A Quotient Property for Matrices with Heavy-Tailed Entries and its Application to Noise-Blind Compressed Sensing

Felix Krahmer∗ Christian Kümmerle∗ Holger Rauhut†

June 13, 2018

Abstract

For a large class of random matrices $A$ with i.i.d. entries we show that the $\ell_1$-quotient property holds with probability exponentially close to 1. In contrast to previous results, our analysis does not require concentration of the entrywise distributions. We provide a unified proof that recovers corresponding previous results for (sub-)Gaussian and Weibull distributions. Our findings generalize known results on the geometry of random polytopes, providing lower bounds on the size of the largest Euclidean ball contained in the centrally symmetric polytope spanned by the columns of $A$.

At the same time, our results establish robustness of noise-blind $\ell_1$-decoders for recovering sparse vectors $x$ from underdetermined, noisy linear measurements $y = Ax + w$ under the weakest possible assumptions on the entrywise distributions that allow for recovery with optimal sample complexity even in the noiseless case. Our analysis predicts superior robustness behavior for measurement matrices with super-Gaussian entries, which we confirm by numerical experiments.

1 Introduction

1.1 Random polytopes

Let $A = (a_{ji})$ be a rectangular $m \times N$ random matrix with independent, symmetric and unit variance entries $a_{ji}$ and $m < N$, and denote by $B^N_p$ the unit ball of the $\ell_p$-norm in $\mathbb{R}^N$. In this paper, we study the geometry of the image $AB^N_1$ under quite general assumptions on the distribution of the entries $a_{ji}$. This object can be also regarded as the random polytope defined by the absolute convex hull of the columns of $A$, i.e.,

$$AB^N_1 = \text{span}\{\pm a_1, \ldots, \pm a_N\},$$

if $a_1, \ldots, a_N$ denote the columns of $A$.

For normally distributed $a_{ji}$, a result due to Gluskin and Kashin quantifies the inclusion of an Euclidean ball $B^N_2$ in $AB^N_1$.

**Theorem 1** ([1, 2, 3]). *If the $a_{ji}$ are independent mean-zero, variance one Gaussian random variables, there exist constants $C, D > 0$ such that if $N \geq 2m$,

$$\mathbb{P}\left(AB^N_1 \supset \frac{\sqrt{\log(eN/m)}}{D} B^N_2\right) \geq 1 - \exp(-Cm).$$

(1)
This statement corresponds to a lower bound on the *inradius* of $AB_1^N$, i.e., the radius of the largest Euclidean ball that is contained in the random polytope $AB_1^N$.

Litvak et al. proved a similar result for $a_{ji}$ that fulfill a concentration property. Below $\|A\|_{2 \rightarrow 2}$ denotes the spectral norm of a matrix $A$.

**Theorem 2** ([4], see also [5, Theorem 11.21.]). If there exist constants $a_1, a_2 > 0$ such that $\mathbb{P}(\|A\|_{2 \rightarrow 2} \geq a_1 \sqrt{N}) \leq \exp(-a_2 N)$ and if the third moments of the $a_{ji}$ are bounded by $\mu$, there exist constants $D(\mu), \tilde{C}(a_1), C(a_2) > 0$ and $c > 0$ such that if $N \geq \tilde{C}(a_1)m$,

$$\mathbb{P} \left( AB_1^N \supset \frac{1}{D(\mu)} \left( \sqrt{\log (eN/m)} B_2^N \cap B_\infty^N \right) \right) \geq 1 - c \exp(-C(a_2)m).$$

(2)

An important instance of distributions fulfilling the assumption of the latter result are sub-Gaussian distributions. By considering symmetric $\pm 1$ random variables $a_{ji}$ (which are sub-Gaussian), it can be seen that the intersection of $\sqrt{\log (eN/m)} B_2^N$ with the unit cube $B_\infty^N$ in the last result is indeed necessary [6] for general sub-Gaussian distributions. In follow-up works, corresponding results have also been obtained for matrices $A$ with dependent entries, most notably the scenario where the vertices of $AB_1^N$ are drawn uniformly from a convex body [7], [8, Chapter 11]. Due to a close connection to log-concave measures, this model can also be seen as a version of a concentration requirement (however, weaker than subgaussianity).

In this paper, we establish results corresponding to Theorem 1 and Theorem 2 for a significantly enlarged class of random matrices $A$. In particular, this class includes heavy-tailed entry-wise distributions which do not fulfill strong concentration properties. Following the arguments in [4], the resulting lower bounds on the inradius have implications on bounds of other geometric quantities of the corresponding random polytopes such as their volume and their mean width. These lower bounds on the volume of random polytopes have also been used in the context of differential privacy [9].

### 1.2 The quotient property in compressive sensing

Our analysis is additionally motivated by the theory of compressive sensing, which studies the recovery of sparse vectors from incomplete linear measurements via efficient methods such as $\ell_1$-minimization [5]. Provably optimal guarantees are available for random matrices. While previous work has mostly considered random matrices with entries obeying strong concentration properties such as Gaussian and subgaussian random random variables, it has recently been shown that concentration is not required for sparse recovery guarantees. More precisely, Lecué and Mendelson [10], see also [11], showed recovery results for random matrices with independent, possibly heavy-tailed entries, requiring only $\log(N)$ finite moments. Their proof establishes the null space property via Mendelson’s small ball method [12, 13].

Our work extends this line of research and yields recovery guarantees for unknown noise levels without requiring concentration on the entries of the measurement matrix. As observed in [3, 6], this problem is closely connected to statements about polytope inclusions as given in (1) and (2), respectively, which in this context are commonly referred to as quotient properties, see Definition 4 below for a precise definition. More precisely, our results imply stable and robust recovery for equality-constrained $\ell_1$-minimization from noisy, random measurements with heavy-tailed matrix entries without requiring an a-priori estimate of the noise level as would be needed for standard noise-aware $\ell_1$-minimization (basis pursuit denoising).

Stated formally, we seek to recover a vector $x \in \mathbb{R}^N$ from noisy, underdetermined measurements

$$y = Ax + w,$$
where $A \in \mathbb{R}^{m \times N}$ with $m < N$ is the so-called measurement matrix and $w \in \mathbb{R}^m$ is a noise vector. If $x$ is $s$-sparse, i.e., $\|x\|_0 := \# \{j : x_j \neq 0\}$, or approximately $s$-sparse in the sense that

$$
\sigma_s(x)_1 := \inf \{ \|x - z\|_1 : z \in \mathbb{R}^N, \|z\|_0 \leq s \}
$$

is small, then we can hope to do so via $\ell_1$-minimization

$$
\min_{z \in \mathbb{R}^N} \|z\|_1 \quad \text{subject to } Az = y.
$$

(3)

In fact, if $A$ is an $m \times N$ matrix with independent standard Gaussian $\mathcal{N}(0, 1)$ random variables,

$$
m \geq Cs \log(eN/s)
$$

and $y = Ax$, then with high probability the minimizer $x^#$ of (3) coincides with $x$ if $\|x\|_0 \leq s$ and more generally [5],

$$
\|x - x^#\|_1 \leq C\sigma_1(x)_s \quad \text{and} \quad \|x - x^#\|_2 \leq C\frac{\sigma_1(x)_s}{\sqrt{s}}.
$$

In the noisy case $y = Ax + w$ with known noise bound $\|w\|_2 \leq \eta$, one commonly considers the constrained $\ell_1$-minimization problem

$$
\min_{z \in \mathbb{R}^N} \|z\|_1 \quad \text{subject to } \|Az - y\|_2 \leq \eta.
$$

(4)

For a Gaussian $m \times N$ matrix with $m \geq Cs \log(eN/s)$, the minimizer $x^#$ of (4) satisfies

$$
\|x - x^#\|_1 \leq C\sigma_1(x)_s + D\sqrt{s}\eta \quad \text{and} \quad \|x - x^#\|_2 \leq C\frac{\sigma_1(x)_s}{\sqrt{s}} + D\eta.
$$

(5)

In practice, however, an accurate noise bound $\eta$ may not be known. If $\eta$ is an underestimation of the true $\|w\|_2$ then the so-called restricted isometry property or the robust null space property as used in the standard proofs [5, Chapters 4, 6] are not sufficient to guarantee the error bounds (5). If $\eta$ is an overestimation of $\|w\|_2$ then the bounds (5) may be very pessimistic as they depend on $\eta$ rather than on the true noise level $\|w\|_2$ (see also Chapter 3 for corresponding numerical experiments).

In order to address this problem, Wojtaszczyk suggested to simply use equality-constrained $\ell_1$-minimization [3] and provided an analysis for Gaussian measurement matrices $A$ based on Theorem 1, which was later adapted to subgaussian matrices [6] using Theorem 2 and also to Weibull matrices [14]. The resulting error bound is of the form

$$
\|x - x^#\|_1 \leq C\sigma_s(x)_s + D\sqrt{s}\|w\| \quad \text{and} \quad \|x - x^#\|_2 \leq C\frac{\sigma_s(x)_s}{\sqrt{s}} + D\|w\|,
$$

where $\|w\|$ is the Euclidean norm for Gaussian and Weibull matrices and an interpolation norm between Euclidean and supremum norm for subgaussian matrices.

Similar results have been recently established in [15] for constrained $\ell_1$-minimization (4), where $\eta$ is possibly underestimated, i.e., $\|w\|_2 \geq \eta$, see also below.

1.3 Outline and contribution of this paper

Our main contribution is twofold: Firstly, we prove a significantly generalized version of Theorem 1 and the Gluskin-type inclusion (1) as compared to the ones for Gaussian [1] or Weibull [14] distributions; namely, our result only requires (independent) matrix entries to be super-Gaussian (for the precise meaning of this concept, we refer to Definition 3), see Theorem 5(b) and Corollary 7(b). Secondly, in Theorem 5(a) and Corollary 7(a), we generalize Theorem 2,
requiring only entrywise distributions with logarithmically many well-behaved moments. In both parts of Theorem 5, our results are expressed in terms of the $\ell_1$-quotient property. All these concepts and results are introduced in detail in Section 2.1.

Based on Theorem 5, we provide, in Section 2.2, new robustness guarantees for noise-blind $\ell_1$-minimization for measurement matrices with quite general entrywise distributions in the regime of optimal sample complexity $m \approx C_s \log(e N/s)$ in Theorem 8. The requirements on the entrywise distributions match the relatively weak moment assumptions of [10] that can be shown to be almost necessary in the regime of optimal sample complexity for sparse recovery even in the noiseless case. Our result covers both the case of equality-blind $\ell_1$-minimization (cf. Remark 9) and the case of quadratically constrained $\ell_1$-minimization with underestimated noise level, as studied in [15].

Notably, we provide a unified proof strategy for our results, which covers all previous results for matrices with independent entries, both on Gluskin-type inclusions and on the robustness and $\ell_1$-minimization for certain heavy-tailed measurement scenarios and underestimating noise level, as studied in [15].

In Section 3, our results are complemented by numerical experiments, confirming the robustness of noise-blind $\ell_1$-minimization for different types of noise.

1.4 Notation

In this section, we recall some of the notation we use in this paper. For $N \in \mathbb{N}$, we write $[N] := \{k \in \mathbb{N} : 1 \leq k \leq N\}$. For a vector $x \in \mathbb{R}^N$, we write $\|x\|_p = (\sum_{j=1}^N |x_j|^p)^{1/p}$, $1 \leq p < \infty$, and $\|x\|_\infty = \sup_{j \in [N]} |x_j|$ for its $\ell_p$-norm, while for a random variable $X$ taking values a normed vector space, we denote by $\|X\|_p = (\mathbb{E}[\|X\|^p])^{1/p}$, $1 \leq p < \infty$, its $p$-th moment. For $1 \leq p \leq \infty$, we denote the unit ball of the $\ell_p$-ball in $\mathbb{R}^N$ as $B_p^N = \{x \in \mathbb{R}^N : \|x\|_p \leq 1\}$. The clipped $\ell_2$-norm with parameter $\alpha$ is defined as $\|\cdot\|^{(\alpha)} := \max\{\|\cdot\|_2, \alpha\|\cdot\|_\infty\}$. A Rademacher sequence $\epsilon = (\epsilon_i)_{i \in [N]}$ is a sequence of independent random variables $\epsilon_i$ taking the values $-1$ and $+1$ with equal probability.

2 Main results

We first state the results about the quotient property and its implication for the geometry of the polytope spanned by the columns of a random matrix. We distinguish two types of assumptions on the entrywise distributions its entries.

Definition 3. Let $X$ be a random variable with $\mathbb{E}[X] = 0$ and unit variance (so that $\|X\|_{L_2} = 1$).

1. $X$ is called a super-Gaussian variable with parameter $\sigma > 0$ if there exists $\sigma > 0$ such that

$$\mathbb{P}(|\sigma g| > t) \leq \mathbb{P}(|X| > t)$$

for all $t > 0$, where $g$ is a standard normal random variable.

2. $X$ is said to fulfill the weak moment assumption of order $k$ with constants $\kappa_1$ and $\gamma \geq 1/2$ if

$$\|X\|_{\ell_p} \leq \kappa_1 p^{\gamma}$$

for all $4 \leq p \leq k$.

2.1 Quotient properties and polytope geometry

The $\ell_1$-quotient property as given in the following definition is a main object of our studies.

Definition 4 ([14, 3]). A matrix $A \in \mathbb{R}^{m \times N}$ is said to possess the $\ell_1$-quotient property with constant $d$ relative to a norm $\|\cdot\|$ on $\mathbb{R}^m$ if, for all $w \in \mathbb{R}^m$, there exists $u \in \mathbb{R}^N$ such that

$$Au = w \quad \text{and} \quad \|u\|_1 \leq ds_s^{1/2} \|w\|,$$

with $s_s = m/\log(e N/m)$. 
We proceed to our main theoretical result. We note that the assumption of identical distributions can be relaxed, but for simplicity we present the theorem under this assumption.

**Theorem 5.** Let \( A = (a_{ij}) \) be an \( m \times N \) random matrix with independent symmetric, unit variance entries \( a_{ij} \sim X \) for all \( j \in [m], i \in [N] \).

(a) If \( X \) fulfills the weak moment assumption of order \( \max\{4, \log(m)\} \) with constants \( \kappa_1 \) and \( \gamma \geq 1/2 \), then there exist an absolute constant \( c_0 \) and constants \( \bar{C} \) and \( D \) such that if \( m \) is large enough such that \( 4 e^{4\gamma} c_0^2 \kappa_1^2 \log(m) \leq m \), then with probability at least \( 1 - 2 \exp(-2m) \),

\[
\text{the matrix } \frac{1}{\sqrt{m}} A \text{ fulfills the } \ell_1\text{-quotient property with constant } D \text{ relative to the clipped } \ell_2\text{-norm } \| \cdot \|_{\sqrt{\log(eN/m)}} := \max\{\| \cdot \|_2, \sqrt{\log(eN/m)} \| \cdot \|_\infty\} \text{ for } N \geq \max\{\tilde{C}m, \log^{2\gamma-1}(m)\}.
\]

(b) If \( X \) is super-Gaussian with parameter \( 0 < \sigma \leq 1 \), there exist constants \( \Bar{C} \) and \( D \) (depending on \( \sigma \)) such that with probability at least \( 1 - 2 \exp(-2m) \), \( \frac{1}{\sqrt{m}} A \) fulfills the \( \ell_1\)-quotient property with constant \( D \) relative to the \( \ell_2\)-norm for \( N \geq \tilde{C}m \).

**Remark 6.**

1. In the first statement of Theorem 5, the constants \( D \) and \( \Bar{C} \) depend on \( \kappa_1 \) and \( \gamma \). In particular, \( D \) can be chosen as

\[
D = 8 e^{-1} \sqrt{1 + (9 + 8\gamma) \log(4) + 8 \log(\kappa_1)} \leq \sqrt{\gamma} + \sqrt{\log(\kappa_1)}
\]

and \( \Bar{C} = \frac{e^{2+8\gamma}}{3} \kappa_1^8 \).

2. The proof given of the second statement works for the constants \( \Bar{C} = 64^2 c_1^{-2} e^{2c_2} \), \( c_1^{-1} = 10 \), \( c_2 = 5220 \) and

\[
D = 8 e^{1/2} \sqrt{1 + 2 \log(64) + 2 \log(c_1^{-1}) + \frac{c_2}{16\sigma^2}} \leq 1 + \sigma^{-1},
\]

which depend on the super-Gaussian parameter \( \sigma \).

We note that in both cases, it was not our objective to find the best possible constants \( \Bar{C} \) and \( D \). By considering the case of \( cm < N < \Bar{C}m \) with much smaller \( c \) than \( \Bar{C} \) separately and analyzing the smallest singular value of \( B \), the range of validity of the theorem can be extended considerably, cf. also [5, Theorem 11.19].

For lower bounding the least singular values under the present random models, results as in [13] are useful tools.

As mentioned before, the \( \ell_1\)-quotient property is closely linked to the geometry of \( AB_N^1 \), which is the polytope defined by the absolute convex hull of the columns of \( A \). We obtain the following corollary by rewriting the definition of the \( \ell_1\)-quotient property, see, e.g., [5, Chapter 11].

**Corollary 7.** Let \( A = (a_{ij}) \) be an \( m \times N \) random matrix with independent symmetric, unit variance entries \( a_{ij} \sim X \) for all \( j \in [m], i \in [N] \).

(a) If \( X \) fulfills the weak moment assumption of order \( \max\{4, \log(m)\} \) with constants \( \kappa_1 \) and \( \gamma \geq 1/2 \), then there exist an absolute constant \( c_0 \) and constants \( \Bar{C} \) and \( D \) such that if \( m \) is large enough such that \( 4 e^{4\gamma} c_0^2 \kappa_1^2 \log(m) \leq m \) and if \( N \geq \max\{\Bar{C}m, \log^{2\gamma-1}(m)\} \),

\[
\mathbb{P} \left( AB_N^1 \supset \frac{1}{D} \left( \sqrt{\log(eN/m)} B_N^2 \cap B_N^\infty \right) \right) \geq 1 - 2 \exp(-2m).
\]

(b) If \( X \) is super-Gaussian with parameter \( 0 < \sigma \leq 1 \), there exist constants \( \Bar{C} \) and \( D \) (depending on \( \sigma \)) such that for every \( m, N \) satisfying \( N \geq \Bar{C}m \), one has

\[
\mathbb{P} \left( AB_N^1 \supset \frac{1}{D} \sqrt{\log(eN/m)} B_N^2 \right) \geq 1 - 2 \exp(-2m).
\]
2.2 Robustness of noise-blind compressed sensing

We will use Theorem 5 to study the robustness of the reconstruction map $\Delta_1$ given by equality-constrained $\ell_1$-minimization

$$\Delta_1(y) := \arg \min_{z \in \mathbb{R}^N} \|z\|_1 \text{ subject to } Az = y,$$

where $A \in \mathbb{R}^{m \times N}$ for $m < N$ and when noise on the measurements $Ax$ of a sparse or approximately sparse vector $x \in \mathbb{R}^N$ is present, i.e., if $y = Ax + w$ with some arbitrary $w \in \mathbb{R}^m$. Our goal is to quantify, for $1 \leq p \leq 2$, the $\ell_p$-error $\|\Delta_1(y) - x\|_{\ell_p}$ of the reconstruction map $\Delta_1(y)$ to $x$. We call the decoder $\Delta_1(y)$ noise-blind since it does not use any information about the noise $w$.

Furthermore, a more canonical reconstruction algorithm in case of noisy observations ($w \neq 0$) is the convex program called quadratically constrained $\ell_1$-minimization

$$\Delta_{1,\eta}(y) := \arg \min_{z \in \mathbb{R}^N} \|z\|_1 \text{ subject to } \|Az - y\|_2 \leq \eta$$

for some $\eta > 0$. The parameter $\eta$ can be chosen in a noise-aware manner such that $\|w\|_2 \leq \eta$, using oracle information about the $\ell_2$-norm of the noise $w$, and error bounds such as (5) have been shown by using a restricted isometry property or robust null space property of $A$ in this case, without the need of using quotient properties.

In the next theorem, we derive error bounds for $\Delta_1(y)$ and for $\Delta_{1,\eta}(y)$ in the case of underestimated noise level such that $\eta < \|w\|_2$. The latter was first studied in [15]. The theorem provides robustness results for measurement matrices drawn from a wide range of i.i.d. entrywise distributions with high probability.

**Theorem 8.** Let $m \leq N$, let $B = (b_{ij})$ be an $m \times N$ random matrix with independent symmetric, unit variance entries $b_{ij} \sim X$ for all $j \in [m]$, $i \in [N]$ and $A := \frac{1}{\sqrt{m}}B$. For $s \in \mathbb{N}$, let $\sigma_s(x)_1 := \inf\{\|x - z\|_1 : z \in \mathbb{R}^N, \|z\|_0 \leq s\}$ be the $\ell_1$-error of the best $s$-term approximation of $x \in \mathbb{R}^m$. Assume $\eta \geq 0$.

(a) Assume that $X$ fulfills the weak moment assumption of order $\max\{4, \log(N)\}$ with constants $\kappa_1$ and $\gamma \geq \frac{1}{2}$. Then there exist constants $\tilde{c}_1, \tilde{c}_2, \tilde{c}_3, \tilde{C}, C, D, E > 0$ depending only on $\kappa_1$ and $\gamma$ such that if $N \geq \tilde{c}_m, \quad m \geq \tilde{c}_3 \log^{\max\{2\gamma - 1, 1\}}(N) \quad$ and $s \leq \tilde{c}_2 s_s := \tilde{c}_2 \frac{m}{\log(e N/m)}$, with probability at least $1 - 3 \exp(-\tilde{c}_0 m)$, the solution of the $\ell_1$-minimization decoder $\Delta_{1,\eta}$ given the measurement matrix $A$ and data vector $y = Ax + w$ fulfills the $\ell_p$-error estimates

$$\|x - \Delta_{1,\eta}(Ax + w)\|_p \leq C \frac{\sigma_s(x)_1}{s^{1 - 1/p}} + s^{1/p - 1/2} \left( D\eta + E \frac{\|w\|_2^{\log(e N/m)}}{\|w\|_2} \max\{\|w\|_2 - \eta, 0\} \right)$$

for $1 \leq p \leq 2$, for all $x \in \mathbb{R}^N$ and all $w \in \mathbb{R}^m$, where we recall that $\|\cdot\|_{\log(e N/m)} := \max\{\|\cdot\|_2, \sqrt{\log(e N/m)} \|\cdot\|_\infty\}$.

(b) If $X$ fulfills the weak moment assumption of order $\max\{4, \log(N)\}$ with constants $\kappa_1$ and $\gamma \geq \frac{1}{2}$ and if $X$ is additionally super-Gaussian with parameter $0 \leq \sigma \leq 1$, then there exist constants $\bar{c}_1, \bar{c}_2, \bar{C}, C, D, E > 0$ depending only on $\kappa_1, \gamma$ and $\sigma$ such that if $N \geq \bar{C}m, \quad m \geq \log^{\max\{2\gamma - 1, 1\}}(N) \quad$ and $s \leq \bar{c}_2 s_s := \bar{c}_2 \frac{m}{\log(e N/m)}$,
with probability at least $1 - 3 \exp(-\tilde{c}_1 m)$, the solution of the equality-constrained \( \ell_1 \)-minimization problem \( \Delta_1(Ax + w) \) fulfills the \( \ell_p \)-error estimates

\[
\|x - \Delta_{1,\eta}(Ax + w)\|_p \leq \frac{C}{s^{1-1/p}} \sigma_s(x)_1 + s_s^{1/p-1/2} \{D\eta + E \max\{\|w\|_2 - \eta, 0\}\}
\]

for \( 1 \leq p \leq 2 \), and for all \( x \in \mathbb{R}^N \), \( w \in \mathbb{R}^m \).

**Remark 9.** We point out that robust recovery guarantees of Theorem 8 can be specified for the noise-blind equality-constrained \( \ell_1 \)-minimization

\[
\Delta_{1,0}(y) = \Delta_1(y) = \arg \min_{z \in \mathbb{R}^N} \|z\|_1 \text{ subject to } Az = y
\]

such that for all \( 1 \leq p \leq 2 \),

\[
\|x - \Delta_1(Ax + w)\|_p \leq \frac{C}{s^{1-1/p}} \sigma_s(x)_1 + s_s^{1/p-1/2} E\|w\|\sqrt{\log e N/m}
\]

and

\[
\|x - \Delta_{1,\eta}(Ax + w)\|_p \leq \frac{C}{s^{1-1/p}} \sigma_s(x)_1 + s_s^{1/p-1/2} \|w\|_2
\]

for all \( x \in \mathbb{R}^N \) and \( w \in \mathbb{R}^m \) in the first and second part of the theorem, respectively.

This means that the existing robust recovery guarantees for this decoder using matrices with i.i.d. sub-Gaussian [5, Theorem 11.10], Gaussian [5, Theorem 11.9] and Weibull [14, Theorem 11] random variables can be considered as special cases of the theorem.

To show Theorem 8, we combine existing results about the robust null space property of matrices with i.i.d. entries drawn from distributions fulfilling weak moment assumptions [11, 10] together with Theorem 5.

Next, we illustrate the generality of the assumptions of Theorems 8 and 5 by enumerating random models which are covered by our theorem, but which mostly have not been covered by the robustness analyses of [4, 3, 14].

**Example 10.** Let \( A \) be a real random matrix with i.i.d. entries \( A_{ji} = \frac{1}{\sqrt{m}} \frac{B_{ji}}{\|B_{ji}\|_2} \), \( j \in [m], i \in [N] \).

(i) Assume that the \( B_{ji} \) are distributed as \( X_\gamma \), where \( X_\gamma \) is a \( \psi_2 \)-random variable with the same distribution as \( \text{sign}(g) |g|^{2\gamma} \), where \( g \) is a standard normal variable and \( \gamma > 0 \). Then, \( X_\gamma \) is of exponential type, i.e., has a probability density function of \( p(x) = c_1 e^{-\frac{|x|^{2\gamma}}{2}} \). If \( 1/2 \leq \gamma \leq 1 \), the assumptions of the second part of Theorem 8 apply, cf. [11, Example V4]. In particular, the \( B_{ji}/\|B_{ji}\|_2 \) are super-Gaussian with a parameter \( \sigma \geq \frac{1}{2} \). We note that the special cases \( \gamma = \frac{1}{2} \) and \( \gamma = 1 \) have been covered already by the existing theory, in the latter case by [14], but not for \( 1/2 < \gamma < 1 \).

For \( \gamma > 1 \), the theorem still applies as \( X_\gamma \) is super-Gaussian with parameter \( \sigma \geq \frac{1}{\gamma} \), but with a worse upper bound \( \tilde{c}_1(\gamma) \frac{m}{\log(N/m)} \) on the sparsity, where \( \tilde{c}_2(\gamma) \) depends on \( \gamma \), cf. [11, Example V4].

(ii) Let \( d \in \mathbb{N} \). If the \( B_{ji} \) are distributed as Student-\( t \) variables \( X_d \) with \( d \) degrees of freedom, they are (after normalization) super-Gaussian with some parameter \( \sigma \geq \frac{1}{\gamma} \) as well, so that Theorem 5(b) applies. If additionally \( d \geq 2 \log(N) \), then also Theorem 8.2 applies.
(iii) If the $B_{ji}$ are distributed as a symmetric Weibull variable $X_r$ with exponent $1 \leq r \leq 2$, recovery guarantees or equality-constrained $\ell_1$-minimization that are robust relative to the $\ell_2$-norm have been shown in the optimal regime of $m$ already in [14]. Since the normalized symmetric Weibull variables $B_{ji}/\|B_{ji}\|_{L_2}$ are super-Gaussian with parameter $\sigma \geq 1/2$ for all $1 \leq r \leq 2$, Theorem 5(b) applies also here.

Interestingly, comparing the two parts of Theorem 8, we see that our analysis suggests that the robustness properties of equality-constrained $\ell_1$-minimization $\Delta_1$ with measurement matrices $A$ with entries drawn from many super-Gaussian distributions are asymptotically better than the ones of measurement matrices whose entries are drawn from certain sub-Gaussian, bounded distributions as the Rademacher distribution with random signs.

3 Numerical experiments

In this section, we show in a case study that the results of Theorem 8 give an appropriate explanation of the empirical robustness behavior of different measurement matrices. In particular, we consider three types of measurement matrices: random matrices with i.i.d. Gaussian, Bernoulli and Student-t entries. The presented numerical experiments have been conducted using MATLAB R2017b on a MacBook Pro with a 2.3 GHz Intel Core i5 processor. The convex optimization problems of our experiments are solved using the CVX package [16].

3.1 Behavior under spherical noise

In our first experiment, we perform simulations for the reconstruction of a $s$-sparse vector $x_0 \in \mathbb{R}^N$ with $\|x_0\|_2 = 1$ from measurements $y = Ax_0 + w$ which are perturbed by a random vector $w \in \mathbb{R}^m$ that is drawn from the uniform distribution on the sphere of radius $\|w\|_2 = 10^{-2}$. To obtain our reconstruction result $\hat{x}$, we use equality-constrained $\ell_1$-minimization (7) as defined...
by $\Delta_1(y)$ and quadratically constrained $\ell_1$-minimization (8)

$$\Delta_{1,\eta}(y) := \arg\min_{z \in \mathbb{R}^N} \|z\|_1 \text{ subject to } \|Az - y\|_2 \leq \eta,$$

where the noise level estimate $\eta$ is chosen such that $\eta \in \{\|w\|_2, 2\|w\|_2, 0.5\|w\|_2\}$, i.e., the noise level $\|w\|_2$ is either estimated accurately or over- or underestimated by a factor of two. The support $S$ of $x_0$ is drawn uniformly among the $\binom{N}{k}$ possibilities, and the non-zero coordinates are drawn uniformly on the sphere $S^{s-1} = \{x \in \mathbb{R}^N : \|x\|_2 = 1, \text{supp}(x) \subseteq S\}$.

In Figure 1, the resulting recovery $\ell_2$-errors $\|\hat{x} - x_0\|_2$ can be observed for the three different random models (in case of Student-t measurements, $k = 9$ degrees of freedoms were used) for the measurement matrix $A$ mentioned above, where the parameters were chosen as $N = 5000$, $s = 10$ and $m \in \{\lfloor kN/20 \rfloor, k = 1, \ldots, 14\}$. The reported errors are averaged over 500 runs of the simulation.

We notice that in the experiment, the recovery error of the equality-constrained algorithm (7) is comparable to the one of quadratically constrained $\ell_1$-minimization (8) with correctly estimated or underestimated noise level $\eta \in \{\|w\|_2, 0.5\|w\|_2\}$, if $A$ has a small number of rows $m \leq 500$. For larger $m$, the robustness of (8) improves further if $\eta = \|w\|_2$, whereas it stagnates for underestimated noise of $\eta = 0.5\|w\|_2$ and it deteriorates slightly for (7).

It can be also observed that an overestimation of the noise level such that $\eta = 2\|w\|_2$ in (8) leads to a significantly worse reconstruction error $\|\hat{x} - x_0\|_2$ than for all the other methods, for all the considered number of measurements $m$.

Importantly, we observe that the robustness behavior of the algorithms does not depend on the choice of Gaussian, Bernoulli or Student-t measurement matrices in this case of presence of spherical noise.

This is precisely in accordance to the result of Theorem 8: Bernoulli variables $X$ fulfill the assumptions of the first part of the theorem, but not of the second part, since they are sub-Gaussian. On the other hand, Gaussian and Student-t variables (with a sufficient number of degrees of freedom) fulfill the assumptions for the stronger statement of Theorem 8.2. In general, Bernoulli measurement matrices entail the weaker statement predicting a reconstruction error of

$$\|x - \Delta_1(y)\|_2 \leq D\|w\|_2(\sqrt{\log e N/m})$$

with a constant $D$ for equality-constrained $\ell_1$-minimization. For spherical noise, though, this coincides with the statement of Theorem 8.2, since $\|w\|_2(\sqrt{\log e N/m}) = \|w\|_2$ with high probability under this noise model.

### 3.2 Behavior under heavy-tailed noise

Next, instead of uniform spherical noise, we consider more heavy-tailed noise such that $w = 10^{-2}\frac{\tilde{w}}{\|\tilde{w}\|_2} \in \mathbb{R}^m$, where $(\tilde{w})_i$ are i.i.d. $\psi_\alpha$ random variables for the parameter $\alpha = 0.2$, cf. also Example 10.(i). Such a noise has most of its mass in just few coordinates, and the size of its largest entry $\|w\|_\infty$ is comparable to its $\ell_2$-norm $\|w\|_2$, i.e. $\|w\|_\infty \approx \|w\|_2 = 10^{-2}$ with high probability.

In this case, the conclusions about the recovery accuracy of equality-constrained $\ell_1$-minimization (7) that can be drawn from Theorem 8.1 and Theorem 8.2 predict a better behavior of Gaussian and Student-t measurements than for Bernoulli measurements, in particular for $m \ll N$: Since then

$$\|w\|_2(\sqrt{\log e N/m}) = \sqrt{\log e N/m} \cdot \|w\|_\infty \approx \sqrt{\log e N/m} \cdot \|w\|_2 = \sqrt{\log e N/m} \cdot 10^{-2}$$
Figure 2: y-axis: Reconstruction errors $\|\tilde{x} - x_0\|_2$ from measurements $y = Ax_0 + \omega$, $\omega = 10^{-2} \cdot \frac{\omega}{\|\omega\|_2}$, where the entries of $\omega$ are i.i.d. $\psi_{0.2}$ random variables, for different measurement matrices $A$. x-axis: number of rows of $m$.

with high probability, Theorem 8 predicts a reconstruction error of

$$\|x - \Delta_1(y)\|_2 \leq D \cdot \sqrt{\frac{e N}{m}} \cdot 10^{-2}$$

for Bernoulli measurements, but a reconstruction error of

$$\|x - \Delta_1(y)\|_2 \leq D \cdot 10^{-2}$$

for the two other, more heavy-tailed measurement models (here, $D$ is some constant).

These predictions can be well confirmed in the experiment illustrated in Figure 2, repeating the experiment from Section 3.1 for this different, heavy-tailed noise model: Unlike before, the reconstruction error of equality-constrained $\ell_1$-minimization (7) for Bernoulli matrices is now consistently worse than for the Gaussian and Student-t measurement matrices if $m \ll N$, i.e., if $m = 250, \ldots, 2000$. It is interesting to note that equality-constrained $\ell_1$-minimization with Student-t matrices (with $k = \lceil \log(N) \rceil = 9$ degrees of freedom) is even slightly more robust than in the case that Gaussian matrices are used, especially if $m$ is small.

On the other hand, the relative performance of Student-t measurements is worse than the one of Gaussian measurements if the noise-aware quadratically constrained $\ell_1$-minimization (8) is used as a reconstruction algorithm.

As for spherical noise, we also note here that overestimating the noise level by a factor of two ($\eta = 2 \cdot \|w\|_2$) in 8 leads to worse reconstructions than the noise-blind usage of (7).

We want to stress two conclusions from these experiments:

- The noise-blind reconstruction algorithm (7) is at least as robust in presence of certain heavy-tailed measurement matrices as in the case of Gaussian measurement matrices, especially if the measurement matrix has few rows $m$. 
While a very precise choice in the noise level estimate \( \eta \) of (8) leads to better reconstructions than using the noise-blind variant (7), the reconstructions deteriorate quickly once \( \eta \) is chosen as an overestimate of the actual noise level. In this sense, it is preferred to choose an underestimated \( \eta \) or even \( \eta = 0 \) (resulting again in (7)) in situations where there is little a priori knowledge about the noise \( w \).

4 Proof of the \( \ell_1 \)-quotient property and of the robustness of noise-blind \( \ell_1 \)-minimization

In this section, we provide proofs of Theorem 5 and Theorem 8. As a first step, we provide characterizations of the clipped \( \ell_2 \)-norm \( \|y\|^{(\alpha)} = \max\{\|y\|_2, \alpha \|y\|_{\infty}\} \) of a vector \( y \in \mathbb{R}^m \) for \( \alpha \geq 1 \) and also of its dual norm \( \|\cdot\|^{(\alpha)} \) in some preliminary lemmas. A sufficient condition for a matrix to \( A \) to fulfill the \( \ell_1 \)-quotient property relative to a general norm is provided in Lemma 14. Then, we present probabilistic arguments for this condition relative to clipped norms using results derived from Mendelson’s small ball method [12, 13, 11] by bounding appropriate quantities related to the distribution in question, which constitutes the main part of the proof.

4.1 Preliminary lemmas

Recall that for \( \alpha \geq 1 \), we defined the clipped \( \ell_2 \)-norm with parameter \( \alpha \) of \( y \in \mathbb{R}^m \) as

\[
\|y\|^{(\alpha)} := \max\{\|y\|_2, \alpha \|y\|_{\infty}\}.
\]

We will use the following two lemmas about its dual norm \( \|y\|^{(\alpha)} = \sup\{\|y\| \leq 1 \mid \langle x, y \rangle\} \) that can be found in [5, Lemma 11.22] and [17, Lemma 2]. They provide an explicit formula for \( \|\cdot\|^{(\alpha)} \) and compare it with the norm \( \|\cdot\|^{(\alpha)} \) defined below, whose advantage will become clear later on.

**Lemma 11.** Let \( T := T(y) := \{i \in [m] : \alpha |y_i| \leq \|y\|_2\} \) and \( T^c = \{i \in [m] : |y_i| > \frac{1}{\alpha} \|y\|_2\} \).

Then

\[
\|y\|_{\alpha}^{(\alpha)} = \|y_T\|_2 + \frac{1}{\alpha} \|y_{T^c}\|_1
\]

for all \( y \in \mathbb{R}^m \).

**Lemma 12.** Assume \( \alpha^2 \) is an integer. Then the dual norm \( \|\cdot\|_{\alpha}^{(\alpha)} \) of \( \|\cdot\|^{(\alpha)} \) is comparable with the norm \( \|\cdot\|_{\alpha^2, \dagger} \) defined by

\[
\|y\|_{\alpha^2, \dagger} := \max\left\{\sum_{\ell=1}^{\alpha^2} \|y_{B_{\ell}}\|_2, B_1, \ldots, B_{\alpha^2} \text{ form a partition of } [m]\right\}
\]

in the sense that

\[
\frac{1}{\alpha} \|y\|_{\alpha^2, \dagger} \leq \|y\|_{\alpha}^{(\alpha)} \leq \frac{\sqrt{2}}{\alpha} \|y\|_{\alpha^2, \dagger}
\]

for all \( y \in \mathbb{R}^m \).

We note that for \( \alpha = 1 \), the clipped norm and its dual norm reduce to the \( \ell_2 \)-norm since \( \|y\|_{\infty} \leq \|y\|_2 \), i.e., \( \|y\|^{(\alpha)} = \|y\|_2 = \|y\|_{\alpha}^{(\alpha)} \) for all \( y \in \mathbb{R}^m \) if \( \alpha = 1 \).

The following two lemmas provide a reformulation of the \( \ell_1 \)-quotient property relative to a norm \( \|\cdot\| \), which will be more convenient to analyze.
Lemma 13 ([5, Lemma 11.17]). A matrix $A \in \mathbb{R}^{m \times N}$ has the $\ell_1$-quotient property with constant $D$ relative to $\| \cdot \|$ if and only if
\[
\|w\|_s \leq ds_{1/2}^q \|A^*w\|_{\infty} \quad \text{for all } w \in \mathbb{R}^m,
\]
where $s_* = m / \log(e N / m)$.

Lemma 14. Let $A = (a_{ji}) \in \mathbb{R}^{m \times N}$ be a matrix with columns $a_i \in \mathbb{R}^m$, $i \in [N]$ and $q = \log(N)$. If
\[
\inf_{w \in S^{\| \cdot \|}} \left( \frac{1}{N} \sum_{i=1}^{N} |\langle a_i, w \rangle|^q \right)^{1/q} \geq \frac{1}{D} \sqrt{\log \left( \frac{e N}{m} \right)},
\]
where $S^{\| \cdot \|} = \{ w \in \mathbb{R}^m : \|w\|_s = 1 \}$ is the unit sphere of the dual norm $\| \cdot \|_s$ of some norm $\| \cdot \|$, then $\frac{1}{\sqrt{m}}A$ fulfills the $\ell_1$-quotient property with constant $D$ relative to the norm $\| \cdot \|$.

Proof. Let $w \in \mathbb{R}^m$. Then with $q = \log(N)$,
\[
\|\left( \frac{1}{\sqrt{m}}A \right)^* w\|_{\infty} \geq \frac{1}{e \sqrt{m}} \|A^*e\|_{\log N} = \frac{1}{e \sqrt{m}} \left( \frac{1}{N} \sum_{i=1}^{N} |\langle a_i, w \rangle|^q \right)^{1/q} = \frac{1}{\sqrt{m}} \left( \frac{1}{N} \sum_{i=1}^{N} |\langle a_i, w \rangle|^q \right)^{1/q},
\]
as $N^{1/q} = N^{1/\log(N)} = e$. It follows from Lemma 13 that the $\ell_1$-quotient property of $\frac{1}{\sqrt{m}}A$ with constant $D$ relative to the norm $\| \cdot \|$ is implied by
\[
\|A^*w\|_{\infty} \geq \frac{1}{D} \sqrt{\log(e N / m)} \|w\|_s \quad \text{for all } w \in \mathbb{R}^m,
\]
where $\| \cdot \|_s$ is the dual norm of the norm $\| \cdot \|$. This implies that
\[
\left( \frac{1}{N} \sum_{i=1}^{N} |\langle a_i, w \rangle|^q \right)^{1/q} \geq \frac{1}{D} \sqrt{\log(e N / m)} \|w\|_s \quad \text{for all } w \in \mathbb{R}^m
\]
is a sufficient condition for the assertion of the lemma. \hfill \square

4.2 Application of the small-ball method

The following result due to [11, Lemma III.1] and [13, Theorem 1.5] will be used to show the sufficient condition of Lemma 14.

Lemma 15. Let $1 \leq q < \infty$, let $S \subset \mathbb{R}^m$ be a set and $b_1, \ldots, b_N$ be i.i.d. copies of a random vector $b$ in $\mathbb{R}^m$. For $u > 0$, define
\[
Q_S(u) := \inf_{w \in S} \mathbb{P}( |\langle b, w \rangle| \geq u )
\]
and
\[
\mathcal{R}_N(S) := \mathbb{E} \left[ \sup_{w \in S} \left| \frac{1}{N} \sum_{i=1}^{N} \epsilon_i(b_i, w) \right| \right],
\]
where $(\epsilon_i)_{i \in [N]}$ is a Rademacher sequence that is independent from $(b_i)$. Then, for $t > 0$, with probability at least $1 - 2e^{-2t^2}$,
\[
\inf_{w \in S} \frac{1}{N} \sum_{i=1}^{N} |\langle b_i, w \rangle|^q \geq u^q \left( Q_S(2u) - \frac{4}{u} \mathcal{R}_N(S) - \frac{t}{\sqrt{N}} \right).
\]
To prove our main results, Lemma 15 is used by choosing the set $S = S^{\|\cdot\|_{\beta}}$, where $S^{\|\cdot\|_{\beta}}$ is the unit sphere of the dual of the clipped norm with $\beta = \sqrt{c_1 \log \left( \frac{\varepsilon N}{m} \right)}$ and $\beta = 1$, respectively, and $c_1 > 0$ being an absolute constant. We show the following lemma.

**Lemma 16.** Let $b_1, \ldots, b_N$ be distributed as i.i.d. vectors in $\mathbb{R}^m$ with independent entries $b_{ij} \sim X$, where $X$ is a variable such that $\|X\|_{L_2} = 1$. Let $S^{\|\cdot\|_{\beta}} = \{ w \in \mathbb{R}^m \mid \|w\|_{\beta} = 1 \}$ be the unit sphere of the norm $\| \cdot \|_{\beta}$ for $\alpha \geq 1$.

1. If $\alpha = 1$, then the complexity parameter $R_N(S^{\|\cdot\|_{\beta}})$ from (12) fulfills
   \[
   R_N(S^{\|\cdot\|_{\beta}}) = R_N(S^{\|\cdot\|_1}) \leq \sqrt{m}. \]

2. Assume $\alpha > 1$. If $X$ fulfills the weak moment assumption of order $\max\{4, \log(m)\}$ with constants $k_1$ and $\gamma$ and if $N \geq (\log(m))^{2\gamma+1}$, then the complexity parameter $R_N(S^{\|\cdot\|_{\beta}})$ from (12) fulfills
   \[
   R_N(S^{\|\cdot\|_{\beta}}) \leq \frac{1}{\sqrt{N}} \left( \sqrt{m} + \alpha e^{2\gamma} c_0 k_1 \log(m) \right)
   \]
   for an absolute constant $c_0 > 0$.

**Proof.** Let $h := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i b_i$, where $(\epsilon_i)_{i=1}^{m}$ is a Rademacher sequence independent of $(b_i)$. Then we obtain
   \[
   R_N(S^{\|\cdot\|_{\beta}}) = \mathbb{E} \left[ \sup_{w \in S^{\|\cdot\|_{\beta}}} \left\| w, \frac{1}{N} \sum_{i=1}^{N} \epsilon_i b_i \right\| \right] = \mathbb{E} \left[ \left\| \frac{1}{N} \sum_{i=1}^{N} \epsilon_i b_i \right\|_{\beta} \right] \tag{14}
   \]
   using the definition of fact that $\| \cdot \|_{\beta}$ is the dual of $\| \cdot \|_{\alpha}$, i.e., that the dual norm $\| \cdot \|_{\beta}$ is again the original norm $\| \cdot \|_{\alpha}$ in the second equality.

Furthermore, using Jensen’s inequality, we estimate
   \[
   \mathbb{E}[\|h\|_2] = \frac{1}{\sqrt{N}} \mathbb{E} \left[ \left\| \sum_{i=1}^{N} \epsilon_i b_i \right\|_2 \right] \leq \frac{1}{\sqrt{N}} \mathbb{E} \left[ \left\| \sum_{i=1}^{N} \epsilon_i b_i \right\|_2^2 \right]^{1/2} \]
   \[
   = \frac{1}{\sqrt{N}} \left( \mathbb{E}_b \left[ \mathbb{E}_e \left[ \sum_{i=1}^{N} \sum_{j=1}^{m} \epsilon_i \epsilon_k b_i b_j \right] \right] \right)^{1/2} = \frac{1}{\sqrt{N}} \left( \mathbb{E}_b \left[ \sum_{i=1}^{N} \sum_{j=1}^{m} b_i b_j \right] \right)^{1/2} \]
   \[
   = \frac{\sqrt{N m}}{\sqrt{N}} \left( \mathbb{E} |X|^2 \right)^{1/2} = \sqrt{m},
   \]
   since $\|X\|_{L_2} = (\mathbb{E} |X|^2)^{1/2} = 1$ by assumption. Assuming that $\alpha = 1$, this shows the first statement, since $\|h\|_{\infty} \leq \|h\|_2$ and therefore
   \[
   R_N(S^{\|\cdot\|_{\beta}}) = \frac{1}{\sqrt{N}} \mathbb{E}[\|h\|_2] \leq \frac{m}{N}.
   \]
To upper bound (14) for \( \alpha > 1 \), we calculate, using the notation \( h_j = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i b_{ij} \) for the \( j \)-th entry of \( h \), that for \( q = \log(m) \)

\[
\mathbb{E}[\|h\|_q^q] = \mathbb{E} \left[ \max_{j=1}^{m} |h_j|^q \right] = \mathbb{E} \left[ (\sum_{j=1}^{m} |h_j|^q)^{1/q} \right] \leq \mathbb{E} \left[ \sum_{j=1}^{m} |h_j|^q \right]
\]

\[
= \sum_{j=1}^{m} \mathbb{E} \left[ |h_j|^q \right] = m \mathbb{E} \left[ \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i b_{i1} \right)^q \right],
\]

where the last inequality holds since the components of \((\sum_{i=1}^{N} \epsilon_i b_{ij})_{j=1}^{m}\) are identically distributed. Since the independent random variables \((\epsilon_i b_{i1})_{i=1}^{N}\) are mean-zero and since they fulfill the weak moment assumption of order \(\log(m)\) moments with constants \(\kappa_1\) and \(\gamma\), the statement of [10, Lemma 2.8.] implies that there exists a constant \(c_0 > 0\) such that

\[
\mathbb{E}[\|h\|_q^q] \leq m (c_0 e^{2\gamma - 1} \kappa_1 \sqrt{q})^q
\]

with \(q = \log(m)\) if \(N \geq (\log(m))^{2\gamma - 1}\). Therefore

\[
\mathbb{E}[\|h\|_\infty] \leq m \frac{1}{\log(m)} c_0 e^{2\gamma - 1} \kappa_1 \sqrt{\log(m)} = c_0 e^{2\gamma} \kappa_1 \sqrt{\log(m)},
\]

and inserting this into (14), we obtain

\[
\mathcal{R}_N(B^{\|\cdot\|_{\alpha}}) = \frac{1}{\sqrt{N}} \mathbb{E} \left[ \max\{\|h\|_2, \alpha \|h\|_\infty\} \right] \leq \frac{1}{\sqrt{N}} \left( \mathbb{E}[\|h\|_2] + \alpha \mathbb{E}[\|h\|_\infty] \right)
\]

\[
\leq \frac{\sqrt{m}}{\sqrt{N}} + e^{2\gamma} c_0 \kappa_1 \frac{\alpha \sqrt{\log(m)}}{\sqrt{N}},
\]

which concludes the proof of the lemma. \(\square\)

Lemma 16 will be helpful to upper bound the term (12) in (13). To achieve a meaningful lower bound of the tail parameter \(Q_S(u)\) in Lemma 15, we will use our next result in Lemma 18 providing a rotation invariant lower bound for super-Gaussian random vectors.

It makes use of the following result by Montgomery-Smith about the distribution of Rademacher sums.

**Lemma 17 ([17]).** There exists a constant \(c > 0\) such that for all \(x \in \mathbb{R}^m\) with \(\|x\|_2 = 1\) and \(\alpha > 0\), we have

\[
\mathbb{P} \left( \langle \epsilon, y \rangle > \alpha \|y\|_{\alpha} \right) \leq e^{-\alpha^2/2}
\]

and

\[
\mathbb{P} \left( \langle \epsilon, y \rangle > \frac{\alpha}{c} \|y\|_{\alpha} \right) \geq c^{-1} e^{-c\alpha^2}
\]

where \(\epsilon = (\epsilon_1, \ldots, \epsilon_m)\) is a vector of i.i.d. Rademacher variables. The constant can be chosen as \(c = 4 \log(12)\).

**Lemma 18.** Assume that \(b = (b_1, \ldots, b_m)\) is a random vector in \(\mathbb{R}^m\) with i.i.d. entries distributed as a symmetric, unit variance random variable \(X\) that is super-Gaussian with parameter \(\sigma\).

Then there exist absolute constants \(c_1, c_2 > 0\) such that

\[
\mathbb{P} \left( \|b, w\| > u \right) > c_1 e^{-c_2 u^2 / 2\sigma^2}
\]

for all \(u \geq \frac{u}{\sqrt{2}}\) and all \(w \in \mathbb{R}^m\) such that \(\|w\|_2 = 1\).
Proof. Let \( t > 0 \). Due to symmetry of the random variables \( b_i, i \in [m] \), we can write \( b_i = \epsilon'_i |b_i| \), where \( \epsilon' = (\epsilon'_i) \) is a Rademacher vector independent of \( (|b_i|) \). By conditioning on \( \{ |b_i| \}_{i=1}^m \), we obtain

\[
\mathbb{P} \left( |\langle b, w \rangle| > t \right) = \mathbb{E} \left[ \mathbb{P} \left( \sum_{i=1}^m \epsilon'_i |b_i| |w_i| > t \{ |b_i| \}_{i=1}^m \right) \right] \\
= \mathbb{E} \left[ \mathbb{P} \left( \sum_{i=1}^m \epsilon'_i \text{sign} w_i |b_i| |w_i| > t \{ |b_i| \}_{i=1}^m \right) \right] \\
= \mathbb{E} \left[ \mathbb{P} \left( \sum_{i=1}^m \epsilon_i |b_i| |w_i| > t \{ |b_i| \}_{i=1}^m \right) \right] ,
\]

where \( \epsilon = (\epsilon_i)_{i \in [m]} \) is again a Rademacher vector independent of \( (|b_i|)_{i \in [m]} \). For \( r, t, f > 0 \), we define the events

\[
\mathcal{E}_{r,t,f}^{[b]} := \left\{ t f \Vert b \Vert \odot \Vert w \Vert^{(ff)} \geq \frac{t}{r} \right\} \quad \text{and} \quad \mathcal{E}_{r,t,f}^{[g]} := \left\{ t f \Vert g \Vert \odot \Vert w \Vert^{(ff)} \geq \frac{t}{r} \right\}
\]

using the dual of the clipped norm of (9), where \( g = (g_1, \ldots, g_m) \) is a vector of standard normal i.i.d. entries and \( |b| \odot |w| = (|b_i| |w_i|)_{i=1}^m \) the entrywise product of the vectors \( |b| \) and \( |w| \). Then

\[
\mathbb{P} \left( |\langle b, w \rangle| > t \right)
= \mathbb{E} \left[ \mathbb{P} \left( \sum_{i=1}^m \epsilon_i |b_i| |w_i| > t \mathcal{E}_{r,t,f}^{[b]} \cap \{ |b_i| \}_{i=1}^m \right) \mathcal{E}_{r,t,f}^{[b]} | \{ |b_i| \}_{i=1}^m \right] \\
+ \mathbb{E} \left[ \mathbb{P} \left( \sum_{i=1}^m \epsilon_i |b_i| |w_i| > t \mathcal{E}_{r,t,f}^{[b]} \cap \{ |b_i| \}_{i=1}^m \right) \mathcal{E}_{r,t,f}^{[b]} | \{ |b_i| \}_{i=1}^m \right] \\
\geq \mathbb{E} \left[ \mathbb{P} \left( \sum_{i=1}^m \epsilon_i |b_i| |w_i| > t \mathcal{E}_{r,t,f}^{[b]} \cap \{ |b_i| \}_{i=1}^m \right) \mathcal{E}_{r,t,f}^{[b]} | \{ |b_i| \}_{i=1}^m \right] \\
\geq \mathbb{E} \left[ \mathbb{P} \left( \sum_{i=1}^m \epsilon_i |b_i| |w_i| > t r f \Vert b \Vert \odot \Vert w \Vert^{(ff)} \mathcal{E}_{r,t,f}^{[b]} \cap \{ |b_i| \}_{i=1}^m \right) \mathcal{E}_{r,t,f}^{[b]} | \{ |b_i| \}_{i=1}^m \right] \\
\times \mathbb{E} \left[ \mathcal{E}_{r,t,f}^{[b]} | \{ |b_i| \}_{i=1}^m \right] \\
\geq \mathbb{E} \left[ \mathcal{E}_{r,t,f}^{[b]} | \{ |b_i| \}_{i=1}^m \right] \\
\geq \mathbb{E} \left[ \mathcal{E}_{r,t,f}^{[b]} | \{ |b_i| \}_{i=1}^m \right]
\]

It follows from Lemma 17 that if we choose \( r = (4 \log(12))^{-1} \), then

\[
\mathbb{P} \left( |\langle b, w \rangle| > t \right) \geq r e^{-\frac{t^2}{4}} \mathbb{E} \left[ \mathcal{E}_{r,t,f}^{[b]} | \{ |b_i| \}_{i=1}^m \right] \\
= r e^{-\frac{t^2}{4}} \mathbb{P} \left( r f \Vert b \Vert \odot \Vert w \Vert^{(ff)} \geq t \right).
\]

Since the \( b_i \) are independent, symmetric, unit variance random variables that fulfill the super-Gaussian assumption (6) for all \( i \in [m] \), we know that there exists a constant \( \sigma > 0 \) and independent standard normal random variables \( (g_i)_{i \in [m]} \) that depend on \( b_i \) for \( i \in [m] \), such that \( |b_i| \geq |\sigma g_i| \) almost surely.

Therefore, since \( |w_i| \geq 0 \) for all \( i \in [m] \), it holds that

\[
\mathbb{P} \left( r f \Vert b \Vert \odot \Vert w \Vert^{(ff)} \geq t \right) \geq \mathbb{P} \left( \sigma r f \Vert g \Vert \odot \Vert w \Vert^{(ff)} \geq t \right)
\]

with the random vector \( |g| = (|g_i|)_{i \in [m]} \). Next, we claim that

\[
\mathbb{P} \left( \mathcal{E}_{\sigma r,t,f}^{[g]} \right) = \mathbb{P} \left( \sigma r f \Vert g \Vert \odot \Vert w \Vert^{(ff)} \geq t \right) \geq e^{-\frac{t^2}{\sigma^2}} \quad \text{for all} \ t \geq t_0 := 1/4.
\]
On the contrary, suppose that this statement does not hold, i.e., that there exists some $t \geq \frac{1}{4}$ such that
\[
\mathbb{P} (\sigma_{t} \|f\|_{G} \cap \|w\|_{(t)} \geq t) < e^{-\frac{125t^2}{\sigma^2}} .
\]
Due to the first statement of Lemma 17 and since the distributions of $\langle g, |w| \rangle$ and $\sum_{i=1}^{m} \varepsilon_{i} |g_{i}| |w_{i}|$ coincide if $(\varepsilon_{i})_{i \in [m]}$ is a Rademacher vector which is independent of $(|g_{i}|)_{i \in [m]}$, we see that
\[
\mathbb{P} (|\langle g, |w| \rangle| > t \|f\|_{G} \cap \|w\|_{(t)})(E_{\sigma_{t}, f})^{c})
\]
\[
= \mathbb{E} \left[ \mathbb{P} \left( \left\{ \sum_{i=1}^{m} \varepsilon_{i} |g_{i}| |w_{i}| \right\} > t \|f\|_{G} \cap \|w\|_{(t)} \} \right) \approx t \|f\|_{G} \cap \|w\|_{(t)})(E_{\sigma_{t}, f})^{c}) \leq e^{-\frac{t^2}{\sigma^2}} .
\]
Furthermore, by the conditioning on the event $(E_{\sigma_{t}, f})^{c} = \{ t \|f\|_{G} \cap \|w\|_{(t)} \leq \frac{t}{\sigma} \}$, it follows that
\[
\mathbb{P} (|\langle g, |w| \rangle| > t \|f\|_{G} \cap \|w\|_{(t)})(E_{\sigma_{t}, f})^{c}) \geq \mathbb{P} (|\langle g, |w| \rangle| > \frac{t}{\sigma}) - \mathbb{P} (\{ |\langle g, |w| \rangle| > \frac{t}{\sigma} \} \cap (E_{\sigma_{t}, f})^{c})
\]
\[
\geq \mathbb{P} (|\langle g, |w| \rangle| > \frac{t}{\sigma}) - \mathbb{P} (E_{\sigma_{t}, f})
\]
\[
\geq \mathbb{P} (|\langle g, |w| \rangle| > \frac{t}{\sigma}) - \mathbb{P} (E_{\sigma_{t}, f}) \geq e^{-\frac{125t^2}{\sigma^2}} - \mathbb{P} (E_{\sigma_{t}, f})
\]
for all $t \geq \frac{1}{4}$, where we use the lower bound of the Gaussian integral
\[
\mathbb{P} \left( \frac{t}{\sigma} \right) = \frac{t}{\sigma} \geq \sqrt{\frac{2}{\pi}} \int_{\frac{t}{\sigma}}^{\infty} e^{-x^2/2} \, dx \geq \sqrt{\frac{2}{\pi}} \frac{2Ct}{\sigma^2} \geq e^{-\frac{125t^2}{\sigma^2}}
\]
in the last inequality, which is true for all $t \geq \frac{1}{4}$ if $r = (4 \log(12))^{-1}$ and $C = 4\sigma r \sqrt{\frac{t}{2}}$. A combination with (18) and (19) yields
\[
e^{-\frac{t^2}{\sigma^2}} > e^{-\frac{125t^2}{\sigma^2}} - e^{-\frac{125t^2}{\sigma^2}} .
\]
Finally, this leads to a contradiction for $f = \sqrt{\frac{2}{\pi}} r$, $r = (4 \log(12))^{-1}$ if $t \geq \frac{1}{4}$, since then the right hand side is strictly larger than $e^{-\frac{t^2}{\sigma^2}} = e^{-\frac{125t^2}{\sigma^2}}$. This shows statement (17).

Inserting (17) into (16), we see that
\[
\mathbb{P} (|\langle b, w \rangle| > \frac{t}{\sigma}) \geq r e^{-\frac{t^2}{\sigma^2}} = r e^{-\frac{125t^2}{\sigma^2}} = \frac{1}{8} (2.25 + 3.252) \approx 1.5220
\]
for all $r \geq \frac{1}{4}$, which concludes the assertion of the lemma with the constants $u_{0} = \frac{1}{4}$, $c_{1} = \frac{1}{4 \log(12)} \approx 1/10$ and $c_{2} = 2000 \log(12) + 250 \approx 5220$.

We note that the proof of the first part of the following lemma is similar to [4, Lemma 4.3].

Lemma 19. Assume that $b$ is a random vector in $\mathbb{R}^{m}$ with i.i.d. entries distributed as a symmetric, unit variance random variable $X$.

1. If $X$ fulfills the weak moment assumption of order 4 with constants $\gamma$ and $\kappa_{1}$, then
\[
Q_{\|f\|_{\ell^{P}}} \left( \frac{1}{4} \sqrt{\frac{\log(eN/m)}{\log(eC)}} \right) = \inf_{w \in \mathbb{S}^{m}_{\|f\|_{\ell^{P}}}} \mathbb{P} (|\langle b, w \rangle| \geq \frac{1}{4} \sqrt{\frac{\log(eN/m)}{\log(eC)}}) \geq 64 \sqrt{\frac{m}{N}}
\]
for $\beta = \sqrt{\log \left( \frac{eN}{m} \right) / \sqrt{\log(eC)}}$ if $N \geq Cm$, where $C := \frac{4^{1+\gamma} \gamma}{3 \gamma \kappa_{1}^{2}}$. 

2. If $X$ is super-Gaussian with parameter $0 < \sigma \leq 1$, then

$$Q_{\tilde{S}^{\perp,1}_{\alpha}}(\frac{1}{4} \sqrt{\frac{\log(eN/m)}{\log(eC)}}) \geq 64 \sqrt{\frac{m}{N}}$$

if $N \geq Cm$ with $C := 64c_1^{-2}e^{\frac{c_2}{6m}}$, where $c_1$ and $c_2$ are the constants from Lemma 18.

Proof. For $\beta \geq 1$ and $u > 0$, we aim at finding a lower bound for $Q_{\tilde{S}^{\perp,1}_{\alpha}}(2u)$. Let $w \in \mathbb{R}^m$. It is clear from the definition of the clipped norm that $\alpha_1 \leq \alpha_2$ implies $\|w\|^{(\alpha_1)} \leq \|w\|^{(\alpha_2)}$ for all $w \in \mathbb{R}^m$. It follows by duality that $\alpha_1 \leq \alpha_2$ implies $\|w\|^{(\alpha_2)} \geq \|w\|^{(\alpha_1)}$. Therefore, using the norm $\| \cdot \|_{\beta^2,1}$ from (10) and the upper inequality of Lemma 12, we estimate that

$$P\left(\|b, w\| \geq 2u\|w\|^{(\beta)}\right) \geq P\left(\|b, w\| \geq 2u\|w\|^{(\sqrt{\beta^2})}\right)$$

$$\geq P\left(\|b, w\| \geq \frac{2^{3/2}u}{\sqrt{\beta^2}}\|w\|_{\beta^2,1}\right).$$

Let now $B_1, \ldots, B_{[\beta^2]}$ be a partition of $[m]$ such that $\|w\|_{\beta^2,1} = \sum_{\ell=1}^{[\beta^2]} \|w_{B_\ell}\|_2$, cf. (10). Then

$$P\left(\|b, w\| \geq \frac{2^{3/2}u}{\sqrt{\beta^2}}\|w\|_{\beta^2,1}\right) \geq P\left(\|b, w\| \geq \frac{2^{3/2}u}{\sqrt{\beta^2}} \sum_{\ell=1}^{[\beta^2]} \|w_{B_\ell}\|_2\right)$$

$$= \prod_{\ell=1}^{[\beta^2]} P\left(\sum_{j \in B_\ell} b_j w_j \geq \frac{2^{3/2}u}{\sqrt{\beta^2}} \|w_{B_\ell}\|_2\right).$$

where we used that the entries $b_1, \ldots, b_m$ of $b$ are independent. Combining this with (20), we obtain

$$P\left(\|b, w\| \geq 2u\|w\|^{(\beta)}\right) = \prod_{\ell=1}^{[\beta^2]} P\left(\sum_{j \in B_\ell} b_j w_j \geq \frac{2^{3/2}u}{\sqrt{\beta^2}} \|w_{B_\ell}\|_2\right).$$

To show the first statement of the lemma, we choose $\beta = \sqrt{\frac{\log(eN/m)}{\log(eC)}}$, where $C \geq 1$, $u = \frac{1}{8} \sqrt{\log(\frac{eN}{m})/\log(eC)}$ and $N \geq Cm$. By the Paley-Zygmund inequality, see, e.g., [5, Lemma 7.17], and using the symmetry of the distribution of the $b_j$, we find the following lower bound for the latter probabilities,

$$P\left(\sum_{j \in B_\ell} b_j w_j \geq \frac{2^{3/2}u}{\sqrt{\beta^2}} \|w_{B_\ell}\|_2\right) = \frac{1}{2} P\left(\sum_{j \in B_\ell} b_j w_j \geq \frac{2^{3/2}u}{\sqrt{\beta^2}} \|w_{B_\ell}\|_2\right)$$

$$= \frac{1}{2} P\left(\sum_{j \in B_\ell} b_j w_j \geq \frac{\sqrt{2}}{4} \frac{\log(eN/m)}{\log(eC)} \|w_{B_\ell}\|_2\right) \geq \frac{1}{2} P\left(\sum_{j \in B_\ell} b_j w_j \geq \frac{1}{2} \|w_{B_\ell}\|_2\right)$$

$$\geq \frac{1}{2} \left(1 - (1/2)^2\right)^2 \geq \frac{9}{16} \frac{\log(eN/m)}{\log(eC)} = \frac{9}{32 \kappa_1^2 4^y}.$$
where we used that $\frac{x}{|x|} \leq 2$ for all $x > 0$ in the second inequality. Combining this lower bound with (22), we obtain
\[
\mathbb{P}\left(|\langle b, w \rangle| \geq 1 \frac{ \log(eN/m) }{ \log(eC) } \|w\|_\alpha \sqrt{ \log(N)} \right) \geq \frac{9}{32}\frac{\log(N/m)}{\log(eC)}\sqrt{\frac{m}{N}}
\]
for $N \geq Cm$ with $C = \frac{32}{e^2} (\kappa_1^4 4^y)^2 = \frac{e^{3+y}}{3} \kappa_1^8$, which concludes the first part of the lemma.

The second statement follows from Lemma 18 and can be considered as the special case of the above strategy for $\beta = 1$: If $\beta = 1$ and if $X$ fulfills the super-Gaussian assumption with parameter $\sigma$, then
\[
\mathbb{P}\left(|\langle b, w \rangle| \geq 1 \frac{ \log(eN/m) }{ \log(eC) } \|w\|_\alpha \right) = \mathbb{P}\left(|\langle b, w \rangle| \geq 1 \frac{ \log(eN/m) }{ \log(eC) } \|w\|_2 \right) \geq c_1 \left( \frac{eN}{m} \right)^{\frac{c_2}{\sigma^2 \log(eC)}} \geq 64\frac{m}{N}
\]
if $N \geq Cm$ with $C := 64^2 c_1^2 e^{\frac{c_2}{\sigma^2}}$, where $c_1$ and $c_2$ are the constants from Lemma 18. □

4.3 Proof of Theorem 5

Proof. By Lemma 14, to prove the first statement of Theorem 5, it suffices to show (11) for the norm $\| \cdot \| = \| \cdot \|^{(\alpha)}$, $\alpha = \sqrt{\log(eN/m)}$ with probability at least $1 - 2 \exp(-2m)$.

Consider the case that $N \geq \max \{ Cm, \log^{2y-1}(m) \}$ with $C = \frac{e^{3+y} \kappa_1^8}{3}$. We use Lemma 15 for the columns $a_1, \ldots, a_N$ of $A$, choosing the set $S$ such that $S = \| \cdot \|^{(\rho)}$ and $\beta = \frac{\log(\frac{m}{2})}{\sqrt{\log(eC)}}$, $q = \log(N)$ and $u = \frac{1}{\sqrt{\log(eC)}}$. Then, further choosing $t = \sqrt{m}$, it follows that with probability at least $1 - 2 e^{-2m}$,
\[
\inf_{w \in S^{\| \cdot \|^{(\rho)}}} \frac{1}{N} \sum_{i=1}^{N} |\langle a_i, w \rangle|^q \geq \frac{\log(eN/m)}{8^q \left( \log(eC) \right)^{q/2}} \left( Q_{S^{\| \cdot \|^{(\rho)}}}(2u) - \frac{4}{u} \mathcal{R}_N(S^{\| \cdot \|^{(\rho)}}) \right) - \sqrt{\frac{m}{N}}
\]
(23)

From Lemma 19, it follows that
\[
Q_{S^{\| \cdot \|^{(\rho)}}}(2u) \geq 64\frac{m}{N}
\]
(24)

For the complexity term $\frac{4}{u} \mathcal{R}_N(S^{\| \cdot \|^{(\rho)}})$, we use Lemma 16 to see that
\[
\frac{4}{u} \mathcal{R}_N(S^{\| \cdot \|^{(\rho)}}) \leq 32 \frac{m}{N} \sqrt{\frac{\log(eC)}{\log(eN/m)}} + 32 e^{2y} \kappa_1 \sqrt{\log(m)} \sqrt{\frac{m}{N}} \leq (32 + 16) \sqrt{\frac{m}{N}} = 48 \sqrt{\frac{m}{N}}
\]
(25)

since $N \geq Cm$ and $4 e^{4y} \kappa_1^2 \log(m) \leq m$ by assumption.

Finally, inserting (24) and (25) in (23), we obtain that with probability at least $1 - 2 e^{-2m}$,
\[
\inf_{w \in S^{\| \cdot \|^{(\rho)}}} \frac{1}{N} \sum_{i=1}^{N} |\langle a_i, w \rangle|^q \geq \frac{\log(eN/m)}{8^q \left( \log(eC) \right)^{q/2}} 15 \sqrt{\frac{m}{N}}
\]
or equivalently,

$$
\inf_{w \in S_{\|\cdot\|}^1} \left( \frac{1}{N} \sum_{i=1}^{N} |\langle a_i, w \rangle|^q \right)^{1/q} \geq \frac{\sqrt{\log(\varepsilon N/m)}}{8\sqrt{\log(e C)}}^{1/2}\frac{15^{1/\log(N)}}{\varepsilon N/m} \geq \frac{\sqrt{\log(\varepsilon N/m)}}{8\sqrt{\log(e C)}}.
$$

since \((m/N)^{1/\log(N)} = \exp(\log(m)/2\log(N) - \frac{1}{2}) > \exp(-1/2), which shows the assertion relative to the norm \(\| \cdot \|^{(\beta)}\) and constant

\[ D = 8 e^{1/2} \sqrt{\log(e C)} = 8 e^{1/2} \sqrt{1 + (9 + 8\gamma) \log(4) + 8 \log(\kappa_1)}. \]

The assertion relative to the norm \(\| \cdot \|^{(\alpha)}\) for \(\alpha = \sqrt{\log(e N/m)}\) follows then trivially since \(\|w\|^{(\beta)} \leq \|w\|^{(\alpha)}\), as \(\beta = \frac{\sqrt{\log(\varepsilon N/m)}}{\log(e C)} \leq \frac{\sqrt{\log(\varepsilon N/m)}}{m} = \alpha\).

Similarly, we show Theorem 5(b). To this end, we apply Lemma 15 for the choice \(S = \mathbb{R}^{1/\|\cdot\|} = \{ w \in \mathbb{R}^m \mid \|w\|_2 = 1 \}, q = \log(N), u = \frac{1}{8} \sqrt{\log(e C)}\) and \(t = \sqrt{m}\). If \(N \geq Cm\) with \(C := 64^2 \varepsilon^{-2} e^{\kappa_2^{1/2}}\), where \(\sigma\) is the super-Gaussian parameter of \(X\) and \(c_1, c_2\) are the constants from Lemma 19, it follows from Lemma 16 and Lemma 19 that with probability at least \(1 - 2 e^{-2m}\),

\[ \inf_{w \in \mathbb{R}^{1/\|\cdot\|}} \frac{1}{N} \sum_{i=1}^{N} |\langle a_i, w \rangle|^q \geq \frac{\log(\varepsilon N/m)^{q/2}}{8q(\log(e C))^{q/2}} \frac{1}{\sqrt{m}}. \]  

(26)

As above, this implies that

\[ \inf_{w \in \mathbb{R}^{1/\|\cdot\|}} \left( \frac{1}{N} \sum_{i=1}^{N} |\langle a_i, w \rangle|^q \right)^{1/q} \geq \frac{\sqrt{\log(\varepsilon N/m)}}{8 e^{1/2} \log(e C)}. \]

which finishes the proof of Theorem 5(b) by Lemma 14, i.e., \(\frac{1}{\sqrt{m}} A\) fulfills the \(\ell_1\)-quotient property with constant

\[ D = 8 e^{1/2} \sqrt{\log(e C)} = 8 e^{1/2} \sqrt{1 + 2 \log(64) + 2 \log(c_1) + \frac{c_2}{16\sigma^2}} \]

relative to \(\| \cdot \|_2\) on the event of (26).

4.4 Proof of Theorem 8

Proof. We first note that under the assumptions of both Theorem 8(a) and Theorem 8(b), it follows from [11, Corollary V.3] that with probability at least \(1 - \exp(-c_1 m)\), \(A\) fulfills the \(\ell_2\)-robust null space property of order \(\tilde{c}_2 s_s\), with \(s_s = m/\log(e N/m)\) and some constant \(0 < c_2 < 1\) and constants \(\rho = 0.9, \tau > 0\) relative to the \(\ell_2\)-norm, i.e.,

\[ \|x_{S'}\|_2 \leq \frac{\rho}{s_{1/2}} \|x_{S^c}\|_1 + \tau \|Ax\|_2 \]

(27)

for all \(x \in \mathbb{R}^m\) and \(S \subset [N]\) with \(|S| = s := \tilde{c}_2 s_s\) and \(S^c = [N] \setminus S\) if \(m \geq (\log(N))^{2\gamma - 1}\), where \(c_1, c_2, \rho, \tau\) depend on \(\kappa_1, \gamma\). We call the event with the latter statement \(E_{NSP}\).
Now we show the statement of Theorem 8(b). Conditional on the event \( E_{\text{NSP}} \), it follows from [15, Theorem II.22] that there exist constants \( C > 1, D > 0 \) that depend only on \( \kappa_1 \) and \( y \) such that
\[
\|x - \Delta_{1,\eta}(Ax + w)\|_p \leq \frac{C}{s^{1-1/p}} \sigma_s(x)_1 + s_1^{1/p-1/2} (D\eta + C(Q_{s_0}(A_1) + \tau) \max\{\|w\|_2 - \eta, 0\})
\]
for all \( 1 \leq p \leq 2 \) and \( s \leq \tilde{c}_2 s_s \) with \( Q_{s_0}(A_1) := \sup_{w \in \mathbb{R}^m \setminus \{0\}} \min_{u \in \mathbb{R}^N, Au = w} \frac{\sqrt{s}\|\|u\|_2}{\|w\|_2} \). By Theorem 5(b), it follows that there exist constants \( \tilde{C} \) and \( \tilde{D} \) such that if \( N \geq \tilde{C}m \), then with probability at least \( 1 - 2 \exp(-2m) \),
\[
Q_{s_0}(A_1) \leq \tilde{D},
\]
and we call this event \( E_{\text{QP-}\ell_2} \). The statement of Theorem 8(b) follows then since there exists a constant \( \tilde{c}_1 > 0 \) such that \( E_{\text{NSP}} \cap E_{\text{QP-}\ell_2} \) occurs with probability at least \( 1 - 3 \exp(-\tilde{c}_1 m) \), and on this event, the assertion follows with the constant \( E := C\tilde{D} + \tau \).

To show Theorem 8(a), we note that in the case \( \|w\|_2 \leq \eta \), the statement follows from the classical result [5, Theorem 4.22] on the event \( E_{\text{NSP}} \), since then there exist constants \( C' > 1, D' > 0 \) such that
\[
\|x - \Delta_{1,\eta}(Ax + w)\|_p \leq \frac{C'}{s^{1-1/p}} \sigma_s(x)_1 + s_1^{1/p-1/2} D'\eta
\]
for \( s = \tilde{c}_2 s_s \), for all \( 1 \leq p \leq 2 \) and all \( x \in \mathbb{R}^N \). Since \( s \mapsto \sigma_s(x)_1/s^{1-1/p} \) is non-increasing we may replace \( s \) also by a smaller value \( s' < s = \tilde{c}_2 s_s \).

Consider now the case \( \|w\|_2 > \eta \) and let \( z \in \mathbb{R}^N \) such that \( \|Az - w\|_2 \leq \eta \). It follows again by [5, Theorem 4.22] that on the event \( E_{\text{NSP}} \), for \( 1 \leq p \leq 2 \), there exist constants \( C' > 1, D' > 0 \) such that
\[
\|x - \Delta_{1,\eta}(Ax + w)\|_p = \|(x + z) - \Delta_{1,\eta}(Ax + z + w - Az) - z\|_p \leq \|(x + z) - \Delta_{1,\eta}(Ax + z + (w - Az))\|_p + \|z\|_p \\
\leq \frac{C'}{s^{1-1/p}} \sigma_s(x)_1 + s_1^{1/p-1/2} D'\eta + \|z\|_p \\
\leq \frac{C'}{s^{1-1/p}} \sigma_s(x)_1 + s_1^{1/p-1/2} D'\eta + \frac{C'}{s^{1-1/p}} \|z\|_1 + \|z\|_p \\
\leq \frac{C'}{s^{1-1/p}} \sigma_s(x)_1 + s_1^{1/p-1/2} D'\eta + C' \left( \frac{\|z\|_1}{s^{1-1/p}} + \|z\|_p \right) \\
\leq \frac{C'}{s^{1-1/p}} \sigma_s(x)_1 + s_1^{1/p-1/2} D'\eta + C s_1^{1/p-1/2} \left( \frac{\|z\|_1}{s^{1/2}} + s_1^{1/2} \|z\|_p \right). \tag{28}
\]
Moreover, if additionally \( N \geq \tilde{C}m \) and \( m \geq 4 e^4 \sqrt{c_0^2 \kappa_1^2 \log(m)} \), where \( \tilde{C} \) and \( c_0 \) are the constants from Theorem 5(a), it follows from this theorem that with probability at least \( 1 - 2 \exp(-2m) \), A fulfills the \( \ell_1 \)-quotient property relative to the norm \( \| \cdot \|_{\left(\frac{\log(e N/m)}{m}\right)} \) with constant \( \tilde{D} \), and we call the corresponding event \( E_{\text{QP-cliped}} \). Consider now on \( E_{\text{QP-cliped}} \cap E_{\text{NSP}} \), which occurs with probability at least \( 1 - 3 \exp(-\tilde{c}_1 m) \), a vector \( \tilde{z} \in \mathbb{R}^N \) such that \( A\tilde{z} = w \) and
\[
\|\tilde{z}\|_{s_{s_0}^{-1/2}} \leq \tilde{D} \left(\frac{\log(e N/m)}{m}\right), \tag{29}
\]
which exists due to the \( \ell_1 \)-quotient property relative to the norm \( \| \cdot \|_{\left(\frac{\log(e N/m)}{m}\right)} \). If \( z \in \mathbb{R}^N \) is chosen such that \( z := (1 - \eta/\|w\|_2)\tilde{z} \), it holds that
\[
\|Az - w\|_2 = \|A\tilde{z} - w - A\tilde{z}\|_2 \leq \eta \left(\frac{\|\tilde{z}\|}{\|w\|_2} - \|A\tilde{z}\|_2 \right) = \eta.
\]
Choose $S$ as an index set of $s$ largest absolute coefficients of $\tilde{z}$. It is known [5, Chapter 4.3] that the $\ell_2$-null space property (27) implies the $\ell_p$-null space property for $1 \leq p \leq 2$ in the form
\[
\|\tilde{z}_S\|_p \leq \frac{\rho}{s^{1-1/p}} \|\tilde{z}_{S^c}\|_p + \tau s^{1/p-1/2} \|A\tilde{z}\|_2.
\]
Together with Stechkin’s estimate, see, e.g., [5, Proposition 2.3], this gives
\[
\|\tilde{z}\|_p \leq \|\tilde{z}_S\|_p + \|\tilde{z}_{S^c}\|_p \leq \frac{\rho}{s^{1-1/p}} \|\tilde{z}_{S^c}\|_1 + \tau s^{1/p-1/2} \|A\tilde{z}\|_2 + \sigma_\delta(\tilde{z})_p
\]
\[
\leq \frac{\rho}{s^{1-1/p}} \|\tilde{z}\|_1 + \frac{1}{s^{1-1/p}} \|\tilde{z}_{S^c}\|_1 + \tau s^{1/p-1/2} \|A\tilde{z}\|_2 = \frac{1 + \rho}{s^{1-1/p}} \|\tilde{z}\|_1 + \tau s^{1/p-1/2} \|w\|_2 \tag{30}
\]
Using (28) and $s = \tilde{c}_2s$, we obtain
\[
\|x - \Delta_{1,\eta}(Ax + w)\|_p \leq C' \frac{s^{1-1/p}}{s^{1-1/p}} \sigma_\delta(x)_1 + (\tilde{c}_2s)_1^{1/p-1/2} \left[ D' \eta + C'(\frac{\|\tilde{z}\|_1}{(\tilde{c}_2s)_1^{1/2}} + (\tilde{c}_2s)_1^{1/2-1/p} \|\tilde{z}\|_p)(1 - \frac{\eta}{\|w\|_2}) \right]
\]
\[
\leq C' \frac{s^{1-1/p}}{s^{1-1/p}} \sigma_\delta(x)_1 + (\tilde{c}_2s)_1^{1/p-1/2} \left[ D' \eta + C'(\frac{(\rho + 2)\tilde{D}}{\tilde{c}_2^{1/2}} + \tau) \|w\|_2 \right](\|\tilde{z}\|_1^{1/2 - 1/p} + \eta)(1 - \frac{\eta}{\|w\|_2})
\]
\[
= C' \frac{s^{1-1/p}}{s^{1-1/p}} \sigma_\delta(x)_1 + s^{1/p-1} \left[ D' \frac{\|\tilde{z}\|_1^{1/2 - 1/p}}{\tilde{c}_2^{1/2}} + \tau \right] \|w\|_2 \left[ \frac{\log(eN/m)}{\|\tilde{z}\|_1^{1/2 - 1/p}} \right]
\]
where we used (29) and (30) in the second inequality. Since $s \mapsto \sigma_\delta(x)_1/s^{1-1/p}$ is non-increasing, again we may replace $s$ by a smaller value $s' < s = \tilde{c}_2s$. This concludes the proof of Theorem 8(a), as the constants can be defined as $C = C'$, $D = D'\tilde{c}_2^{1/2 - 1/p}$ and $E = \tilde{c}_2^{1/2 - 1/p}C'((\rho + 2)\tilde{D} + \tilde{c}_2^{1/2} \tau)$. \qed

Acknowledgements
The three authors acknowledge the support of the German-Israeli Foundation (GIF) through the grant G-1266-304.6/2015 (Analysis of structured random matrices in recovery problems).

References
[1] E. D. Gluskin, Extremal properties of orthogonal parallelepipeds and their applications to the geometry of banach spaces, Sb. Math. 64 (1) (1989) 85. doi:10.1070/SM1989v064n01ABEH003295.
[2] B. S. Kashin, On properties of random sections of an n-dimensional cube, Vestnik Moskov. Univ. Ser. I Mat. Mekh. (3) (1983) 8–11, English transl. in Moscow Univ. Math. Bull. 38 (1983).
[3] P. Wojtaszczyk, Stability and instance optimality for Gaussian measurements in compressed sensing, Found. Comput. Math. 10 (1) (2010) 1–13. doi:10.1007/s10208-009-0946-4.
[4] A. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, Smallest singular value of random matrices and geometry of random polytopes, Adv. Math. 195 (2) (2005) 491 – 523. doi:10.1016/j.aim.2004.08.004.
[5] S. Foucart, H. Rauhut, A Mathematical Introduction to Compressive Sensing. Applied and Numerical Harmonic Analysis. Springer New York, 2013. doi:10.1007/978-0-8176-4948-7.

[6] R. DeVore, G. Petrova, P. Wojtaszczyk, Instance-optimality in probability with an $\ell_1$-minimization decoder, Appl. Comput. Harmon. Anal. 27 (3) (2009) 275 – 288. doi: 10.1016/j.acha.2009.05.001.

[7] N. Dafnis, A. Giannopoulos, A. Tsolomitis, Asymptotic shape of a random polytope in a convex body, J. Funct. Anal. 257 (9) (2009) 2820 – 2839. doi:10.1016/j.jfa.2009.06.027.

[8] S. Brazitikos, A. Giannopoulos, P. Valettas, B.-H. Vritsiou, Geometry of isotropic convex bodies, American Mathematical Society, Providence, Rhode Island, 2014.

[9] M. Hardt, K. Talwar, On the geometry of differential privacy, in: Proceedings of the forty-second ACM symposium on Theory of computing, ACM Press, New York, USA, 2010, pp. 705–714. doi:10.1145/1806689.1806786.

[10] S. Mendelson, G. Lecué, Sparse recovery under weak moment assumptions, J. Eur. Math. Soc. 19 (3) (2017) 881–904. doi:10.4171/JEMS/682.

[11] S. Dirksen, G. Lecue, H. Rauhut, On the gap between restricted isometry properties and sparse recovery conditions, IEEE Trans. Inform. Theory to appear (2016) 1–1. doi:10.1109/TIT.2016.2570244.

[12] S. Mendelson, Learning without concentration, J. ACM 62 (3) (2015) 21:1–21:25. doi:10.1145/2699439.

[13] V. Koltchinskii, S. Mendelson, Bounding the smallest singular value of a random matrix without concentration, Int. Math. Res. Notices 2015 (23) (2015) 12991–13008. doi:10.1093/imrn/rnv096.

[14] S. Foccart, Stability and robustness of $\ell_1$-minimizations with weibull matrices and redundant dictionaries, Linear Algebra Appl. 441 (2014) 4 – 21, special issue on Sparse Approximate Solution of Linear Systems. doi:10.1016/j.laa.2012.10.003.

[15] S. Brugiapaglia, B. Adcock, Robustness to unknown error in sparse regularization, IEEE Trans. Inform. Theory to appear (2018) 1–1. doi:10.1109/TIT.2017.2788445.

[16] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1, http://cvxr.com/cvx (Mar. 2014).

[17] S. J. Montgomery-Smith, The distribution of Rademacher sums, Proc. Amer. Math. Soc. 109 (2) (1990) 517–522. doi:10.2307/2048015.