SIAM J. COMPUT.
Vol. 53, No. 2, pp. 315–345

FAST METRIC EMBEDDING INTO THE HAMMING CUBE*

SJOERD DIRKSEN†, SHAHAR MENDELSON‡, AND ALEXANDER STOLLENWERK§

Abstract. We consider the problem of embedding a subset of $\mathbb{R}^n$ into a low-dimensional Hamming cube in an almost isometric way. We construct a simple, data-oblivious, and computationally efficient map that achieves this task with high probability; we first apply a specific structured random matrix, which we call the *double circulant matrix*; using that a matrix requires linear storage and matrix-vector multiplication that can be performed in near-linear time. We then binarize each vector by comparing each of its entries to a random threshold, selected uniformly at random from a well-chosen interval. We estimate the number of bits required for this encoding scheme in terms of two natural geometric complexity parameters of the set: its Euclidean covering numbers and its localized Gaussian complexity. The estimate we derive turns out to be the best that one can hope for, up to logarithmic terms. The key to the proof is a phenomenon of independent interest: we show that the double circulant matrix mimics the behavior of the Gaussian matrix in two important ways. First, it maps an arbitrary set in $\mathbb{R}^n$ into a set of well-spread vectors. Second, it yields a fast near-isometric embedding of any finite subset of ℓ_2^n into ℓ_1^m. This embedding achieves the same dimension reduction as the Gaussian matrix in near-linear time, under an optimal condition—up to logarithmic factors—on the number of points to be embedded. This improves a well-known construction due to Ailon and Chazelle.

Key words. dimension reduction, Johnson–Lindenstrauss embeddings, Hamming cube, circulant matrices, Gaussian width

MSC codes. 68R12, 60B20

DOI. 10.1137/22M1520220

1. Introduction.

In modern data analysis one is frequently confronted with sets that contain a large number of points, and each point is represented by a high-dimensional vector. This high-dimensionality causes significant storage consumption and comes at a high computational cost. In an attempt at addressing those issues, dimension reduction techniques have been used, for example, in clustering schemes [37], computational geometry [22], and numerical linear algebra [44, 46] (see, e.g., [6] and the references therein for many more examples). The idea is to map the given set into a lower-dimensional space, while preserving its key features. And obviously, what counts as a key feature changes according to the application one has in mind.

The Gaussian random matrix $A \in \mathbb{R}^{m\times n}$, whose entries are independent, standard Gaussian random variables, is a surprisingly powerful and versatile tool that is frequently used in dimension reduction methods. The most basic result of that flavor is the (Gaussian formulation of the) Johnson–Lindenstrauss lemma [31]: if $f : \mathbb{R}^n \to \mathbb{R}^m$ is defined by $f(x) = \frac{1}{\sqrt{m}}Ax$, then for any finite set T and $\epsilon > 0$,

*Received by the editors September 7, 2022; accepted for publication (in revised form) October 18, 2023; published electronically March 14, 2024.
https://doi.org/10.1137/22M1520220
Funding: The first author was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under SPP 1798 (COSIP - Compressed Sensing in Information Processing) through project CoCoMiMo. The second author acknowledges support by the Fonds de la Recherche Scientifique - FNRS under Grant T.0136.20 (Learn2Sense).

†Utrecht University, Mathematical Institute, Utrecht, Netherlands (s.dirksen@uu.nl).
‡University of Warwick, Department of Statistics, Coventry, England and The Australian National University, Centre for Mathematics and its Applications, Canberra, Australia (shahar.mendelson@gmail.com).
§UCLouvain, ICTEAM Institute, Louvain-la-Neuve, Belgium (alexander.stollenwerk@uclouvain.be).

$$(1.1)\qquad (1-\epsilon)\|x-y\|_2^2 \le \|f(x)-f(y)\|_2^2 \le (1+\epsilon)\|x-y\|_2^2 \qquad \text{for all } x,y\in T$$

with high probability, provided that $m \gtrsim \epsilon^{-2}\log|T|$. Here, and throughout this article, $|T|$ denotes the number of points in T, $a \lesssim b$ means that $a \le cb$ for an absolute constant $c>0$, and $a \sim b$ means $a \lesssim b$ and $b \lesssim a$.

The Johnson–Lindenstrauss lemma is remarkable in at least two respects. First, the map f is *data-oblivious,* i.e., it is constructed without any prior information on the set one wishes to embed. This property is crucial in certain applications, e.g., one-pass streaming applications [9] and data structural problems such as nearest neighbor search [19]. Second, the Johnson–Lindenstrauss embedding is in general *optimal*: Larsen and Nelson [35] showed that if $\epsilon > \min\{n,|T|\}^{-0.49}$, then any map $f: T \to \mathbb{R}^m$ that satisfies (1.1) must also satisfy that $m \gtrsim \epsilon^{-2}\log|T|$.

Despite this general optimality, the embedding dimension achieved by the Gaussian matrix can be substantially lower. Indeed, for a set T, let

$$(1.2)\qquad T' = \left\{ \frac{x-y}{\|x-y\|_2} : x\neq y,\ x,y\in T \right\},$$

denote by G the standard Gaussian random vector, and let

$$(1.3)\qquad \ell_*(S) = \mathbb{E}\sup_{x\in S}|\langle G,x\rangle|$$

be the *Gaussian mean width* of a set S. A result due to Gordon [18] shows that for $T\subset\mathbb{R}^n$, f satisfies (1.1) with high probability if $m \gtrsim \epsilon^{-2}\ell_*^2(T')$. Gordon's result is an *instance-adaptive* version of the Johnson–Lindenstrauss lemma: if T' is as in (1.2), then $\ell_*^2(T')$ is always upper bounded by $c\log|T|$ for an absolute constant $c>0$, and it may be substantially lower if T has a low-complexity structure, e.g., if it consists of sparse vectors or if it belongs to a low-dimensional subspace or manifold.

It is also known that the Gaussian random matrix can be used to define dimension reduction schemes that go beyond the Euclidean setting. Most relevant to this work is the fact that it is possible to embed subsets of ℓ_2^n into the Hamming cube $\{-1,1\}^m$ in an almost isometric way—by combining the Gaussian matrix with a straightforward binarization scheme (see [39, 42] when $T\subset S^{n-1}$ and [14] for $T\subset\mathbb{R}^n$).

At the same time, using the Gaussian matrix in dimension reduction schemes is problematic from a computational perspective. First, a typical realization of the matrix is fully populated and unstructured; thus, simply storing it requires plenty of memory ($O(mn)$). Second, and more importantly, it takes significant time ($O(mn)$) to compute a matrix-vector product Ax. It is therefore highly desirable to find an alternative to the Gaussian matrix: specifically, some structured random matrix that requires less storage space and supports fast matrix-vector multiplication. Obviously, one would want that matrix to be as effective in dimension reduction as the Gaussian matrix, resulting in the best of both worlds: an optimal data-oblivious embedding that is computationally efficient.

Our main result achieves this goal for binary embeddings: we identify a computationally friendly replacement for the Gaussian matrix that leads to a near-isometric embedding of an arbitrary subset of $\mathbb{R}^n$ into a low-dimensional Hamming cube.

The heart of the proof of this result is to show that the matrix we define—the double circulant matrix—mimics the behavior of the Gaussian matrix in two important ways: it yields an almost isometric embedding of any subset of ℓ_2^n into ℓ_1^m and, at

the same time, it maps an arbitrary set in $\mathbb{R}^n$ into a set of well-spread vectors. We will make these statements precise in section 1.2. This behavior is remarkable because the double circulant matrix has limited randomness and its entries are strongly dependent. Although this may be somewhat speculative, we believe that the Gaussian behavior exhibited by the double circulant matrix will have many additional applications—well beyond the scope of binary embeddings.

1.1. Main result. Before stating our main result, we recall a binary embedding that is based on the Gaussian matrix and was studied in [14]. For a matrix $A \in \mathbb{R}^{m\times n}$ we consider the map $f : \mathbb{R}^n \to \{-1,1\}^m$ defined by

$$f(x) = \operatorname{sign}(Ax+\tau), \tag{1.4}$$

where τ is uniformly distributed in $[-\lambda,\lambda]^m$ and is independent of A, and the sign-function is applied componentwise. In what follows, $\ell_*(T)$ denotes the Gaussian mean width of a set T (as in (1.3)) and $\mathcal{N}(T,\theta)$ is the Euclidean covering number of T at scale θ, i.e., the smallest number of open Euclidean balls of radius θ needed to cover the set T. Our starting point is the following fact, which was established in [14]. Here and throughout, d_H denotes the Hamming distance on $\{-1,1\}^m$.

THEOREM 1.1. *There exist absolute constants $c_0,\dots,c_3$ such that the following holds. Let $T \subset \mathbb{R}^n$ and put $R = \sup_{t\in T}\|t\|_2$. Set $0<\delta\le\frac{R}{2}$, $u\ge 1$, and let*

$$0<\theta\le c_0\frac{\delta}{\sqrt{\log(e\lambda/\delta)}}, \qquad \lambda \ge c_1 R\sqrt{\log(R/\delta)}.$$

Suppose that

$$m \ge c_2\left(\lambda^2\frac{\log\mathcal{N}(T,\theta)}{\delta^2}+\lambda\frac{\ell_*^2((T-T)\cap\theta B_2^n)}{\delta^3}\right). \tag{1.5}$$

If $A\in\mathbb{R}^{m\times n}$ is the standard Gaussian matrix and τ is uniformly distributed in $[-\lambda,\lambda]^m$ and independent of A, then with probability at least $1-2\exp(-c_3\delta^2 m/\lambda^2)$, the map $f(t)=\operatorname{sign}(At+\tau)$ satisfies

$$\sup_{x,y\in T}\left|\frac{\sqrt{2\pi}\lambda}{m}d_H(f(x),f(y))-\|x-y\|_2\right|\le\delta. \tag{1.6}$$

Although the bound on the dimension m in (1.5) seems unnatural, it is, in fact, optimal. We refer the reader to [14] for the proof of this surprising fact.

In what follows we show that a version of Theorem 1.1 is true for a certain computationally friendly matrix—the double circulant matrix. To define that matrix, let $I\subset[n]$ with $|I|=m$ and set $R_I x=\sum_{i\in I}x_i e_i$. For vectors $x,y\in\mathbb{R}^n$, let $\Gamma_x y = x\circledast y$; thus $\Gamma_x\in\mathbb{R}^{n\times n}$ is the discrete convolution operator with x. Let $D_x=\operatorname{diag}(x_1,\dots,x_n)\in\mathbb{R}^{n\times n}$ be the diagonal matrix defined by x, let $G\in\mathbb{R}^n$ be the standard Gaussian vector, and set $\varepsilon'',\varepsilon',\varepsilon\in\mathbb{R}^n$ to be Rademacher vectors, i.e., vectors consisting of independent random variables taking values 1 and -1 with equal probability. We assume throughout that $G,\varepsilon'',\varepsilon'$, and ε are independent.

DEFINITION 1.2. *The* double circulant matrix $A\in\mathbb{R}^{m\times n}$ *is defined by*

$$A=\frac{1}{\sqrt{n}}R_I\Gamma_G D_{\varepsilon''}\Gamma_{\varepsilon'}D_\varepsilon. \tag{1.7}$$

Clearly, A requires $O(n)$ storage capacity, and it is well known that matrix-vector multiplication for a circulant matrix can be carried out in time $O(n\log n)$ by exploiting the fast Fourier transform.

Our main result is that the binary embedding endowed by the double circulant matrix performs as well as the Gaussian embedding (up to logarithmic factors in n and with a worse success probability).

THEOREM 1.3. *For any $\gamma \geq 1$, there exist $\tilde{c}_0, \dots, \tilde{c}_3$ that depend only (polynomially) on $\log(n)$ and γ such that the following holds. Fix $0 < \delta < R/2$, let $T \subset RB_2^n$, and set*

$$\theta = \frac{\delta}{\tilde{c}_0 \sqrt{\log(e\lambda/\delta)}}, \qquad \lambda \geq \tilde{c}_1 R \sqrt{\log(R/\delta)}.$$

Suppose that $n \geq \tilde{c}_2 m$ and

$$m \geq \tilde{c}_3 \left(\lambda^2 \frac{\log \mathcal{N}(T, \theta)}{\delta^2} + \lambda \frac{\ell_*^2((T-T) \cap \theta B_2^n)}{\delta^3} \right). \tag{1.8}$$

Let $A \in \mathbb{R}^{m \times n}$ be the double circulant matrix. If τ is uniformly distributed in $[-\lambda, \lambda]^m$ and independent of A, then with probability at least $1 - n^{-\gamma}$, the map $f(t) = \operatorname{sign}(At + \tau)$ satisfies

$$\sup_{x,y \in T} \left| \frac{\sqrt{2\pi}\lambda}{m} d_H(f(x), f(y)) - \|x - y\|_2 \right| \leq \delta.$$

This result significantly improves on earlier attempts to create a computationally efficient, oblivious near-isometric embedding into the Hamming cube- see section 1.4 for details.

1.2. The Gaussian behavior of the double circulant matrix. The proof of Theorem 1.1 is based on two well-known properties of the standard Gaussian matrix.

First, if $A \in \mathbb{R}^{m \times n}$ is the Gaussian matrix, then for any $T \subset \mathbb{R}^n$, with high probability, A is *an embedding of T into ℓ_1^m*, in the following sense: for any $u > 0$

$$\mathbb{P}\left(\sup_{z \in T} \left| \frac{1}{m} \sqrt{\frac{\pi}{2}} \|Az\|_1 - \|z\|_2 \right| \leq \frac{4\ell_*(T)}{\sqrt{m}} + u \right) \leq 2e^{-mu^2/(2\mathcal{R}(T)^2)}, \tag{1.9}$$

where $\mathcal{R}(T) = \sup_{x \in T} \|x\|_2$. The proof of this fact can be found in [42] (see Lemma 2.1 there).

Second, the Gaussian matrix *maps any T into a set of "well-spread" vectors*: for any $1 \leq k \leq m$ and $u \geq 1$, with probability at least $1 - 2\exp(-cu^2 k \log(em/k))$,

$$\sup_{z \in T} \|Az\|_{[k]} \leq C \left(\ell_*(T) + u\mathcal{R}(T) \sqrt{k \log(em/k)} \right), \tag{1.10}$$

where

$$\|x\|_{[k]} = \sup_{|I|=k} \left(\sum_{i \in I} x_i^2 \right)^{1/2}.$$

The proof can be found, for example, in [14] (see Theorem 2.5 there).

Together with a generic binary embedding result, stated in Theorem 2.1, these two facts imply Theorem 1.1.

With that in mind, the heart of the proof of Theorem 1.3 is to show that the double circulant matrix "acts as the Gaussian matrix"; specifically, that it satisfies suitable versions of (1.9) and (1.10). This behavior is surprising in view of the limited

randomness in the double circulant matrix and the strong dependence of its entries. As a result, the approaches used to prove (1.9) and (1.10) fail in the case of the double circulant matrix. We will develop an entirely new approach to establish those properties.

We first develop a general recipe for constructing a matrix that maps an arbitrary set to a collection of vectors that are well spread. The key feature that we introduce for this purpose is the notion of *strong regularity.* Intuitively, a matrix $B \in \mathbb{R}^{m\times n}$ is strongly regular if it acts as a Euclidean almost-isometry on sparse vectors, and also maps sparse vectors into well-spread ones (see Definition 3.1 for a formulation of the strong regularity property). We prove that the matrix BD_ε, obtained by randomizing the column signs of a strongly regular matrix B, satisfies an estimate similar to (1.10) for an arbitrary set T and with high probability with respect to ε. The accurate formulation of this statement can be found in section 3.

Next, with the notion of strong regularity in mind, the second component of the proof of Theorem 1.3 is to show that

$$B = \frac{1}{\sqrt{mn}} R_I \Gamma_G D_{\varepsilon''} \Gamma_{\varepsilon'}$$

is strongly regular (obviously, $A = \sqrt{m} B D_\varepsilon$). That fact is established in section 4 by using known (but nontrivial) tools, developed in [33] and [13].

Combining those facts, it follows that a typical realization of the double circulant matrix A maps an arbitrary set to a collection of well-spread vectors, thus leading to an estimate as in (1.10).

Once a version of (1.10) is established, we turn our attention to (1.9): showing that just like the Gaussian matrix, the double circulant matrix satisfies a uniform ℓ_1-concentration phenomenon. To that end, we first prove that for a fixed vector y, the random variable $\|R_I \Gamma_G y\|_1$ concentrates sharply around its mean $m\sqrt{\frac{2}{\pi}}\|y\|_2$ if the discrete Fourier transform of y is well-spread- see section 5. The exact notion of well-spread needed here is clarified in what follows. Next, recalling that

$$A = \frac{1}{\sqrt{n}} R_I \Gamma_G D_{\varepsilon''} \Gamma_{\varepsilon'} D_\varepsilon,$$

we prove that $\|At\|_1$ concentrates by showing that the discrete Fourier transform of $D_{\varepsilon''}\Gamma_{\varepsilon'} D_\varepsilon t$ is well-spread. The concentration for any fixed vector $t \in T$ is sufficiently strong to derive a uniform concentration estimate in a straightforward manner- see section 6. Note that in this final part of the argument we make essential use of the fact that A is a *double* circulant matrix- simplifying the construction to a "single" circulant matrix of the form $R_I \Gamma_G D_{\varepsilon''}$, say, would require a nontrivial effort (if it is possible at all).

1.3. Fast ℓ_2-ℓ_1 dimension reduction of finite sets. As a side product, our analysis yields a result of independent interest: a new fast embedding of any finite set of points in ℓ_2^n into ℓ_1^m. The embedding has runtime $O(n \log n)$ and achieves the same dimension reduction as the Gaussian matrix. This improves a well-known construction by Ailon and Chazelle [2]; most significantly, we remove a strong restriction on the cardinality of the set that one wishes to embed. The condition obtained here is optimal up to a polylogarithmic factor if the goal is maximal dimension reduction- see the next section for a detailed discussion. Note that the embedding dimension $|I|$ lies between $m/2$ and $3m/2$, say, with probability at least $1 - e^{-cm}$.

THEOREM 1.4. *For any $\gamma \geq 1$, there exist $\tilde{c}_0, \tilde{c}_1, \tilde{c}_2$ that depend only (polynomially) on γ such that the following holds. Let $m \leq n/(\tilde{c}_0 \log^4(n))$ and consider the scaled double circulant matrix $C = \frac{1}{m}\sqrt{\frac{\pi}{2}}A$, where I is chosen using random selectors: $I = \{i \in [n] \,:\, \theta_i = 1\}$, where $\theta_1, \dots, \theta_n$ are independent and $1 - \mathbb{P}(\theta_i = 0) = \mathbb{P}(\theta_i = 1) = m/n$. Let $T \subset \mathbb{R}^n$ be finite and let $\epsilon \leq \log^{-5/2}(n)$. Then, C satisfies*

$$(1-\epsilon)\|x-y\|_2 \leq \|Cx - Cy\|_1 \leq (1+\epsilon)\|x-y\|_2 \qquad \text{for all } x, y \in T \tag{1.11}$$

with probability at least $1 - n^{-\gamma}$ if $m \geq \tilde{c}_1 \epsilon^{-2} \log|T|$ and $|T| \leq \exp(\epsilon^2 n/(\tilde{c}_2 \log^6(n)))$.

Remark 1.5. If T is finite, then one can similarly improve Theorem 1.3 by selecting I using independent and identically distributed (i.i.d.) random selectors. In this case, one can show that this result remains true under the optimal condition $m \gtrsim \delta^{-2} \log|T|$ if $n \geq \tilde{c}\delta^{-2}\log|T|$ and $\tilde{c}$ depends only (polynomially) on $\log(n)$ and γ. We omit the details of this argument.

1.4. Related work.

Fast ℓ_2-ℓ_2 dimension reduction with circulant matrices. Numerous works have proposed and analyzed computationally efficient random matrices for ℓ_2-ℓ_2 dimension reduction (see, e.g., [1, 2, 3, 6, 10, 16, 17, 20, 30, 32, 34, 45] and the references therein). We will only highlight the successful use of random circulant matrices for this task- an approach that was first considered by Hinrichs and Vybíral [20]. They proved that the matrix $C = \frac{1}{\sqrt{m}} R_I \Gamma_{\varepsilon'} D_\varepsilon$ satisfies (1.1) with large probability if $m \sim \epsilon^{-2}\log^3(|T|)$. This was later improved to $m \sim \epsilon^{-2}\log^2(|T|)$ [45] and shown not to be improvable further if $m|T| \leq n$ [17]- however, it is possible to improve the scaling in terms of $|T|$ to $\log|T|$ at the expense of additional logarithmic factors in the dimension (by combining [33] and [34]). In fact, the main result of [40] (see also Theorem 3.2 from [38] below) shows that C satisfies (1.1) with high probability under the refined condition $m \gtrsim \epsilon^{-2}\ell_*^2(T')\log^4(n)$. These results show that C can serve as a computationally efficient replacement of the Gaussian matrix for ℓ_2-ℓ_2 dimension reduction.

Fast ℓ_2-ℓ_1 dimension reduction. In their groundbreaking paper [2], Ailon and Chazelle initiated the study of fast Johnson–Lindenstrauss transforms that use structured random matrices that support fast matrix-vector multiplication. Although most subsequent works have focused on alternative fast transforms and improved guarantees for ℓ_2-ℓ_2 dimension reduction (see, e.g., [16, 20, 30, 34, 45]), the original work [2] also studied computationally efficient ℓ_2-ℓ_1 dimension reduction of finite point sets, motivated by approximate nearest neighbor search.

Ailon and Chazelle's original transform takes the form $C = \frac{1}{m}\sqrt{\frac{\pi}{2}} PHD_\varepsilon$, where D_ε is as before, $H \in \mathbb{R}^{n\times n}$ is a normalized Hadamard matrix, and $P \in \mathbb{R}^{m\times n}$ is a sparsified Gaussian matrix: it has i.i.d. entries which are equal to 0 with probability $1-q$ and equal to a Gaussian with mean zero and variance $1/q$ with probability q, where q needs to be picked appropriately. It satisfies (1.11) if $q \sim \min\{\log(|T|)/(n\epsilon), 1\}$ and $m \sim \epsilon^{-2}\log(|T|)$. The runtime is

$$O(n\log(n) + \min\{n\epsilon^{-2}\log(|T|), \epsilon^{-3}\log^2(|T|)\})$$

so that it achieves $O(n\log n)$ runtime under an appropriate restriction on the number of points in T: $|T|$ must be $O_\epsilon(\exp(n^{1/2}))$. Essentially the same bottleneck appears in the alternative construction in [3]. This bottleneck was recently improved to $O_\epsilon(\exp(n^{1-\zeta}))$ by the method of [5], but this improvement applies only for the related problem of *simultaneously* embedding a subset T' of T in time $O(|T'|n\log n)$, where T' has at least polynomial size $n^{g(\zeta)}$ (for a certain function g). We improve these results in two steps.

In the case of ℓ_2-ℓ_2 dimension reduction, it is well known that one can construct structured random matrices with improved runtime $O(n \log n)$ (without a restriction on the number of points) if one is willing to raise the embedding dimension by a polylogarithmic factor in $|T|$ and/or n (see, e.g., [20, 30, 34, 45]). As a first step, we make an analogous contribution for fast ℓ_2-ℓ_1 dimension reduction: we show that the scaled double circulant matrix $C = \frac{1}{m}\sqrt{\frac{\pi}{2}}A$ satisfies (1.11) with high probability if $m \gtrsim \epsilon^{-2} \log(|T|) \log^6(n)$ (see Corollary 6.2 for T' as in (1.2)).

Theorem 1.4 improves this further by using a double circulant matrix with I picked using *random selectors*: it achieves the same dimension reduction as the Gaussian matrix ($m \sim \epsilon^{-2} \log |T|$) under the significantly relaxed restriction that $|T| = O_\epsilon(\exp(n/\log^6(n)))$- which is nearly optimal if one is interested in dimension reduction (i.e., ensuring that $m \leq n$). Note that beyond the dimension reduction setting, Indyk [23] showed that for any distortion $\epsilon > 1/\log n$ there is an embedding of *all* of ℓ_2^n into ℓ_1^m with $m = O(n^{1+o(1)})$ and runtime $O(n^{1+o(1)})$.

Finally, let us mention for completeness a less closely related result on fast ℓ_2-ℓ_1 dimension reduction derived in the context of one-bit compressed sensing [12]. It was shown there that the matrix $C = \frac{1}{m}\sqrt{\frac{\pi}{2}} R_I \Gamma_G$, where I is picked using i.i.d. random selectors, satisfies (1.11) with high probability on the set of all s-sparse vectors if $m \gtrsim \epsilon^{-2} s \log(n/s\epsilon)$, provided that the sparsity level s is small enough (e.g., if $s \leq \epsilon\sqrt{n}$).

Binary encoding of Euclidean distances. Theorem 1.3 fits into a more general line of work [15, 21, 24, 25, 26, 27, 28, 29, 41, 47, 48, 49] that strives to create an *efficient binary encoding* of all Euclidean distances in a given set T (see also [4], which focuses on inner products and *squared* distances). This task consists in constructing a computationally efficient embedding map $f : T \to \{-1, 1\}^m$ and reconstruction map $d : \{-1, 1\}^m \times \{-1, 1\}^m \to \mathbb{R}$ such that for any pair $x, y \in T$, $d(f(x), f(y))$ is an accurate proxy of $\|x - y\|_2$. The binary encoding in Theorem 1.3 stands out in comparison to the aforementioned works in achieving all of the following properties simultaneously:

(i) The embedding map f is a metric embedding into the Hamming cube, i.e., the reconstruction map d is (a constant multiple of) the natural Hamming distance, as in [15, 41, 47, 48]. Consequently, the number of bit operations needed to compute $d(f(x), f(y))$, is minimal—one only needs to directly compare two bit strings.

(ii) The embedding time $O(n \log n)$ of our construction, i.e., the time needed to compute $f(x)$ for a given x, is on par with the best existing results [4, 15, 21, 41, 47, 48].

(iii) The construction is data-oblivious, as in [4, 15, 21, 26, 27, 28, 29, 41, 47, 48, 49].

(iv) The bit complexity of our encoding, i.e., the number of bits (or dimension of the Hamming cube) required to encode the Euclidean distances within the given set of points, is optimal for finite sets (when picking I using random selectors; see Remark 1.5). Indeed, any oblivious random binary encoding scheme (f, d) that embeds, with some given probability, any given finite set of N points into $\{-1, 1\}^m$ with an additive error of δ, must satisfy $m \geq c\delta^{-2} \log N$ (see [14], whose proof is based on [4]). Similar (near-)optimal bit complexity estimates were achieved in [15, 21, 26, 27, 28, 47, 49] (see also [25] and [4] for methods with optimal bit complexity for the related tasks of encoding distances up to a multiplicative error and encoding squared distances up to an additive error, respectively).

(v) The bit complexity estimate is in terms of more refined complexity measures of the dataset than the cardinality of the set which, in particular, makes the result applicable to arbitrary infinite datasets in $\mathbb{R}^n$ (as in [21, 26, 27, 28]).

Let us comment in more detail on prior works [15, 41, 47, 48] that have the same goal of constructing a fast, data-oblivious metric embedding into the Hamming cube. These works all considered finite sets of vectors on the unit sphere and strived to find a structured random matrix $A \in \mathbb{R}^{m\times n}$ that supports fast-matrix vector multiplication so that with high probability

$$|d_H(\operatorname{sign}(Ax), \operatorname{sign}(Ay)) - d_{S_{n-1}}(x,y)| \leq \delta \qquad \text{for all } x, y \in T, \tag{1.12}$$

where $d_{S_{n-1}}(x,y)$ is the normalized geodesic distance on the sphere. If A is standard Gaussian, then it is straightforward to see that this holds under the optimal condition $m \gtrsim \delta^{-2}\log|T|$. As was remarked in [15, 47], by simply applying a fast or sparse Johnson–Lindenstrauss transform before applying the Gaussian matrix, one obtains (1.12) under the same condition with runtime $O(n\log n)$ if the number of points is small (e.g., if $\log|T| \lesssim \delta^2\sqrt{n}$ in the case of a fast Johnson–Lindenstrauss transform). It was observed empirically in [48] that $A = R_I\Gamma_G D_\varepsilon$ performs well and that performance deteriorates if D_ε is left out. The latter was rigorously established in [15]: it exhibits a two-point set for which (1.12) fails with positive probability if $A = R_I\Gamma_G$. The best result in the former direction stems from [41]: it considers $A = R_I\Gamma_G D_{G_1} H D_{G_2}$, where I consists of m indices selected uniformly at random and G, G_1, G_2 are independent standard Gaussians, and shows that (1.12) holds if $m \gtrsim \delta^{-3}\log|T|$ and $\log|T| \lesssim \delta^2 n^{1/3}$. Even though the embedding time of this transform is $O(n\log n)$ without restrictions, the guarantee is worse than for the simple combination of the standard Gaussian matrix and a fast or sparse Johnson–Lindenstrauss transform. Finally, let us mention that [15] (which fixed a proof gap in [47]) established a binary encoding with optimal bit complexity and runtime $O(n\log n)$ under the relaxed condition $\log|T| \lesssim \delta\sqrt{n}/\sqrt{\log(1/\delta)}$. However, this encoding is not a metric embedding, as it no longer uses the Hamming metric as the reconstruction map. Nevertheless, it illustrates that it is nontrivial to overcome the bottleneck on the number of points to be embedded.

Our work improves over these earlier attempts to create a fast metric embedding into the Hamming cube. We do not require an artificial restriction on the number of points to be embedded, achieve an embedding time $O(n\log n)$, and achieve a bit complexity that is optimal for finite datasets. Moreover, our bit complexity estimate is in terms of more refined complexity measures than the cardinality of the dataset and in particular allows for the embedding of infinite sets. This bit complexity estimate matches (up to logarithmic factors) the one for the Gaussian embedding in Theorem 1.1, which is known to be optimal for the Gaussian embedding [14]. These improvements are made possible by the Gaussian behavior of the double circulant matrix established in this work (uniform ℓ_1-concentration and mapping into well-spread vectors), together with the use of the uniformly random shifts in (1.4).

Let us finally mention two binary encodings that improve over our construction at the expense of some of the listed properties (i)–(v). First, the dependence of m on the additive error parameter δ can be improved (beyond the lower bound mentioned under (iv)) if an additional relative error term is present: this was achieved in [21] using a binary encoding that combines a fast Johnson–Lindenstrauss embedding with a so-called noise shaping method (see also [49]). A downside of this encoding is that it is not a metric embedding into the Hamming cube. Second, it is possible to preserve distances up to (only) a multiplicative error. This is the goal of a different binary encoding scheme developed by Indyk and Wagner [24, 25], which achieves the minimal bit complexity for a finite set for this setting. This scheme is, however, not a metric

embedding, data-adaptive rather than oblivious, and has a higher computational complexity.

2. A generic binary embedding result. The starting point of the proof of Theorem 1.3 is a generic embedding result from [14], which we now outline. Let $A \in \mathbb{R}^{m\times n}$ be a matrix and for a parameter $\lambda > 0$, let τ be uniformly distributed in $[-\lambda, \lambda]^m$. Consider $f : \mathbb{R}^n \to \{-1, 1\}^m$, defined by

$$f(x) = \operatorname{sign}(Ax + \tau), \tag{2.1}$$

where the sign-function is applied componentwise, and denote the normalized Hamming distance on $\{-1, 1\}^m$ by

$$\tilde{d}(x, y) = \frac{2\lambda\kappa}{m} d_H(x, y).$$

The constant κ turns out to be an absolute constant in our application, and its value is specified in what follows.

Let $0 < \theta < \delta$, and set $T_\theta \subset T$ to be a θ-net of T of minimal cardinality. Finally, assume that A "acts well" on T in the following sense:

(a) *A satisfies uniform ℓ_1-concentration on T_θ:*

$$\sup_{x,y\in T_\theta} \left| \frac{\kappa}{m} \|A(x - y)\|_1 - \|x - y\|_2 \right| \le \delta. \tag{2.2}$$

(b) *A maps T to 'well-spread' vectors:* For $k = \lfloor \delta m/\lambda \rfloor$ we have that

$$\frac{1}{\sqrt{k}} \sup_{x\in T_\theta} \|Ax\|_{[k]} \le \lambda \tag{2.3}$$

and

$$\frac{1}{\sqrt{k}} \sup_{x\in (T-T)\cap\theta B_2^n} \|Ax\|_{[k]} \le \delta. \tag{2.4}$$

Then one has the following:

THEOREM 2.1. [14] *There exist absolute constants c_1, c_2, and c_3 such that the following holds. Let*

$$m \ge c_1\lambda^2\kappa^2 \frac{\log \mathcal{N}(T, \theta)}{\delta^2}$$

and assume that A satisfies (2.2)*,* (2.3)*, and* (2.4)*. Then with probability at least $1 - 2\exp(-c_2\delta^2 m/\lambda^2\kappa^2)$,*

$$\sup_{x,y\in T} \left| \tilde{d}(f(x), f(y)) - \|x - y\|_2 \right| \le c_3(\kappa + 1)\delta.$$

If A is the standard Gaussian matrix and the conditions of Theorem 1.1 hold, then (2.2), (2.3), and (2.4) (with $\kappa = \sqrt{\pi/2}$) are immediate outcomes of (1.9) and (1.10).

Thanks to Theorem 2.1, it is clear that proving the analogs of (2.2), (2.3), and (2.4) would yield a binary embedding estimate- and the proof of Theorem 1.3. The rest of the article is devoted to the proof of those estimates for the double circulant matrix.

3. Strong regularity. For $x \in \mathbb{R}^n$, set $\|x\|_0 = |\mathrm{supp}(x)|$. The vector x is called s-sparse if $\|x\|_0 \leq s$. Denote by $U_s \subset S^{n-1}$ the set of all s-sparse vectors on the Euclidean unit sphere and let

$$\Sigma_{s,n} = \{x \in \mathbb{R}^n \ : \ \|x\|_0 \leq s, \ \|x\|_2 \leq 1\}$$

be the set of s-sparse vectors in the Euclidean unit ball.

A matrix $B \in \mathbb{R}^{m\times n}$ satisfies the (s,δ)-*Restricted Isometry Property* if

$$\sup_{x\in U_s} \big| \ \|Bx\|_2^2 - \|x\|_2^2 \ \big| \leq \delta,$$

and we denote this property by $\mathrm{RIP}(s,\delta)$.

The following definition is of crucial importance in the context of the proof.

DEFINITION 3.1. *Let $\rho > 0$ and set $s_\rho = \lceil \rho^{-2} \rceil$. A matrix $B \in \mathbb{R}^{m\times n}$ is* ρ-regular *if it satisfies* $\mathrm{RIP}(r, \rho\sqrt{r})$ *for all $1 \leq r \leq s_\rho$. It is* ρ-strongly regular *if it is ρ-regular and, in addition, satisfies*

$$\sup_{x\in\Sigma_{r,n}} \|Bx\|_{[r]} \leq \rho\sqrt{r} \quad \text{for all } 1 \leq r \leq s_\rho.$$

In other words, thinking of the case where ρ is small, B is regular if it is an almost Euclidean isometry on sufficiently sparse unit vectors, and it is strongly regular if additionally, for any sufficiently sparse unit vector x, Bx is relatively well-spread: if x is r-sparse, then $1 - \rho\sqrt{r} \leq \|Bx\|_2^2 \leq 1 + \rho\sqrt{r}$ but the contribution of the r largest coordinates of Bx to its Euclidean norm is at most $\rho\sqrt{r}$.

To put this notion in perspective, consider the standard Gaussian matrix $A \in \mathbb{R}^{m\times n}$. Then for any $u \geq 1$, with probability at least $1 - 2\exp(-c_0 u^2 r \log(em/r))$,

$$\sup_{x\in U_r} \left| \ \frac{1}{m}\|Ax\|_2^2 - \|x\|_2^2 \ \right| \leq c_1 u \sqrt{\frac{r\log(em/r)}{m}}$$

and

$$\sup_{x\in\Sigma_{r,n}} \frac{1}{\sqrt{m}}\|Ax\|_{[r]} \leq c_1 u \sqrt{\frac{r\log(em/r)}{m}}.$$

Hence, by taking the union bound over $1 \leq r \leq m$ we find that with nontrivial probability, the Gaussian matrix $\frac{1}{\sqrt{m}}A$ is strongly ρ-regular for $\rho \sim \sqrt{\frac{\log m}{m}}$.

The main result of this section is that by randomizing the columns of a strongly regular matrix B using independent random signs, one obtains a matrix that is "well-behaved" on an arbitrary set. To formulate this claim, let D_ε be a diagonal matrix whose diagonal entries are independent, symmetric, random signs $\varepsilon_1, \ldots, \varepsilon_n$.

THEOREM 3.2. *There exist absolute constants c_1 and c_2 such that the following holds. Let $0 < \rho < 1/\sqrt{\log(m+n)}$ and consider $1 \leq k \leq m$. Assume that $B \in \mathbb{R}^{m\times n}$ is ρ-strongly regular. If $T \subset \mathbb{R}^n$, $R \geq \mathcal{R}(T)$, and $u \geq 1$ then with probability at least*

$$1 - 2\exp\left(-c_1 u^2 \left[\ell_*^2(T) + R^2 k \log(em/k)\right]/R^2\right)$$

with respect to ε, we have that

$$\sup_{x\in T} \|BD_\varepsilon x\|_{[k]} \leq c_2 u^2 \left[\rho\left(\ell_*(T) + R\cdot\sqrt{k\log(em/k)}\right) + \rho^2 R^{-1}\left(\ell_*^2(T) + R^2\cdot k\log(em/k)\right)\right].$$

The proof of Theorem 3.2 is based on the following analogous result on column randomization of a regular matrix, which was established in [38]. In this result we use

$$d^*(T) = \left(\frac{\ell_*(T)}{\mathcal{R}(T)}\right)^2 \tag{3.1}$$

to denote the *Dvoretzky–Milman dimension* (or *stable dimension*) of a set $T \subset \mathbb{R}^n$.

THEOREM 3.3. *There exist absolute constants $c_1, c_2 > 0$ such that the following holds. Let $0 < \rho < 1/\sqrt{\log n}$. Assume that $B \in \mathbb{R}^{m\times n}$ is ρ-regular. Then for any $T \subset \mathbb{R}^n$ and $u \geq 1$, with probability at least $1 - 2\exp(-c_1 u^2 d^*(T))$ with respect to ε,*

$$\sup_{x\in T} \left| \|BD_\varepsilon x\|_2^2 - \|x\|_2^2 \right| \leq c_2 u^2 \mathcal{R}^2(T) \cdot (\rho\sqrt{d^*(T)} + \rho^2 d^*(T)).$$

The following simple lemma is the key to the proof of Theorem 3.2. In what follows, $\mathrm{Id}_m \in \mathbb{R}^{m\times m}$ denotes the identity matrix.

LEMMA 3.4. *Let $B \in \mathbb{R}^{m\times n}$, $\rho > 0$. If B is ρ-strongly regular, then $\begin{bmatrix} B & \mathrm{Id}_m \end{bmatrix} \in \mathbb{R}^{m\times(n+m)}$ is 3ρ-regular.*

Proof. Let $1 \leq r \leq \lceil (3\rho)^{-2} \rceil \leq \lceil \rho^{-2} \rceil$. Clearly,

$$\begin{aligned}
&\sup_{x\in\Sigma_{r,n}} \sup_{y\in\Sigma_{r,m}} \left| \left\| \begin{bmatrix} B & \mathrm{Id}_m \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right\|_2^2 - \left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\|_2^2 \right| \\
&\quad \leq \sup_{x\in\Sigma_{r,n}} \left| \|Bx\|_2^2 - \|x\|_2^2 \right| + 2 \sup_{x\in\Sigma_{r,n}} \sup_{y\in\Sigma_{r,m}} |\langle Bx, y\rangle| \\
&\quad = \sup_{x\in\Sigma_{r,n}} \left| \|Bx\|_2^2 - \|x\|_2^2 \right| + 2 \sup_{x\in\Sigma_{r,n}} \|Bx\|_{[r]} \leq 3\rho\sqrt{r},
\end{aligned}$$

and the result follows. ∎

Proof of Theorem 3.2. Set $R \geq \mathcal{R}(T) = \sup_{t\in T} \|t\|_2$ and consider $T' = \{t/R : t \in T\}$. We shall use the fact that $\Sigma_{k,m}$ is an unconditional set: for any choice of signs $\zeta_1, \ldots, \zeta_m$, $D_\zeta \Sigma_{k,m} = \Sigma_{k,m}$. This allows one to introduce additional randomness. Indeed, let $\zeta_1, \ldots, \zeta_m$ be independent, symmetric random signs that are also independent of $(\varepsilon_i)_{i=1}^n$. Then

$$\sup_{x\in T} \|BD_\varepsilon x\|_{[k]} = R \sup_{x\in T'} \|BD_\varepsilon x\|_{[k]} \tag{3.2}$$

and

$$\sup_{x\in T'} \|BD_\varepsilon x\|_{[k]} = \sup_{x\in T'} \sup_{y\in\Sigma_{k,m}} |\langle BD_\varepsilon x, y\rangle| = \sup_{x\in T'} \sup_{y\in\Sigma_{k,m}} |\langle BD_\varepsilon x, D_\zeta y\rangle|.$$

Moreover, by the polarization identity,

$$\begin{aligned}
|\langle BD_\varepsilon x, D_\zeta y\rangle| &= \frac{1}{4} \left| \|BD_\varepsilon x + D_\zeta y\|_2^2 - \|BD_\varepsilon x - D_\zeta y\|_2^2 \right| \\
&= \frac{1}{4} \left| \left\| \begin{bmatrix} BD_\varepsilon & D_\zeta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right\|_2^2 - \left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\|_2^2 - \left(\left\| \begin{bmatrix} BD_\varepsilon & D_\zeta \end{bmatrix} \begin{bmatrix} x \\ -y \end{bmatrix} \right\|_2^2 - \left\| \begin{bmatrix} x \\ -y \end{bmatrix} \right\|_2^2 \right) \right|.
\end{aligned}$$

Clearly,

$$\begin{bmatrix} BD_\varepsilon & D_\zeta \end{bmatrix} = \begin{bmatrix} B & \mathrm{Id}_m \end{bmatrix} \begin{bmatrix} D_\varepsilon & 0 \\ 0 & D_\zeta \end{bmatrix}$$

and hence

$$\sup_{x\in T'}\|BD_\varepsilon x\|_{[k]} \le \frac{1}{2}\sup_{(x,y)^T\in\tilde{T}}\left|\left\|\begin{bmatrix}B & \mathrm{Id}_m\end{bmatrix}\begin{bmatrix}D_\varepsilon & 0\\ 0 & D_\zeta\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}\right\|_2^2 - \left\|\begin{bmatrix}x\\ y\end{bmatrix}\right\|_2^2\right|, \tag{3.3}$$

where $\tilde{T} = T'\times\Sigma_{k,m}$. Lemma 3.4 implies that $\begin{bmatrix}B & \mathrm{Id}_m\end{bmatrix}$ is 3ρ-regular, and one may therefore invoke Theorem 3.3 for the set $\tilde{T}$. The result follows thanks to the straightforward observations that $\ell_*(\tilde{T})\sim\ell_*(T')+\ell_*(\Sigma_{k,m})$, $\ell_*(\Sigma_{k,m})\sim\sqrt{k\log(em/k)}$, and $\mathcal{R}(\tilde{T}) = \sup_{t\in\tilde{T}}\|t\|_2\sim 1$. ∎

Remark 3.5. It is useful to present the estimate from Theorem 3.2 in terms of the following parameter. For a set $T\subset\mathbb{R}^n$, $R\ge\mathcal{R}(T)$, and $1\le k\le m$ define

$$Q_k(T,R) = \rho\left(\ell_*^2(T) + R^2 k\log(em/k)\right)^{1/2}. \tag{3.4}$$

Theorem 3.2 implies that with probability at least $1-2\exp(-c_1u^2[\ell_*^2(T)+R^2k\log(em/k)]/R^2)$

$$\sup_{x\in T}\|BD_\varepsilon x\|_{[k]} \le c_2u^2\max\{Q_k(T,R), R^{-1}Q_k^2(T,R)\}. \tag{3.5}$$

In particular, if $Q_k(T,R)\le R$, as will be the case in the situations that interest us, the dominating term in the estimate from Theorem 3.2 is $\sim u^2Q_k(T,R)$.

3.1. Strong regularity and Theorem 2.1. Let us return to the last two conditions that are required in the generic embedding result from Theorem 2.1. If T_θ is a θ-net of T of minimal cardinality and $k=\lfloor\delta m/\lambda\rfloor$, one has to show that

$$\sup_{x\in T_\theta}\|Ax\|_{[k]} \le \lambda\sqrt{k} \tag{3.6}$$

and

$$\sup_{x\in(T-T)\cap\theta B_2^n}\|Ax\|_{[k]} \le \delta\sqrt{k} \tag{3.7}$$

hold for the matrix

$$A = R_I\Gamma_G D_{\varepsilon''}\frac{1}{\sqrt{n}}\Gamma_{\varepsilon'}D_\varepsilon.$$

Set

$$B = \frac{1}{\sqrt{m}}R_I\Gamma_G D_{\varepsilon''}\frac{1}{\sqrt{n}}\Gamma_{\varepsilon'}$$

and observe that $A=\sqrt{m}BD_\varepsilon$. We will spend considerable effort in showing that B is ρ-strongly regular, where $\Upsilon := \rho m^{1/2}$ is independent of m and is at most polylogarithmic in n (see Theorem 4.1). Before going down that long road, let us first show its benefit: combined with the following result, it will establish (3.6) and (3.7).

THEOREM 3.6. *For $u\ge 1$ there exist constants $\tilde{c}_1,\tilde{c}_2$ that depend only (polynomially) on u and absolute constants c_3,c_4 such that the following holds. Let $0<\rho<\frac{1}{\sqrt{\log(m+n)}}$ be such that $\Upsilon=\rho m^{1/2}\ge 1$. Let $B\in\mathbb{R}^{m\times n}$ be ρ-strongly regular, set $0<\delta\le\mathcal{R}(T)$, and consider*

$$\lambda \ge \tilde{c}_1\max\left\{\delta\Upsilon^2\log(e\Upsilon), \Upsilon\mathcal{R}(T)\log^{1/2}\left(\frac{\Upsilon\mathcal{R}(T)}{\delta}\right)\right\}.$$

Put

$$0 < \theta \le \frac{\delta}{\tilde{c}_2 \Upsilon} \log^{-1/2}\left(\frac{\lambda}{\delta}\right).$$

If

$$m \ge c_3 \frac{\lambda}{\log\left(\frac{e\lambda}{\delta}\right)} \max\left\{ \frac{\ell_*^2(T_\theta)}{\delta \mathcal{R}^2(T)}, \frac{\ell_*^2((T-T)\cap \theta B_2^n)}{\delta\theta^2} \right\}, \tag{3.8}$$

then with probability at least $1 - 2\exp(-c_4 u^2 m\delta/\lambda)$*, the matrix* $A = \sqrt{m} B D_\varepsilon$ *satisfies* (3.6) *and* (3.7).

Remark 3.7. To put the estimate (3.8) in Theorem 3.6 in a more familiar form, observe that

$$\ell_*^2(T_\theta) \lesssim \mathcal{R}^2(T_\theta) \log|T_\theta| \lesssim \mathcal{R}^2(T) \log \mathcal{N}(T,\theta).$$

Hence, by choosing θ as large as possible, the second term on the right-hand side of (3.8) leads to the ℓ_*^2/δ^3 term in (1.8) from Theorem 1.3 (up to polylogarithmic factors in n), while the first term is dominated by the entropic term in (1.8).

Proof. Set $k = \lfloor \delta m/\lambda \rfloor$ and define $K_\theta = (T-T)\cap \theta B_2^n$. We will apply the estimate in Remark 3.5 for the sets T_θ and K_θ. To that end, observe that $\mathcal{R}(T_\theta) \le \mathcal{R}(T)$ and $\mathcal{R}(K_\theta) \le \theta$. Let use write $Q(T_\theta) := Q_k(T_\theta, \mathcal{R}(T))$ and $Q(K_\theta) := Q_k(K_\theta, \theta)$ and observe that

$$Q(T_\theta) \sim \rho \mathcal{R}(T) \left(\frac{\ell_*^2(T_\theta)}{\mathcal{R}^2(T)} + \frac{\delta}{\lambda} m \log\left(\frac{e\lambda}{\delta}\right)\right)^{1/2}$$

and

$$Q(K_\theta) \sim \rho\theta \left(\frac{\ell_*^2(K_\theta)}{\theta^2} + \frac{\delta}{\lambda} m \log\left(\frac{e\lambda}{\delta}\right)\right)^{1/2}.$$

We begin by imposing conditions that ensure that

$$Q(T_\theta) \le \mathcal{R}(T) \quad \text{and} \quad Q(K_\theta) \le \theta, \tag{3.9}$$

so that both $Q(T_\theta)$ and $Q(K_\theta)$ are the dominant terms in the estimate (3.5) for the sets T_θ and K_θ, respectively. Observe that if (3.8) holds with $c_3 \ge 1$, then

$$\max\left\{ \frac{\ell_*^2(T_\theta)}{\mathcal{R}^2(T)}, \frac{\ell_*^2(K_\theta)}{\theta^2} \right\} \le \frac{\delta}{\lambda} m \log\left(\frac{e\lambda}{\delta}\right)$$

and, hence, (3.9) holds if we ensure that

$$\rho\sqrt{m} \cdot \left(\frac{\delta}{\lambda} \log\left(\frac{e\lambda}{\delta}\right)\right)^{1/2} \le \frac{1}{4}.$$

Since $\Upsilon = \rho\sqrt{m}$, the latter condition is satisfied if

$$\frac{\lambda}{\delta} \ge c_1 \Upsilon^2 \log(e\Upsilon). \tag{3.10}$$

To summarize, if (3.8) and (3.10) hold, then Remark 3.5 implies that

$$\sup_{x\in T_\theta}\|Ax\|_{[k]} = \sup_{x\in T_\theta}\sqrt{m}\|BD_\varepsilon x\|_{[k]} \lesssim u^2\sqrt{m}Q(T_\theta)$$

and

$$\sup_{x\in K_\theta}\|Ax\|_{[k]} \lesssim u^2\sqrt{m}Q(K_\theta).$$

Thus, to establish (3.6) and (3.7), all that remains is to show that

$$u^2\sqrt{m}Q(T_\theta)\le\lambda\sqrt{k} \quad\text{and}\quad u^2\sqrt{m}Q(K_\theta)\le\delta\sqrt{k},$$

i.e., that

$$u^2\Upsilon\mathcal{R}(T)\left(\frac{\delta}{\lambda}\log\left(\frac{e\lambda}{\delta}\right)\right)^{1/2}\le\sqrt{\delta}\sqrt{\lambda} \tag{3.11}$$

and that

$$u^2\Upsilon\theta\left(\frac{\delta}{\lambda}\log\left(\frac{e\lambda}{\delta}\right)\right)^{1/2}\le\delta\sqrt{\frac{\delta}{\lambda}}. \tag{3.12}$$

A straightforward computation shows that (3.11) holds if

$$\lambda\ge c_1(u)\Upsilon\mathcal{R}(T)\log^{1/2}\left(\frac{\Upsilon\mathcal{R}(T)}{\delta}\right), \tag{3.13}$$

and (3.12) holds if

$$\theta\le\frac{\delta}{c_2(u)\Upsilon}\log^{-1/2}\left(\frac{e\lambda}{\delta}\right). \tag{3.14}$$

This completes the proof. ∎

4. Strong regularity features of the double circulant matrix. Recall that the double circulant matrix is

$$A = R_I\Gamma_G D_{\varepsilon''}\frac{1}{\sqrt{n}}\Gamma_{\varepsilon'}D_\varepsilon = \sqrt{m}BD_\varepsilon,$$

where $I\subset\{1,\dots,n\}$ is a fixed set of indices of cardinality m, G is the standard Gaussian random vector in $\mathbb{R}^n$, and ε, ε', and ε'' are uniformly distributed in $\{-1,1\}^n$. Moreover, G, ε, ε', and ε'' are all independent. Also, for every $x\in\mathbb{R}^n$, we have that

$$\Gamma_G x = G\circledast x = \Gamma_x G$$

is the discrete circular convolution of G and x. And, denoting by $\mathcal{F}\in\mathbb{C}^{n\times n}$ the discrete Fourier matrix, we have that $\Gamma_x G=\sqrt{n}UD_{Wx}OG$, where

$$O = W = \frac{\mathcal{F}}{\sqrt{n}} \quad\text{and}\quad U=\frac{\mathcal{F}^{-1}}{\sqrt{n}} = W^*. \tag{4.1}$$

In particular, U, W, and O are *Hadamard-type matrices,* i.e., they are unitary and all their entries are bounded by $\frac{1}{\sqrt{n}}$.

In light of Theorem 2.1, a key part of the analysis of the embedding procedure is to show that A maps an arbitrary set T into "well-spread vectors", specifically, that (3.6) and (3.7) hold. By invoking Theorem 3.6, that can be established by proving that B possesses the following strong regularity property:

THEOREM 4.1. *For $\gamma\ge1$ there are constants $\tilde{c}_1$ and $\tilde{c}_2$ that depend only (polynomially) on γ such that the following holds. If $m\le n/(\tilde{c}_1\log^4 n)$, then with probability*

at least $1-n^{-\gamma}$ *with respect to* $G\otimes\varepsilon'\otimes\varepsilon''$*, the random matrix* B *is* ρ*-strongly regular for*

$$\rho=\tilde{c}_2\frac{\log^{5/2}n}{\sqrt{m}}.$$

Note that $\Upsilon=\rho\sqrt{m}$ is polylogarithmic in n, as required.

The idea behind the proof of Theorem 4.1 is outlined in the next section. The full proof is presented in Appendix B.

4.1. Highlights of the proof of Theorem 4.1. The proof is based on two well-known facts that are formulated in what follows. The first fact is an outcome of [33], on the behavior of second-order chaos processes (see also Theorem 6.5 in [11] for the refinement that is used here). The bound is based on Talagrand's γ_2-functional. For a detailed exposition on chaining methods and the γ-functionals in a general setup, we refer the reader to [43].

DEFINITION 4.2. *Let* $\mathcal{A}$ *be a subset of a normed space. An admissible sequence of* $\mathcal{A}$ *is a collection of sets* $\mathcal{A}_\ell\subset\mathcal{A}$*, where* $|\mathcal{A}_0|=1$ *and for* $\ell\geq 1$*,* $|\mathcal{A}_\ell|\leq 2^{2^\ell}$*. For* $a\in\mathcal{A}$*, let* $\pi_\ell a\in\mathcal{A}_\ell$ *be a nearest point to* a *in* $\mathcal{A}_\ell$ *with respect to the underlying norm. Define*

$$\gamma_2(\mathcal{A},\|\cdot\|)=\inf\sup_{a\in\mathcal{A}}\left(\|\pi_0 a\|+\sum_{\ell\geq 1}2^{\ell/2}\|\pi_\ell a-\pi_{\ell-1}a\|\right),$$

where the infimum is taken with respect to all admissible sequences of $\mathcal{A}$*.*

In the setup we focus on here, $\mathcal{A}$ is a class of matrices. Denote by $\|\cdot\|_{2\to 2}$ the standard operator norm, let $\|\cdot\|_{HS}$ be the Hilbert–Schmidt norm, and put

$$d_{HS}(\mathcal{A})=\sup_{A\in\mathcal{A}}\|A\|_{HS},\qquad d_{2\to 2}(\mathcal{A})=\sup_{A\in\mathcal{A}}\|A\|_{2\to 2}.$$

Let $\gamma_2(\mathcal{A})\equiv\gamma_2(\mathcal{A},\|\cdot\|_{2\to 2})$ be the γ_2-functional of $\mathcal{A}$ with respect to the operator norm.

Recall that a centered random variable ξ is L-subgaussian if for every $p\geq 2$, $\|\xi\|_{L_p}\leq L\sqrt{p}\|\xi\|_{L_2}$. A random vector X is L-subgaussian if it is centered and for any $t\in\mathbb{R}^n$, $\langle X,t\rangle$ is an L-subgaussian random variable.

THEOREM 4.3. *There is an absolute constant* $C>0$ *such that the following holds. Let* ξ *be a random vector whose coordinates* $(\xi_i)_{i=1}^n$ *are independent, mean-zero, variance* 1 *random variables that are also* L*-subgaussian. Then for any* $u\geq 1$*, with probability at least* $1-2\exp(-u)$*,*

$$\sup_{A\in\mathcal{A}}\big|\ \|A\xi\|_2^2-\mathbb{E}\|A\xi\|_2^2\ \big|\leq CL^2\Big(\gamma_2^2(\mathcal{A})+d_{HS}(\mathcal{A})\gamma_2(\mathcal{A})+\sqrt{u}d_{HS}(\mathcal{A})d_{2\to 2}(\mathcal{A})+ud_{2\to 2}^2(\mathcal{A})\Big).$$

To see why Theorem 4.3 is useful for establishing regularity, let us return to the random operator

$$B=\frac{1}{\sqrt{m}}R_I\Gamma_G D_{\varepsilon''}\frac{1}{\sqrt{n}}\Gamma_{\varepsilon'}.$$

Set

$$\Psi=D_{\varepsilon''}\frac{1}{\sqrt{n}}\Gamma_{\varepsilon'},$$

then, for every $x \in \mathbb{R}^n$,

$$Bx = \frac{1}{\sqrt{m}} R_I \Gamma_G \Psi x = \frac{1}{\sqrt{m}} R_I \Gamma_{\Psi x} G.$$

Therefore,

(4.2)
$$\sup_{x \in \Sigma_{r,n}} \left| \, \|Bx\|_2^2 - \|x\|_2^2 \, \right| \leq \sup_{x \in \Sigma_{r,n}} \left| \, \|Bx\|_2^2 - \mathbb{E}_G \|Bx\|_2^2 \, \right| + \sup_{x \in \Sigma_{r,n}} \left| \, \mathbb{E}_G \|Bx\|_2^2 - \|x\|_2^2 \, \right|,$$

and by a straightforward computation,

$$\mathbb{E}_G \|Bx\|_2^2 = \mathbb{E}_G \left\| \frac{1}{\sqrt{m}} R_I \Gamma_{\Psi x} G \right\|_2^2 = \frac{1}{n} \|\Gamma_{\varepsilon'} x\|_2^2 = \frac{1}{n} \|\Gamma_x \varepsilon'\|_2^2. \tag{4.3}$$

Hence, (4.2) becomes

$$\sup_{x \in \Sigma_{r,n}} \left| \, \|Bx\|_2^2 - \|x\|_2^2 \, \right|$$

(4.4)
$$\leq \sup_{x \in \Sigma_{r,n}} \left| \left\| \frac{1}{\sqrt{m}} R_I \Gamma_{\Psi x} G \right\|_2^2 - \mathbb{E}_G \left\| \frac{1}{\sqrt{m}} R_I \Gamma_{\Psi x} G \right\|_2^2 \right| + \sup_{x \in \Sigma_{r,n}} \left| \left\| \frac{1}{\sqrt{n}} \Gamma_x \varepsilon' \right\|_2^2 - \|x\|_2^2 \right|$$

and

$$\|x\|_2^2 = \mathbb{E}_{\varepsilon'} \left\| \frac{1}{\sqrt{n}} \Gamma_x \varepsilon' \right\|_2^2.$$

The two terms in (4.4) are exactly in the form that is dealt with in Theorem 4.3. Taking into account the regularity estimates we are looking for, the classes of matrices of interest are

$$\left\{ \frac{1}{\sqrt{m}} R_I \Gamma_{\Psi x} : x \in \Sigma_{r,n} \right\} \quad \text{for } r \leq m \tag{4.5}$$

and

$$\left\{ \frac{1}{\sqrt{n}} \Gamma_x : x \in \Sigma_{r,n} \right\} \quad \text{for } r \leq n. \tag{4.6}$$

The key estimate in the case of (4.6) was established in [33]:

THEOREM 4.4. *For $L, \gamma \geq 1$ there are constants $\tilde{c}_0$ and $\tilde{c}_1$ depending only (polynomially) on L and γ, respectively, such that the following holds. Let ξ be as in Theorem* 4.3 *and set $1 \leq r \leq n$. Then with probability at least $1 - 2\exp(-\gamma \log^4(n)/\tilde{c}_0)$,*

$$\sup_{x \in \Sigma_{r,n}} \left| \, \frac{1}{n} \|\Gamma_x \xi\|_2^2 - \|x\|_2^2 \, \right| \leq \rho \sqrt{r}$$

for

$$\rho = \tilde{c}_1 \frac{\log^2 n}{\sqrt{n}}.$$

Obtaining a similar estimate for the class (4.5) is technically more involved but is based on similar ideas. The details are presented in Appendix B.

Next, one has to establish the strong regularity of B. That is based on the following fact, which is a straightforward generalization of Theorem 3.4 from [13]. To formulate the claim, consider two unitary matrices $\mathcal{O},\mathcal{U} \in \mathbb{C}^{n\times n}$, and let $\mathcal{W} \in \mathbb{C}^{n\times n}$. Set

$$d_{\mathcal{U}} = \sqrt{n} \max_{1\le i,j\le n} |\mathcal{U}_{ij}| \quad \text{and} \quad d_{\mathcal{W}} = \sqrt{n} \max_{1\le i,j\le n} |\mathcal{W}_{ij}|,$$

and assume further that

$$\sup_{x\in\Sigma_{r,n}} \|\mathcal{W}x\|_2 \le 2.$$

Note, for example, that if $\mathcal{U}$ and $\mathcal{W}$ are Hadamard-type matrices then $d_{\mathcal{U}} \le 1$, $d_{\mathcal{W}} \le 1$, and the condition on $\mathcal{W}$ is trivially satisfied.

THEOREM 4.5. *For $L \ge 1$ there exist constants $\tilde{c}_1$ and $\tilde{c}_2$ that depend only (polynomially) on L such that the following holds. Let ζ be an L-subgaussian random vector, set $1 \le r \le n$, and let $u \ge 1$. Then with probability at least $1 - e^{-u/\tilde{c}_1}$,*

$$\sup_{x\in\Sigma_{r,n}} \|\mathcal{U}D_{\mathcal{W}x}\mathcal{O}\zeta\|_{[r]} \le \tilde{c}_2 \sqrt{\frac{r}{n}} \max\{d_{\mathcal{U}}, d_{\mathcal{W}}\} \left(\log(n)\log(r) + \sqrt{u}\right).$$

The proof of Theorem 4.5 is based on a straightforward modification of the argument used to prove Theorem 3.4 in [13]; its details are omitted.

Let us return to the random matrix B. To establish ρ-strong regularity, one has to estimate $\sup_{x\in\Sigma_{r,n}} \|Bx\|_{[r]}$ for every $1 \le r \le m$. Since

$$Bx = \frac{1}{\sqrt{m}} R_I \Gamma_{\Psi x} G$$

and $\|R_I\Gamma_{\Psi x}G\|_{[r]} \le \|\Gamma_{\Psi x}G\|_{[r]}$, it suffices to control

$$\|\Gamma_{\Psi x}G\|_{[r]} = \sqrt{n}\|UD_{W\Psi x}OG\|_{[r]}$$

for the Hadamard-type matrices U, W, and O defined in (4.1). Thus, a high probability (with respect to G) estimate on $\sup_{x\in\Sigma_{r,n}} \|UD_{W\Psi x}OG\|_{[r]}$ follows from Theorem 4.5 once one shows that

$$\sup_{x\in\Sigma_{r,n}} \|W\Psi x\|_2 = \sup_{x\in\Sigma_{r,n}} \left\| \frac{1}{\sqrt{n}} D_{\varepsilon''}\Gamma_{\varepsilon'}x \right\|_2 \le 2. \tag{4.7}$$

The proof of (4.7) is straightforward, thanks to Theorem 4.4. Indeed,

$$\left\| \frac{1}{\sqrt{n}} D_{\varepsilon''}\Gamma_{\varepsilon'}x \right\|_2 = \left\| \frac{1}{\sqrt{n}} \Gamma_x \varepsilon' \right\|_2,$$

and by Theorem 4.4, with high probability with respect to ε',

$$\sup_{x\in\Sigma_{r,n}} \left| \, \|\Gamma_x\varepsilon'\|_2^2 - \mathbb{E}\|\Gamma_x\varepsilon'\|_2^2 \, \right|$$

is well-behaved.

5. Concentration of the Gaussian convolution operator. Next, let us turn to the second ingredient needed for the application of the generic embedding result, formulated in Theorem 2.1: proving the ℓ_1-concentration phenomenon for the double circulant matrix.

Let G be the standard Gaussian random vector in $\mathbb{R}^n$ and for a fixed $x \in \mathbb{R}^n$ and $I \subset \{1,\dots,n\}$, consider the partial convolution $R_I(x \circledast G)$. The first order of business is to study the concentration of $\|R_I(x \circledast G)\|$ around its mean. An important ingredient in the analysis is the following immediate consequence of the Gaussian concentration theorem for Lipschitz functions (see, e.g., [36]).

THEOREM 5.1. *Let $S \subset \mathbb{R}^n$ and set $\mathcal{R}(S) = \sup_{x\in S} \|x\|_2$. Then for $u > 0$,*

$$\mathbb{P}\left(\left| \sup_{x\in S} \langle G, x\rangle - \mathbb{E}\sup_{x\in S} \langle G, x\rangle \right| \geq u\mathcal{R}(S) \right) \leq 2\exp(-cu^2),$$

where c is an absolute constant.

To formulate the concentration estimate for convolutions, recall that $W = \mathcal{F}/\sqrt{n}$, where $\mathcal{F}$ is the discrete Fourier matrix. Below we will consider a general norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ and use $\mathcal{B}$ to denote the corresponding unit ball. Recall that the dual norm is defined by

$$\|x\|_* = \sup_{y\in\mathcal{B}} \langle x, y\rangle$$

and that for any $x \in \mathbb{R}^n$

$$\|x\| = \sup_{y\in\mathcal{B}^\circ} \langle x, y\rangle,$$

where $\mathcal{B}^\circ = \{x \in \mathbb{R}^n \,:\, \|x\|_* \leq 1\}$ is the dual unit ball (see, e.g., [7, Appendix A.1.6]). In the application below we consider the ℓ_1-norm, in which case the dual norm is the ℓ_∞-norm.

THEOREM 5.2. *There is an absolute constant $c > 0$ such that the following holds. For any $I \subset \{1,\dots,n\}$ and $u > 0$, with probability at least $1 - 2\exp(-cu^2)$,*

$$\big|\, \|R_I(x \circledast G)\| - \mathbb{E}\|R_I(x \circledast G)\| \,\big| \leq \sqrt{n}\mathcal{R}(R_I\mathcal{B}^\circ) \inf_{\{y,z : x=y+z\}} (u\|Wz\|_\infty + 2\|WG\|_\infty \|y\|_2).$$

Proof. Fix a decomposition $x = y + z$ and observe that

$$x \circledast G = y \circledast G + z \circledast G = y \circledast G + \sqrt{n} U D_{Wz} O G,$$

where, as before, $O = W$ and $U = W^*$. Clearly,

$$y \circledast G = W^*W(y \circledast G) = \sqrt{n} W^* \left(\sum_{i=1}^n (Wy)_i (WG)_i e_i \right),$$

$R_I^* = R_I$, and, hence, almost surely,

$$\begin{aligned} \|R_I(y \circledast G)\| &= \sqrt{n} \left\| R_I W^* \left(\sum_{i=1}^n (Wy)_i (WG)_i e_i \right) \right\| \\ &\leq \sqrt{n} \sup_{a \in B_2^n} \|R_I W^* a\| \cdot \left\| \sum_{i=1}^n (Wy)_i (WG)_i e_i \right\|_2 \\ &\leq \sqrt{n}\mathcal{R}(R_I\mathcal{B}^\circ) \cdot \|WG\|_\infty \|y\|_2; \end{aligned} \tag{5.1}$$

in particular, this estimate also holds for $\mathbb{E}\|R_I(y \circledast G)\|$.

Next, observe that

$$\|R_I(z \circledast G)\| = \left\|\sqrt{n} R_I U D_{Wz} O G\right\| = \sqrt{n} \sup_{t \in B^\circ} \langle G, O^* D_{Wz}^* U^* R_I^* t \rangle$$

and

$$\sup_{t \in B^\circ} \|O^* D_{Wz}^* U^* R_I^* t\|_2 \le \|Wz\|_\infty \cdot \mathcal{R}(R_I \mathcal{B}^\circ).$$

Hence, by Theorem 5.1, for $u > 0$, with probability at least $1 - 2\exp(-cu^2)$,

$$| \, \|R_I(z \circledast G)\| - \mathbb{E}\|R_I(z \circledast G)\| \, | \le u\sqrt{n} \cdot \mathcal{R}(R_I \mathcal{B}^\circ) \cdot \|Wz\|_\infty. \tag{5.2}$$

The claim follows by combining (5.1) and (5.2). ∎

To apply this result for the ℓ_1^n-norm, note that for every $x \in \mathbb{R}^n$,

$$\mathbb{E}\|R_I(x \circledast G)\|_1 = \mathbb{E}|g| \cdot m\|x\|_2 = \sqrt{\frac{2}{\pi}} m \|x\|_2,$$

and $\mathcal{R}(R_I \mathcal{B}^\circ) = \sup_{t \in B_\infty^I} \|t\|_2 = \sqrt{m}$. Hence, for $u > 0$, with probability at least $1 - 2\exp(-cu^2)$,

$$\left| \, \|R_I(x \circledast G)\|_1 - \sqrt{\frac{2}{\pi}} m \|x\|_2 \, \right| \le \sqrt{nm} \inf_{x=y+z} \left(u\|Wz\|_\infty + 2\|WG\|_\infty \|y\|_2\right). \tag{5.3}$$

With (5.3) in mind, let us show that when the Fourier transform of x is well-spread, $\|R_I(x \circledast G)\|_1$ exhibits a sharp concentration around its mean. One useful notion of "being well-spread" is that there is some $1 \le r \le n$ and $1 \le \Lambda \le \sqrt{n}$ such that

$$\|Wx\|_{[r]} \le \frac{\Lambda}{\sqrt{n}} \|x\|_2. \tag{5.4}$$

COROLLARY 5.3. *There is an absolute constant $c > 0$ such that the following holds. Let x satisfy* (5.4)*. For any $\delta > 0$, with probability at least*

$$1 - 2\exp\left(-c\delta^2 \frac{mr}{\Lambda^2}\right),$$

we have that

$$\left| \, \|R_I(x \circledast G)\|_1 - \sqrt{\frac{2}{\pi}} m \|x\|_2 \, \right| \le \sqrt{\frac{2}{\pi}} m \|x\|_2 \left(\delta + \sqrt{2\pi} \frac{\Lambda}{\sqrt{m}} \|WG\|_\infty\right). \tag{5.5}$$

Proof. Let J be the set of indices corresponding to the largest r coordinates of $(|(Wx)_i|)_{i=1}^n$. Using the invertibility of W, one may write $x = y + z$, where $Wy = R_J W x$ and $Wz = R_{J^c} W x$. By (5.4),

$$\|y\|_2 = \|Wy\|_2 = \|Wx\|_{[r]} \le \frac{\Lambda}{\sqrt{n}} \|x\|_2 \quad \text{and} \quad \|Wz\|_\infty \le \frac{1}{\sqrt{r}} \|Wx\|_{[r]} \le \frac{\Lambda}{\sqrt{rn}} \|x\|_2.$$

Hence, (5.3) implies that with probability at least $1 - 2\exp(-cu^2)$,

$$\begin{aligned} \left| \, \|R_I(x \circledast G)\|_1 - \sqrt{\frac{2}{\pi}} m \|x\|_2 \, \right| &\le \sqrt{nm} \left(u\|Wz\|_\infty + 2\|WG\|_\infty \|y\|_2\right) \\ &\le \|x\|_2 \sqrt{m} \left(u \frac{\Lambda}{\sqrt{r}} + 2\|WG\|_\infty \Lambda\right). \end{aligned} \tag{5.6}$$

The claim follows by setting $u = \sqrt{\frac{2}{\pi}} \delta(\sqrt{mr}/\Lambda)$. ∎

It is straightforward to verify that a similar estimate holds if instead of (5.4) we only have an upper estimate on $\|Wx\|_{[r]}$. That is the form we use in what follows.

COROLLARY 5.4. *There is an absolute constant c such that the following holds. Let x satisfy that $\|Wx\|_{[r]} \leq \frac{\Lambda}{\sqrt{n}}$. Then for $\delta > 0$, with probability at least*

$$1 - 2\exp\left(-c\delta^2 \frac{mr}{\Lambda^2}\right),$$

we have that

$$\left| \|R_I(x \circledast G)\|_1 - \sqrt{\frac{2}{\pi}} m \|x\|_2 \right| \leq \sqrt{\frac{2}{\pi}} m \left(\delta + \sqrt{2\pi}\frac{\Lambda}{\sqrt{m}} \|WG\|_\infty\right). \tag{5.7}$$

We will apply Corollary 5.4 to a finite set that is in a 'good position' - namely, consisting of vectors for which there is enough control on $\|W \cdot\|_{[r]}$.

To put Corollary 5.4 in some perspective, let us consider our benchmark- the Gaussian matrix.

5.1. The Gaussian benchmark. Let $\Gamma = n^{-1/2}\sum_{i=1}^n \langle G_i, \cdot\rangle\, e_i$ be the normalized standard Gaussian matrix. Consider $T \subset \mathbb{R}^n$ and let us explore the outcome of Corollary 5.4 for points in the set ΓT. To that end, pick the largest $1 \leq r \leq n$ such that

$$r \log\left(\frac{en}{r}\right) \leq \left(\frac{\ell_*(T)}{\mathcal{R}(T)}\right)^2 = d^*(T).$$

Assuming that $d^*(T) \geq \log n$, r is well-defined and satisfies $r \sim d^*(T)/\log(en/d^*(T))$. By the proof of [14, Theorem 2.5], for any $u \geq 1$, we have

$$\|W\Gamma t\|_{[r]} \leq Cu\frac{\ell_*(T)}{\sqrt{n}} \tag{5.8}$$

with probability at least $1 - 2\exp(-cu^2 d^*(T))$. Denote that event by Ω_u, and let us invoke Corollary 5.4, where for every $t \in T$ we set

$$r \sim \frac{d^*(T)}{\log(en/d^*(T))} \quad \text{and} \quad \Lambda \sim \ell_*(T).$$

It follows that conditioned on Ω_u, for $\delta > 0$ and every $t \in T$, with probability at least

$$1 - 2\exp\left(-c_0\delta^2 \frac{md^*(T)}{\ell_*^2(T)\log(en/d^*(T))}\right) = 1 - 2\exp\left(-c_0\delta^2 \frac{m}{\mathcal{R}^2(T)\log(en/d^*(T))}\right)$$

with respect to G we have that

$$\left| \|R_I(\Gamma t \circledast G)\|_1 - \sqrt{\frac{2}{\pi}} m \|\Gamma t\|_2 \right| \leq c_1 m \left(\delta + \frac{\ell_*(T)}{\sqrt{m}} \|WG\|_\infty\right).$$

In the next section we will show that the double circulant matrix A satisfies an almost identical inequality.

6. Uniform ℓ_1-concentration for the double circulant matrix. Fix $1 \le r \le n$. Corollary 5.4 implies that for any $t \in \mathbb{R}^n$, if

$$\frac{1}{\sqrt{n}} \|W D_{\varepsilon''} \Gamma_{\varepsilon'} D_{\varepsilon} t\|_{[r]} \le \frac{\Lambda}{\sqrt{n}}, \tag{6.1}$$

then with high probability with respect to G, $\|At\|_1$ concentrates around its mean. Thus, the key question is whether, with high probability, for every $t \in T$, (6.1) holds for suitable values of Λ and r. We will show that with high probability, one may set

$$r \sim \frac{d^*(T)}{\log\left(\frac{en}{d^*(T)}\right)} \quad \text{and} \quad \Lambda \sim \Upsilon \ell_*(T), \text{where } \Upsilon \sim \log^{5/2} n, \tag{6.2}$$

where $d^*(T)$ is defined as in (3.1). Up to the factor of Υ, this is the same as in the Gaussian case.

THEOREM 6.1. *For $\gamma \ge 1$ there exist constants $\tilde{c}_1, \tilde{c}_2$ depending only (polynomially) on γ, and an event Ω_1 with probability at least $1 - n^{-\gamma}$ with respect to $\varepsilon' \otimes \varepsilon''$ such that the following holds. Let $T \subset \mathbb{R}^n$ and assume that $\log n \le d^*(T) \le n/\log^5 n$. There is an event $\Omega_{2,T}$ with probability at least $1 - 2\exp(-\gamma d^*(T))$ with respect to ε, such that, conditioned on Ω_1 and $\Omega_{2,T}$, for every $t \in T$, with probability at least*

$$1 - 2\exp\left(-\delta^2 \frac{m}{\tilde{c}_1 \mathcal{R}^2(T) \Upsilon^2 \log\left(\frac{en}{d^*(T)}\right)}\right)$$

with respect to G,

$$\left| \frac{1}{m}\|At\|_1 - \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{n}} \|\Gamma_{\varepsilon'} D_{\varepsilon} t\|_2 \right| \le \tilde{c}_2 \left(\delta + \frac{\|WG\|_\infty \Upsilon \ell_*(T)}{\sqrt{m}}\right), \tag{6.3}$$

where $\Upsilon \sim \log^{5/2} n$.

Therefore, up to the factors of Υ in the probability estimate and in (6.3), Theorem 6.1 shows that the ℓ_1-concentration phenomenon exhibited by the double circulant matrix is the same as exhibited by the standard Gaussian matrix. The proof of Theorem 6.1 is presented in the next section.

COROLLARY 6.2. *Let $\gamma \ge 1$. There exists a constant $\tilde{c}_1$ depending only (polynomially) on γ such that the following holds. Let T' be a finite set. If $\log n \le d^*(T') \le c_0 n/\log^5 n$, $m \le n$, and*

$$m \ge \tilde{c}_1 \mathcal{R}^2(T') \frac{\log |T'|}{\delta^2} \cdot \log^6 n$$

then with probability at least $1 - n^{-\gamma}$ with respect to $G \otimes \varepsilon \otimes \varepsilon' \otimes \varepsilon''$,

$$\sup_{t \in T'} \left| \frac{1}{m}\|At\|_1 - \sqrt{\frac{2}{\pi}} \|t\|_2 \right| \le \delta.$$

Proof. By Hoeffding's inequality and a union bound, with probability at least $1 - n^{-\gamma}$ with respect to G,

$$\|WG\|_\infty \le c_0 \sqrt{\gamma \log n}.$$

Hence, by Theorem 6.1, we only need to establish uniform concentration of $\frac{1}{\sqrt{n}}\|\Gamma_{\varepsilon'}D_\varepsilon t\|_2$ around $\|t\|_2$ for all $t\in T'$. Set $\tilde{T}=\{t/\|t\|_2 \ : \ t\in T'\}$. Using that $|a-1|=\frac{|a^2-1|}{a+1}\leq |a^2-1|$ if $a\geq 0$, we find

$$\sup_{t\in T'}\Big|\frac{1}{\sqrt{n}}\|\Gamma_{\varepsilon'}D_\varepsilon t\|_2-\|t\|_2\Big| = \sup_{t\in T'}\|t\|_2\Big|\frac{1}{\sqrt{n}}\Big\|\Gamma_{\varepsilon'}D_\varepsilon\frac{t}{\|t\|_2}\Big\|_2-1\Big| \leq \mathcal{R}(T')\ \sup_{t\in\tilde{T}}\Big|\frac{1}{n}\|\Gamma_{\varepsilon'}D_\varepsilon t\|_2^2-1\Big|.$$

By Theorem 4.4, with probability at least $1-2\exp(-c_1\gamma\log^4 n)$ with respect to ε', the matrix $n^{-1/2}\Gamma_{\varepsilon'}$ is ρ-regular for

$$\rho = c_2(\gamma)\frac{\log^2 n}{\sqrt{n}}.$$

On that event, Theorem 3.3 implies that with probability at least $1-\exp(-c_3u^2d^*(\tilde{T}))$ with respect to ε,

$$\sup_{t\in\tilde{T}}\Big|\frac{1}{n}\|\Gamma_{\varepsilon'}D_\varepsilon t\|_2^2-1\Big| \leq c_4(\gamma)u^2\mathcal{R}(\tilde{T})\left(\rho\sqrt{d^*(\tilde{T})}+\rho^2 d^*(\tilde{T})\right) \sim c_5(\gamma)u^2\left(\frac{\log^2(n)\sqrt{\log|T'|}}{\sqrt{n}}+\frac{\log^4(n)\log|T'|}{n}\right),$$

as $\mathcal{R}(\tilde{T})=1$ and $d^*(\tilde{T})=\ell_*^2(\tilde{T})\lesssim\log|T'|$. Setting $u^2=c_6\gamma\log(n)$, we conclude that

$$\sup_{t\in T'}\Big|\frac{1}{\sqrt{n}}\|\Gamma_{\varepsilon'}D_\varepsilon t\|_2-\|t\|_2\Big|\leq\delta$$

with probability at least $1-n^{-\gamma}$ under the assumed bound on m (which then also holds for n). ∎

Remark 6.3. The caveat that $\log(n)\leq d^*(T)\leq n/\log^5 n$ is there only for the sake of a simpler presentation. In any case, since we are making no attempt of obtaining a result that is accurate at the logarithmic level, that is not a real issue. Indeed, the condition $\log(n)\leq d^*(T)$ can be ensured by replacing T by $T\cup\{e_i, i=1,\dots,n\}$- this only leads to an additional logarithmic factor. If $d^*(T)\geq n/\log^5 n$ then replacing T by the Euclidean ball $\mathcal{R}(T)B_2^n$ comes at most at a logarithmic price, and the latter case can be analyzed directly, by noting that $Q_k(rB_2^n,r)\sim\rho r\sqrt{n}$.

Corollary 6.2 is the final ingredient needed in the proof of Theorem 1.3.

Proof of Theorem 1.3. Recall the three conditions required in Theorem 2.1: let T_θ be a θ-net of T of minimal cardinality and set $k=\lfloor\delta m/\lambda\rfloor$. One has to show that

(1) $\sup_{x\in T_\theta}\|Ax\|_{[k]}\leq\lambda\sqrt{k}$;
(2) $\sup_{x\in(T-T)\cap\theta B_2^n}\|Ax\|_{[k]}\leq\delta\sqrt{k}$;
(3) $\sup_{x\in(T_\theta-T_\theta)}\big|\frac{\kappa}{m}\|Ax\|_1-\|x\|_2\ \big|\leq\delta$

for the matrix

$$A=R_I\Gamma_G D_{\varepsilon''}\frac{1}{n^{1/2}}\Gamma_{\varepsilon'}D_\varepsilon=\sqrt{m}BD_\varepsilon.$$

The proof that the three conditions are satisfied with the wanted probability is an immediate outcome of Theorem 3.6 combined with Theorem 4.1 - used to establish (1) and (2); and Corollary 6.2, which implies (3) (for $\kappa=\sqrt{\pi/2}$). ∎

6.1. Proof of Theorem 1.4. Let T' be the set of normalized differences defined in (1.2). Let $A_I = \sqrt{\frac{\pi}{2}}A$ be the rescaled double circulant matrix with index set $I \subset [n]$. Consider independent selectors $\theta_1, \dots, \theta_n$ satisfying $\mathbb{P}(\theta_i = 1) = 1 - \mathbb{P}(\theta_i = 0) = \frac{m}{n}$ and let $I_\theta = \{i \in [n] \,:\, \theta_i = 1\}$ be the set of selected indices. Clearly,

$$\frac{1}{m}\mathbb{E}_\theta \|A_{I_\theta} z\|_1 = \frac{1}{n}\|A_{[n]} z\|_1$$

for any $z \in \mathbb{R}^n$ and hence

$$\sup_{z \in T'} \left| \frac{1}{m}\|A_{I_\theta} z\|_1 - 1 \right| \le \sup_{z \in T'} \left| \frac{1}{m}\|A_{I_\theta} z\|_1 - \frac{1}{m}\mathbb{E}_\theta \|A_{I_\theta} z\|_1 \right| + \sup_{z \in T'} \left| \frac{1}{n}\|A_{[n]} z\|_1 - 1 \right|.$$

By Corollary 6.2, if $n \gtrsim c(\gamma)\epsilon^{-2} \log |T| \log^6 n$, then

$$\sup_{z \in T'} \left| \frac{1}{n}\|A_{[n]} z\|_1 - 1 \right| \le \epsilon$$

with probability at least $1 - n^{-\gamma}$.

To control the first term, set $X_{\theta,z} = \frac{1}{m}\|A_{I_\theta} z\|_1 - \frac{1}{m}\mathbb{E}_\theta \|A_{I_\theta} z\|_1$ and let θ' be an independent copy of θ. By a symmetrization argument (see (6.3) in [36]),

$$\mathbb{P}_\theta \left(\sup_{z \in T'} |X_{\theta,z}| \ge 4\epsilon \right) \le \mathbb{P}_{\theta,\theta'} \left(\sup_{z \in T'} |X_{\theta,z} - X_{\theta',z}| \ge 2\epsilon \right) + \sup_{z \in T'} \mathbb{P}_\theta (|X_{\theta,z}| \ge 2\epsilon). \tag{6.4}$$

Observe that $X_{\theta,z} - X_{\theta',z}$ has the same distribution as

$$\frac{1}{m}\sum_{i=1}^n \zeta_i(\theta_i - \theta_i')|\langle a_i, z\rangle|,$$

where the a_i are the rows of $A_{[n]}$ and ζ is a Rademacher vector that is independent of all other random variables. Hence, it is standard to verify that

$$\mathbb{P}_{\theta,\theta'} \left(\sup_{z \in T'} |X_{\theta,z} - X_{\theta',z}| \ge 2\epsilon \right) \le 2\mathbb{P}_{\theta,\zeta} \left(\sup_{z \in T'} \frac{1}{m}\sum_{i=1}^n \zeta_i \theta_i |\langle a_i, z\rangle| \ge \epsilon \right).$$

Observe that the event $\{m/2 \le |I_\theta| \le 3m/2\}$ holds with θ-probability at least $1 - e^{-cm}$. On this event, Theorem 4.1 shows that with probability at least $1 - n^{-\gamma}$ with respect to $G \otimes \varepsilon' \otimes \varepsilon''$

$$B_{I_\theta} = \frac{1}{\sqrt{|I_\theta|}} R_{I_\theta} \Gamma_G D_{\varepsilon''} \frac{1}{\sqrt{n}} \Gamma_{\varepsilon'}$$

is ρ-regular for $\rho = c(\gamma)\log^{5/2}(n)/\sqrt{m}$. Since $A_{I_\theta} = \sqrt{|I_\theta|} B_{I_\theta} D_\varepsilon$, we can use Theorem 3.3 to conclude that with probability at least $1 - n^{-\gamma}$ with respect to $G \otimes \varepsilon \otimes \varepsilon' \otimes \varepsilon''$,

$$\sup_{z \in T'} \frac{1}{m}\sum_{i=1}^n \theta_i |\langle a_i, z\rangle|^2 = \sup_{z \in T'} \frac{1}{m}\|A_{I_\theta} z\|_2^2 \le c,$$

where c is an absolute constant, provided that

$$m \ge c(\gamma) \log |T| \log^5(n). \tag{6.5}$$

Clearly, (6.5) is satisfied when $m \geq c(\gamma)\epsilon^{-2}\log|T|$ and $\epsilon \leq \log^{-5/2}(n)$. In particular, all of the above events hold simultaneously with probability at least $1-e^{-cm}-n^{-\gamma}$ with respect to $\theta \otimes G \otimes \varepsilon \otimes \varepsilon' \otimes \varepsilon''$. Conditioned on that event, Hoeffding's inequality implies that

$$\mathbb{P}_\zeta\left(\frac{1}{m}\sum_{i=1}^n \zeta_i\theta_i|\langle a_i, z\rangle| \geq \epsilon\right) \leq 2e^{-cm\epsilon^2}$$

for any $z \in T'$. Recalling that $m \gtrsim \epsilon^{-2}\log|T|$, a union bound yields

$$\mathbb{P}_\zeta\left(\sup_{z\in T'}\frac{1}{m}\sum_{i=1}^n \zeta_i\theta_i|\langle a_i, z\rangle| \geq \epsilon\right) \leq 2e^{-c'm\epsilon^2}.$$

In a similar, but simpler manner one can estimate the second term on the right-hand side of (6.4), which completes the proof.

6.2. Proof of Theorem 6.1. As we noted previously, to prove Theorem 6.1 it remains to show that (6.1) holds with the parameters specified in (6.2). The random vectors ε, ε', and ε'' each play a different role in the argument, leading to the somewhat cumbersome formulation of Theorem 6.1.

First, we show that for a typical realization of $\varepsilon' \otimes \varepsilon''$, the matrix $n^{-1/2}WD_{\varepsilon''}\Gamma_{\varepsilon'}$ is ρ-strongly regular for ρ of the order of $n^{-1/2}\log^{5/2} n$. Clearly, that fact has nothing to with the identity of the set T, but rather only with the way in which the matrix $n^{-1/2}WD_{\varepsilon''}\Gamma_{\varepsilon'}$ acts on sparse vectors. Second, given a set T, Theorem 3.3 and Theorem 3.2 imply that (conditioned on the fact that $n^{-1/2}WD_{\varepsilon''}\Gamma_{\varepsilon'}$ is ρ-strongly regular), with high probability with respect to ε, $n^{-1/2}WD_{\varepsilon''}\Gamma_{\varepsilon'}D_\varepsilon$ "acts well" on T. These two steps are formulated in Theorem 6.4 and Corollary 6.5.

THEOREM 6.4. *For $\gamma \geq 1$ there is a constant $\tilde{c}$ that depends only (polynomially) on γ such that the following holds. With probability at least $1-n^{-\gamma}$ with respect to $\varepsilon' \otimes \varepsilon''$, the matrix $n^{-1/2}WD_{\varepsilon''}\Gamma_{\varepsilon'}$ is ρ-strongly regular for*

$$\rho = \tilde{c}\frac{\log^{5/2} n}{\sqrt{n}}.$$

We prove Theorem 6.4 in the next section. We first show how it implies (6.1) and, hence, Theorem 6.1 as well.

COROLLARY 6.5. *Set $\Upsilon = \rho\sqrt{n}$ and let $T \subset \mathbb{R}^n$. Assume that $\log n \leq d^*(T) \leq c_1 n/\Upsilon^2$ and set r to satisfy that $d^*(T) \sim r\log(en/r)$. Then on the event from Theorem 6.4, with probability at least $1-2\exp(-c_2u^2d^*(T))$ with respect to ε, we have that for any $t \in T$,*

$$\|n^{-1/2}WD_{\varepsilon''}\Gamma_{\varepsilon'}D_\varepsilon t\|_{[r]} \leq c_3u^2\Upsilon\frac{\ell_*(T)}{\sqrt{n}}$$

Proof. On the event from Theorem 6.4, Theorem 3.2 and Remark 3.5 imply that for $T \subset \mathbb{R}^n$, with probability at least $1-2\exp(-c_0u^2[d^*(T)+r\log(en/r)])$ with respect to ε, for every $t \in T$,

$$\|n^{-1/2}WD_{\varepsilon''}\Gamma_{\varepsilon'}D_\varepsilon t\|_{[r]} \leq c_1u^2\max\{Q(T), \mathcal{R}^{-1}(T)Q^2(T)\},$$

where

$$Q(T) := Q_r(T, \mathcal{R}(T)) = \rho\mathcal{R}(T)(d^*(T)+r\log(en/r))^{1/2}.$$

By setting r to satisfy that $d^*(T) \sim r\log(en/r)$, we have that $Q(T) \sim \rho\mathcal{R}(T)\sqrt{d^*(T)}$. Note that $Q(T) \geq \mathcal{R}^{-1}(T)Q^2(T)$ provided that $Q(T) \leq \mathcal{R}(T)$. Since $\Upsilon = \rho\sqrt{n}$, this condition is satisfied if $d^*(T) \leq n/\Upsilon^2$. This yields the asserted estimate. ∎

6.3. Proof of Theorem 6.4. The proof follows two steps:

Step 1: *Proof of regularity.*

The first step in the proof is to establish that with high probability with respect to $\varepsilon' \otimes \varepsilon''$, the matrix $n^{-1/2}WD_{\varepsilon''}\Gamma_{\varepsilon'}$ is regular, i.e., it acts in a norm preserving way on sparse vectors. Observe that for any $t \in \mathbb{R}^n$ and any realization of ε'',

$$\|n^{-1/2}WD_{\varepsilon''}\Gamma_{\varepsilon'}\|_2 = \|n^{-1/2}\Gamma_{\varepsilon'}\|_2.$$

Thus, it suffices to establish the regularity of $n^{-1/2}\Gamma_{\varepsilon'}$. This follows immediately from Theorem 4.4, by which implies:

With probability at least $1 - 2\exp(-c_1\gamma\log^4 n)$ with respect to ε', the matrix $n^{-1/2}\Gamma_{\varepsilon'}$ is ρ-regular for

$$\rho = c_2(\gamma)\frac{\log^2 n}{\sqrt{n}}. \tag{6.6}$$

Step 2: *Proof of strong regularity.*

Recall that for any $x \in \mathbb{R}^n$ we have $\Gamma_{\varepsilon'}x = \Gamma_x\varepsilon'$ and $\Gamma_x = \sqrt{n}UD_{Wx}O$, where U, W, and O are as in (4.1). Therefore,

$$n^{-1/2}WD_{\varepsilon''}\Gamma_{\varepsilon'}x = WD_{\varepsilon''}UD_{Wx}O\varepsilon'.$$

Note that this is a random vector of the form $\mathcal{U}D_{\mathcal{W}x}\mathcal{O}\xi$ for

$$\mathcal{U} = WD_{\varepsilon''}U, \quad \mathcal{W} = W, \quad \mathcal{O} = O, \quad \text{and} \quad \xi = \varepsilon'.$$

Therefore, to establish strong regularity, we invoke Theorem 4.5. To that end, observe that the matrix $\mathcal{W} = W$ is of Hadamard-type; in particular, $d_{\mathcal{W}} = 1$ and trivially, $\sup_{x\in\Sigma_{r,n}} \|\mathcal{W}x\|_2 \leq 2$. To estimate

$$d_{\mathcal{U}} = \sqrt{n}\max_{1\leq i,j\leq n} |(WD_{\varepsilon''}U)_{ij}|,$$

we use the following fact.

LEMMA 6.6. *There is an absolute constant $c > 0$ such that the following holds for any $\gamma \geq 1$: for any $V_1, V_2 \in \mathbb{C}^{n\times n}$, with probability at least $1 - n^{-\gamma}$*

$$\max_{1\leq i,j\leq n} |(V_1D_{\varepsilon''}V_2)_{ij}| \leq c\sqrt{\gamma\log n}\max_{1\leq i\leq n} \|V_1^*e_i\|_2 \max_{1\leq i,j\leq n} |(V_2)_{ij}|.$$

The proof of Lemma 6.6 is standard and is presented in Appendix A.

Since W and U are Hadamard-type matrices, Lemma 6.6 shows that there is an event with probability at least $1 - n^{-\gamma}$ (with respect to ε'') such that

$$d_{\mathcal{U}} \leq c_1\sqrt{\gamma\log n}.$$

On that event, it follows from Theorem 4.5 that there is an absolute constant c_2 such that with probability at least $1-2\exp(-c_2u)$ with respect to ε',

$$\sup_{x\in\Sigma_{r,n}} \|n^{-1/2}WD_{\varepsilon''}\Gamma_{\varepsilon'}x\|_{[r]} \le c_3\sqrt{\frac{r}{n}}\cdot\sqrt{\log n}\left(\log(n)\log(r)+\sqrt{u}\right). \tag{6.7}$$

In particular, setting $u\sim\gamma\log^4 n$ and taking the union bound over $1\le r\le n$, we have that with probability at least $1-2\exp(-c_2\gamma\log^4 n)$ with respect to ε',

$$\sup_{x\in\Sigma_{r,n}} \|n^{-1/2}WD_{\varepsilon''}\Gamma_{\varepsilon'}x\|_{[r]} \le c_4(\gamma)\sqrt{\frac{r}{n}}\log^{5/2} n \quad \text{for every} \quad 1\le r\le n. \tag{6.8}$$

In particular:

> On an event with probability at least $1-n^{-\gamma}-\exp(-c_5\gamma\log^4 n)$ with respect to $\varepsilon'\otimes\varepsilon''$, $n^{-1/2}WD_{\varepsilon''}\Gamma_{\varepsilon'}$ is ρ-strongly regular for
>
> $$\rho = c_4(\gamma)\frac{\log^{5/2} n}{\sqrt{n}}.$$

That completes the proof of Theorem 6.4.

Appendix A. Proof of Lemma 6.6. For any $1\le i,j\le n$,

$$(V_1D_{\varepsilon''}V_2)_{ij} = \langle V_1D_{\varepsilon''}V_2e_j,e_i\rangle = \langle\varepsilon'', D^*_{V_2e_j}V_1^*e_i\rangle.$$

Moreover,

$$\|D^*_{V_2e_j}V_1^*e_i\|_2 \le \|D_{V_2e_j}\|_{2\to2}\|V_1^*e_i\|_2 = \|V_2e_j\|_\infty\|V_1^*e_i\|_2 \le \max_{1\le i\le n}\|V_1^*e_i\|_2\max_{1\le i,j\le n}|(V_2)_{ij}|.$$

By Hoeffding's inequality, for any $u>0$,

$$\mathbb{P}\Big(|\langle\varepsilon'', D^*_{V_2e_j}V_1^*e_i\rangle| \ge u\max_{1\le i\le n}\|V_1^*e_i\|_2\max_{1\le i,j\le n}|(V_2)_{ij}|\Big) \le 2\exp(-u^2/2).$$

The result now follows by taking a union bound over all i and j.

Appendix B. Proof of Theorem 4.1. The goal is to show that with probability at least $1-n^{-\gamma}$ with respect to $G\otimes\varepsilon'\otimes\varepsilon''$, the matrix $B=m^{-1/2}R_I\Gamma_GD_{\varepsilon''}n^{-1/2}\Gamma_{\varepsilon'}$ is ρ-strongly regular for $\rho\sim c(\gamma)m^{-1/2}\log^{5/2} n$.

Recall that $\Psi=n^{-1/2}D_{\varepsilon''}\Gamma_{\varepsilon'}$, and thus, for every $x\in\mathbb{R}^n$,

$$Bx = \frac{1}{\sqrt{m}}R_I\Gamma_G\Psi x = \frac{1}{\sqrt{m}}R_I\Gamma_{\Psi x}G.$$

B.1. Regularity of B. As we have noted in (4.4), to establish regularity it suffices to estimate

$$\sup_{x\in\Sigma_{r,n}} \big|\ \|Bx\|_2^2-\|x\|_2^2\ \big|$$

$$\le \sup_{x\in\Sigma_{r,n}}\left|\ \left\|\frac{1}{\sqrt{m}}R_I\Gamma_{\Psi x}G\right\|_2^2 - \mathbb{E}_G\left\|\frac{1}{\sqrt{m}}R_I\Gamma_{\Psi x}G\right\|_2^2\ \right| + \sup_{x\in\Sigma_{r,n}}\left|\left\|\frac{1}{\sqrt{n}}\Gamma_x\varepsilon'\right\|_2^2-\|x\|_2^2\right|. \tag{B.1}$$

As it happens, the proof that $n^{-1/2}\Gamma_{\varepsilon'}$ is ρ-regular for $\rho \sim c(\gamma)\frac{\log^2 n}{\sqrt{n}}$ is standard and was used previously in this presentation (see (6.6)). Hence, all that remains is to estimate

$$\sup_{x\in\Sigma_{r,n}} \left| \left\| \frac{1}{\sqrt{m}} R_I \Gamma_{\Psi x} G \right\|_2^2 - \mathbb{E}_G \left\| \frac{1}{\sqrt{m}} R_I \Gamma_{\Psi x} G \right\|_2^2 \right|$$

for $1 \le r \le m$. To that end we invoke Theorem 4.3 for the class of matrices

$$\mathcal{A}_r = \{ R_I \Gamma_{\Psi x} : x \in \Sigma_{r,n} \}$$

in the case $\xi = G$. To estimate the quantities in Theorem 4.3 we follow an almost identical argument to the one used in [33]. We will therefore only outline it here. Let $x, y \in \mathbb{R}^n$. Then

$$\|R_I \Gamma_{\Psi x} - R_I \Gamma_{\Psi y}\|_{2\to 2} \le \|\Gamma_{\Psi(x-y)}\|_{2\to 2} = \sqrt{n}\|\Psi(x-y)\|_\infty$$

and, in particular,

$$\sup_{x\in\Sigma_{r,n}} \|R_I \Gamma_{\Psi x}\|_{2\to 2} \le \sup_{x\in\Sigma_{r,n}} \sqrt{n}\|\Psi x\|_\infty.$$

Set $\|x\| := \sqrt{n}\|\Psi x\|_\infty$. Estimating the γ_2-functional by an entropy integral (see, e.g., [43]), we find that for an absolute constant c_0,

$$\gamma_2(\mathcal{A}_r, \|\cdot\|_{2\to 2}) \le \gamma_2(\Sigma_{r,n}, \|\cdot\|) \le c_0 \int_0^\infty \log^{1/2} \mathcal{N}(\Sigma_{r,n}, \|\cdot\|, u)\, du. \tag{B.2}$$

To estimate the right-hand side, we use two entropic estimates. The first is based on Maurey's lemma and is essentially due to Carl [8] (see also [33] for a proof).

LEMMA B.1. *There exists an absolute constant c such that the following holds. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Let $U \subset \mathbb{R}^n$ be a finite set and assume that for every $1 \le k \le |U|$ and every subset $\{u_1, \dots, u_k\} \subset U$ of cardinality k, $\mathbb{E}_\varepsilon \|\sum_{i=1}^k \varepsilon_i u_i\| \le \alpha\sqrt{k}$. Then for every $t > 0$,*

$$\log \mathcal{N}(\mathrm{conv}(U), \|\cdot\|, t) \le c\left(\frac{\alpha}{t}\right)^2 \log|U|.$$

In our case, $\Sigma_{r,n} \subset \mathrm{conv}(U)$, where $U = \{\pm 2\sqrt{r} e_i : 1 \le i \le n\}$, and $\|x\| = \sqrt{n}\|\Psi x\|_\infty$. Note that for any $x_1, \dots, x_k \in \mathbb{R}^n$

$$\mathbb{E}\left\| \sum_{i=1}^k \varepsilon_i x_i \right\|_\infty \le c_0 \sqrt{\log n} \left(\sum_{i=1}^k \|x_i\|_\infty^2 \right)^{1/2}.$$

Hence, if $J \subset \{1, \dots, n\}$ and $|J| = k$, we have that

$$\begin{aligned} \mathbb{E}_\varepsilon \left\| \sum_{i\in J} \varepsilon_i e_i \right\| = \mathbb{E}_\varepsilon \left\| \sum_{i\in J} \varepsilon_i \sqrt{n}\Psi e_i \right\|_\infty &\le c_0 \sqrt{\log n} \left(\sum_{i\in J} \|\sqrt{n}\Psi e_i\|_\infty^2 \right)^{1/2} \\ &\le c_0 \sqrt{\log n} \max_{1\le i,j\le n} \sqrt{n}|\Psi_{ij}| \cdot \sqrt{k}. \end{aligned}$$

Now, condition on the event on which

$$\max_{1\le i,j\le n}\sqrt{n}|\Psi_{ij}|\le c_1\sqrt{\gamma\log n}. \tag{B.3}$$

Since $\Psi = n^{-1/2}D_{\varepsilon''}\Gamma_{\varepsilon'}$, Lemma 6.6 (with $V_1 = \mathrm{Id}_n$ and $V_2 = \frac{1}{\sqrt{n}}\Gamma_{\varepsilon'}$) shows that this event holds with probability at least $1-n^{-\gamma}$ with respect to ε''. On that event, Lemma B.1 with

$$\alpha = c_0\sqrt{r}\sqrt{\log n}\max_{1\le i,j\le n}\sqrt{n}|\Psi_{ij}|\le c_2(\gamma)\sqrt{r}\log n$$

implies that for every $t>0$

$$\log\mathcal{N}(\Sigma_{r,n},\|\cdot\|,t)\le c_3(\gamma)\frac{r}{t^2}\log^3 n.$$

The second entropic estimate we require is volumetric: $\Sigma_{r,n}$ is the union of the Euclidean unit balls B_2^J that are supported on sets $J\subset\{1,\dots,n\}$, where $|J|=r$. If $x,y\in B_2^J$ then

$$\begin{aligned}\|x-y\| &= \sqrt{n}\|\Psi(x-y)\|_\infty \le \max_{1\le i\le n}|\sqrt{n}\,\langle\Psi^* e_i, x-y\rangle| \le \max_{1\le i,j\le n}\sqrt{n}|\Psi_{ij}|\|x-y\|_1\\ &\le \max_{1\le i,j\le n}\sqrt{n}|\Psi_{ij}|\sqrt{r}\|x-y\|_2 \le c_4(\gamma)\sqrt{\log n}\sqrt{r}\|x-y\|_2,\end{aligned}$$

where we again used (B.3). Hence, for any $t>0$

$$\log\mathcal{N}(\Sigma_{r,n},\|\cdot\|,t)=\log\mathcal{N}\left(\cup_{|J|=r}B_2^J,\|\cdot\|,t\right)\le\max_{|J|=r}\log\mathcal{N}(B_2^J,\|\cdot\|_2,t')+r\log(en/r),$$

where

$$t'=\frac{t}{c_4(\gamma)\sqrt{\log n}\sqrt{r}}.$$

Now the entropy estimate follows from a standard volumetric argument.

Conditioned on the event (B.3) and using these two entropic estimates for large and small u, respectively, in the entropy integral in (B.2), it follows from a standard computation that for any $1\le r\le n$,

$$\gamma_2(\mathcal{A}_r,\|\cdot\|_{2\to 2})\le c_5(\gamma)\sqrt{r}\log^{5/2} n.$$

A similar argument shows that $d_{2\to 2}(\mathcal{A}_r)\le c_6(\gamma)\sqrt{r}\sqrt{\log n}$.

Finally, let us estimate

$$d_{HS}(\mathcal{A}_r)=\sup_{x\in\Sigma_{r,n}}\|R_I\Gamma_{\Psi x}\|_{HS}\le\sqrt{m}\sup_{x\in\Sigma_{r,n}}\|\Psi x\|_2.$$

Observe that

$$\sup_{x\in\Sigma_{r,n}}\|\Psi x\|_2=\frac{1}{\sqrt{n}}\sup_{x\in\Sigma_{r,n}}\|\Gamma_{\varepsilon'}x\|_2.$$

We have that $\mathbb{E}\|\Gamma_x\varepsilon'\|_2^2 = n\|x\|_2^2\le n$, and if

$$a:=\sup_{x\in\Sigma_{r,n}}\left|\,\|\Gamma_x\varepsilon'\|_2^2-\mathbb{E}\|\Gamma_x\varepsilon'\|_2^2\,\right|,$$

then

$$\frac{1}{\sqrt{n}} \sup_{x\in\Sigma_{r,n}} \|\Gamma_{\varepsilon'} x\|_2 \le \frac{1}{\sqrt{n}} (a+n)^{1/2}.$$

Using Theorem 4.4 it follows that with probability at least $1-2\exp(-c_7\gamma\log^4(n))$ with respect to ε', for all $1\le r\le m$

$$\sup_{x\in\Sigma_{r,n}} \left| \|\Gamma_x\varepsilon'\|_2^2 - \mathbb{E}\|\Gamma_x\varepsilon'\|_2^2 \right| \le n,$$

provided that $n\ge c_8(\gamma)m\log^4(n)$, implying that

$$\sup_{x\in\Sigma_{r,n}} \|\Psi x\|_2 = \frac{1}{\sqrt{n}} \sup_{x\in\Sigma_{r,n}} \|\Gamma_{\varepsilon'} x\|_2 \le 2. \tag{B.4}$$

Therefore, conditioned on the above event it follows that

$$d_{HS}(\mathcal{A}_r) \le 2\sqrt{m}.$$

Combining these estimates with Theorem 4.3 and a union bound over all $1\le r\le m$, we have that with probability at least $1-2\exp(-c_9\gamma\log^4 n)$ with respect to G, the matrix B is ρ-regular for $\rho = c_{10}(\gamma)\frac{\log^{5/2} n}{\sqrt{m}}$.

B.2. Strong regularity of B. To complete the proof of Theorem 4.1 one has to show that for every $1\le r\le m$,

$$\sup_{x\in\Sigma_{r,n}} \left\| \frac{1}{\sqrt{m}} R_I \Gamma_{\Psi x} G \right\|_{[r]} \le \rho\sqrt{r}.$$

To that end, it suffices to prove that

$$\sup_{x\in\Sigma_{r,n}} \left\| \frac{1}{\sqrt{m}} \Gamma_{\Psi x} G \right\|_{[r]} = \sqrt{\frac{n}{m}} \sup_{x\in\Sigma_{r,n}} \|U D_{W\Psi x} O G\|_{[r]} \le \rho\sqrt{r}.$$

Thus, the proof of Theorem 4.1 is completed once the following lemma is established.

LEMMA B.2. *For $\gamma>1$ there exist constants $\tilde{c}_0$ and $\tilde{c}_1$ depending only (polynomially) on γ such that the following holds. Let $n\ge \tilde{c}_0 m\log^4 n$. With probability at least $1-n^{-\gamma}$ with respect to $G\otimes\varepsilon'\otimes\varepsilon''$, for every $1\le r\le m$,*

$$\sup_{x\in\Sigma_{r,n}} \|U D_{W\Psi x} O G\|_{[r]} \le \tilde{c}_1 \sqrt{\frac{r}{n}} \log^{5/2} n.$$

Proof. Using the notation of Theorem 4.5,

$$U D_{W\Psi x} O G = \mathcal{U} D_{\mathcal{W}x} \mathcal{O}\xi,$$

where

$$\mathcal{U} = U, \quad \mathcal{W} = W\Psi, \quad \mathcal{O} = O, \quad \text{and} \quad \xi = G.$$

Just as in (B.4), with probability at least $1-2\exp(-c_0\gamma\log^4(n))$ with respect to ε', for all $1\le r\le m$,

$$\sup_{x\in\Sigma_{r,n}} \|\mathcal{W}x\|_2 = \sup_{x\in\Sigma_{r,n}} \|W\Psi x\|_2 = \frac{1}{\sqrt{n}} \sup_{x\in\Sigma_{r,n}} \|\Gamma_{\varepsilon'} x\|_2 \le 2.$$

Moreover, by Lemma 6.6 (with $V_1 = W$ and $V_2 = \frac{1}{\sqrt{n}}\Gamma_{\varepsilon'}$)

$$d_{\mathcal{W}} = \max_{1\le i,j\le n} \sqrt{n}|(W\Psi)_{ij}| \le c_1\sqrt{\gamma \log n}$$

with probability at least $1-n^{-\gamma}$ with respect to ε''. The result now follows by applying Theorem 4.5 conditioned on those two events. ∎

REFERENCES

[1] D. Achlioptas, *Database-friendly random projections: Johnson-Lindenstrauss with binary coins*, J. Comput. System Sci., 66 (2003), pp. 671–687.

[2] N. Ailon and B. Chazelle, *The Fast Johnson–Lindenstrauss transform and approximate nearest neighbors*, SIAM J. Comput., 39 (2009), pp. 302–322.

[3] N. Ailon and E. Liberty, *Fast dimension reduction using Rademacher series on dual BCH codes*, Discrete Comput. Geom., 42 (2009), pp. 615–630.

[4] N. Alon and B. Klartag, *Optimal compression of approximate inner products and dimension reduction*, in 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society, Los Alamitos, CA, 2017, pp. 639–650.

[5] S. Bamberger and F. Krahmer, *Optimal fast Johnson–Lindenstrauss embeddings for large data sets*, Sampl. Theory Signal Process. Data Anal., 19 (2021), pp. 1–23.

[6] J. Bourgain, S. Dirksen, and J. Nelson, *Toward a unified theory of sparse dimensionality reduction in Euclidean space*, Geom. Funct. Anal., 25 (2015), pp. 1009–1088.

[7] S. P. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[8] B. Carl, *Inequalities of Bernstein-Jackson-type and the degree of compactness of operators in Banach spaces*, in Ann. Inst. Fourier, Vol. 35, 1985, pp. 79–118.

[9] K. L. Clarkson and D. P. Woodruff, *Numerical linear algebra in the streaming model*, in Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, ACM, New York, 2009, pp. 205–214.

[10] A. Dasgupta, R. Kumar, and T. Sarlós, *A sparse Johnson-Lindenstrauss transform*, in Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC), ACM, New York, 2010, pp. 341–350.

[11] S. Dirksen, *Tail bounds via generic chaining*, Electron. J. Probab., 20 (2015), pp. 1–29.

[12] S. Dirksen, H. C. Jung, and H. Rauhut, *One-bit compressed sensing with partial Gaussian circulant matrices*, Inf. Inference, 9 (2020), pp. 601–626.

[13] S. Dirksen and S. Mendelson, *Robust one-bit compressed sensing with partial circulant matrices*, Ann. Appl. Probab., 33 (2023), pp. 1874–1903.

[14] S. Dirksen, S. Mendelson, and A. Stollenwerk, *Sharp estimates on random hyperplane tessellations*, SIAM J. Math. Data Sci., 4 (2022), pp. 1396–1419.

[15] S. Dirksen and A. Stollenwerk, *Fast binary embeddings with Gaussian circulant matrices: Improved bounds*, Discrete Comput. Geom., 60 (2018), pp. 599–626.

[16] O. N. Fandina, M. M. Høgsgaard, and K. G. Larsen, *The fast Johnson-Lindenstrauss transform is even faster*, in Proceedings of the 40th International Conference on Machine Learning (PMLR), 202 (2023), pp. 9689–9715.

[17] C. B. Freksen and K. G. Larsen, *On using Toeplitz and circulant matrices for Johnson–Lindenstrauss transforms*, Algorithmica, 82 (2020), pp. 338–354.

[18] Y. Gordon, *On Milman's inequality and random subspaces which escape through a mesh in* $\mathbb{R}^n$, in Geometric Aspects of Functional Analysis, Springer, Berlin, 1988, pp. 84–106.

[19] S. Har-Peled, P. Indyk, and R. Motwani, *Approximate nearest neighbor: Towards removing the curse of dimensionality*, Theory Comput., 8 (2012), pp. 321–350.

[20] A. Hinrichs and J. Vybíral, *Johnson-Lindenstrauss lemma for circulant matrices*, Random Structures Algorithms, 39 (2011), pp. 391–398.

[21] T. Huynh and R. Saab, *Fast binary embeddings and quantized compressed sensing with structured matrices*, Comm. Pure Appl. Math., 73 (2020), pp. 110–149.

[22] P. Indyk, *Algorithmic applications of low-distortion geometric embeddings*, in Proceedings of the 42nd Annual Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society, Los Alamitos, CA, 2001, pp. 10–33.

[23] P. Indyk, *Uncertainty principles, extractors, and explicit embeddings of l2 into l1*, in Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing, ACM, New York, 2007, pp. 615–620.

[24] P. Indyk and T. Wagner, *Near-optimal (Euclidean) metric compression*, in Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, 2017, pp. 710–723.
[25] P. Indyk and T. Wagner, *Optimal (Euclidean) metric compression*, SIAM J. Comput., 51 (2022), pp. 467–491.
[26] L. Jacques, *A quantized Johnson-Lindenstrauss lemma: The finding of Buffon's needle*, IEEE Trans. Inform. Theory, 61 (2015), pp. 5012–5027.
[27] L. Jacques, *Small width, low distortions: Quantized random embeddings of low-complexity sets*, IEEE Trans. Inform. Theory, 63 (2017), pp. 5477–5495.
[28] L. Jacques and V. Cambareri, *Time for dithering: Fast and quantized random embeddings via the restricted isometry property*, Inf. Inference, 6 (2017), pp. 441–476.
[29] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, *Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors*, IEEE Trans. Inform. Theory, 59 (2013), pp. 2082–2102.
[30] V. Jain, N. S. Pillai, A. Sah, M. Sawhney, and A. Smith, *Fast and memory-optimal dimension reduction using Kac's walk*, Ann. Appl. Probab., 32 (2022), pp. 4038–4064.
[31] W. B. Johnson and J. Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, Contemp. Math., 26 (1984), pp. 189–206.
[32] D. M. Kane and J. Nelson, *Sparser Johnson-Lindenstrauss transforms*, J. ACM, 61 (2014), 4.
[33] F. Krahmer, S. Mendelson, and H. Rauhut, *Suprema of chaos processes and the restricted isometry property*, Comm. Pure Appl. Math., 67 (2014), pp. 1877–1904.
[34] F. Krahmer and R. Ward, *New and improved Johnson–Lindenstrauss embeddings via the restricted isometry property*, SIAM J. Math. Anal., 43 (2011), pp. 1269–1281.
[35] K. G. Larsen and J. Nelson, *Optimality of the Johnson-Lindenstrauss lemma*, in 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society, Los Alamitos, CA, 2017, pp. 633–638.
[36] M. Ledoux and M. Talagrand, *Probability in Banach Spaces*, Springer, Berlin, 1991.
[37] K. Makarychev, Y. Makarychev, and I. Razenshteyn, *Performance of Johnson-Lindenstrauss transform for k-means and k-medians clustering*, in Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, ACM, New York, 2019, pp. 1027–1038.
[38] S. Mendelson, *Column randomization and almost-isometric embeddings*, Inf. Inference, 12 (2023), pp. 1–25.
[39] S. Oymak and B. Recht, *Near-Optimal Bounds for Binary Embeddings of Arbitrary Sets*, preprint, arXiv:1512.04433, 2015.
[40] S. Oymak, B. Recht, and M. Soltanolkotabi, *Isometric sketching of any set via the restricted isometry property*, Inf. Inference, 7 (2018), pp. 707–726.
[41] S. Oymak, C. Thrampoulidis, and B. Hassibi, *Near-optimal sample complexity bounds for circulant binary embedding*, in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Piscataway, NJ, 2017, pp. 6359–6363.
[42] Y. Plan and R. Vershynin, *Dimension reduction by random hyperplane tessellations*, Discrete Comput. Geom., 51 (2014), pp. 438–461.
[43] M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*, Ergeb. Math. Grenzgeb. (3) 60, Springer, Heidelberg, 2014.
[44] J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher, *Streaming low-rank matrix approximation with an application to scientific simulation*, SIAM J. Sci. Comput., 41 (2019), pp. A2430–A2463.
[45] J. Vybíral, *A variant of the Johnson-Lindenstrauss lemma for circulant matrices*, J. Funct. Anal., 260 (2011), pp. 1096–1105.
[46] D. P. Woodruff, *Sketching as a tool for numerical linear algebra*, Found. Trends Theor. Comput. Sci., 10 (2014), pp. 1–157.
[47] X. Yi, C. Caramanis, and E. Price, *Binary embedding: Fundamental limits and fast algorithm*, in Proceedings of the 32nd International Conference on Machine Learning (PMLR), 37 (2015), pp. 2162–2170.
[48] F. X. Yu, A. Bhaskara, S. Kumar, Y. Gong, and S.-F. Chang, *On binary embedding using circulant matrices*, J. Mach. Learn. Res., 18 (2018), pp. 1–30.
[49] J. Zhang and R. Saab, *Faster Binary Embeddings for Preserving Euclidean Distances*, preprint, arXiv:2010.00712, 2020.