Column randomization and almost-isometric embeddings

Shahar Mendelson *

March 10, 2021

Abstract

The matrix $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is $(\delta, k)$-regular if for any $k$-sparse vector $x$,

$$\|Ax\|_2^2 - \|x\|_2^2 \leq \delta \sqrt{k} \|x\|_2^2.$$ 

We show that if $A$ is $(\delta, k)$-regular for $1 \leq k \leq 1/\delta^2$, then by multiplying the columns of $A$ by independent random signs, the resulting random ensemble $A_\varepsilon$ acts on an arbitrary subset $T \subset \mathbb{R}^n$ (almost) as if it were gaussian, and with the optimal probability estimate: if $\ell_*(T)$ is the gaussian mean-width of $T$ and $d_T = \sup_{t \in T} \|t\|_2$, then with probability at least $1 - \frac{2}{\sqrt{\log n}}$, for every $t \in T$,

$$\sup_{t \in T} \left( \|A_\varepsilon t\|_2^2 - \|t\|_2^2 \right) \leq C \left( \Lambda d_T \ell_*(T) + (\delta \ell_*(T))^2 \right),$$

where $\Lambda = \max\{1, \delta^2 \log(n\delta^2)\}$. This estimate is optimal for $0 < \delta \leq 1/\sqrt{\log n}$.

1 Introduction

Linear operators that act in an almost-isometric way on subsets of $\mathbb{R}^n$ are of obvious importance. Although approximations of isometries are the only operators that almost preserve the Euclidean norm of any point in $\mathbb{R}^n$, one may consider a more flexible alternative: a random ensemble of operators $\Gamma$ such that, for any fixed $T \subset \mathbb{R}^n$, with high probability, $\Gamma$ “acts well” on every element of $T$. Such random ensembles have been studied extensively over the years, following the path paved by the celebrated work of Johnson and Lindenstrauss in [5]. Here we formulate the Johnson-Lindenstrauss Lemma in one of its gaussian versions:

**Theorem 1.1.** There exist absolute constants $c_0$ and $c_1$ such that the following holds. Let $1 \leq m \leq n$ and set $\Gamma : \mathbb{R}^n \rightarrow \mathbb{R}^m$ to be a random matrix whose entries are independent, standard gaussian random variables. Let $T \subset S^{n-1}$ be of cardinality at most $\exp(c_0m)$. Then for $m^{-1/2} \sqrt{\log |T|} < \rho < 1$, with probability at least $1 - 2 \exp(-c_1 \rho^2 m)$, for every $t \in T$,

$$\left| \left\| m^{-1/2} \Gamma t \right\|_2^2 - 1 \right| \leq \rho.$$ 

The scope of Theorem 1.1 can be extended to more general random ensembles than the gaussian one, e.g., to a random matrix whose rows are iid copies of a centred random vector

*Centre for Mathematics and its Applications, The Australian National University. Email: shahar.mendelson@anu.edu.au
that exhibits suitable decay properties (see, e.g. \[4, 8\]). It is far more challenging to construct a random ensemble that, on the one hand, satisfies a version of Theorem 1.1 and on the other is based on “few random bits” or is constructed using a heavy-tailed random vector.

A significant breakthrough towards more general “Johnson-Lindenstrauss transforms” came in \[7\], where it was shown that a matrix that satisfies a suitable version of the restricted isometry property, can be converted to the wanted random ensemble by multiplying its columns by random signs. More accurately, let \(\varepsilon_1, \ldots, \varepsilon_n\) be independent, symmetric \([-1, 1]\)-valued random variables. Set \(D_\varepsilon = \text{diag}(\varepsilon_1, \ldots, \varepsilon_n)\) and for a matrix \(A : \mathbb{R}^n \to \mathbb{R}^m\) define

\[A_\varepsilon = AD_\varepsilon.\]

From here on we denote by \(\Sigma_k\) the subset of \(S^{n-1}\) consisting of vectors that are supported on at most \(k\) coordinates.

**Definition 1.2.** A matrix \(A : \mathbb{R}^n \to \mathbb{R}^m\) satisfies the restricted isometry property of order \(k\) and level \(\delta \in (0, 1)\) if

\[
\sup_{x \in \Sigma_k} \left| \Vert Ax \Vert_2^2 - 1 \right| \leq \delta.
\]

**Theorem 1.3.** \([7]\) There are absolute constants \(c_0\) and \(c_1\) such that the following holds. Let \(\lambda > 0\) and \(\rho \in (0, 1)\). Consider \(T \subset \mathbb{R}^n\) and let \(k \geq c \log(e |T| / \lambda)\). If \(A\) satisfies the restricted isometry property of order \(k\) and at level \(\delta < \frac{\rho}{4}\), then with probability at least \(1 - \lambda\), for every \(t \in T\),

\[
(1 - \rho)\|t\|_2^2 \leq \|A_\varepsilon t\|_2^2 \leq (1 - \rho)\|t\|_2^2.
\]

While Theorem 1.3 does not recover the probability estimate from Theorem 1.1 it does imply at the constant probability level that \(A_\varepsilon\) is an almost isometry in the random ensemble sense: if \(A\) is a matrix that \(1 \pm \delta\)-preserves the norms of vectors that are \(c \log |T|\) sparse, then a typical realization of the random ensemble \(A_\varepsilon\), \(1 \pm c'\delta\) preserves the norms of all the elements in \(T\).

Various extensions of Theorem 1.1 that hold for arbitrary subsets of \(\mathbb{R}^n\) have been studied over the years. In such extensions the “complexity parameter” \(\log |T|\) is replaced by more suitable counterparts. A rather general version of Theorem 1.1 follows from a functional Bernstein inequality (see, e.g., \[3, 8, 2\]), and to formulate that inequality in the gaussian case we require the following definition.

**Definition 1.4.** Let \(g_1, \ldots g_n\) be independent, standard gaussian random variables. For \(T \subset \mathbb{R}^n\) set

\[
\ell_s(T) = \mathbb{E} \sup_{t \in T} \left| \sum_{i=1}^n g_i t_i \right| \quad \text{and} \quad d_T = \sup_{t \in T} \|t\|_2.
\]

Let

\[
\left( \frac{\ell_s(T)}{d_T} \right)^2
\]

be the critical dimension of the set \(T\).

The critical dimension appears naturally when studying the geometry of convex sets—for example, in the context of the Dvoretzky-Milman Theorem (see \[1\] and references therein for
more details). It is the natural alternative to log \(|T|\)—which was suitable for finite subsets of sphere \(S^{n-1}\).

Let \(G = (g_i)_{i=1}^n\) be the standard gaussian random vector in \(\mathbb{R}^n\), set \(G_1, ..., G_m\) to be independent copies of \(G\) and put
\[
\Gamma = \sum_{i=1}^m \langle G_i, \cdot \rangle e_i
\]
to be the random ensemble used in Theorem 1.1.

**Theorem 1.5.** There exist absolute constants \(c_0, c_1\) and \(C\) such that the following holds. If \(T \subset \mathbb{R}^n\) and \(u \geq c_0\) then with probability at least \(1 - 2 \exp\left(-c_1 u^2 \left(\frac{\ell_s(T)}{d_T}\right)^2\right)\),

for every \(t \in T\),
\[
\left\|m^{-1/2} \Gamma t\right\|_2^2 - \|t\|_2^2 \leq C \left(u d_T \frac{\ell_s(T)}{\sqrt{m}} + u^2 \left(\frac{\ell_s(T)}{\sqrt{m}}\right)^2\right).
\]

(1.1)

One may use Theorem 1.5 to ensure that the uniform error in (1.1) is at most \(\max\{\rho, \rho^2\} d_T^2\).

Indeed, if \(\ell_s(T)/d_T \sim \rho\), then with probability at least \(1 - 2 \exp(-c_3 \rho^2 m)\),
\[
\sup_{t \in T} \left\|m^{-1/2} \Gamma t\right\|_2^2 - \|t\|_2^2 \leq \max\{\rho, \rho^2\} d_T^2,
\]
(1.2)

which is a natural counterpart of Theorem 1.1 once \(\log |T|\) is replaced by \((\ell_s(T)/d_T)^2\).

As it happens, a version of Theorem 1.3 that is analogous to (1.2) was proved in [9], using the notion of a multi-level RIP.

**Definition 1.6.** Let \(L = \lceil \log_2 n \rceil\). For \(\delta > 0\) and \(s \geq 1\) the matrix \(A\) satisfies a multi-scale RIP with distortion \(\delta\) and sparsity \(s\) if, for every \(1 \leq \ell \leq L\) and every \(x \in \Sigma_{2^\ell s}\), one has
\[
\|Ax\|_2^2 - \|x\|_2^2 \leq \max\left\{2^{\ell/2} \delta, 2^{\ell/2} \delta^2\right\}.
\]

Definition 1.6 implies that if \(k \geq s\) then
\[
\sup_{x \in \Sigma_k} \|Ax\|_2^2 - \|x\|_2^2 \leq \max\{\sqrt{k} \delta, k \delta^2\}.
\]

**Example 1.7.** Let \(\Gamma : \mathbb{R}^n \to \mathbb{R}^m\) be a gaussian matrix as above and set \(A = m^{-1/2} \Gamma\). It is standard to verify (using, for example, Theorem 1.5 and a well-known estimate on \(\ell_s(\Sigma_k)\)) that with probability at least \(1 - 2 \exp(-ck \log(en/k))\),
\[
\sup_{x \in \Sigma_k} \|Ax\|_2^2 - 1 \leq C \sqrt{\frac{k \log(en/k)}{m}}.
\]
By the union bound over \(k\) it follows that with a nontrivial probability, \(A\) satisfies a multi-scale RIP with \(s = 1\) and \(\delta \sim m^{-1/2} \sqrt{\log(en)}\). Observe that the second term in the multi-scale RIP—namely \(k \delta^2\)—is not needed here.
Remark 1.8. Example 1.7 gives a good intuition on the role $\delta$ has in well-behaved situations: it should scale (roughly) like $1/\sqrt{m}$, where $m$ is the number of rows of the matrix $A$.

The following theorem is the starting point of this note: an estimate on the error a typical realization of the random ensemble $A_\varepsilon = AD_\varepsilon$ has when acting on an arbitrary $T \subset \mathbb{R}^n$, given that $A$ satisfies an appropriate multi-scale RIP.

**Theorem 1.9.** [9] There are absolute constants $c$ and $C$ such that the following holds. Let $\eta, \rho > 0$ and $A : \mathbb{R}^n \to \mathbb{R}^m$ that satisfies a multi-scale RIP with sparsity level $s = c(1 + \eta)$ and distortion

$$
\delta = C \frac{\rho d_T}{\max\{\ell_*(T), d_T\}}. 
$$

(1.3)

Then for $T \subset \mathbb{R}^n$, with probability at least $1 - \eta$,

$$
\sup_{t \in T} \left| \left| A_\varepsilon t \right|_2^2 - \|t\|_2^2 \right| \leq \max\{\rho^2, \rho\} d_T^2.
$$

(1.4)

To put Theorem 1.9 in some context, if the belief is that $A_\varepsilon$ should exhibit the same behaviour as the gaussian matrix $m^{-1/2}\Gamma$, then (keeping in mind that $\delta$ should scale like $1/\sqrt{m}$), “a gaussian behaviour” as in Theorem 1.5 is that with high probability,

$$
\sup_{t \in T} \left| \left| A_\varepsilon t \right|_2^2 - \|t\|_2^2 \right| \leq c \left( \delta d_T \ell_*(T) + (\delta \ell_*(T))^2 \right).
$$

Observe that $\ell_*(T) \gtrsim d_T$, implying by (1.3) that $\rho \sim \delta \ell_*(T)/d_T$. Hence, the error in (1.4) in terms of $\delta$ is indeed

$$
\sim d_T \delta \ell_*(T) + (\delta \ell_*(T))^2.
$$

However, despite the “gaussian error”, the probability estimate in Theorem 1.9 is far weaker than in Theorem 1.5—it is just at the constant level.

Our main result is that using a modified, seemingly less restrictive version of the multi-scale RIP, $A_\varepsilon$ acts on $T$ as if it were a gaussian operator: achieving the same distortion and probability estimate as in Theorem 1.5.

**Definition 1.10.** Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a matrix. For $\delta > 0$ let $1 \leq k^\ast \leq n$ be the largest such that for every $1 \leq k \leq k^\ast$, $A$ is a $(\delta, k)$ regular; that is, for every $1 \leq k \leq k^\ast$

$$
\sup_{x \in \Sigma_k} \left| \left| Ax \right|_2^2 - \|x\|_2^2 \right| \leq \delta \sqrt{k}.
$$

**Theorem 1.11.** There exist absolute constants $c$ and $C$ such that the following holds. Let $\delta > 0$ and set $\Lambda = \max\{1, \delta^2 \log(n\delta^2)\}$. If $k^\ast \geq 1/\delta^2$, $T \subset \mathbb{R}^n$ and $u \geq 1$, then with probability at least

$$
1 - 2 \exp \left( -cu^2 \left( \frac{\ell_*(T)}{d_T} \right)^2 \right),
$$

we have that

$$
\sup_{t \in T} \left| \left| A_\varepsilon t \right|_2^2 - \|t\|_2^2 \right| \leq Cu^2 \left( \Lambda \cdot d_T \delta \ell_*(T) + (\delta \ell_*(T))^2 \right).
$$

(1.5)
Remark 1.12. The sub-optimality in Theorem 1.11 lies in the factor \( \Lambda \)—in the range where \( \delta \) is relatively large: at least \( 1/\log n \). For \( \delta \leq 1/\log n \) we have that \( \Lambda = 1 \) and Theorem 1.11 recovers the functional Bernstein inequality for \( u \sim 1 \); that holds despite the fact that \( A_\varepsilon \) is based only on \( n \) “random bits”.

Moreover, for the error in (1.5) to have a chance of being a nontrivial two-sided estimate, i.e., that for some \( 0 < \rho < 1 \) and every \( t \in T \),

\[
\|A_\varepsilon t\|_2^2 - \|t\|_2^2 \leq \rho \delta T^2,
\]

\( \delta \) has to be smaller than \( \sim d_T/\ell_*(T) \). In particular, if the critical dimension of \( T \), \((\ell_*(T)/d_T)^2\) is at least \( \log n \), a choice of \( \delta \leq (d_T/\ell_*(T)) \) leads to \( \Lambda = 1 \) and thus to an optimal outcome in Theorem 1.11.

Theorem 1.11 clearly improves the probability estimate from Theorem 1.9. The other (virtual) improvement is that the matrix \( A \) need only be \((\delta, k_*)\)-regular for \( k \leq 1/\delta^2 \), and the way \( A \) acts on \( \Sigma_k \) for \( k > 1/\delta^2 \) is of no importance. The reason for calling that improvement “virtual” is the following observation:

**Lemma 1.13.** If \( k^* \geq 1/\delta^2 \) then for any \( 1 \leq s \leq n \),

\[
\sup_{x \in \Sigma_s} \|Ax\|_2^2 - 1 \leq 4 \max\{\delta \sqrt{s}, \delta^2 s\}.
\]

In other words, the second term in the multi-scale RIP condition follows automatically from the first one and the fact that \( k^* \) is sufficiently large.

**Proof.** Let \( x \in \Sigma_s \) for \( s \geq k_* \), and let \((J_i)_{i=1}^\ell\) be a decomposition of the support of \( x \) to coordinate blocks of cardinality \( k_*/2 \). Set \( y_i = P_{J_i} x \), that is, the projection of \( x \) onto \( \text{span}\{e_m : m \in J_i\} \) and write \( x = \sum_{i=1}^\ell y_i \). Note that \( \ell \leq 4s/k_* \) and that

\[
\|Ax\|_2^2 = \sum_{i=1}^\ell \|Ay_i\|_2^2 = \sum_{i=1}^\ell \|Ay_i\|_2^2 + \sum_{i \neq j} \langle Ay_i, Ay_j \rangle.
\]

The vectors \( y_i \) are orthogonal and so \( \|x\|_2^2 = \sum_{i=1}^\ell \|y_i\|_2^2 \). Therefore,

\[
\|Ax\|_2^2 - \|x\|_2^2 \leq \sum_{i=1}^\ell \|Ay_i\|_2^2 - \|y_i\|_2^2 + \sum_{i \neq j} \langle Ay_i, Ay_j \rangle.
\]

For the first term, as each \( y_i \) is supported on at most \( k_*/2 \) coordinates, it follows from the regularity condition that

\[
\sum_{i=1}^\ell \|Ay_i\|_2^2 - \|y_i\|_2^2 = \delta \sqrt{k_*}/2 = \delta \sqrt{k_*}/2 \leq \delta \sqrt{s}.
\]

As for the second term, since \( y_i \) and \( y_j \) are orthogonal, \( \|y_i + y_j\|_2 = \|y_i - y_j\|_2 \) and

\[
\langle Ay_i, Ay_j \rangle = \frac{1}{4} \left( \|A(y_i + y_j)\|_2^2 - \|A(y_i - y_j)\|_2^2 \right) = \frac{1}{4} \left( \|A(y_i + y_j)\|_2^2 - \|y_i + y_j\|_2^2 \right) - \frac{1}{4} \left( \|y_i - y_j\|_2^2 - \|A(y_i - y_j)\|_2^2 \right).
\]
Thus, by the regularity of $A$ and as $|\text{supp}(y_i \pm y_j)| \leq k_s$, 

$$\| \langle Ay_i, Ay_j \rangle \| \leq \frac{1}{4} \left( c \sqrt{k_s} \| y_i + y_j \|_2 + \sqrt{k_s} \| y_i - y_j \|_2 \right)$$

$$\leq \frac{1}{2} \sqrt{k_s} \left( \| y_i \|_2^2 + \| y_j \|_2^2 \right).$$

Taking the sum over all pairs $i \neq j$, $i, j \leq \ell$, each factor $\| y_i \|_2^2$ appears at most $2\ell$ times, and $2\ell \leq 8s/k_s$. Hence, using that $1/\delta^2 \leq k_s$

$$\sum_{i \neq j} | \langle Ay_i, Ay_j \rangle | \leq \frac{1}{2} \delta \sqrt{k_s} \cdot \frac{8s}{k_s} \sum_{i=1}^{\ell} \| y_i \|_2^2 \leq 2\delta \frac{s}{\sqrt{k_s}} \| x \|_2^2 \leq 4\delta^2 s.$$

Clearly, Theorem 1.11 implies a suitable version of Theorem 1.9.

**Corollary 1.14.** There exist absolute constants $c$ and $c_1$ such that the following holds. Let $A$ be as above, set $T \subset \mathbb{R}^n$ and $0 < \delta < 1/\sqrt{\log n}$. Let $\rho = c\delta \ell_s(T)/d_T$. Then with probability at least $1 - 2\exp(-c_1 \rho^2/\delta^2)$,

$$\sup_{t \in T} \| A_s t \|_2^2 - \| t \|_2^2 \leq \rho d_T^2.$$

**Remark 1.15.** Recalling the intuition that $m \sim 1/\delta^2$, the outcome of Corollary 1.14 coincides with the estimate in (1.2).

In Section 3 we present one simple application of Theorem 1.11. We show that column randomization of a typical realization of a Bernoulli circulant matrix (complete or partial) exhibits an almost gaussian behaviour (conditioned on the generating vector). In particular, only $2n$ random bits ($n$ from the generating Bernoulli vector and $n$ from the column randomization) are required if one wishes to create a random ensemble that is, effectively, an almost isometry.

The proof of Theorem 1.11 is based on a chaining argument. For more information on the generic chaining mechanism, see Talagrand’s treasured manuscript [10]. We only require relatively basic notions from generic chaining theory, as well as the celebrated majorizing measures theorem.

**Definition 1.16.** Let $T \subset \mathbb{R}^n$. A collection of subsets of $T$, $(T_s)_{s \geq 0}$, is an admissible sequence if $|T_0| = 1$ and for $s \geq 1$, $|T_s| \leq 2^s$. For every $t \in T$ denote by $\pi_s t$ a nearest point to $t$ in $T_s$ with respect to the Euclidean distance. Set $\Delta_s t = \pi_{s+1} t - \pi_s t$ for $s \geq 1$ and let $\Delta_0 t = \pi_0 t$.

The $\gamma_2$ functional with respect to the $\ell_2$ metric is defined by 

$$\gamma_2(T, \| \cdot \|_2) = \inf_{(T_s)} \sup_{t \in T} \sum_{s \geq 0} 2^{s/2} \| \Delta_s t \|_2,$$

where the infimum is taken with respect to all admissible sequences of $T$.

An application of Talagrand’s majorizing measures theorem to the gaussian process $t \rightarrow \sum_{i=1}^n g_i t_i$ shows that $\gamma_2(T, \| \cdot \|_2)$, $\ell_s(T)$ and $\ell_s(t)$ are equivalent:

**Theorem 1.17.** There are absolute constants $c$ and $C$ such that for every $T \subset \mathbb{R}^n$, 

$$c \gamma_2(T, \| \cdot \|_2) \leq \ell_s(T) \leq C \gamma_2(T, \| \cdot \|_2).$$

The proof of Theorem 1.17 can be found, for example, in [10].
2 Proof of Theorem 1.11

We begin the proof with a word about notation: throughout, absolute constants, that is, positive numbers that are independent of all the parameters involved in the problem, are denoted by \(c, c_1, C, \) etc. Their value may change from line to line.

As noted previously, the proof is based on a chaining argument. Let \((T_s)_{s \geq 0}\) be an optimal admissible sequence of \(T\). Set \(s_0\) to satisfy that \(2^{s_0}\) is the critical dimension of \(T\), i.e.,

\[
2^{s_0} = \left( \frac{\ell_s(T)}{d_T} \right)^2
\]

(without loss of generality we may assume that equality holds). Let \(s_0 \leq s_1\) to be named in what follows and observe that

\[
\|A_\varepsilon t\|_2^2 = \|A_\varepsilon (t - \pi_{s_1} t) + A_\varepsilon \pi_{s_1} t\|_2^2 = \|A_\varepsilon (t - \pi_{s_1} t)\|_2^2 + 2 \langle A_\varepsilon (t - \pi_{s_1} t), A_\varepsilon \pi_{s_1} t \rangle + \|A_\varepsilon \pi_{s_1} t\|_2^2.
\]

Writing \(t - \pi_{s_1} t = \sum_{s \geq s_1} \Delta_s t\), it follows that

\[
\|A_\varepsilon (t - \pi_{s_1} t)\|_2 \leq \sum_{s \geq s_1} \|A_\varepsilon \Delta_s t\|_2,
\]

and

\[
|\langle A_\varepsilon (t - \pi_{s_1} t), A_\varepsilon \pi_{s_1} t \rangle| \leq \|A_\varepsilon (t - \pi_{s_1} t)\|_2 \cdot \|A_\varepsilon \pi_{s_1} t\|_2;
\]

Therefore, setting

\[
\Psi^2 = \sup_{t \in T} \|A_\varepsilon \pi_{s_1} t\|_2^2 - \|\pi_{s_1} t\|_2^2 \quad \text{and} \quad \Phi = \sum_{s \geq s_1} \|A_\varepsilon \Delta_s t\|_2^2
\]

we have that for every \(t \in T\),

\[
\|A_\varepsilon \pi_{s_1} t\|_2 \leq \sqrt{\Psi^2 + d_T^2},
\]

and

\[
|\langle A_\varepsilon (t - \pi_{s_1} t), A_\varepsilon \pi_{s_1} t \rangle| \leq \Phi \cdot \sqrt{\Psi^2 + d_T^2}.
\]

Hence,

\[
\sup_{t \in T} \|A_\varepsilon t\|_2^2 - \|t\|_2^2 \leq \Psi^2 + 2\Phi \cdot \sqrt{\Psi^2 + d_T^2} + \Phi^2 + \sup_{t \in T} \|\pi_{s_1} t\|_2^2 - \|t\|_2^2.
\]

(2.3)

To estimate the final term, note that for every \(t \in T\),

\[
\|t\|_2^2 - \|\pi_{s_1} t\|_2^2 \leq \|t - \pi_{s_1} t\|_2^2 + 2 \|t - \pi_{s_1} t, \pi_{s_1} t\| \leq \|t - \pi_{s_1} t\|_2^2 + 2 \|t - \pi_{s_1} t\| \cdot \|\pi_{s_1} t\|_2 \leq \left( \sum_{s \geq s_1} \|\Delta_s t\|_2^2 \right)^2 + 2d_T \sum_{s \geq s_1} \|\Delta_s t\|_2.
\]

By the definition of the \(\gamma_2\) functional and the majorizing measures theorem, for every integer \(s\) and every \(t \in T\),

\[
\|\Delta_s t\|_2 \leq 2^{-s/2} \gamma_2(T, \|2\| \leq c_1 2^{-s/2} \ell_s(T).
\]
Thus,
\[
\sum_{s \geq s_1} \| \Delta_s t \|_2 \leq c_1 \ell_s(T) \sum_{s \geq s_1} 2^{-s/2} \leq c_2 2^{-s_1/2} \ell_s(T),
\]
and
\[
\| \pi_{s_1} t \|_2^2 - \| t \|_2^2 \leq c_3 \left( 2^{-s_1} \ell_s^2(T) + d_T 2^{-s_1/2} \ell_s(T) \right). \tag{2.4}
\]

Equation (2.4) and the wanted estimate in Theorem 1.11 hint on the identity of $2^{s_1}$: it should be larger than $1/\delta^2$. Recalling that $s_0 \leq s_1$ and that $2^{s_0} = (\ell_s(T)/d_T)^2$, set
\[
2^{s_1} = \max \left\{ \frac{1}{\delta^2}, \left( \frac{\ell_s(T)}{d_T} \right)^2, \log (1 + n\delta^2) \right\}.
\]
The reason behind the choice of the third term will become clear in what follows.

With that choice of $s_1$,
\[
\sup_{t \in T} \| \pi_{s_1} t \|_2^2 - \| t \|_2^2 \leq c_3 \left( \delta^2 \ell_s^2(T) + d_T \delta \ell_s(T) \right), \tag{2.5}
\]
and the nontrivial part of the proof is to control $\Phi$ and $\Psi$ with high probability that would lead to the wanted estimate on (2.3).

### 2.1 A decoupling argument

For every $t \in \mathbb{R}^n$,
\[
\| A \varepsilon t \|_2^2 = \sum_{i,j} \langle A e_i, A e_j \rangle \varepsilon_i \varepsilon_j t_i t_j = \sum_{i=1}^n \| A e_i \|_2^2 t_i^2 + \sum_{i \neq j} \langle A e_i, A e_j \rangle \varepsilon_i \varepsilon_j t_i t_j. \tag{2.6}
\]

By the assumption that $A$ is $(\delta, 1)$-regular,
\[
\max_{1 \leq i \leq n} \| A e_i \|_2^2 - 1 \leq \delta,
\]
and noting that $d_T \leq c \ell_s(T)$,
\[
\left| \sum_{i=1}^n \| A e_i \|_2^2 t_i^2 - \| t \|_2^2 \right| = \sum_{i=1}^n \left( \| A e_i \|_2^2 - 1 \right) t_i^2 \leq \delta \| t \|_2^2 \leq \delta d_T^2 \leq c d_T \delta \ell_s(T).
\]

Next, we turn to the “off-diagonal” term in (2.6). For $t \in \mathbb{R}^n$ let
\[
Z_t = \sum_{i \neq j} \varepsilon_i \varepsilon_j t_i t_j \langle A e_i, A e_j \rangle
\]
and let us obtain high probability estimates on $\sup_{u \in U} | Z_u |$ for various sets $U$.

The first step in that direction is decoupling: let $\eta_1, \ldots, \eta_n$ be independent $\{0, 1\}$-valued random variables with mean $1/2$. Set $I = \{ i : \eta_i = 1 \}$ and observe that for every $(\varepsilon_i)_{i=1}^n$,
\[
\sup_{u \in U} | Z_u | \leq 4 \sup_{u \in U} \mathbb{E}_u \left| A \left( \sum_{i \in I} \varepsilon_i u_i e_i \right), A \left( \sum_{j \in I^c} \varepsilon_j u_j e_j \right) \right|. \tag{2.7}
\]
Indeed, for every \((\varepsilon_i)_{i=1}^n\),
\[
\sup_{u \in U} \left| \sum_{i \neq j} \langle A e_i, A e_j \rangle \varepsilon_i \varepsilon_j u_i u_j \right| = 4 \sup_{u \in U} \left| \sum_{i \neq j} \langle A e_i, A e_j \rangle E_\eta (1 - \eta_j) \varepsilon_i \varepsilon_j u_i u_j \right|
\leq 4 \sup_{u \in U} E_\eta \left| \sum_{i \in I, j \in I^c} \langle A e_i, A e_j \rangle \varepsilon_i \varepsilon_j u_i u_j \right|
\]
and for every \(u \in \mathbb{R}^n\),
\[
\sum_{i \in I, j \in I^c} \langle A e_i, A e_j \rangle \varepsilon_i \varepsilon_j u_i u_j = \left\langle A \left( \sum_{j \in I^c} \varepsilon_j u_j e_j \right), A \left( \sum_{i \in I} \varepsilon_i u_i e_i \right) \right\rangle.
\]

Equation (2.7) naturally leads to the following definition:

**Definition 2.1.** For \(v \in \mathbb{R}^n\) and \(I \subset \{1, \ldots, n\}\), set
\[
W_{v,I} = A^\ast A \left( \sum_{i \in I} \varepsilon_i v_i e_i \right).
\]

Recall that \(\pi_{s+1} t\) is the nearest point to \(t\) in \(T_{s+1}\) and \(\Delta_s t = \pi_{s+1} t - \pi_s t\).

**Lemma 2.2.** For every \(t\) and every \((\varepsilon_i)_{i=1}^n\),
\[
|Z_{\pi_s t}| \leq 4 \sum_{s=1}^{s_1} \left( E_\eta \left| \sum_{j \in I^c} \varepsilon_j (\Delta_s t)_j (W_{\pi_{s+1}, I})_j \right| + E_\eta \left| \sum_{i \in I} \varepsilon_i (\Delta_s t)_i (W_{\pi_{s+1}, I})_i \right| \right)
\leq 4 E_\eta \left| \sum_{j \in I^c} \varepsilon_j (\pi_{s_0} t)_j (W_{\pi_{s_0}, I})_j \right|.
\]

**Proof.** Fix an integer \(s\). With the decoupling argument in mind, fix \(I \subset \{1, \ldots, n\}\) and observe that
\[
\left\langle A \left( \sum_{i \in I} \varepsilon_i (\pi_{s+1} t)_i e_i \right), A \left( \sum_{j \in I^c} \varepsilon_j (\pi_{s+1} t)_j e_j \right) \right\rangle
= \left\langle A \left( \sum_{i \in I} \varepsilon_i (\pi_{s+1} t)_i e_i \right), A \left( \sum_{j \in I^c} \varepsilon_j (\Delta_s t)_j e_j \right) \right\rangle
+ \left\langle A \left( \sum_{i \in I} \varepsilon_i (\pi_{s+1} t)_i e_i \right), A \left( \sum_{j \in I^c} \varepsilon_j (\pi_s t)_j e_j \right) \right\rangle
+ \left\langle A \left( \sum_{i \in I} \varepsilon_i (\pi_{s+1} t)_i e_i \right), A \left( \sum_{j \in I^c} \varepsilon_j (\pi_s t)_j e_j \right) \right\rangle
+ \left\langle A \left( \sum_{i \in I} \varepsilon_i (\pi_s t)_i e_i \right), A \left( \sum_{j \in I^c} \varepsilon_j (\pi_s t)_j e_j \right) \right\rangle.
\]

Moreover,
\[
\left\langle A \left( \sum_{i \in I} \varepsilon_i (\pi_{s+1} t)_i e_i \right), A \left( \sum_{j \in I^c} \varepsilon_j (\Delta_s t)_j e_j \right) \right\rangle = \left\langle A^\ast A \left( \sum_{i \in I} \varepsilon_i (\pi_{s+1} t)_i e_i \right), \left( \sum_{j \in I^c} \varepsilon_j (\Delta_s t)_j e_j \right) \right\rangle
= \sum_{j \in I^c} \varepsilon_j (\Delta_s t)_j (W_{\pi_{s+1}, t})_j.
\]
and
\[ \left\langle A\left( \sum_{i \in I} \varepsilon_i(\Delta_{st}^i), \varepsilon_i \right), A\left( \sum_{j \in I^c} \varepsilon_j(\pi_{st}^j), \varepsilon_j \right) \right\rangle = \sum_{i \in I} \varepsilon_i(\Delta_{st}^i)(W_{\pi_{st},t^c})_i. \]

Combining these observations, for every \( t \in T \) and \( (\varepsilon_i)_{i=1}^n \),
\[ \frac{1}{4}Z_{\pi_{st},t} = \mathbb{E}_\eta \left( A\left( \sum_{i \in I} \varepsilon_i(\pi_{st},t), \varepsilon_i \right), A\left( \sum_{j \in I^c} \varepsilon_j(\pi_{st},t), \varepsilon_j \right) \right) \]
\[ = \mathbb{E}_\eta \sum_{s=s_0}^{s_1-1} \sum_{i \in I} \varepsilon_i(\Delta_{st}^i)(W_{\pi_{st},t^c})_i \]
\[ + \mathbb{E}_\eta \sum_{s=s_0}^{s_1-1} \sum_{j \in I^c} \varepsilon_j(\Delta_{st}^j)(W_{\pi_{st+1},t})_j \]
\[ + \mathbb{E}_\eta \sum_{j \in I^c} \varepsilon_j(\pi_{s_0,t})(W_{\pi_{s_0,t},t})_j, \]
from which the claim follows immediately. \( \blacksquare \)

As part of the decoupling argument and to deal with the introduction of the random variables \((\eta_i)_{i=1}^n\) in (2.7), we will use the following elementary fact:

**Lemma 2.3.** Let \( f \) be a function of \((\varepsilon_i)_{i=1}^n\) and \((\eta_i)_{i=1}^n\). If \( \mathbb{E}_\eta|f|^q \leq \kappa^q \) then with \((\varepsilon_i)_{i=1}^n\) - probability at least \( 1 - \exp(-q) \),
\[ \mathbb{E}_\eta(|f| |(\varepsilon_i)_{i=1}^n) \leq e\kappa. \quad (2.10) \]

**Proof.** For a nonnegative function \( h \) we have that point-wise, \( 1_{\{h \geq t\}} \leq h^q/t^q \). Let \( h = \mathbb{E}_\eta(|f| |(\varepsilon_i)_{i=1}^n) \) and note that
\[ Pr_{\varepsilon} (\mathbb{E}_\eta(|f| |(\varepsilon_i)_{i=1}^n) \geq u\kappa) = \mathbb{E}_\varepsilon 1_{\{h \geq u\kappa\}} \leq (u\kappa)^{-q}(\mathbb{E}_\varepsilon h^q). \]

By Jensen’s inequality followed by Fubini’s Theorem,
\[ \mathbb{E}_\varepsilon h^q = \mathbb{E}_\varepsilon (\mathbb{E}_\eta(|f| |(\varepsilon_i)_{i=1}^n) \geq \mathbb{E}_\varepsilon |f|^q \leq \kappa^q, \]
and setting \( u = e \) proves (2.10). \( \blacksquare \)

We shall use Lemma 2.3 in situations where we actually have more information—namely that for any \((\eta_i)_{i=1}^n\), \( \mathbb{E}_\varepsilon(|f|^q |(\eta_i)_{i=1}^n) \leq \kappa^q \) for a well chosen \( \kappa \). As a result,
\[ \mathbb{E}_\eta(|f| |(\varepsilon_i)_{i=1}^n) \leq e\kappa \text{ with probability at least } 1 - \exp(-q). \]

Taking into account Lemma 2.2 and Lemma 2.3 it follows that if one wishes to estimate \( \sup_{t \in T}|Z_{\pi_{st},t}| \) using a chaining argument, it suffices to obtain, for every \( I \subset \{1, \ldots, n\} \), bounds on moments of random variables of the form
\[ \sum_{j \in I^c} \varepsilon_j(\Delta_{st}^j)(W_{\pi_{s+1},t})_j, \sum_{i \in I} \varepsilon_i(\Delta_{st}^i)(W_{\pi_{st},t^c})_i, \text{ and } \sum_{j \in I^c} \varepsilon_j(\pi_{s_0,t})(W_{\pi_{s_0,t},t})_j; \quad (2.11) \]
as that results in high probability estimates on each of the terms in (2.7). And as there are at most \( 2^{2s+3} \) random variables involved in this chaining argument at the \( s \)-stage, the required moment in (2.11) is \( q \sim 2^{s} \) for the first two terms and \( q \sim 2^{s_0} \) for the third one.
2.2 Preliminary estimates

For \( J \subset \{1,...,n\} \) let \( P_J x = \sum_{j \in J} x_j e_j \) be the projection of \( x \) onto \( \text{span}(e_j)_{j \in J} \). The key lemma in the proof of Theorem 1.11 is:

**Lemma 2.4.** There exists an absolute constant \( c \) such that the following holds. Let \( I \subset \{1,...,n\} \) and \( W_{v,I} \) be as in (2.3). Set \( J \subset I^c \) such that \( |J| \leq k_* \). Then for \( q \geq |J| \),

\[
(\mathbb{E}\|P_J W_{v,I}\|_2^{q/2})^{1/q} \leq c\sqrt{q}\delta\|v\|_2.
\]

**Proof.** Let \( S^J \) be the Euclidean unit sphere in the subspace \( \text{span}(e_j)_{j \in J} \) and let \( U \) be a maximal 1/10-separated subset of \( S^J \). By a volumetric estimate (see, e.g. [1]), there is an absolute constant \( c_0 \) such that \( |U| \leq \exp(c_0|J|) \). Moreover, a standard approximation argument shows that for every \( y \in \mathbb{R}^n \),

\[
\|P_J y\|_2 = \sup_{x \in S^J} \langle y, x \rangle \leq c_1 \max_{x \in U} \langle y, x \rangle,
\]

where \( c_1 \) is a suitable absolute constant. Therefore,

\[
\sup_{x \in S^J} \langle W_{v,I}, x \rangle \leq c_1 \max_{x \in U} \langle W_{v,I}, x \rangle,
\]

and it suffices to control, with high probability,

\[
\max_{x \in U} \left( \sum_{i \in I} \varepsilon_i v_i e_i, A^* A x \right) = \max_{x \in U} \sum_{i \in I} \varepsilon_i v_i (A^* A x)_i.
\]

Fix \( x \in U \), recall that \( A \) is \((\delta,k)\)-regular for \( 1 \leq k \leq k_* \) and we first explore the case \( 1 \leq q \leq k_*/2 \).

Denote by \( I' \subset I \) the set of indices corresponding to the \( q \) largest values of \( |(A^* A x)_i|, \ i \in I \). If \( |I| \leq q \) then set \( I' = I \).

It is straightforward to verify (e.g., using Höfﬁng’s inequality) that there is an absolute constant \( c_2 \) such that

\[
\left\| \sum_{i \in I} \varepsilon_i v_i (A^* A x)_i \right\|_{L_q} \leq \sum_{i \in I'} |v_i (A^* A x)_i| + c_2 \sqrt{q} \left( \sum_{i \in I \setminus I'} v_i^2 (A^* A x)_i^2 \right)^{1/2}
\]

\[
\leq \|v\|_2 \|P_{I'} (A^* A x)\|_2 + c_2 \sqrt{q} \cdot \frac{\|P_{I'} (A^* A x)\|_2}{\sqrt{q}} \|v\|_2
\]

\[
\leq c_3 \|v\|_2 \|P_{I'} (A^* A x)\|_2,
\]

where we used that fact that for \( i \in I \setminus I' \),

\[
|(A^* A x)_i| \leq \frac{\|P_{I'} (A^* A x)\|_2}{\sqrt{q}}.
\]

Therefore,

\[
\left\| \sum_{i \in I} \varepsilon_i v_i (A^* A x)_i \right\|_{L_q} \leq c_3 \|v\|_2 \max_{I' \subset I, |I'| = q} \sup_{z \in S^{I'}} \langle A^* A x, z \rangle.
\]
Note that $x$ is supported in $J \subset I^c$, while each ‘legal’ $z$ is supported in a subset of $I$; in particular, $x$ and $z$ are orthogonal, implying that for every such $z$,

$$\|x + z\|_2 = \|x - z\|_2 \leq 2 \quad \text{and} \quad |\text{supp}(x + z)|, |\text{supp}(x - z)| \leq 2q.$$ 

Thus, by the $(\delta, 2q)$-regularity of $A$ (as $2q \leq k_*$),

$$|\langle A^* Ax, z \rangle| = |\langle Az, Ax \rangle| = \frac{1}{4} \|\langle A(x + z) \rangle_2 - \|A(x - z)\|_2^2\| \leq \frac{1}{4} \|\langle A(x + z) \rangle_2 - \|x + z\|_2^2\| - \|\langle A(x - z) \rangle_2 - \|x - z\|_2^2\|\| \leq \frac{\delta}{4} \sqrt{2q} \max\{|\langle x + z \rangle_2^2, |\langle x - z \rangle_2^2|\} \leq \delta \sqrt{q},$$

and it follows that

$$\left\| \sum_{i \in I} \varepsilon_i v_i (A^* Ax)_i \right\|_{L_q} \leq c_4 \|v\|_{2\delta \sqrt{q}}.$$

Turning to the case $q \geq k_* / 2$, let $I'$ be the set of indices corresponding to the $k_* / 2$ largest coordinates of $(\langle A^* Ax \rangle_i)_{i \in I}$. Therefore,

$$\left\| \sum_{i \in I} \varepsilon_i v_i (A^* Ax)_i \right\|_{L_q} \leq \sum_{i \in I'} \v_i (A^* Ax)_i + c_2 \sqrt{q} \left( \sum_{i \in I' \setminus I'} v_i^2 (A^* Ax)_i^2 \right)^{1/2} \leq \|v\|_2 \cdot \|P_I A^* Ax\|_2 \left( 1 + c_5 \sqrt{\frac{q}{k_*}} \right), \quad (2.12)$$

using that for $i \in I \setminus I'$,

$$|\langle A^* Ax \rangle_i| \leq \frac{\|P_I A^* Ax\|_2}{\sqrt{k_* / 2}}.$$ 

Recall that $|\text{supp}(x)| = |J| \leq k_*$ and that $|I'| = k_* / 2$. The same argument used previously shows that

$$\|P_I A^* Ax\|_2 \leq c_6 \delta \sqrt{k_*};$$ 

hence,

$$\left\| \sum_{i \in I} \varepsilon_i v_i (A^* Ax)_i \right\|_{L_q} \leq c_7 \|v\|_{2\delta \sqrt{q}},$$

and the estimate holds for each $q \geq |J|$ for that fixed $x$.

Setting $u \geq 1$, it follows from Chebychev’s inequality that with probability at least $1 - 2 \exp(-c_8 u^2 q)$,

$$\left\| \sum_{i \in I} \varepsilon_i v_i (A^* Ax)_i \right\| \leq c_9 u \|v\|_{2\delta \sqrt{q}},$$

and by the union bound, recalling that $q \geq |J|$, the same estimate holds uniformly for every $x \in U$, provided that $u \geq c_{10}$. Thus, with probability at least $1 - 2 \exp(-c u^2 q)$,

$$\|P_J W_{u,I}\|_2 \leq c' \max_{x \in U} \left\| \sum_{i \in I} \varepsilon_i v_i (A^* Ax)_i \right\| \leq C u \|v\|_{2\delta \sqrt{q}},$$

and the wanted estimate follows from tail integration. □
The next observation deals with more refined estimates on random variables of the form
\[ X_{a,b} = \sum_{i \in I^c} \varepsilon_i a_i b_i. \]

Once again, we use the fact that for any \( J \subset I^c \)
\[ \|X_{a,b}\|_{L_q} \leq \sum_{j \in J} |a_i b_i| + c\sqrt{q} \left( \sum_{j \in I^c \setminus J} a_i^2 b_i^2 \right)^{1/2}. \] (2.13)

As one might have guessed, the choice of \( b \) will be vectors of the form \( W_{\pi,t,I} \). These are random vectors that are independent of the Bernoulli random variables involved in the definition of \( X \). At the same time, \( a \) will be deterministic.

Without loss of generality assume that \( a_i, b_i \geq 0 \) for every \( i \). Let \( J_1 \) be the set of indices corresponding to the \( k_1 \) largest coordinates of \( (a_i)_{i \in I^c} \), \( J_2 \) is the set corresponding to the following \( k_2 \) largest coordinates, and so on. The choice of \( k_1, k_2, \ldots \), will be specified in what follows.

Note that for any \( \ell > 1 \),
\[ \|P_{J,1} a\|_\infty \leq \|P_{J,2} a\|_2 \frac{\|P_{J,2} a\|_2}{\sqrt{|J_0|}}. \]

Set \( J = J_1 \), and by (2.13),
\[ \|X_{a,b}\|_{L_q} \leq \sum_{j \in J_1} |a_i b_i| + c\sqrt{q} \left( \sum_{\ell \geq 1} \sum_{j \in J_{\ell+1}} a_i^2 b_i^2 \right)^{1/2} \]
\[ \leq \|P_{J,1} a\|_2 \|P_{J,2} b\|_2 + c\sqrt{q} \left( \sum_{\ell \geq 1} \|P_{J,1} a\|_2^2 \|P_{J,2} b\|_2^2 \right)^{1/2} \]
\[ \leq \|P_{J,1} a\|_2 \|P_{J,2} b\|_2 + c\sqrt{q} \left( \sum_{\ell \geq 1} \frac{\|P_{J,1} a\|_2^2}{|J_\ell|} \|P_{J,2} b\|_2 \right)^{1/2} \]
\[ \leq \|P_{J,1} a\|_2 \|P_{J,2} b\|_2 + c\sqrt{q} \max_{\ell \geq 2} \frac{\|P_{J,2} b\|_2}{\sqrt{k_{\ell-1}}} \left( \sum_{\ell \geq 2} \|P_{J,2} a\|_2^2 \right)^{1/2} \]
\[ \leq \|a\|_2 \left( \|P_{J,2} b\|_2 + c\sqrt{q} \max_{\ell \geq 2} \frac{\|P_{J,2} b\|_2}{\sqrt{k_{\ell-1}}} \right). \] (2.14)

And, in the case where \( |J_\ell| = k \) for every \( \ell \), it follows that
\[ \|X_{a,b}\|_{L_q} \leq \|a\|_2 \left( \|P_{J,2} b\|_2 + c_1 \sqrt{q} \max_{\ell \geq 2} \|P_{J,2} b\|_2 \right). \] (2.15)

### 2.3 Estimating \( \Phi \)

Recall that
\[ \Phi = \sup_{t \in T} \sum_{s \geq s_1} \|A_s \Delta_s t\|_2 \]
and that for every \( t \in \mathbb{R}^n \),
\[ Z_t = \sum_{i \neq j} \varepsilon_i \varepsilon_j t_i t_j \langle Ae_i, Ae_j \rangle. \]
Theorem 2.5. There are absolute constants $c$ and $C$ such that for $u > 1$, with probability at least $1 - 2\exp(-cu^22^{s_1})$,
\[ \Phi^2 \leq Cu^2\delta^2\ell_s^2(T). \]

Proof. Let $s \geq s_1$, and as noted previously, for every $t \in T$,
\[
\|A_s\Delta_s t\|^2 = \sum_{i=1}^n \|Ae_i\|^2(\Delta_s t)^2 + \sum_{i \neq j} \langle Ae_i, Ae_j \rangle \varepsilon_i \varepsilon_j (\Delta_s t)_i(\Delta_s t)_j \\
\leq (1 + \delta)\|\Delta_s t\|^2 + \alpha \sum_{i \neq j} \langle Ae_i, Ae_j \rangle \varepsilon_i \varepsilon_j (\Delta_s t)_i(\Delta_s t)_j \\
\leq 2\|\Delta_s t\|^2 + |Z_{\Delta,s}t|.
\]

Following the decoupling argument, one may fix $I \subset \{1, \ldots, n\}$. The core of the argument is to obtain satisfactory estimates on moments of the random variables
\[ V_{\Delta,s,t,I} = \sum_{j \in I^c} \varepsilon_j (\Delta_s t)_j(W_{\Delta,s,t,I})_j \]
for that (arbitrary) fixed choice of $I$. And, since $|I| = 2^s$, for a uniform estimate that holds for every random variable of the form $V_{\Delta,s,t,I}$, $t \in T$, it is enough to control the $q$-th moment of each $V_{\Delta,s,t,I}$ for $q \sim 2^s$.

With that in mind, denote by $E_{I^c}$ the expectation taken with respect to $(\varepsilon_i)_{i \in I^c}$, and set $E_I$ the expectation taken with respect to $(\varepsilon_i)_{i \in I}$.

Let us apply (2.15), with the choice
\[ k = 1/\delta^2 \leq k_s, \quad a = \Delta_s t \quad \text{and} \quad b = W_{\Delta,s,t,I}. \]

Thus, for every $(\varepsilon_i)_{i \in I}$,
\[ (E_{I^c}|V_{\Delta,s,t,I}|^q)^{1/q} \leq \|\Delta_s t\|_2 \left( \|P_{I^c}W_{\Delta,s,t,I}\|_2 + c_0\delta\sqrt{q} \cdot \max_{\ell \geq 2} \|P_{\ell}W_{\Delta,s,t,I}\|_2 \right). \]

The sets $J_\ell$ are all of cardinality $1/\delta^2$ and so there are at most $\max\{\delta^2n, 1\}$ of them. By Lemma 2.4 for each one of the sets $J_\ell \subset I^c$ and $q \geq |J_\ell| = 1/\delta^2$,
\[ (E_I\|P_{J_\ell}W_{\Delta,s,t,I}\|_2^q)^{1/q} \leq c_1\delta\sqrt{q}\|\Delta_s t\|_2, \]

implying that
\[ \left( E_{I^c} \max_{\ell} \|P_{J_\ell}W_{\Delta,s,t,I}\|_2^q \right)^{1/q} \leq \left( \sum_{\ell} E_{I^c}\|P_{J_\ell}W_{\Delta,s,t,I}\|_2^q \right)^{1/q} \leq \ell^{1/q} \cdot c_1\delta\sqrt{q}\|\Delta_s t\|_2 \leq c_2\delta\sqrt{q}\|\Delta_s t\|_2 \quad (2.16) \]
as long as $\max\{\delta^2n, 1\} \leq \exp(q)$. In particular, since $2^s_1 \geq \log(e(1 + n\delta^2))$, it suffices that $q \geq 2^s_1$ to ensure that (2.16) holds.
Hence, for every $I \subset \{1, \ldots, n\}$,

$$(\mathbb{E}|V_{\Delta_s,t}|^q)^{1/q} = (\mathbb{E}_I \mathbb{E}_c|V_{\Delta_s,t}|^q)^{1/q}$$

$$\leq \|\Delta_s t\|_2 \left( \mathbb{E}_I \left( \|P_{\ell_1} W_{\Delta_s,t}\|_2 + c_0 \delta \sqrt{q} \cdot \max_{\ell \geq 2} \|P_{\ell_1} W_{\Delta_s,t}\|_2 \right) \right)^{1/q}$$

$$\leq \|\Delta_s t\|_2 \cdot 2 \left( \mathbb{E}_I \max_{\ell \geq 2} \|P_{\ell_1} W_{\Delta_s,t}\|_2^{1/q} \right)$$

$$\leq c_3 \|\Delta_s t\|_2^2 (\delta \sqrt{q} + \delta^2 q)$$

$$\leq c_4 \delta^2 \|\Delta_s t\|_2^2,$$

because $\delta^2 q \geq \delta^2 2s_1 \geq 1$.

By Jensen’s inequality, for every $t \in T$,

$$(\mathbb{E}|Z_{\Delta_s,t}|^q)^{1/q} \leq 4 \left( \mathbb{E}_q \mathbb{E}_c \left| \sum_{t \in I_c} \varepsilon_i (\Delta_s t)_{i/(W_{\Delta_s,t})_{i}} \right|^q \right)^{1/q}$$

$$= 4 \left( \mathbb{E}_q \mathbb{E}_c |V_{\Delta_s,t}|^q \right)^{1/q} \leq c_5 q \delta^2 \|\Delta_s t\|_2^2.$$

Setting $q = u^2$ for $u > 1$ and $s > s_1$, Chebychev’s inequality implies that with $(\varepsilon_i)_{i=1}^n$ probability at least $1 - 2 \exp(-c_0 u^2 2^s)$,

$$\|A_{\varepsilon} \Delta_s t\|_2^2 = 2\|\Delta_s t\|_2^2 + |Z_{\Delta_s,t}| \leq 2\|\Delta_s t\|_2^2 + c_7 u^2 2^s \delta^2 \|\Delta_s t\|_2^2$$

$$\leq c_8 u^2 2^s \delta^2 \|\Delta_s t\|_2^2.$$

By the union bound, this estimate holds for every $t \in T$ and $s \geq s_1$. Thus, there are absolute constants $c$ and $C$ such that with probability at least $1 - 2 \exp(-c u^2 2^s)$,

$$\sum_{s \geq s_1} \|A_{\varepsilon} \Delta_s t\|_2 \leq C u \delta \sum_{s \geq s_1} 2^s \|\Delta_s t\|_2 \leq C u \delta \ell^*_s(T),$$

as claimed.

\[\square\]

### 2.4 Estimating $\Psi$

Next, recall that

$$\Psi^2 = \sup_{t \in T} \left| \|A_{\varepsilon} \pi_{s_1,t}\|_2^2 - \|\pi_{s_1,t}\|_2^2 \right|.$$ 

Expanding as in (2.16), the “diagonal term” is at most $\delta^2 d_T^2$, and one has to deal with the “off-diagonal” term

$$\sup_{t \in T} |Z_{\pi_{s_1,t}}|.$$ 

As observed in Lemma 2.22 combined with Lemma 2.23 it suffices to obtain, for every fixed $I \subset \{1, \ldots, n\}$, estimates on the moments of

$$\sum_{j \in I_c} \varepsilon_j (\Delta_s t)_{j/(W_{\pi s_1,t})_{j}} \sum_{i \in I} \varepsilon_i (\Delta_s t)_{i/(W_{\pi s_1,t})_{i}}; \quad \text{and} \quad \sum_{j \in I_c} \varepsilon_j (\pi_{s_0,t})_{j/(W_{\pi s_0,t})_{j}}.$$ 

(2.17)
Remark 2.6. We shall assume throughout that \((\ell_e(T)/d_T)^2 \leq 1/\delta^2\); the required modifications when the reverse inequality holds are straightforward and are therefore omitted.

We begin with a standard observation:

**Lemma 2.7.** There is an absolute constant \(c\) such that the following holds. Let \((X_\ell)_{\ell \geq 0}\) be nonnegative random variables, let \(q \geq 1\) and set \(q_\ell = q^{2\ell}\). If there is some \(\kappa\) such that for every \(\ell\), \(\|X_\ell\|_{L_q} \leq \kappa\), then

\[
\|\max_\ell X_\ell\|_{L_q} \leq c\kappa.
\]

**Proof.** By Chebychev’s inequality, for every \(\ell \geq 1\) and \(u \geq 2\),

\[
Pr (|X_\ell| \geq u\kappa) \leq \frac{E|X_\ell|^{q_\ell}}{(u\kappa)^{q_\ell}} \leq u^{-q_\ell}.
\]

Therefore,

\[
Pr (\exists \ell \geq 2 : |X_\ell| \geq u\kappa) \leq \sum_{\ell \geq 2} u^{-q_\ell} \leq c_1 u^{-4q},
\]

implying that

\[
E \max_{\ell \geq 2} |X_\ell|^q \leq \int_0^\infty qu^{q-1} Pr \left( \max_{\ell \geq 1} |X_\ell| > u \right) du \leq (c_1 \kappa)^q.
\]

Hence,

\[
\|\max_\ell X_\ell\|_{L_q} \leq \|X_0\|_{L_q} + \|X_1\|_{L_q} + \|\max_{\ell \geq 2} X_\ell\|_{L_q} \leq c_2 \kappa.
\]

The analysis is split into two cases.

**Case I:** \(\log(c(1 + \delta^2 n)) \leq 1/\delta^2\).

In this case, \(2^{s_1} = 1/\delta^2\). For every \(s_0 \leq s \leq s_1\) we invoke (2.11) for sets \(J_\ell\) of cardinality \(2^{s+\ell}\) when \(s + \ell \leq s_1\), and of cardinality \(1/\delta^2\) when \(s + \ell > s_1\).

Set \(a = \Delta s\) and \(b = W_{\pi_{s+1}, I}\); the treatment when \(b = W_{\pi_{s+1}, I}\) is identical and is omitted. Let \(q \geq 2^s\) and put \(q_\ell = q^{2\ell}\) if \(s + \ell \leq s_1\). Finally, set \(p \geq 1/\delta^2\).

Consider \(\ell\) such that \(s + \ell \leq s_1\) and observe that \(q_\ell \geq |J_\ell| = 2^{s+\ell}\). By Lemma 2.4

\[
|J_{\ell-1}|^{-1/2} \left( E \|P_{J_\ell} W_{\pi_{s+1}, I}\|_2^{q_\ell} \right)^{1/q_\ell} \leq c_1 \delta \|\pi_{s+1} I\|_2 \sqrt{\frac{q_\ell}{2^{s+\ell}}} \leq c_1 \sqrt{\frac{q}{2^s}} \delta d_T. \tag{2.18}
\]

Therefore, by Lemma 2.7

\[
\left( E \max_{\{\ell:s+\ell \leq s_1\}} |J_{\ell-1}|^{-1/2} \|P_{J_\ell} W_{\pi_{s+1}, I}\|_2^q \right)^{1/q} \leq c_2 \sqrt{\frac{q}{2^s}} \delta d_T.
\]

Next, if \(s + \ell > s_1\) then \(p \geq |J_\ell| = 1/\delta^2\) and by Lemma 2.4

\[
|J_{\ell-1}|^{-1/2} \left( E \|P_{J_\ell} W_{\pi_{s+1}, I}\|_2^p \right)^{1/p} \leq c_3 \delta \|\pi_{s+1} I\|_2 \sqrt{\frac{p}{1/\delta^2}} \leq c_3 \sqrt{p \delta^2} \cdot \delta d_T. \tag{2.19}
\]
There are at most \( n\delta^2 \) sets \( J_\ell \) in that range, and because \( n\delta^2 \leq \exp(p) \), it is evident that
\[
\left( \mathbb{E} \max_{\ell:s+\ell > s_1} \left( |J_{\ell-1}|^{-1/2} \|P_{\ell} W_{\pi_{s+1:t},l}\|_2 \right)^p \right)^{1/p} \leq c_4 \sqrt{p\delta^2 \cdot \delta T};
\]
therefore, as \( q \leq p \),
\[
\left( \mathbb{E} \max_{\ell:s+\ell > s_1} \left( |J_{\ell-1}|^{-1/2} \|P_{\ell} W_{\pi_{s+1:t},l}\|_2 \right)^q \right)^{1/q} \leq c_4 \sqrt{p\delta^2 \delta T};
\]
Thus, for every fixed \( I \subset \{1, \ldots, n\} \),
\[
\left( \mathbb{E} \max_{\ell} \left( |J_{\ell-1}|^{-1/2} \|P_{\ell} W_{\pi_{s+1:t},l}\|_2 \right)^q \right)^{1/q} \leq c_5 \max \left\{ \sqrt{\frac{q}{2^s}}, \sqrt{p\delta^2} \right\} \cdot \delta T. \quad (2.20)
\]
Now, by (2.15),
\[
\left( \mathbb{E}_{\ell^c} \left| \sum_{j \in \ell^c} \varepsilon_j(\Delta_s t) j(W_{\pi_{s+1:t},l})_j \right|^q \right)^{1/q} \leq c_7 \sqrt{q} \|\Delta_s t\|_2 \cdot \left( 1 + \max \left\{ \sqrt{\frac{q}{2^s}}, \sqrt{p\delta^2} \right\} \right) \cdot \delta T. \quad (2.22)
\]
Clearly, there are at most \( 2^{2s+3} \) random variables as in (2.22). With that in mind, set \( u \geq 1 \) and let \( q = u2^s \), \( p = u/\delta^2 \). By Lemma 2.2 and Lemma 2.3 followed by the union bound, we have that with probability at least \( 1 - 2\exp(-c_8 u^2 2^s) \), for every \( t \in T \),
\[
\mathbb{E}_n \left| \sum_{j \in \ell^c} \varepsilon_j(\Delta_s t) j(W_{\pi_{s+1:t},l})_j \right| \leq c_9 u^2 2^s/2 \|\Delta_s t\|_2 \cdot \delta T. \quad (2.23)
\]
And, by the union bound and recalling that \( 2^{s_0} = \ell_s(T)/d_T \), (2.23) holds uniformly for every \( t \in T \) and \( s_0 \leq s \leq s_1 \) with probability at least
\[
1 - 2\exp \left( -c_{10} u^2 \left( \frac{\ell_s(T)}{d_T} \right)^2 \right).
\]
An identical argument shows that with probability at least
\[
1 - 2\exp(-c_{11} u^2 2^{s_0}) = 1 - 2\exp(-c_{11} u^2 (\ell_s(T)/d_T)^2),
\]
for every \( t \in T \),
\[
\mathbb{E}_n \left| \sum_{i \in I^c} \varepsilon_i(\pi_{s_0} t) i(W_{\pi_{s_0} t},l)_i \right| \leq c_{12} u^2 2^{s_0}/2 d_T \cdot \delta d_T \leq c_{13} d_T \delta \ell_s(T).
\]
Finally, invoking Lemma 2.2 we have that with probability at least \( 1 - 2\exp(-c u^2 (\ell_s(T)/d_T)^2) \), for every \( t \in T \),
\[
|Z_t| \leq C u^2 d_T \delta \ell_s(T),
\]
as required.
**Case II:** $\log(e(1+\delta^2 n)) > 1/\delta^2$.

The necessary modifications are minor and we shall only sketch them. In this range, $2^{s_1} = \log(e(1+\delta^2 n))$, and the problem is that for each vector $\Delta x_t$, the number of blocks $J_\ell$ of cardinality $1/\delta^2$—namely, $n\delta^2$, is larger than $\exp(2^s)$ when $s \leq s_1$. Therefore, setting $|J_\ell| = 1/\delta^2$ for every $\ell$, the uniform estimate on
\[
\max_\ell \|P_{J_t+1}W_{\pi_{s+1}t},I\|_2 \leq \delta \max_\ell \|P_{J_t}W_{\pi_{s+1}t},I\|_2
\]
can be obtained by bounding $(E\|P_{J_t}W_{\pi_{s+1}t},I\|_2^q)^{1/q}$ for $q \geq \log(\delta^2 n)$. Indeed, by Lemma 2.4, for every $I \subset \{1,\ldots,n\}$ and every $t \in T$, we have that
\[
\delta \left(E \max_\ell \|P_{J_t}W_{\pi_{s+1}t},I\|_2^q\right)^{1/q} \leq \delta^2 \sqrt{qd_T}.
\]
The rest of the argument is identical to the previous one and is omitted.

**Proof of Theorem 1.11.** Using the estimates we established, it follows that for $u \geq c_0$, with probability at least
\[
1 - 2\exp\left(-c_1 u^2 \left(\frac{\ell_*(T)}{d_T}\right)^2\right),
\]
\[
\Phi \leq Cu\delta\ell_*(T) \quad \text{and} \quad \Psi^2 \leq Cu^2d_T\delta\ell_*(T) \left(1 + \delta \sqrt{\log(e(1+n\delta^2))}\right),
\]
noting that $\delta d_T^2 \leq c_2d_T\delta\ell_*(T)$. Since
\[
\sup_{t \in T} \|A_t\|_2^2 - \|t\|_2^2 \leq \Psi^2 + 2\Phi \sqrt{\Psi^2 + d_T^2 + \Phi^2} + C' \left((\delta\ell_*(T))^2 + d_T\delta\ell_*(T)\right)
\]
the claim follows from a straightforward computation, by separating to the cases $\Phi \leq \Psi$ and $\Psi \leq \Phi$.

**3 Application - A circulant Bernoulli matrix**

Let $(\varepsilon_i)_{i=1}^n$ and $(\varepsilon_i')_{i=1}^n$ be independent Bernoulli vectors. Set $\xi = (\varepsilon_i')_{i=1}^n$ and put $D_{\varepsilon} = \text{diag}(\varepsilon_1,\ldots,\varepsilon_n)$. Let $\Gamma$ be the circulant matrix generated by the random vector $\xi$; that is, $\Gamma$ is the matrix whose $j$-th row is the shifted vector $\tau_j \xi$, where for every $v \in \mathbb{R}^n$, $\tau_j v = (v_{j-i})_{i=1}^n$.

Fix $I \subset \{1,\ldots,n\}$ of cardinality $m$ and let
\[
A = \sqrt{\frac{1}{m}P_I} : \mathbb{R}^n \to \mathbb{R}^m
\]
to be the normalized partial circulant matrix. It follows from Theorem 3.1 in [6] and the estimates in Section 4 there that for a typical realization of $\xi$, the matrix $A$ is $(\delta,k)$-regular for $\delta \sim m^{-1/2}\log^2 n$: 

---

18
Theorem 3.1. [6] There exist absolute constants $c$ and $C$ such that the following holds. For $x > 0$ with probability at least

$$1 - 2 \exp\left(-c \min\left\{x^2 \frac{m}{k}, x \sqrt{\frac{m}{k} \log^2 n}\right\}\right)$$

we have that

$$\sup_{t \in \Sigma_k} \left| \|At\|_2^2 - 1 \right| \leq C(1 + x) \sqrt{\frac{k}{m} \log^2 n}.$$ 

By Theorem 3.1 and the union bound for $1 \leq k \leq 1/\delta^2$, there is an event $\Omega$ with probability at least $1 - 2 \exp(-c' \log^4 n)$ with respect to $(\varepsilon'_i)_{i=1}^n$, on which, for every $k \leq 1/\delta^2$,

$$\sup_{t \in \Sigma_k} \left| \|At\|_2^2 - 1 \right| \leq \delta \sqrt{k}.$$ 

This verifies the assumption needed in Theorem 1.1 on the event $\Omega$. Now fix $(\varepsilon'_i)_{i=1}^n \in \Omega$ and let $A$ be the resulting partial circulant matrix. Set $A_\varepsilon = AD_\varepsilon$ and let $T \subset \mathbb{R}^n$. By Theorem 1.1 with probability at least

$$1 - 2 \exp(-c'(\ell_*(T)/d_T)^2)$$

with respect to $(\varepsilon_i)_{i=1}^n$, we have that

$$\sup_{t \in T} \left| \|A_\varepsilon t\|_2^2 - \|t\|_2^2 \right| \leq C' \left( \Lambda d_T \frac{\ell_*(T)}{\sqrt{m}} \cdot \log^2 n + \left( \frac{\ell_*(T)}{\sqrt{m}} \right)^2 \cdot \log^4 n \right)$$

where

$$\Lambda \leq c'' \max \left\{ 1, \frac{\log^5 n}{m} \right\}.$$ 

Thus, a random matrix generated by $2n$ independent random signs is a good embedding of an arbitrary subset of $\mathbb{R}^n$ with the same accuracy (up to logarithmic factors) as a gaussian matrix. Moreover, conditioned on the choice of the circulant matrix $A$, the probability estimate coincides with the estimate in the gaussian case.

References

[1] Shiri Artstein-Avidan, Apostolos Giannopoulos, and Vitali D. Milman. Asymptotic geometric analysis. Part I, volume 202 of Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2015.

[2] Withold Bednorz. Concentration via chaining method and its applications. arXiv:1405.0676.

[3] Sjoerd Dirksen. Tail bounds via generic chaining. Electron. J. Probab., 20:no. 53, 29, 2015.

[4] Sjoerd Dirksen. Dimensionality reduction with subgaussian matrices: a unified theory. Found. Comput. Math., 16(5):1367–1396, 2016.
[5] William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In Conference in modern analysis and probability (New Haven, Conn., 1982), volume 26 of Contemp. Math., pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[6] Felix Krahmer, Shahar Mendelson, and Holger Rauhut. Suprema of chaos processes and the restricted isometry property. Comm. Pure Appl. Math., 67(11):1877–1904, 2014.

[7] Felix Krahmer and Rachel Ward. New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. SIAM J. Math. Anal., 43(3):1269–1281, 2011.

[8] Shahar Mendelson. Upper bounds on product and multiplier empirical processes. Stochastic Process. Appl., 126(12):3652–3680, 2016.

[9] Samet Oymak, Benjamin Recht, and Mahdi Soltanolkotabi. Isometric sketching of any set via the restricted isometry property. Inf. Inference, 7(4):707–726, 2018.

[10] Michel Talagrand. Upper and lower bounds for stochastic processes. Springer, 2014.