mathbb{E}\left[\epsilon_i \epsilon_j \epsilon_k \epsilon_l\right]$ and we may thus infer

$$\begin{aligned} \mathbb{E}\left[|\langle \mathbf{z}, \mathbf{a}_{oa}\rangle|^4\right] &= \sum_{i,j,k,l=1}^{n} \mathbb{E}\left[\epsilon_i \epsilon_j \epsilon_k \epsilon_l\right] \bar{z}_i \bar{z}_j z_k z_l \\ &= \sum_{i=1}^{n} \mathbb{E}\left[\epsilon_i^4\right] |z_i|^4 + \sum_{i\neq j} \mathbb{E}\left[\epsilon_i^2\right] \mathbb{E}\left[\epsilon_j^2\right] \left(\bar{z}_i^2 z_j^2 + 2|z_i|^2 |z_j|^2\right) \\ &= 2\|\mathbf{z}\|_{\ell_2}^4 + \sum_{i\neq j} \bar{z}_i^2 z_j^2 - \|\mathbf{z}\|_{\ell_4}^4 \leq 3\|\mathbf{z}\|_{\ell_2}^4. \end{aligned}$$

In summary: $\mathbf{a}_{oa}$ meets all the requirements of Theorem 8. Theorem 1 then follows from the fact that the complex nullspace property ensures uniform recovery of all s-sparse signals simultaneously.

2.5. Recovery Guarantee for Mutually Unbiased Bases

Suppose that $\mathbf{a}_{mub} \in \mathbb{C}^n$ is chosen uniformly from a maximal set of n mutually unbiased bases (excluding the standard basis) whose elements are re-normalized to length $\sqrt{n}$. Random time-frequency shift of the Alltop sequence (6) are a concrete example for such a sampling procedure, provided that the dimension $n \geq 5$ is prime.

The vector $\mathbf{a}_{mub}$ is chosen from a union of n bases that are all mutually unbiased with respect to the standard basis, see Equation (5). Together with normalization to length $\sqrt{n}$, this readily establishes incoherence: $\max_{1\leq k\leq n} |\langle \mathbf{e}_k, \mathbf{a}\rangle|^2 = \frac{n}{n} = 1$ with probability one. Next, by assumption $\mathbf{a}_{mub}$ is chosen uniformly from a union of n re-scaled orthonormal bases. Let us

denote each of them by $\mathbf{b}_1^{(l)}, \dots, \mathbf{b}_n^{(l)}$, where $1 \leq l \leq n$ labels the different basis. Then,

$$\begin{aligned}\mathbb{E}\left[|\langle \mathbf{a}_{mub}, \mathbf{z}\rangle|^2\right] &= \frac{1}{n^2}\sum_{l=1}^{n}\sum_{i=1}^{n}|\sqrt{n}\langle \mathbf{b}_i^{(l)}, \mathbf{z}\rangle|^2 = \frac{1}{n}\sum_{l=1}^{n}\|\mathbf{z}\|_{\ell_2}^2 \\ &= \|\mathbf{z}\|_{\ell_2}^2 \quad \text{for all} \quad \mathbf{z} \in \mathbb{C}^n\end{aligned}$$

which implies isotropy. Finally, a maximal set of $(n+1)$ mutually unbiased bases – including the standard basis which we denote by $\mathbf{b}_k^{(n+1)} = \mathbf{e}_k$ – forms a 2-design according to Definition 3. For any $\mathbf{z} \in \mathbb{C}^n$ this property ensures

$$\begin{aligned}\mathbb{E}\left[|\langle \mathbf{a}_{mub}, \mathbf{z}\rangle|^4\right] &= \sum_{l=1}^{n+1}\sum_{i=1}^{n}|\langle \mathbf{b}_i^{(l)}, \mathbf{z}\rangle|^4 - \sum_{k=1}^{n}|\langle \mathbf{e}_k, \mathbf{z}\rangle|^4 \\ &= 2\|\mathbf{z}\|_{\ell_2}^4 - \|\mathbf{z}\|_{\ell_4}^4 \leq 2\|\mathbf{z}\|_{\ell_2}^4\end{aligned}$$

which establishes the 4th moment bound. In summary, the random vector $\mathbf{a}_{mub} \in \mathbb{C}^n$ meets the requirements of Theorem 8. Theorem 2 then readily follows form the implications of the nullspace property for s-sparse recovery.

3. EXTENSION TO NOISY MEASUREMENTS

The nullspace property may be generalized to address two imperfections in s-sparse recovery simultaneously: (i) the vector $\mathbf{x} \in \mathbb{C}^d$ may only be approximately sparse in the sense that it is well-approximated by an s-sparse vector, (ii) the measurements may be corrupted by additive noise: $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{s}$ with $\mathbf{s} \in \mathbb{C}^m$.

To state this generalization, we need some additional notation. For $\mathbf{z} \in \mathbb{C}^n$ and $s \in [n]$, let $\mathbf{z}_s \in \mathbb{C}^n$ be the vector that only contains the s largest entries in modulus. All other entries are set to zero. Likewise, we write $\mathbf{z}_{\bar{s}} = \mathbf{z} - \mathbf{z}_s$ to denote the remainder. In particular, $\sigma_s(\mathbf{z}) = \|\mathbf{z}_{\bar{s}}\|_{\ell_1}$. An $m \times n$ matrix $\mathbf{A}$ obeys the *robust nullspace property* of order s with parameters $\rho \in (0,1)$ and $\tau > 0$ if

$$\|\mathbf{z}_s\|_{\ell_2} \leq \frac{\rho}{\sqrt{s}}\|\mathbf{z}_{\bar{s}}\|_{\ell_1} + \tau\|\mathbf{A}\mathbf{z}\|_{\ell_2} \quad \text{for all} \quad \mathbf{z} \in \mathbb{S}^{n-1},$$

see e.g., Foucart and Rauhut [3, Definition 4.21]. This extension of the nullspace property is closely related to stable s-sparse recovery via basis pursuit denoising:

$$\begin{aligned}&\text{minimize} \quad \|\mathbf{z}\|_{\ell_1} \\ &\text{subject to} \quad \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_{\ell_2} \leq \eta.\end{aligned} \tag{16}$$

Here, $\eta > 0$ denotes an upper bound on the strength of the noise corruption: $\|\mathbf{s}\|_{\ell_2} \leq \eta$. [3, Theorem 4.22] draws the following precise connection: suppose that $\mathbf{A}$ obeys the robust nullspace property with parameters ρ and τ. Then, the solution $\mathbf{z}^\sharp \in \mathbb{C}^n$ to (16) is guaranteed to obey

$$\|\mathbf{z}^\sharp - \mathbf{x}\|_{\ell_2} \leq \frac{D_1}{\sqrt{s}}\sigma_s(\mathbf{x}) + D_2\eta, \tag{17}$$

where $D_1 = (1+\rho)^2/(1-\rho)$ and $D_2 = (3+\rho)\tau/(1-\rho)$. The first term on the r.h.s. vanishes if $\mathbf{x}$ is exactly s-sparse and remains small if $\mathbf{x}$ is well approximated by a s-sparse vector. The second term scales linearly in the noise bound $\eta \geq \|\mathbf{s}\|_{\ell_2}$ and vanishes in the absence of noise corruption.

In the previous section, we have established the classical nullspace property for measurements that are chosen independently from a vector distribution that is isotropic, incoherent and obeys a bound on the 4th moments. This argument may readily be extended to establish the robust nullspace property with relatively little extra effort. To this end, define the set

$$T_{\rho,s} = \left\{\mathbf{z} \in \mathbb{S}^{n-1} : \ \|\mathbf{z}_s\|_{\ell_2} > \frac{\rho}{\sqrt{s}}\|\mathbf{z}_{\bar{s}}\|_{\ell_1}\right\}.$$

A moment of thought reveals that the matrix $\mathbf{A}$ obeys the robust nullspace property with parameters ρ, τ if

$$\inf_{\mathbf{z}\in T_{\rho,s}}\|\mathbf{A}\mathbf{z}\|_{\ell_2} \geq \frac{1}{\tau}. \tag{18}$$

What is more, the following inclusion formula is also valid:

$$T_{\rho,s} \subset \frac{3}{\rho}\mathrm{conv}\left(\Sigma_s\right),$$

see Kabanava and Rauhut [35, Lemma 3] and Rudelson and Vershynin [14, Lemma 4.5]. This ensures that the bounds on the parameters in Mendelson's small ball method generalize in a rather straightforward fashion. Isotropy, incoherence and the 4th moment bound ensure

$$Q_{2\xi}(\mathbf{a}, T_{\rho,s}) \geq \frac{(1-2\xi^2)^2}{C_4} \quad \text{and} \quad W_m(\mathbf{a}, T_{\rho,s}) \leq \frac{12}{\rho}\sqrt{2s\log(2n)}.$$

Now, suppose that $\mathbf{A}$ subsumes $m \geq C\rho^{-2}s\log(2n)$ independent copies of the random vector $\mathbf{a} \in \mathbb{C}^n$. Then, Theorem 7 readily asserts that with probability at least $1 - 2\mathrm{e}^{-\tilde{c}m}$

$$\inf_{\mathbf{z}\in T_{\rho,s}}\|\mathbf{A}\mathbf{z}\|_{\ell_2} \geq \frac{c}{\rho}\sqrt{m}, \tag{19}$$

where $c > 0$ is another constant. Previously, we employed Mendelson's small ball method to merely assert that a similar infimum is not equal to zero. Equation (19) provides a strictly positive lower bound with comparable effort. Comparing this relation to Equation (18) highlights that this is enough to establish the robust nullspace property with parameters ρ and $\tau = \frac{\rho}{c\sqrt{m}}$. In turn, a stable generalization of the main recovery guarantee follows from Equation (17).

Theorem 9. *Fix $\rho \in (0,1)$ and $s \in \mathbb{N}$. Suppose that we sample $m \geq C\rho^{-2}s\log(n)$ independent copies of an isotropic, incoherent random vector $\mathbf{a} \in \mathbb{C}^n$ that also obeys the 4th moment bound. Then, with probability at least $1 - 2\mathrm{e}^{-\tilde{c}m}$, the resulting measurement matrix $\mathbf{A}$ allows for stable, uniform recovery of*

(approximately) s-sparse vectors. More precisely, the solution $\mathbf{z}^\sharp$ *to (16) is guaranteed to obey*

$$\|\mathbf{x}-\mathbf{z}^\sharp\|_{\ell_2} \leq \frac{D_1}{\sqrt{s}}\sigma_s(\mathbf{x}) + D_2\frac{\eta}{\sqrt{m}},$$

where $D_1, D_2 > 0$ *depend only on* ρ.

4. NUMERICAL EXPERIMENTS

In this part we demonstrate the performance which can be achieved with our proposed derandomized constructions and we compare this to generic measurement matrices (Gaussian, signed Bernoulli). However, since the orthogonal array construction is more involved we first provide additional details relevant for numerical experiments.

Details on orthogonal arrays: An orthogonal array $\mathrm{OA}(\lambda\sigma^k, n, \sigma, k)$ of strength k, with n factors and σ levels are an $\lambda\sigma^k \times n$ array of σ different symbols such that in any k columns every ordered σ^k-tuple occurs in exactly λ rows. Arrays with $\lambda = 1$ are called simple. A comprehensive treatment can be found in the book [16]. Known arrays are listed in several libraries[4]. Often the symbol alphabet is not relevant, but we use the set $\mathbb{Z}_\sigma = \{0, \dots, \sigma-1\}$ for concreteness. Such arrays can be represented as a matrix in $\mathbb{Z}_\sigma^{\lambda\sigma^k \times n}$. For $\sigma = q^p$ with q prime the simple orthogonal array $\mathrm{OA}(\sigma^k, n, \sigma, k)$ is *linear* if the q^{pt} rows of the matrix form a vector space over $\mathbb{F}_q$. The runs of an orthogonal array (the rows of the corresponding matrix) can also be interpreted as codewords of a code and vice versa. The array is linear if and only if the corresponding code is linear [16, Chapter 4]. This relationship allows to employ classical code constructions to construct orthogonal arrays.

In this work we propose to generate sampling matrices $\mathbf{A} \in \mathbb{Z}_\sigma^{m\times n}$ by selecting $m \leq M = \lambda\sigma^k$ rows at random from an orthogonal array $\mathrm{OA}(\lambda\sigma^4, n, \sigma, 4)$ of strength $k = 4$ and with n factors. Intuitively, $m\log_2(M)$ bits are then required to specify such a matrix $\mathbf{A}$. For $k = 4$, a classical lower bound due to Rao [42] demands

$$M = \lambda\sigma^4 \geq 1 + n + \binom{n}{2} = \Omega(n^2). \tag{20}$$

Arrays that saturate this bound are called tight (or complete). In summary, an order of $s\log^2(n)$ bits are required to sample a $m\times n$ matrix $\mathbf{A}$ with $m \geq Cs\log(n)$ rows according to this procedure.

Strength-4 Constructions: For compressed sensing applications, we want arrays with a large number of factors n since this corresponds to the ambient dimension of the sparse vectors to recover. On the other hand the run size M should scale "moderately" to describe the random matrices only with few bits. Most constructions use an existing orthogonal array as a *seed* to construct larger arrays. Known binary arrays of strength 4 are for example the simple array $\mathrm{OA}(16, 5, 2, 4)$, or $\mathrm{OA}(80, 6, 2, 4)$. Pat [43] proposes an algorithm that uses a linear orthogonal array $\mathrm{OA}(N, n, \sigma, k)$ as a seed to construct a linear orthogonal array $\mathrm{OA}(N^2, n^2+2n, \sigma, k)$. This procedure may then be iterated.

Numerical results for orthogonal arrays: Figure 1 summarizes the empirical performance of basis pursuit (1) from independent orthogonal array measurements. We consider real-valued signals $\mathbf{x}$ and quantify the performance in terms of the normalized ℓ_2-recovery error (NMSE) $\|\mathbf{z}^\sharp - \mathbf{x}\|_{\ell_2}/\|\mathbf{x}\|_{\ell_2}$ where $\mathbf{z}^\sharp$ is the solution of (1). To construct the orthogonal array, algorithm [43] is applied twice $\mathrm{OA}(16, 5, 2, 4) \to \mathrm{OA}(256, 35, 2, 4) \to \mathrm{OA}(65536, 1295, 2, 4)$. The 323 rows are uniformly samples from this array, i.e., the sampling matrix $\mathbf{A}$ has ± 1 entries (instead of binary entries) and size 323×1295. But note that, in the case of non-negative sparse vectors, the corresponding binary 0/1-matrices may be used instead directly to recover with non-negative least-squares [44]. The sparsity of the unknown vector has been varied between $1 \dots 180$. For each sparsity many experiments are performed to compute NMSE. In each run, the support of the unknown vector has been chosen uniformly at random and the values are independent instances of a standard Gaussian random variable. For comparison, we have also included the corresponding performances of a generic sampling matrix (signed Bernoulli) of the same size. Numerically, the partially derandomized orthogonal array construction achieves essentially the same performance as its generic counterpart.

Numerical results for the Alltop design: Figure 1 shows the NMSE achieved for measurement matrices based on subsampling from an Alltop-design (6). The data is obtained in the same way as above, but the sparse vectors are generated as iid. complex-normal distributed on the support. For comparison the results for a (complex) standard Gaussian sampling matrix are included as well. Again, the performance of random Alltop-design measurements essentially matches its generic (Gaussian) counterpart.

5. ADDITIONAL PROOFS

5.1. Proof of Theorem 7

The proof is based on rather straightforward modifications of Tropp's proof for Mendelson's small ball method [37]. Let $\mathbf{a} \in \mathbb{C}^n$ be a complex-valued random vector. Suppose that $\mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbb{C}^n$ are independent copies of $\mathbf{a}$ and let $\mathbf{A}$ be the $m \times n$ matrix whose i-th row is given by $\mathbf{a}_i$. The goal is to obtain a lower bound on $\inf_{\mathbf{z}\in E}\|\mathbf{A}\mathbf{z}\|_{\ell_2}$, where $E \subset \mathbb{C}^n$ is an arbitrary set. First, note that ℓ_1 and ℓ_2 norms on $\mathbb{R}^{2m}$ are related via $\|\mathbf{v}\|_{\ell_1} \leq \sqrt{2m}\|\mathbf{v}\|_{\ell_2}$. For any $\mathbf{z} \in E$ this relation ensures

$$\begin{aligned}
\|\mathbf{A}\mathbf{z}\|_{\ell_2} &= \sqrt{\sum_{i=1}^m |\langle \mathbf{a}_i, \mathbf{z}\rangle|^2} = \sqrt{\sum_{i=1}^m \mathrm{Re}\left(\langle \mathbf{a}_i, \mathbf{z}\rangle\right)^2 + \sum_{i=1}^m \mathrm{Im}\left(\langle \mathbf{a}_i, \mathbf{z}\rangle\right)^2} \\
&\geq \frac{1}{\sqrt{2m}}\left(\sum_{i=1}^m |\mathrm{Re}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| + \sum_{i=1}^m |\mathrm{Im}(\langle \mathbf{a}_i, \mathbf{z}\rangle)|\right) \\
&= \frac{1}{\sqrt{2m}}\sum_{i=1}^m \left(|\mathrm{Re}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| + |\mathrm{Im}(\langle \mathbf{a}_i, \mathbf{z}\rangle)|\right).
\end{aligned}$$

[4] For example http://neilsloane.com/oadir/ or http://pietereendebak.nl/oapage/

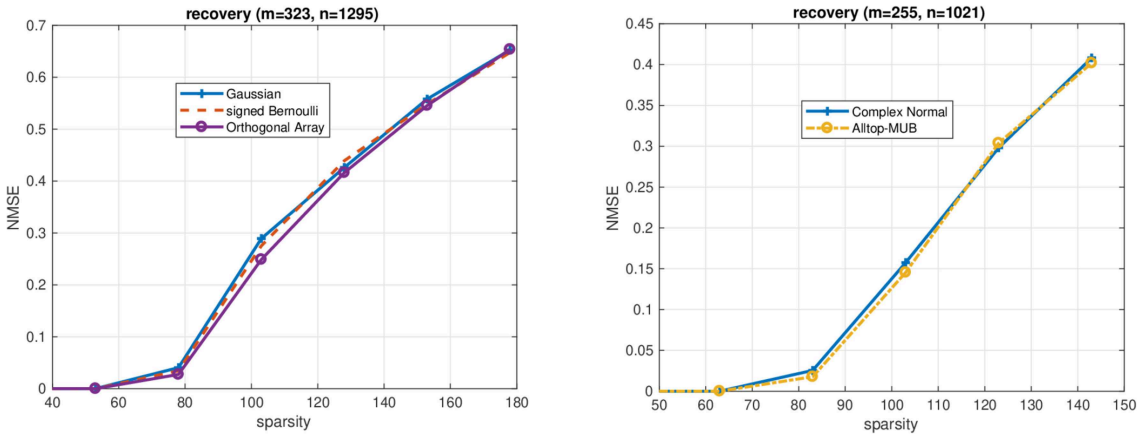

FIGURE 1 | (**Left**) Performance of basis pursuit for $m = 323$ and $n = 1295$ (real-valued signals) from random orthogonal array measurements and their generic counterpart (signed Bernoulli). (**Right**) Performance of basis pursuit for $m = 255$ and $n = 1021$ (complex-valued signals) from random time-frequency shifted Alltop sequences and their generic counterpart (standard complex Gaussian vectors).

Next, fix $\xi > 0$ and introduce the indicator function $\mathbb{I}\{x \geq \xi\}$ which obeys $x \geq \xi \mathbb{I}\{x \geq \xi\}$ for all $x \geq 0$. Consequently,

$$\|\mathbf{Az}\|_{\ell_2} \geq \frac{\xi}{\sqrt{2m}} \sum_{i=1}^{m} \left(\mathbb{I}\left\{|\mathrm{Re}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq \xi\right\} + \mathbb{I}\left\{|\mathrm{Im}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq \xi\right\}\right). \tag{21}$$

Also, note that the expectation value of each summand obeys

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{I}\left\{|\mathrm{Re}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq \xi\right\}\right] + \mathbb{E}\left[\mathbb{I}\left\{|\mathrm{Im}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq \xi\right\}\right] \\
=&\Pr\left[|\mathrm{Re}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq \xi\right] + \Pr\left[|\mathrm{Im}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq \xi\right] \\
\geq&\Pr\left[|\mathrm{Re}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq \xi \ \vee \ |\mathrm{Im}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq \xi\right] \\
\geq&\Pr\left[|\langle \mathbf{a}_i, \mathbf{z}\rangle| \geq \sqrt{2}\xi\right],
\end{aligned}$$

according to the union bound. The last line follows from a simple observation. Let $z = a + ib$ be a complex number. Then, $|z| = \sqrt{a^2 + b^2} \geq \sqrt{2}\xi$ necessarily implies either $|a| \geq \xi$, or $|b| \geq \xi$ (or both). Next, define

$$Q_{2\xi}(\mathbf{z}) = \Pr\left[|\mathrm{Re}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq 2\xi\right] + \Pr\left[|\mathrm{Im}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq 2\xi\right]$$

and note that the estimate from above ensures

$$\inf_{\mathbf{z}\in E} Q_{2\xi}(\mathbf{z}) \geq \inf_{\mathbf{z}\in E} \Pr\left[|\langle \mathbf{a}, \mathbf{z}\rangle| \geq 2^{3/2}\xi\right] = Q_{2^{3/2}\xi}(\mathbf{a}, E). \tag{22}$$

Adding and subtracting $\xi\sqrt{m/2}Q_{2\xi}(\mathbf{z})$ to Equation (21) and taking the infimum yields

$$\begin{aligned}
&\inf_{\mathbf{z}\in E} \|\mathbf{Az}\|_{\ell_2} \\
&\geq \inf_{\mathbf{z}\in E} \left(\xi\sqrt{\frac{m}{2}}Q_{2\xi}(\mathbf{z}) - \xi\sqrt{\frac{m}{2}}Q_{2\xi}(\mathbf{z})\right. \\
&\quad \left. + \frac{\xi}{\sqrt{2m}} \sum_{i=1}^{m} \left(\mathbb{I}\left\{|\mathrm{Re}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq \xi\right\} + \mathbb{I}\left\{|\mathrm{Im}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq \xi\right\}\right)\right) \\
&\geq \xi\sqrt{\frac{m}{2}}Q_{2^{3/2}\xi}(\mathbf{a}, E) - \frac{\xi}{\sqrt{2m}} \sup_{\mathbf{z}\in E} \left(mQ_{2\xi}(\mathbf{z})\right. \\
&\quad \left. - \sum_{i=1}^{m} \left(\mathbb{I}\left\{|\mathrm{Re}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq \xi\right\} + \mathbb{I}\left\{|\mathrm{Im}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq \xi\right\}\right)\right).
\end{aligned} \tag{23}$$

Here we have applied Equation (22) to bound the contribution of the first term. Since $Q_{2\xi}(\mathbf{z})$ features both a real and imaginary part and we can split up the remaining supremum accordingly. The suprema over real and complex parts individually correspond to

$$R(E, \mathbf{a}) = \sup_{\mathbf{z}\in E} \sum_{i=1}^{m} \left(\Pr\left[|\mathrm{Re}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq 2\xi\right] - \mathbb{I}\left\{|\mathrm{Re}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq \xi\right\}\right),$$

$$I(E, \mathbf{a}) = \sup_{\mathbf{z}\in E} \sum_{i=1}^{m} \left(\Pr\left[|\mathrm{Im}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq 2\xi\right] - \mathbb{I}\left\{|\mathrm{Im}(\langle \mathbf{a}_i, \mathbf{z}\rangle)| \geq \xi\right\}\right),$$

and the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$ are independent copies of a single random vector $\mathbf{a} \in \mathbb{C}^n$. The bounded difference inequality [45, section 6.1] asserts that both expressions concentrate around their expectation. More precisely, for any $t > 0$

$$\Pr\left[R(E, \mathbf{a}) \geq \mathbb{E}R(E, \mathbf{a}) + t\sqrt{m}\right] \leq \mathrm{e}^{-t^2/2} \quad \text{and}$$
$$\Pr\left[I(E, \mathbf{a}) \geq \mathbb{E}I(E, \mathbf{a}) + t\sqrt{m}\right] \leq \mathrm{e}^{-t^2/2}.$$

Therefore, the union bound grants a transition from $R(E, \mathbf{a}) + I(E, \mathbf{a})$ to $\mathbb{E}R(E, \mathbf{a}) + \mathbb{E}I(E, \mathbf{a}) + 2\sqrt{m}t$ with probability at least $1 - 2\mathrm{e}^{-t^2/2}$. These expectation values can be further simplified. Define the soft indicator function

$$\psi_\xi(s) = \begin{cases} 0 & |s| \leq \xi, \\ (|s| - \xi)/\xi & \xi \leq |s| \leq 2\xi, \\ 1 & |s| \geq 2\xi \end{cases}$$

which admits the following bounds: $\mathbb{I}\{|s| \geq 2\xi\} \leq \psi_\xi(s) \leq \mathbb{I}\{|s| \geq \xi\}$ for all $s \in \mathbb{R}$. Moreover, $\xi\psi_\xi(s)$ is a contraction, i.e., a

real-valued function with Lipschitz constant one that also obeys $\xi\,\psi_\xi(0) = 0$. Rademacher symmetrization [3, Lemma 8.4] and the Rademacher comparison principle [46, Equation (4.20)] yield

$$\begin{aligned}\mathbb{E}\, R(E,\mathbf{a}) =& \mathbb{E}\sup_{\mathbf{z}\in E}\sum_{i=1}^{m}\left(\mathbb{E}\mathbb{I}\left\{|\mathrm{Re}(\mathbf{a}_i,\mathbf{z})|\geq 2\xi\right\} - \mathbb{I}\left\{|\mathrm{Re}(\langle \mathbf{a}_i,\mathbf{z}\rangle)|\geq \xi\right\}\right)\\ \leq& \mathbb{E}\sup_{\mathbf{z}\in E}\sum_{i=1}^{m}\left(\mathbb{E}\psi_\xi(\mathrm{Re}(\langle \mathbf{a}_i,\mathbf{z}\rangle)) - \psi_\xi(\mathrm{Re}(\langle \mathbf{a}_i,\mathbf{z}\rangle))\right)\\ \leq& 2\mathbb{E}\sum_{i=1}^{m}\epsilon_i\psi_\xi(\mathrm{Re}(\langle \mathbf{a}_i,\mathbf{z}\rangle) \leq \frac{2}{\xi}\mathbb{E}\sup_{\mathbf{z}\in E}\sum_{i=1}^{m}\epsilon_i\mathrm{Re}(\langle \mathbf{a}_i,\mathbf{z}\rangle)\\ \leq& \frac{2\sqrt{m}}{\xi}\mathbb{E}\sup_{\mathbf{z}\in E}|\langle \mathbf{z},\mathbf{h}\rangle|,\end{aligned}$$

where $\mathbf{h} = \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\epsilon_i\mathbf{a}_i \in \mathbb{C}^n$. A completely analogous bound holds true for $\mathbb{E}I(E,\mathbf{a})$. Inserting both bounds into Equation (23) establishes

$$\begin{aligned}\inf_{\mathbf{z}\in E}\|\mathbf{A}\mathbf{z}\|_{\ell_2} \geq& \xi\sqrt{\frac{m}{2}}Q_{2^{3/2}\xi} - \frac{\xi}{\sqrt{2m}}\left(\frac{4\sqrt{m}}{\xi}\mathbb{E}\sup_{\mathbf{z}\in E}|\langle \mathbf{z},\mathbf{h}\rangle| + 2\sqrt{mt}\right)\\ =& \xi\sqrt{\frac{m}{2}}Q_{2^{3/2}\xi} - 2^{3/2}\mathbb{E}\sup_{\mathbf{z}\in E}|\langle \mathbf{z},\mathbf{h}\rangle| - \sqrt{2}\xi t\end{aligned}$$

with probability at least $1 - 2\mathrm{e}^{-t^2/2}$. Setting $W_m(E,\mathbf{z}) = \mathbb{E}\sup_{\mathbf{z}\in E}|\langle \mathbf{z},\mathbf{h}\rangle|$ establishes the claim.

5.2. Proof of Lemma 2

The inclusion $T_s \subset 2\mathrm{conv}(\Sigma_s)$ remains valid in the complex case. Moreover, every $\mathbf{z}\in\mathrm{conv}(\Sigma_s)$ necessarily obeys

$$\max_{\mathbf{z}\in\mathrm{conv}(\Sigma_s)}\|\mathbf{z}\|_{\ell_1} \leq \max_{\mathbf{z}\in\Sigma_s}\|\mathbf{z}\|_{\ell_1} = \sqrt{s},$$

because the maximum value of a convex function is achieved at the boundary. Hoelder's inequality therefore implies

$$\begin{aligned}W_m(\mathbf{a},T_s) =& \mathbb{E}\sup_{\mathbf{z}\in T_s}|\langle \mathbf{z},\mathbf{h}\rangle| \leq 2\mathbb{E}\sup_{\mathbf{z}\in\mathrm{conv}(\Sigma_s)}\|\mathbf{z}\|_{\ell_1}\|\mathbf{h}\|_{\ell_\infty}\\ \leq& 2\sqrt{s}\,\mathbb{E}\|\mathbf{h}\|_{\ell_\infty},\end{aligned} \tag{24}$$

where $\mathbf{h} = \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\epsilon_i\mathbf{a}_i \in \mathbb{C}^n$. Moreover,

$$\begin{aligned}\mathbb{E}\|\mathbf{h}\|_{\ell_\infty} = \mathbb{E}\max_{1\leq k\leq n}|\langle \mathbf{e}_k,\mathbf{h}\rangle| \leq& \mathbb{E}\max_{1\leq k\leq n}\left|\mathrm{Re}(\langle \mathbf{e}_k,\mathbf{h}\rangle)\right|\\ &+\mathbb{E}\max_{1\leq k\leq n}|\mathrm{Im}(\langle \mathbf{e}_k,\mathbf{h}\rangle|\end{aligned}$$

and we may bound both expressions on the r.h.s. independently. For the first term, fix $\theta > 0$ and use Jensen's inequality (the logarithm is a concave function) to obtain

$$\begin{aligned}\mathbb{E}\max_{1\leq k\leq n}|\mathrm{Re}(\langle \mathbf{e}_k,\mathbf{h}\rangle)| =& \mathbb{E}\max_{1\leq k\leq n}\max_{\sigma=\pm}\sigma\,\mathrm{Re}(\langle \mathbf{e}_k,\mathbf{h}\rangle)\\ \leq& \frac{1}{\theta}\log\left(\mathbb{E}\exp\left(\max_{1\leq k\leq n}\max_{\sigma=\pm}\theta\sigma\,\mathrm{Re}\left(\langle \mathbf{e}_k,\mathbf{h}\rangle\right)\right)\right).\end{aligned}$$

Monotonicity and non-negativity of the exponential function then imply

$$\begin{aligned}&\mathbb{E}\exp\left(\max_{1\leq k\leq n}\max_{\sigma=\pm}\theta\sigma\,\mathrm{Re}\left(\langle \mathbf{e}_k,\mathbf{h}\rangle\right)\right)\\ &\quad\leq \sum_{k=1}^{n}\sum_{\sigma=\pm}\mathbb{E}\exp\left(\theta\sigma\,\mathrm{Re}\left(\langle \mathbf{e}_k,\mathbf{h}\rangle\right)\right)\\ &\quad= \sum_{k=1}^{n}\sum_{\sigma=\pm}\prod_{i=1}^{m}\mathbb{E}\exp\left(\frac{\theta\sigma}{\sqrt{m}}\epsilon_i\mathrm{Re}\left(\langle \mathbf{e}_k,\mathbf{a}_i\rangle\right)\right),\end{aligned}$$

where we have also used that all ϵ_i's and $\mathbf{a}_i$'s are independent. The remaining moment generating functions can be bounded individually. Fix $1 \leq k \leq n$, $\sigma \in \{\pm 1\}$ and $1 \leq i \leq m$ and exploit the Rademacher randomness to infer

$$\begin{aligned}\mathbb{E}\exp\left(\frac{\theta\sigma}{\sqrt{m}}\epsilon_i\mathrm{Re}\left(\langle \mathbf{e}_k,\mathbf{a}_i\rangle\right)\right) =& \mathbb{E}\cosh\left(\frac{\theta\sigma}{\sqrt{m}}\mathrm{Re}\left(\langle \mathbf{e}_k,\mathbf{a}_i\rangle\right)\right)\\ \leq& \mathbb{E}\exp\left(\frac{\theta^2\sigma^2}{2m}\mathrm{Re}\left(\langle \mathbf{e}_k,\mathbf{a}_i\rangle\right)^2\right)\\ =& \mathbb{E}\exp\left(\frac{\theta^2}{2m}\mathrm{Re}\left(\langle \mathbf{e}_k,\mathbf{a}_i\rangle\right)^2\right),\end{aligned}$$

because $\sigma^2 = 1$. Incoherence moreover ensures $(\mathrm{Re}(\langle \mathbf{e}_k,\mathbf{a}_i\rangle)^2 \leq |\langle \mathbf{e}_k,\mathbf{a}_i\rangle|^2 \leq 1$. This ensures that the remaining expectation value is upper-bounded by $\exp\left(\frac{\theta^2}{2m}\right)$. Inserting these individual bounds into the original expression yields

$$\begin{aligned}\mathbb{E}\max_{1\leq k\leq n}\left|\mathrm{Re}(\langle \mathbf{e}_k,\mathbf{h}\rangle)\right| \leq& \frac{1}{\theta}\log\left(\sum_{k=1}^{n}\sum_{\sigma=\pm}\prod_{i=1}^{m}\mathbb{E}\exp\left(\frac{\theta\sigma}{\sqrt{m}}\epsilon_i\mathrm{Re}\left(\langle \mathbf{e}_k,\mathbf{a}_i\rangle\right)\right)\right)\\ \leq& \frac{1}{\theta}\log\left(\sum_{k=1}^{n}\sum_{\sigma=\pm}\prod_{i=1}^{m}\exp\left(\frac{\theta^2}{2m}\right)\right)\\ =& \frac{1}{\theta}\log\left(2n\exp\left(\frac{\theta^2}{2}\right)\right) = \frac{\log(2n)}{\theta} + \frac{\theta}{2}\end{aligned}$$

for any $0 < \theta \leq \sqrt{2m}$. Choosing $\theta = \sqrt{2\log(2n)}$ minimizes this upper bound and is feasible, by assumption. A completely analogous bound can be derived for the expected maximum absolute value of the imaginary part. Combining both yields

$$\mathbb{E}\|\mathbf{h}\|_{\ell_\infty} \leq \sqrt{2\log(2n)} + \sqrt{2\log(2n)} = 2\sqrt{2\log(2n)}$$

and inserting this bound into (24) ensures

$$W_m(\mathbf{a},T_s) \leq 4\sqrt{2s\log(2n)}.$$

AUTHOR CONTRIBUTIONS

RK developed the technical aspects of the paper. PJ is responsible for the numerical aspects and the discussion of

orthogonal arrays. Some of the main conceptual ideas are due to DM. These in particular include the idea of employing orthogonal arrays to partially derandomize signed Bernoulli matrices and the insight that a partial de-randomization of Gordon's escape through a mesh is essential for achieving the results. All authors contributed to the introduction and general presentation.

FUNDING

PJ is supported by the DFG grant JU 2795/3 and DAAD grant 57417688. RK was in part supported by Joel A. Tropp under the ONR Award No. N00014-17-12146 and also acknowledges funding provided by the Institute of Quantum Information and Matter, an NSF Physics Frontiers Center (NSF Grant PHY-1733907). DM was partially supported by AFOSR FA9550-18-1-0107, NSF DMS 1829955, and the Simons Institute of the Theory of Computing. We acknowledge support by the German Research Foundation and the Open Access Publication Fund of TU Berlin.

ACKNOWLEDGMENTS

The authors want to thank David Gross for providing the original ideas that ultimately led to this publication.

REFERENCES

1. Donoho DL. Compressed sensing. *IEEE Trans Inform Theory*. (2006) **52**:1289–306. doi: 10.1109/TIT.2006.871582
2. Candes EJ, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inform Theory*. (2006) **52**:489–509. doi: 10.1109/TIT.2005.862083
3. Foucart S, Rauhut H. *A Mathematical Introduction to Compressive Sensing. Applied and Numerical Harmonic Analysis.* New York, NY: Birkhäuser/Springer (2013). doi: 10.1007/978-0-8176-4948-7
4. Eldar YC, Kutyniok G. *Compressed Sensing: Theory and Applications.* Cambridge, UK: Cambridge University Press (2012).
5. Candès EJ, Strohmer T, Voroninski V. Phaselift: exact and stable signal recovery from magnitude measurements via convex programming. *Commun Pure Appl Math*. (2013) **66**:1241–74. doi: 10.1002/cpa.21432
6. Candès E, Li X. Solving quadratic equations via PhaseLift when there are about as many equations as unknowns. *Found Comput Math*. (2013). p. 1–10. doi: 10.1007/s10208-013-9162-z
7. Gross D, Krahmer F, Kueng R. A partial derandomization of phaseLift using spherical designs. *J Fourier Anal Appl*. (2015) **21**:229–66. doi: 10.1007/s00041-014-9361-2
8. Kueng R, Rauhut H, Terstiege U. Low rank matrix recovery from rank one measurements. *Appl Comput Harmon Anal*. (2017) **42**:88–116. doi: 10.1016/j.acha.2015.07.007
9. Kabanava M, Kueng R, Rauhut H, Terstiege U. Stable low-rank matrix recovery via null space properties. *Inform Inference*. (2016) **5**:405–41. doi: 10.1093/imaiai/iaw014
10. Kueng R, Gross D, Krahmer F. Spherical designs as a tool for derandomization: the case of PhaseLift. In: *2015 International Conference on Sampling Theory and Applications* (Washington, DC: SampTA) (2015). p. 192–6. doi: 10.1109/SAMPTA.2015.7148878
11. Kueng R, Zhu H, Gross D. Low rank matrix recovery from Clifford orbits. *arXiv* [Preprint] (2016). arXiv:161008070.
12. Tropp JA. A mathematical introduction to compressive sensing [Book Review]. *Bull Amer Math Soc*. (2017) **54**:151–65. doi: 10.1090/bull/1546
13. Bandeira AS, Fickus M, Mixon DG, Wong P. The road to deterministic matrices with the restricted isometry property. *J Fourier Anal Appl*. (2013) **19**:1123–49. doi: 10.1007/s00041-013-9293-2
14. Rudelson M, Vershynin R. On sparse reconstruction from Fourier and Gaussian measurements. *Commun Pure Appl Math*. (2008) **61**:1025–45. doi: 10.1002/cpa.20227
15. Cheraghchi M, Guruswami V, Velingker A. Restricted isometry of fourier matrices and list decodability of random linear codes. *SIAM J Comput*. (2013) **42**:1888–914. doi: 10.1137/120896773
16. Hedayat AS, Sloane NJA, Stufken J. *Orthogonal Arrays: Theory and Applications* New York, NY: Springer-Verlag (1999). doi: 10.1007/978-1-4612-1478-6
17. Delsarte P, Goethals JM, Seidel JJ. Spherical codes and designs. *Geometriae Dedicata*. (1997) **6**:363–88. doi: 10.1007/BF03187604
18. Bajnok B. Construction of spherical *t*-designs. *Geom Dedicata*. (1992) **43**:167–79. doi: 10.1007/BF00147866
19. Korevaar J, Meyers JLH. Chebyshev-type quadrature on multidimensional domains. *J Approx Theory*. (1994) **79**:144–64. doi: 10.1006/jath.1994.1119
20. Hayashi A, Hashimoto T, Horibe M. Reexamination of optimal quantum state estimation of pure states. *Phys Rev A*. (2005) **72**:032325. doi: 10.1103/PhysRevA.72.032325
21. Bandyopadhyay S, Boykin PO, Roychowdhury V, Vatan F. A new proof for the existence of mutually unbiased bases. *Algorithmica*. (2002) **34**:512–8. doi: 10.1007/s00453-002-0980-7
22. Klappenecker A, Rotteler M. Mutually unbiased bases are complex projective 2-designs. In: A. Klappenecker, M. Roetteler Editors. *Proceedings International Symposium on Information Theory, 2005*. (Adelaide, AU: ISIT) (2005). p. 1740–4. doi: 10.1109/ISIT.2005.1523643
23. Klappenecker A, Rötteler M. Constructions of Mutually Unbiased Bases. In: Mullen GL, Poli A, Stichtenoth H, editors. *Finite Fields and Applications*. Berlin; Heidelberg: Springer (2004). p. 137–44. doi: 10.1007/978-3-540-24633-6-10
24. Alltop W. Complex sequences with low periodic correlations. *IEEE Trans Inform Theory*. (1980) **26**:350–4. doi: 10.1109/TIT.1980.1056185
25. Bourgain J, Dilworth S, Ford K, Konyagin S, Kutzarova D. Explicit constructions of RIP matrices and related problems. *Duke Math J*. (2011) **159**:145–85. doi: 10.1215/00127094-1384809
26. Mixon DG. In: Boche H, Calderbank R, Kutyniok G, Vybíral J, editors. *Explicit Matrices With the Restricted Isometry Property: Breaking the Square-Root Bottleneck*. Cham: Springer International Publishing (2015). p. 389–417. doi: 10.1007/978-3-319-16042-9-13
27. Bandeira AS, Mixon DG, Moreira J. A conditional construction of restricted isometries. *Int Math Res Notices*. (2017) **2017**:372–81. doi: 10.1093/imrn/rnv385
28. Bandeira AS, Fickus M, Mixon DG, Moreira J. Derandomizing restricted isometries via the legendre symbol. *Constr Approx*. (2016) **43**:409–24. doi: 10.1007/s00365-015-9310-6
29. Baraniuk R, Davenport M, DeVore R, Wakin M. A simple proof of the restricted isometry property for random matrices. *Constr Approx*. (2008) **28**:253–63. doi: 10.1007/s00365-007-9003-x
30. Krahmer F, Ward R. New and improved johnson-Lindenstrauss embeddings via the restricted isometry property. *SIAM J Math Anal*. (2011) **43**:1269–81. doi: 10.1137/100810447
31. Kane DM, Nelson J. A derandomized sparse Johnson-Lindenstrauss transform. *arXiv* [Preprint] (2010). arXiv:10063585.
32. Duarte MF, Davenport MA, Takhar D, Laska JN, Sun T, Kelly KF, et al. Single-pixel imaging via compressive sampling. *IEEE Signal Process Magaz*. (2008) **25**:83–91. doi: 10.1109/MSP.2007.914730
33. Herman MA, Strohmer T. High-resolution radar via compressed sensing. *IEEE Trans Signal Process*. (2009) **57**:2275–84. doi: 10.1109/TSP.2009.2014277
34. Tropp JA, Dhillon IS, Heath RW, Strohmer T. CDMA signature sequences with low peak-to-average-power ratio via alternating projection. In: *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers*. Pacific Grove, CA (2003). p. 475–9.

35. Kabanava M, Rauhut H. Analysis ℓ_1-recovery with frames and gaussian measurements. *Acta Appl Math.* (2015) **140**:173–95. doi: 10.1007/s10440-014-9984-y
36. Dirksen S, Lecue G, Rauhut H. On the gap between restricted isometry properties and sparse recovery conditions. *IEEE Trans Inform Theory.* (2016) **99**:1.
37. Tropp JA. Convex Recovery of a Structured Signal from Independent Random Linear Measurements. In: Pfander EG, editor. *Sampling Theory, a Renaissance: Compressive Sensing and Other Developments.* Birkhäuser: Springer International Publishing (2015). p. 67–101. doi: 10.1007/978-3-319-19749-4-2
38. Gordon Y. On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$. In: Lindenstrauss J, Milman VD, editors. *Geometric Aspects of Functional Analysis.* Berlin; Heidelberg: Springer Berlin Heidelberg (1988). p. 84–106. doi: 10.1007/BFb0081737
39. Dudley R. Sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J Funct Anal.* (1967) **1**:290–330. doi: 10.1016/0022-1236(67)90017-1
40. Mendelson S. Learning without concentration. *J ACM.* (2015) **62**:21:25. doi: 10.1145/2699439
41. Koltchinskii V, Mendelson S. Bounding the smallest singular value of a random matrix without concentration. *Int Math Res Notices.* (2015) **2015**:12991–3008. doi: 10.1093/imrn/rnv096
42. Rao CR. Factorial experiments derivable from combinatorial arrangements of arrays. *J R Stat Soc.* (1947) **9**:128–39. doi: 10.2307/2983576
43. Pat A. On Construction of a Class of Orthogonal Arrays. Ph.D. Thesis, Department of Mathematics, Indian Institute of Technology, Kharagpur, India. (2012).
44. Kueng R, Jung P. Robust nonnegative sparse recovery and the nullspace property of 0/1 measurements. *IEEE Trans Inform Theory.* (2018) **64**:689–703. doi: 10.1109/TIT.2017.2746620
45. Boucheron S, Lugosi G, Massart P. *Concentration Inequalities: A Nonasymptotic Theory of Independence.* Oxford: Oxford University Press (2013). doi: 10.1093/acprof:oso/9780199535255.001.0001
46. Ledoux M, Talagrand M. *Probability in Banach Spaces: Isoperimetry and Processes.* Berlin; Heidelberg: Springer Science & Business Media (2013).

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.