The Annals of Probability
2013, Vol. 41, No. 6, 4287–4316
DOI: 10.1214/12-AOP771

NONCOMMUTATIVE BENNETT AND ROSENTHAL INEQUALITIES

BY MARIUS JUNGE[1] AND QIANG ZENG

University of Illinois at Urbana-Champaign

In this paper we extend the Bernstein, Prohorov and Bennett inequalities to the noncommutative setting. In addition we provide an improved version of the noncommutative Rosenthal inequality, essentially due to Nagaev, Pinelis and Pinelis, Utev for commutative random variables. We also present new best constants in Rosenthal's inequality. Applying these results to random Fourier projections, we recover and elaborate on fundamental results from compressed sensing, due to Candes, Romberg and Tao.

0. Introduction. Rosenthal's inequality [42] was initially discovered to construct some new Banach spaces. However, Rosenthal's inequality gives a very nice bound for the p-norm of independent random variables and has found many generalizations and applications. The martingale version of Rosenthal's inequality was discovered almost simultaneously by Burkholder [4]. Since then, the order of the constants in these inequalities has been studied extensively, in particular by Johnson, Schechtman and Zinn [20]. The correct order in the martingale version has been established by Hitczenko [19], based on fundamental work of Kwapień and Woyczyński [27]. Nowadays, easy proofs of Rosenthal inequalities can be found with the help of Bernstein, Prohorov and Bennett's inequalities; see [3, 39] and the references therein. Historically, Bernstein's inequality was first established in the 1920s, according to the references in [3]. Later on, Prohorov improved Bernstein's inequality in [39]. Then, Bennett, who seemed to be unaware of Prohorov's work, strengthened Bernstein's results directly in [3], which provided an even more precise bound than Prohorov's inequality. We will extend Bennett's inequalities to the noncommutative setting, and then obtain the noncommutative Bernstein and Prohorov inequalities as consequences.

Let us recall that the classical Rosenthal inequality says that for independent mean 0 random variables, we have

$$\left(\mathbb{E}\left|\sum_{k=1}^{n} f_k\right|^p\right)^{1/p} \le c(p)\left(\left(\sum_{k=1}^{n}\mathbb{E}|f_k|^2\right)^{1/2} + \left(\sum_{k=1}^{n}\mathbb{E}|f_k|^p\right)^{1/p}\right). \tag{0.1}$$

Received November 2011; revised March 2012.

[1]Supported in part by NSF Grant DMS-09-01457.

MSC2010 subject classifications. Primary 46L53, 60E15; secondary 46L52, 60F10, 94A12.

Key words and phrases. (Noncommutative) Bennett inequality, (noncommutative) Rosenthal inequality, (noncommutative) Bernstein inequality, (noncommutative) Prohorov inequality, noncommutative L_p spaces, compressed sensing, large deviation, Cramér's theorem.

According to [20], the order of the best constant here is $c(p) = p/(1 + \log p)$. In this paper we separate the two terms and ask for

$$(0.2)\quad \left(\mathbb{E}\left|\sum_{k=1}^{n} f_k\right|^p\right)^{1/p} \le A(p)\left(\sum_{k=1}^{n} \mathbb{E}|f_k|^2\right)^{1/2} + B(p)\left(\sum_{k=1}^{n} \mathbb{E}|f_k|^p\right)^{1/p}.$$

The central limit theorem immediately implies $A(p) \ge c\sqrt{p}$ for every choice of $B(p)$. Problem (0.2) is by no means new. Nagaev and Pinelis [32] obtained a very precise bound on the tail behavior of $S_n = \sum_{k=1}^n X_k$ which implies that $(A(p), B(p)) = C(\sqrt{p}, p)$ is possible. Pinelis and Utev showed that in some sense $A(p) = C\sqrt{p}$ and $B(p) = Cp$ are also best. In Section 3, we will revisit this problem and show that assuming $A(p) \le Cp^m$ for some $m > 1/2$, we must have

$$B(p) \ge c\frac{p}{1 + \log p}.$$

This is exactly consistent with $(A(p), B(p)) = C(p/(1 + \log p), p/(1 + \log p))$. Moreover, we show that the worst case is obtained for independent random selectors $f_k = (\delta_k - \lambda)$ with expectation $\lambda > 0$.

We will prove a vast generalization of (0.2) in the noncommutative setting for conditionally independent random variables with $A(p) = c\sqrt{p}$ and $B(p) = Cp$. This improves the corresponding results from [25] of the form $A(p) = B(p) = Cp$. Our new results are motivated by applications in compressed sensing for random selectors with matrix valued coefficients. More precisely, we have to consider rank-one operators

$$a_j = [\bar{x}_j(l)x_j(r)]_{1\le l,r\le n}$$

such that $|x_k(j)| \le D$. Then the aim is to estimate

$$(0.3)\quad \left\|\frac{1}{k}\sum_{j=1}^{n} \delta_j f a_j f - f\right\|_{B(\ell_2^n)} \le ?$$

for independent selectors $\delta_j \in \{0, 1\}$ with $\mathbb{E}\delta_j = k/n$ and a projection f. As in the foundational paper on compressed sensing by Candes, Romberg and Tao [6], it is tempting to use moment estimates, or equivalently, estimates of the Schatten p-norm of these matrices. In fact, the improved Rosenthal inequality allows us to recover the famous estimates in [6].

Let us recall that the noncommutative L_p space associated with the trace on $B(\ell_2)$ is given by

$$\|x\|_p = [\mathrm{tr}(|x|^p)]^{1/p} = \left(\sum_j s_j(x)^p\right)^{1/p},$$

where the singular number $s_j(x) = \lambda_j(|x|)$, that is, the eigenvalues of the positive matrix $|x| = \sqrt{x^*x}$. Thus a good estimate of (0.3) can certainly be obtained from an estimate of the form

$$\left(\mathbb{E}\left\|\sum_j \delta_j f a_j f - kf\right\|_p^p\right)^{1/p} \le C\sqrt{p}\left\|\sum_j \mathbb{E}(\delta_j^2) f a_j f^2\right\|_{p/2}^{1/2} + Cp\left(\sum_j \|f a_j f\|_p^p\right)^{1/p}. \tag{0.4}$$

Let us now describe the more general setup which allows us to prove results in noncommutative probability which includes all the statements above. Indeed, we assume that $\mathcal{M}$ is a von Neumann algebra equipped with a normal faithful tracial state $\tau : \mathcal{M} \to \mathbb{C}$, that is, $\tau(1) = 1$ and $\tau(xy) = \tau(yx)$. Then $L_p(\mathcal{M}, \tau)$ is the completion of $\mathcal{M}$ with respect to $\|x\|_p = [\tau(|x|^p)]^{1/p}$. It is well known (see, e.g., [15, 38]) that $\|\cdot\|_p$ is a norm for $1 \le p \le \infty$. In particular, $\|\cdot\|_\infty = \|\cdot\|$. Here and in the following, $\|\cdot\|$ will always denote the operator norm. Let $\mathcal{N} \subset \mathcal{M}$ be a von Neumann subalgebra. Then there exists a unique conditional expectation $E_\mathcal{N} : \mathcal{M} \to \mathcal{N}$ such that $E_\mathcal{N}(1) = 1$ and

$$E_\mathcal{N}(axb) = aE_\mathcal{N}(x)b, \qquad a, b \in \mathcal{N} \text{ and } x \in \mathcal{M}.$$

We say that two subalgebras $\mathcal{N} \subset A, B \subset \mathcal{M}$ are *independent* over $\mathcal{N}$ if

$$E_\mathcal{N}(ab) = E_\mathcal{N}(a)E_\mathcal{N}(b), \qquad a \in A, b \in B.$$

In particular, we say that $x, y \in \mathcal{M}$ are independent if the algebras they generate, respectively, are independent over $\mathbb{C}$. A sequence of subalgebra $A_1, \ldots, A_n$ are called successively independent over $\mathcal{N}$ if A_{k+1} is independent of the algebra $\mathcal{M}(k)$ generated by $A_1, \ldots, A_k$. Our noncommutative Bennett inequality reads as follows.

THEOREM 0.1. *Let $\mathcal{N} \subset A_j \subset \mathcal{M}$ be successively independent over $\mathcal{N}$ and $a_j \in A_j$ be self-adjoint such that*:

(i) $E_\mathcal{N}(a_j) = 0$; (ii) $E_\mathcal{N}(a_j^2) \le \sigma_j^2$; (iii) $\|a_j\| \le M_j$.

Then for $t \ge 0$,

$$\tau\left(1_{[t,\infty)}\left(\sum_{j=1}^n a_j\right)\right) \le \exp\left(-\frac{\sum_{j=1}^n \sigma_j^2}{\sup_{j=1,\ldots,n} M_j^2}\phi\left(\frac{t \sup_{j=1,\ldots,n} M_j}{\sum_{j=1}^n \sigma_j^2}\right)\right),$$

where $\phi(x) = (1+x)\log(1+x) - x$.

Here we used $1_I(a) = \int_I dE_t$ for the spectral projection given by the spectral decomposition $a = \int t\, dE_t$. We should mention that the key new ingredient in this theorem is the Golden–Thompson inequality, which has already played a crucial role in Ahlswede and Winter's paper [1], Gross's paper [16] and Oliveira's paper [33]. The best constants for random matrices probability inequalities so far are due to Tropp [47] by using Lieb's theorem [28]. However, it seems Lieb's theorem does not apply to the fully noncommutative setting. In our approach we allow general randomness via independence not necessarily given by classical filtrations. Indeed, all the other works we mentioned only considered the semicommutative case or the random matrix case where operators with classical randomness act on a finite-dimensional Hilbert space. We invite the reader to rewrite the inequality for conditionally independent copies x_j with $\sigma = \sigma_j$, $M_j = M$. Note that in the commutative context,

$$\tau\big(1_{[t,\infty)}(a)\big) = \mathrm{Prob}(a \geq t).$$

In the future we will simply take this formula as a definition. Then our Bernstein and Prohorov inequalities for noncommutative random variables reads as follows.

COROLLARY 0.2. *Under the same hypothesis of Theorem 0.1, we have*

$$\mathrm{Prob}\left(\sum_{j=1}^n a_j \geq t\right) \leq \exp\left(-\frac{t^2}{2\sum_{j=1}^n \sigma_j^2 + (2t/3)\sup_{j=1,\dots,n} M_j}\right) \tag{0.5}$$

and

$$\begin{aligned}&\mathrm{Prob}\left(\sum_{j=1}^n a_j \geq t\right)\\ &\quad\leq \exp\left(-\frac{t}{2\sup_{j=1,\dots,n} M_j}\operatorname{arcsinh}\left(\frac{t\sup_{j=1,\dots,n} M_j}{2\sum_{j=1}^n \sigma_j^2}\right)\right).\end{aligned} \tag{0.6}$$

It is now rather standard to derive Rosenthal's inequality from Bernstein's inequality (0.5).

COROLLARY 0.3. *Let $2 \leq p < \infty$ and a_j satisfy the hypothesis of Theorem 0.1. Then*

$$\left\|\sum_{j=1}^n a_j\right\|_p \leq C\left(\left(p\sum_{j=1}^n \sigma_j^2\right)^{1/2} + p\sup_{j=1,\dots,n} M_j\right).$$

For unbounded operators and fixed p, we can prove a similar inequality. Here we have to make a slightly stronger assumption. Let us recall that $(A_j)_{j=1}^n$ are fully independent over $\mathcal{N}$ if for every subset $I \subset \{1,\dots,n\}$ the algebra $\mathcal{M}(I)$ generated by $\bigcup_{i\in I} A_i$ is independent from $\mathcal{M}(I^c)$ over $\mathcal{N}$.

THEOREM 0.4. *Let (A_i) be fully independent over $\mathcal{N}$, $1 \le p < \infty$, $x_i \in L_p(A_i)$ with $E_{\mathcal{N}}(x_i) = 0$. Then*

$$\left\| \sum_{j=1}^n x_j \right\|_p \le C \max\left\{ \sqrt{p} \left\| \left(\sum_{j=1}^n E_{\mathcal{N}}(x_j x_j^* + x_j^* x_j) \right)^{1/2} \right\|_p, \; p \left(\sum_{j=1}^n \|x_j\|_p^p \right)^{1/p} \right\}. \tag{0.7}$$

If moreover, $p \ge 2.5$, then

$$\left\| \sum_{j=1}^n x_j \right\|_p \le C' \max\left\{ \sqrt{p} \left\| \left(\sum_{j=1}^n E_{\mathcal{N}}(x_j x_j^* + x_j^* x_j) \right)^{1/2} \right\|_p, \; p \|(x_j)\|_{L_p(\ell_\infty)} \right\}. \tag{0.8}$$

According to [36] and [21], the norm of (x_j) in $L_p(\ell_\infty)$ is given by $\inf\{\|a\|_{2p} \times \|b\|_{2p}\}$ such that

$$x_j = a y_j b \qquad \text{with } \|y_j\|_\infty \le 1.$$

Clearly, the orders $\sqrt{p}$ and p in the above theorem are optimal because they are already optimal in commutative probability. Note that in this version Theorem 0.4 improves on Corollary 0.3 for p large enough. The passage from first assertion to the second follows from an argument in [25]. After we put this paper on arXiv.org and submitted it for publication, S. Dirksen, being aware of our work, showed us his different proof of (0.7) and (0.9) with slightly better constants (private communication). Two months later, J. A. Tropp informed us that he obtained a particular case (i.e., the random matrix version) of (0.7) with several coauthors independently by using a different method in a later paper [31]. In fact, Rosenthal inequalities in the noncommutative setting have been successively explored in [23, 24] and [25]. The martingale situation is completely settled due to the work of [40] which shows that for noncommutative martingales,

$$\left\| \sum_j d_j \right\|_p \le Cp \left(\left\| \left(\sum_k E_{k-1}(d_k d_k^* + d_k^* d_k) \right)^{1/2} \right\|_p + \left(\sum_k \|d_k\|_p^p \right)^{1/p} \right),$$

where (d_k) is a sequence of martingale differences given by $E_k(x) = E_{\mathcal{N}_k}(x)$ and $d_k = d_k(x) = E_k(x) - E_{k-1}(x)$ for a filtration $(\mathcal{N}_k) \subset \mathcal{M}$. As observed in [24], the constant Cp gives the correct order.

Let us return to the situation in compressed sensing. Here we obtain the following result.

COROLLARY 0.5. *Let $x_j \in \mathcal{N}$ be positive operator, τ a normalized trace such that*:

(i) $\frac{1}{m}\sum_{j=1}^m x_j = 1$; (ii) $\|x_j\| \le r$.

Let δ_j be independent selectors such that $\mathbb{E}\delta_j = k/m$. Then for $p \ge 2.5$,

$$\left(\mathbb{E}\left\|\frac{1}{k}\sum_{j=1}^m \delta_j x_j - 1\right\|_{L_p(\tau)}^p\right)^{1/p} \le C\max\left\{\sqrt{\frac{pr}{k}}, \frac{pr}{k}\right\}. \tag{0.9}$$

Moreover, if tr *is a trace on $\mathcal{N}$ such that*

$$\|x\|_{L_\infty(\mathrm{tr})} \le \|x\|_{L_p(\mathrm{tr})}$$

and $r/k = \varepsilon^2$, then, for $t^2 \ge 2.5C^2e$ and $t \ge 2.5Ce\varepsilon$, we have

$$\mathrm{Prob}\left(\left\|\frac{1}{k}\sum_{j=1}^m \delta_j x_j - 1\right\|_{L_\infty(\mathrm{tr})} > t\varepsilon\right) \le \mathrm{tr}(1)\begin{cases} e^{-t^2/(2C^2e)}, & \text{if } t\varepsilon \le C,\\ e^{-t/(2Ce\varepsilon)}, & \text{if } t\varepsilon \ge C.\end{cases} \tag{0.10}$$

Here C is an absolute constant.

These results are closely related to the matrix Bernstein inequality from Tropp's paper [47] and operator Bernstein inequality from [16]. Their application to problem in compressed sensing will be explained in Section 4. Section 1 provides the proof of the Bennett's inequality and its consequences. An application to large deviation inequalities and how noncommutative Gaussian random variables may violate the classical equalities are discussed in Section 2. The improved Rosenthal inequality is proved in Section 3.

1. Noncommutative Bennett inequality. Let us first recall some background. For a self-adjoint operator $a \in \mathcal{M}$, we have the spectral decomposition $a = \int t\,dE_t$, where E_t is the spectral measure of a. For any Borel set $A \subset \mathbb{R}$, we define $\mu(A) = \tau(E(A))$. Then μ is a scalar-valued spectral measure for a and $\mu(\mathbb{R}) = 1$. By the measurable functional calculus (see, e.g., [11], Section IX.8), there exists a $*$-homomorphism $\pi : L^\infty(\mu) \to \mathcal{M}$ depending on a such that for all $f \in L^\infty(\mu)$, $\pi(f) = f(a)$ and

$$\tau\big(f(a)\big) = \int f(t)\mu(dt). \tag{1.1}$$

In particular, for $f = 1_{[t,\infty)}$, we have the exponential Chebyshev inequality

$$\tau\big(1_{[t,\infty)}(a)\big) = \mathrm{Prob}(a \ge t) \le e^{-t}\tau\big(e^a\big). \tag{1.2}$$

Our proof of Bennett's inequality relies on the well-known Golden–Thompson inequality. For the usual trace on $B(H)$ we may refer to Simon's book [45]. The fully general case is due to Araki [2]. A transparent proof for semifinite von Neumann algebras can be found in Ruskai's paper ([44], Theorem 4).

LEMMA 1.1 (Golden–Thompson inequality). *Suppose that a, b are self-adjoint operators, bounded above and that $a + b$ are essentially self-adjoint* (*i.e.*, *the closure of $a + b$ is self-adjoint*). *Then*

$$\tau(e^{a+b}) \le \tau(e^{a/2}e^b e^{a/2}).$$

Furthermore, if $\tau(e^a) < \infty$ or $\tau(e^b) < \infty$, then

$$\tau(e^{a+b}) \le \tau(e^a e^b). \tag{1.3}$$

Note that if $a, b \in \mathcal{M}$ are self-adjoint, the hypotheses in Lemma 1.1 are automatically satisfied. Therefore we have (1.3). With the help of (1.2) and (1.3), we can prove the noncommutative Bennett inequality following the commutative case given in [3].

PROOF OF THEOREM 0.1. (1.2) implies for $\lambda \ge 0$,

$$\operatorname{Prob}\left(\sum_{i=1}^n a_i \ge t\right) \le e^{-\lambda t}\tau\left(e^{\lambda\sum_{i=1}^n a_i}\right). \tag{1.4}$$

Since (a_i) are successively independent, we deduce from (1.3) that

$$\begin{aligned}\tau\left(e^{\lambda\sum_{i=1}^n a_i}\right) &\le \tau\left(e^{\lambda\sum_{i=1}^{n-1} a_i}e^{\lambda a_n}\right) = \tau\left(E_{\mathcal{N}}\left(e^{\lambda\sum_{i=1}^{n-1} a_i}e^{\lambda a_n}\right)\right)\\ &= \tau\left(E_{\mathcal{N}}\left(e^{\lambda\sum_{i=1}^{n-1} a_i}\right)E_{\mathcal{N}}\left(e^{\lambda a_n}\right)\right).\end{aligned} \tag{1.5}$$

Expanding, we obtain

$$\begin{aligned}E_{\mathcal{N}}(e^{\lambda a_n}) &= E_{\mathcal{N}}\left(\sum_{k=0}^\infty \frac{(\lambda a_n)^k}{k!}\right) = \sum_{k=0}^\infty \frac{\lambda^k}{k!}E_{\mathcal{N}}(a_n^k)\\ &= 1 + \sum_{k=2}^\infty \frac{\lambda^k}{k!}E_{\mathcal{N}}(a_n^2 a_n^{k-2}) \le 1 + \sum_{k=2}^\infty \frac{\lambda^k}{k!}M_n^{k-2}\sigma_n^2\\ &= 1 + \frac{\sigma_n^2}{M_n^2}(e^{\lambda M_n} - 1 - \lambda M_n) \le \exp\left(\frac{\sigma_n^2}{M_n^2}(e^{\lambda M_n} - 1 - \lambda M_n)\right).\end{aligned}$$

Note that the function $f(x) := \exp(x^{-2}(e^{\lambda x} - 1 - \lambda x))$ is increasing for $x > 0$. It follows that

$$E_{\mathcal{N}}(e^{\lambda a_n}) \le \exp\left(\frac{\sigma_n^2}{C^2}(e^{\lambda C} - 1 - \lambda C)\right),$$

where $C = \sup_{i=1,\dots,n} M_i$. Iterating $n-2$ times, we obtain

$$\tau\left(e^{\lambda\sum_{i=1}^n a_i}\right) \le \exp\left(\frac{\sum_{i=1}^n \sigma_i^2}{C^2}(e^{\lambda C} - 1 - \lambda C)\right).$$

This yields

$$\operatorname{Prob}\left(\sum_{i=1}^{n} a_i \geq t\right) \leq \exp\left(-\lambda t + \frac{\sum_{i=1}^{n} \sigma_i^2}{C^2}\left(e^{\lambda C} - 1 - \lambda C\right)\right). \tag{1.6}$$

By differentiating we find the minimizing value $\lambda = C^{-1}\log(1 + tC/(\sum_{i=1}^{n}\sigma_i^2))$. Then (1.6) yields the assertion. □

PROOF OF COROLLARY 0.2. Note that $\phi(x) \geq x^2/(2+2x/3)$ and that $\phi(x) \geq (x/2)\operatorname{arcsinh}(x/2)$ for $x \geq 0$. Then the corollary follows by relaxing the bound in Bennett's inequality. □

In the following we use Corollary 0.2 to prove Corollary 0.3. Let $a \in \mathcal{M}$ be positive. Recall that $\operatorname{Prob}(a > t)$ is an analog of the classical distribution function of a. In particular, we may use it to compute the L_p norm of a. Indeed, by the same argument as commutative case, for $p > 0$ and positive $a \in \mathcal{M}$, we have

$$\|a\|_p^p = p\int_0^\infty t^{p-1}\operatorname{Prob}(a > t)\,dt. \tag{1.7}$$

Recall that the Gamma function is defined as $\Gamma(p) = \int_0^\infty e^{-r} r^{p-1}\,dr$, and the incomplete Gamma function is defined as $\Gamma(\alpha, p) = \int_p^\infty e^{-t} t^{\alpha-1}\,dt$. We need an elementary estimate for $\Gamma(\alpha, p)$. Note that for $t \geq p \geq 2(\alpha - 1)$, we have

$$(e^{-t}t^{\alpha-1})' = -e^{-t}t^{\alpha-1}\left(1 - \frac{\alpha-1}{t}\right) \leq -\frac{1}{2}e^{-t}t^{\alpha-1}.$$

This gives the following lemma.

LEMMA 1.2. *If $p \geq 2\alpha - 2$, then $\Gamma(\alpha, p) \leq 2e^{-p}p^{\alpha-1}$.*

PROOF OF COROLLARY 0.3. First note that symmetry and Corollary 0.2 imply

$$\operatorname{Prob}\left(\left|\sum_{i=1}^{n} a_i\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\sum_{i=1}^{n}\sigma_i^2 + (2t/3)\sup_{1\leq i\leq n} M_i}\right).$$

Put $S = \sum_{i=1}^{n}\sigma_i^2$ and $R = \sup_{i=1,\dots,n} M_i$. By (1.7), we have

$$\begin{aligned}
\left\|\sum_{i=1}^{n} a_i\right\|_p^p &\leq 2p\int_0^\infty \exp\left(-\frac{t^2}{2S + 2tR/3}\right)t^{p-1}\,dt \\
&= 2p\int_0^{3S/R} \exp\left(-\frac{t^2}{2S + 2tR/3}\right)t^{p-1}\,dt \\
&\quad + 2p\int_{3S/R}^\infty \exp\left(-\frac{t^2}{2S + 2tR/3}\right)t^{p-1}\,dt \\
&= 2p(I + II),
\end{aligned}$$

where

$$I = \int_0^{3S/R} \exp\Big(-\frac{t^2}{2S + 2tR/3}\Big) t^{p-1}\, dt$$

and

$$II = \int_{3S/R}^{\infty} \exp\Big(-\frac{t^2}{2S + 2tR/3}\Big) t^{p-1}\, dt.$$

We first estimate I. Since $t \leq 3S/R$, we have

$$I \leq \int_0^{3S/R} e^{-t^2/(4S)} t^{p-1}\, dt = 2^{p-1} S^{p/2} \int_0^{9S/(4R^2)} e^{-r} r^{p/2-1}\, dr.$$

For $9S/(4R^2) \leq p$, we have $I \leq 2^{p-1} S^{p/2} \int_0^p e^{-r} r^{p/2-1}\, dr \leq 2^p S^{p/2} p^{p/2-1}$. For $9S/(4R^2) > p$, we have

$$\begin{aligned} I &\leq 2^{p-1} S^{p/2} \Big(\int_0^p e^{-r} r^{p/2-1}\, dr + \int_p^{9S/(4R^2)} e^{-r} r^{p/2-1}\, dr \Big) \\ &\leq 2^p S^{p/2} p^{p/2-1} + I_2, \end{aligned}$$

where $I_2 = 2^{p-1} S^{p/2} \int_p^{\infty} e^{-r} r^{p/2-1}\, dr$, and by Lemma 1.2, $I_2 \leq 2^p S^{p/2} p^{p/2-1} \times e^{-p}$. Hence, we obtain

$$I \leq 2^{p+1} S^{p/2} p^{p/2-1}.$$

To estimate II, since $2S < 2tR/3$, we have

$$\begin{aligned} II &\leq \int_{3S/R}^{\infty} e^{-3t/(4R)} t^{p-1}\, dt = \Big(\frac{4}{3}R\Big)^p \int_{9S/(4R^2)}^{\infty} e^{-r} r^{p-1}\, dr \\ &\leq \Big(\frac{4}{3}R\Big)^p \Gamma(p) \leq \Big(\frac{4}{3}Rp\Big)^p. \end{aligned}$$

Combining all the inequalities together, we find $\| \sum_{i=1}^n a_i \|_p^p \leq 2^{p+2} S^{p/2} p^{p/2} + 2(4R/3)^p p^{p+1}$. Hence, we obtain

$$\Big\| \sum_{i=1}^n a_i \Big\|_p \leq 4\sqrt{Sp} + \frac{4\sqrt{2}}{3} e^{1/e} Rp \leq 4(\sqrt{Sp} + Rp). \qquad \square$$

We remark that the constant in the above inequality is explicit and quite small, which may be good for numerical purpose.

2. Large deviation principle. Bennett's inequality is a large deviation type inequality giving an upper bound for the tail probability. In the commutative setting lower bounds have been analyzed intensively in large deviation theory. Despite the fact that our arguments in the previous section are almost commutative, lower bounds for noncommutative random variables are very different. Let us start with Cramér's theorem. We consider a sequence of fully independent and identically distributed (i.i.d.) τ-measurable (see, e.g., [15]) noncommutative random variables $(a_i)_{i\in I}$.

Let $\Lambda(\lambda) = \log \tau(e^{\lambda a_1})$. Following [12] we define the Fenchel–Legendre transform of $\Lambda(\lambda)$ for $x \in \mathbb{R}$

$$\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}}\bigl[\lambda x - \Lambda(\lambda)\bigr]. \tag{2.1}$$

If (a_i) is a commutative i.i.d. sequence, then Cramér's theorem ([12], Theorem 2.2.3) says that (a_i) satisfies the large deviation principle (LDP) with rate function Λ^*, which implies [12], Corollary 2.2.19,

$$\limsup_{n\to\infty} \frac{1}{n} \log \operatorname{Prob}\Biggl(\sum_{i=1}^{n} a_i \geq nt\Biggr) = -\inf_{s\geq t} \Lambda^*(s). \tag{2.2}$$

The upper bound remains valid in the noncommutative setting.

PROPOSITION 2.1. *Let $(a_i)_{i\geq 1}$ be an i.i.d. sequence in $(\mathcal{M}, \tau)$ such that $\tau(a_i) = 0$ for all $i \geq 1$. Then for any $t > 0$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \operatorname{Prob}\Biggl(\sum_{i=1}^{n} a_i \geq nt\Biggr) \leq -\inf_{s\geq t} \Lambda^*(s).$$

PROOF. Thanks to the Golden–Thompson inequality, we can follow the proof in the commutative case in [12]. Using (1.4) and (1.5), we obtain

$$\operatorname{Prob}\Biggl(\sum_{i=1}^{n} a_i \geq nt\Biggr) \leq e^{-\lambda nt} \prod_{i=1}^{n} \tau\bigl(e^{\lambda a_i}\bigr) = e^{-n(\lambda t - \Lambda(\lambda))}.$$

This implies

$$\frac{1}{n} \log \operatorname{Prob}\Biggl(\sum_{i=1}^{n} a_i \geq nt\Biggr) \leq -\Lambda^*(t) \leq -\inf_{s\geq t} \Lambda^*(s).$$ □

REMARK 2.2. Although we assumed a_i's are in $(\mathcal{M}, \tau)$, using truncation and approximation, we can also prove the previous proposition for symmetric Gaussians. To be more precise, for independent symmetric Gaussian random variables

a and b, let $a_N = a1_{\{|a|<N\}}$ and $b_N = b1_{\{|b|<N\}}$. Then the monotone convergence theorem implies that $\tau(e^{a_N}) \to \tau(e^a)$, $\tau(e^{b_N}) \to \tau(e^b)$. Since the symmetric Gaussian random variable is in $\bigcap_{p\geq 1} L_p(\mathcal{M}, \tau)$, the triangle inequality implies $\tau((a_N + b_N)^p) \to \tau((a+b)^p)$. By symmetry, we have

$$\tau\bigl(e^{a_N+b_N}\bigr) \to \tau\bigl(e^{a+b}\bigr).$$

In the following we give two examples which violate the LDP for noncommutative random variables.

EXAMPLE 2.3 (Noncommutative semicircular law [48]). Recall that the *semicircular law* centered at $a \in \mathbb{R}$ and of radius $r > 0$ is the distribution $\gamma_{a,r} : \mathbb{C}[X] \to \mathbb{C}$ defined by

$$\gamma_{a,r}(P) = \frac{2}{\pi r^2} \int_{a-r}^{a+r} P(t)\sqrt{r^2 - (t-a)^2}\, dt.$$

Here $\mathbb{C}[X]$ is the algebra of complex polynomials in one variable.

Let us recall that copies of semicircular random variables can be constructed on the full Fock space; see, for example, [48], Section 2.6. We find a sequence of the so-called free (thus fully independent) Gaussian random variables $\{s_i\}_{i\in I}$ with the identical distribution $\gamma_{0,2}$. By rotation invariance of the free functor, we deduce from [48], Section 3.4, that

$$\hat{s}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i \sim \gamma_{0,2}, \tag{2.3}$$

which means that the distribution of $\hat{s}_n$ is $\gamma_{0,2}$. Since $\gamma_{0,2}$ is supported in $[-2, 2]$, for any $t > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \operatorname{Prob}\Biggl(\sum_{i=1}^{n} s_i \geq nt\Biggr) = \lim_{n\to\infty} \frac{1}{n} \log \operatorname{Prob}(\hat{s}_n \geq \sqrt{n}t) = -\infty.$$

On the other hand, by the integral representation of the modified Bessel function I_1 ([46], (9.46)), the moment generating function of $\gamma_{0,2}$ is given by

$$M(\lambda) = \frac{1}{2\pi} \int_{-2}^{2} e^{\lambda t} \sqrt{4 - t^2}\, dt = \frac{I_1(2\lambda)}{\lambda}.$$

Using the series representation of I_1 ([46], (9.28)), we have for $\lambda > 0$,

$$M(\lambda) = \sum_{n=0}^{\infty} \frac{\lambda^{2n}}{(n+1)!n!} \geq \sum_{n=0}^{\infty} \frac{\lambda^{2n}}{2(2n)!} = \frac{e^{\lambda} + e^{-\lambda}}{4} \geq \frac{1}{4} e^{\lambda}.$$

We find $\Lambda(\lambda) = \log M(\lambda) \geq \lambda - \log 4$. Since $\tau(a_1) = 0$, by [12], Lemma 2.2.5, for $x \geq 0$,

$$\Lambda^*(x) = \sup_{\lambda\geq 0} \bigl[\lambda x - \Lambda(\lambda)\bigr].$$

Therefore,

$$\Lambda^*(1) = \sup_{\lambda \geq 0} [\lambda - \Lambda(\lambda)] \leq \log 4 < \infty,$$

which shows that the sequence (s_i) violates the LDP lower bound in (2.2). We have proved the following result.

PROPOSITION 2.4. *The semicircular sequence* $(s_n)_{n\in\mathbb{N}}$ *does not satisfy LDP* (2.2).

The counterexample works in free probability because s_1 is bounded. In order to motivate the next example, we first clarify the relationship between the logarithmic moment generating function Λ and the rate function I of the LDP.

REMARK 2.5. Suppose that an i.i.d. sequence (a_n) satisfies the LDP with rate function $I(x)$ and that $\Lambda(\lambda)$ is well defined. Then the Fenchel–Legendre transform of $I(x)$ coincides with $\Lambda(\lambda)$, that is,

$$I^*(\lambda) = \Lambda(\lambda).$$

Indeed, by Hölder's inequality $\Lambda(\lambda)$ is convex, and by Fatou's lemma for τ-measurable operators ([15], Theorem 3.5), $\Lambda(\lambda)$ is lower semicontinuous. Then Cramér's theorem and the duality lemma ([12], Lemma 4.5.8) yield the assertion. In particular, if (a_n) satisfies the LDP with rate function $I(x)$ and $\Lambda(\lambda)$ exists, then $I(x) = x^2/2$ implies $\Lambda(\lambda) = I^*(\lambda) = \lambda^2/2$; that is, the sequence (a_n) follows standard normal distribution. This means in classical probability *the distribution of an i.i.d. sequence can be recovered from the rate function* given by the LDP. The next proposition will show that this is no longer the case in the noncommutative setting. Therefore, a literal translation of the LDP is not to be expected in noncommutative probability.

PROPOSITION 2.6 (Gaussian family). *Let* $\theta \in (0, 1)$. *There exists an i.i.d. sequence* $(\xi_n)_{n\geq 1}$ *of noncommutative Gaussian random variables with logarithmic moment generating function* $\Lambda_\theta(\lambda)$ *such that*:

(i) (ξ_n) *satisfies the LDP with rate function* $I_\theta(x) = x^2/2$;

(ii) $|\Lambda_\theta(\lambda) - \frac{\lambda^2}{2} - \log(1-\theta)| \leq \frac{\theta}{1-\theta} e^{2\lambda - \lambda^2/2}$.

In particular, $I_\theta^*(\lambda) = \lambda^2/2 \neq \Lambda_\theta(\lambda)$. *Therefore, the law of* (ξ_n) *cannot be recovered from the LDP rate function*.

Before going to the proof, we remark that the failure of recovering the law $\Lambda_\theta(\cdot)$ from rate function $I_\theta(\cdot)$ is because Cramér's theorem is no longer true in the noncommutative setting. Indeed, since $\Lambda_0(\lambda) = \lambda^2/2$ is the logarithmic moment generating function of standard normal distribution and $\Lambda_0^*(x) = x^2/2$, if

Cramér's theorem were true, we would have $\Lambda_\theta^*(x) = I_\theta(x) = x^2/2 = \Lambda_0^*(x)$. But $\Lambda_\theta(\lambda) \neq \Lambda_0(\lambda)$ as stated above, this contradicts the injectivity of Fenchel–Legendre transform.

PROOF OF PROPOSITION 2.6. For $\theta \in (0, 1)$, given a noncommutative standard Gaussian random variable g_0 (with probability density function $e^{-x^2/2}/\sqrt{2\pi}$) and a noncommutative semicircular random variable $g_1 \sim \gamma_{0,2}$, there exists a noncommutative random variable g_θ such that

$$\tau(g_\theta^k) = (1-\theta)\tau(g_0^k) + \theta\tau(g_1^k).$$

This implies by approximation (see [22])

$$\tau\big(f(g_\theta)\big) = (1-\theta)\tau\big(f(g_0)\big) + \theta\tau\big(f(g_1)\big)$$

for all measurable function f. In particular, for any Borel set $A \subset \mathbb{R}$,

$$\tau\big(1_A(g_\theta)\big) = (1-\theta)\tau\big(1_A(g_0)\big) + \theta\tau\big(1_A(g_1)\big) \tag{2.4}$$

and for all $\lambda \in \mathbb{R}$,

$$\tau\big(e^{\lambda g_\theta}\big) = (1-\theta)\tau\big(e^{\lambda g_0}\big) + \theta\tau\big(e^{\lambda g_1}\big). \tag{2.5}$$

Moreover, for every real Hilbert space H there exists an algebra $\mathcal{N}_\theta(H)$, together with a map $u : H \to \mathcal{N}_\theta(H)$ and a family of trace preserving automorphisms $\alpha_o : \mathcal{N}_\theta(H) \to \mathcal{N}_\theta(H)$ indexed by the contractions o of H such that

$$\alpha_o\big(u(h)\big) = u\big(o(h)\big).$$

We apply this for $H = \ell_2(\mathbb{N})$ and define $\xi_i = u(e_i)$ where ξ_1 has the same distribution as g_θ. Using the permutations, we see that $\xi_i = \alpha_{(i1)}(\xi_1)$, and hence these variables are identical distributed. Using the conditional expectations onto $N_\theta(\ell_2(I))$, $I \subset \mathbb{N}$, we see that (ξ_i) is a fully independent sequence. Using a real unitary which maps e_1 to $\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i$, we deduce that ξ_1 and $\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i$ have the same distribution, that is,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i \overset{D}{=} \xi_1 \overset{D}{=} g_\theta; \tag{2.6}$$

see [10, 17, 18] for more details. Following [12], Section 2.2, we define $S_n = \frac{1}{n}\sum_{k=1}^n \xi_k$ and $\mu_n(A) = \tau(1_A(S_n))$. By the invariance property (2.6), we have $\mu_n(A) = \tau(1_{\sqrt{n}A}(\sqrt{n}S_n)) = \tau(1_{\sqrt{n}A}(g_\theta))$. Using (2.4), we find

$$\mu_n(A) = \tau\big(1_{\sqrt{n}A}(g_\theta)\big) = (1-\theta)\tau\big(1_{\sqrt{n}A}(g_0)\big) + \theta\tau\big(1_{\sqrt{n}A}(g_1)\big). \tag{2.7}$$

We aim to establish an LDP for (μ_n). Let A be a Borel set and $I_\theta(x) = x^2/2$. Note that the support of the distribution of g_1 is $[-2, 2]$. We consider the following two cases:

(1) $0 \in \mathrm{cl}(\mathrm{int}(A))$, the closure of interior of A. If there exists an interval $(-\delta, \delta) \subset \mathrm{cl}(\mathrm{int}(A))$, then $\lim_{n\to\infty} \tau(1_{\sqrt{n}A}(g_1)) = 1$. If no such interval exists, 0 is a boundary point of $\mathrm{cl}(\mathrm{int}(A))$, then $\lim_{n\to\infty} \tau(1_{\sqrt{n}A}(g_1)) = 1/2$. In any case, we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_n(A) = 0 = -\inf_{x\in\mathrm{int}(A)} I_\theta(x) = -\inf_{x\in\mathrm{cl}(A)} I_\theta(x).$$

(2) $0 \notin \mathrm{cl}(\mathrm{int}(A))$. In this case, $\mathrm{int}(\sqrt{n}A \cap [-2, 2])$ will eventually be empty for n large enough. Then we have $\lim_{n\to\infty} \tau(1_{\sqrt{n}A}(g_1)) = 0$. First we assume $\mathrm{int}(A) \neq \varnothing$ and without loss of generality, we assume $\mathrm{int}(A) \subset \mathbb{R}_+$. Let $x = \inf\{\mathrm{int}\, A\}$ and (x, T) be an interval contained in A. Then we have

$$\int_{\sqrt{n}x}^{\sqrt{n}T} e^{-t^2/2}\, dt \le \int_{\sqrt{n}A} e^{-t^2/2}\, dt \le \int_{\sqrt{n}x}^{\infty} e^{-t^2/2}\, dt.$$

Since $\tau(1_{\sqrt{n}A}(g_0)) = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{n}A} e^{-t^2/2}\, dt$, straightforward computation shows that

$$-\frac{x^2}{2} \le \liminf_{n\to\infty} \frac{1}{n} \log \tau\big(1_{\sqrt{n}A}(g_0)\big) \le \limsup_{n\to\infty} \frac{1}{n} \log \tau\big(1_{\sqrt{n}A}(g_0)\big) \le -\frac{x^2}{2}.$$

This fact together with (2.7) yields

$$-\inf_{x\in\mathrm{int}(A)} I_\theta(x) \le \lim_{n\to\infty} \frac{1}{n} \log \mu_n(A) \le -\inf_{x\in\mathrm{cl}(A)} I_\theta(x). \tag{2.8}$$

Note that if $\mathrm{int}(A) = \varnothing$, (2.8) is trivial.

According to [12], (1.2.4), we have shown that (μ_n) or (ξ_n) satisfies the LDP with rate function $I_\theta(x) = x^2/2$. On the other hand, if we put $\Lambda_\theta(\lambda) = \log \tau(e^{\lambda g_\theta})$ and let ν denote the probability measure of g_1, then (2.5) implies

$$\begin{aligned}
\Lambda_\theta(\lambda) &= \log\Big((1-\theta)e^{\lambda^2/2} + \theta \int_{-2}^{2} e^{\lambda t} \nu(dt)\Big) \\
&= \log\Big((1-\theta)e^{\lambda^2/2}\Big(1 + \frac{\theta e^{-\lambda^2/2}}{1-\theta} \int_{-2}^{2} e^{\lambda t} \nu(dt)\Big)\Big) \\
&\le \log\Big((1-\theta)e^{\lambda^2/2}\Big(1 + \frac{\theta e^{-\lambda^2/2} e^{2\lambda}}{1-\theta} \int_{-2}^{2} \nu(dt)\Big)\Big) \\
&\le \log(1-\theta) + \frac{\lambda^2}{2} + \log\Big(1 + \frac{\theta}{1-\theta} e^{2\lambda - \lambda^2/2}\Big)
\end{aligned}$$

and similarly,

$$\Lambda_\theta(\lambda) \ge \log(1-\theta) + \frac{\lambda^2}{2} + \log\Big(1 + \frac{\theta}{1-\theta} e^{-2\lambda - \lambda^2/2}\Big).$$

Combining these two inequalities, we obtain

$$
\begin{aligned}
\left|\Lambda_\theta(\lambda) - \frac{\lambda^2}{2} - \log(1-\theta)\right| &\le \log\left(1 + \frac{\theta}{1-\theta} e^{2\lambda - \lambda^2/2}\right) \\
&\le \frac{\theta}{1-\theta} e^{2\lambda-\lambda^2/2},
\end{aligned}
\tag{2.9}
$$

which implies $\lim_{\lambda\to\infty} \Lambda_\theta(\lambda) - \lambda^2/2 = \log(1-\theta)$. In particular, $\Lambda_\theta(\lambda) \neq \lambda^2/2$. Since $I_\theta^*(\lambda) = \lambda^2/2 \neq \Lambda_\theta(\lambda)$, we have proved that the law $\Lambda_\theta(\cdot)$ of (ξ_n) cannot be recovered from the LDP rate function $I_\theta(\cdot)$. □

3. Improved noncommutative Rosenthal's inequality. We prove the improved noncommutative Rosenthal inequality and show that the coefficients cannot be improved in this section. In order to prove Theorem 0.4, we will follow and refine the standard iteration procedure given in [25], used before by Lust-Piquard [29] and Pisier, Gilles and Xu [38].

PROOF OF THEOREM 0.4. Instead of proving (0.7) directly, we prove the following equivalent inequality:

$$
\begin{aligned}
\left\|\sum_{j=1}^n x_j\right\|_p \le D_p \max\Bigg\{ &\sqrt{p}\left\|\left(\sum_{j=1}^n E_{\mathcal{N}}(x_j^* x_j)\right)^{1/2}\right\|_p, \\
&\sqrt{p}\left\|\left(\sum_{j=1}^n E_{\mathcal{N}}(x_j x_j^*)\right)^{1/2}\right\|_p, p\left(\sum_{j=1}^n \|x_j\|_p^p\right)^{1/p}\Bigg\},
\end{aligned}
\tag{3.1}
$$

and we assume at the moment that D_p is the best constant which may depend on the range of p. By [25], Theorem 2.1, (3.1) is true for $1 \le p \le 4$. This is the starting point of our iteration argument. Assume $p > 2$. We only need to show "$p \Rightarrow 2p$." Let $x_i \in L_{2p}(\mathcal{M}, \tau)$. Write the conditional expectation operator $E = E_{\mathcal{N}}$ in the following proof. Put

$$
A = \sqrt{2p}\left\|\left(\sum_{i=1}^n E(x_i^* x_i)\right)^{1/2}\right\|_{2p} \quad \text{and} \quad B = 2p\left(\sum_{i=1}^n \|x_i\|_{2p}^{2p}\right)^{1/(2p)}.
$$

Using [23], Lemma 1.2, and the noncommutative Khintchine inequality in [36] with the right order of best constant, we have

$$
\left\|\sum_{i=1}^n x_i\right\|_{2p} \le 2\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i x_i\right\|_{2p} \le c\sqrt{p}\max\left\{\left\|\sum_{i=1}^n x_i^* x_i\right\|_p^{1/2}, \left\|\sum_{i=1}^n x_i x_i^*\right\|_p^{1/2}\right\},
$$

where (ε_i) is a sequence of Rademacher random variables, and $\mathbb{E}$ denotes the corresponding expectation. Let $y_i = x_i^* x_i - E(x_i^* x_i)$. Then

$$\left\| \sum_{i=1}^n x_i^* x_i \right\|_p \le 2 \max\left\{ \left\| \sum_{i=1}^n y_i \right\|_p, \left\| \sum_{i=1}^n E(x_i^* x_i) \right\|_p \right\}.$$

Applying the induction hypothesis, we obtain

$$\left\| \sum_{i=1}^n y_i \right\|_p \le D_p \max\left\{ \sqrt{p} \left\| \left(\sum_{i=1}^n E(y_i^2) \right)^{1/2} \right\|_p, p \left(\sum_{i=1}^n \|y_i\|_p^p \right)^{1/p} \right\}.$$

Note that

$$E(y_i^2) = E(|x_i|^4) - (E(|x_i|^2))^2 \le E(|x_i|^4).$$

By [23], Lemma 5.2, we obtain

$$\begin{aligned} \left\| \sum_{i=1}^n E(|x_i|^4) \right\|_{p/2} &\le \left\| \sum_{i=1}^n E(|x_i|^2) \right\|_p^{(p-2)/(p-1)} \left(\sum_{i=1}^n \|x_i\|_{2p}^{2p} \right)^{1/(p-1)} \\ &= (A^2/2p)^{(p-2)/(p-1)} (B/2p)^{2p/(p-1)} \\ &= A^{(2p-4)/(p-1)} B^{2p/(p-1)} (2p)^{-(3p-2)/(p-1)}. \end{aligned}$$

On the other hand, since E is a contraction on $L_p(\mathcal{M}, \tau)$, we have

$$\begin{aligned} \left(\sum_{i=1}^n \|y_i\|_p^p \right)^{1/p} &= \left(\sum_{i=1}^n \|x_i^* x_i - E(x_i^* x_i)\|_p^p \right)^{1/p} \\ &\le 2 \left(\sum_{i=1}^n \|x_i^* x_i\|_p^p \right)^{1/p} = 2 \left(\sum_{i=1}^n \|x_i\|_{2p}^{2p} \right)^{1/p} = \frac{B^2}{2p^2}. \end{aligned}$$

This gives

$$\begin{aligned} \left\| \sum_{i=1}^n y_i \right\|_p &\le D_p \max\left\{ \sqrt{p} A^{(p-2)/(p-1)} B^{p/(p-1)} (2p)^{-(3p-2)/(2p-2)}, \frac{pB^2}{2p^2} \right\} \\ &\le D_p \max\left\{ 2^{-(3p-2)/(2p-2)} A^{(p-2)/(p-1)} B^{p/(p-1)} p^{-1-1/(2p-2)}, \frac{B^2}{2p} \right\}. \end{aligned}$$

Hence, we find

$$\begin{aligned} \left\| \sum_{i=1}^n x_i^* x_i \right\|_p \le \max\Big\{ &2^{-p/(2p-2)} D_p A^{(p-2)/(p-1)} B^{p/(p-1)} p^{-1-1/(2p-2)}, \\ &\frac{D_p B^2}{p}, \frac{A^2}{p} \Big\}. \end{aligned} \tag{3.2}$$

Young's inequality for products implies

$$A^{(p-2)/(2p-2)}B^{p/(2p-2)} \le \frac{(p-2)A}{2p-2} + \frac{pB}{2p-2} \le \max\{A, B\}. \tag{3.3}$$

Note that $2^{-p/(4p-4)} \le 2^{-1/4}$ and $p^{-1/(4p-4)} \le 1$. Equations (3.2) and (3.3) yield

$$\begin{aligned}
&\sqrt{p}\left\|\sum_{i=1}^{n} x_i^* x_i\right\|_p^{1/2} \\
&\quad \le \max\{2^{-1/4}\sqrt{D_p}\max\{A, B\}, \sqrt{D_p}B, A\} \le \sqrt{D_p}\max\{A, B\} \\
&\quad = \sqrt{D_p}\max\left\{\sqrt{2p}\left\|\left(\sum_{i=1}^{n} E(x_i^* x_i)\right)^{1/2}\right\|_{2p}, 2p\left(\sum_{i=1}^{n}\|x_i\|_{2p}^{2p}\right)^{1/(2p)}\right\}.
\end{aligned}$$

Applying the same argument to $x_i x_i^*$, we obtain

$$\begin{aligned}
&\sqrt{p}\left\|\sum_{i=1}^{n} x_i x_i^*\right\|_p^{1/2} \\
&\quad \le \sqrt{D_p}\max\left\{\sqrt{2p}\left\|\left(\sum_{i=1}^{n} E(x_i x_i^*)\right)^{1/2}\right\|_{2p}, 2p\left(\sum_{i=1}^{n}\|x_i\|_{2p}^{2p}\right)^{1/(2p)}\right\}.
\end{aligned}$$

Hence, (3.1) is true for $2p$ with constant $c\sqrt{D_p}$. It follows that

$$D_{2p} \le c\sqrt{D_p},$$

and thus $D_p \le c^2$ which is independent of p. Therefore, the iteration argument is complete, and we have proved the first assertion. As mentioned in the Introduction of this paper, the interpolation argument from [25], Section 4, shows that the first assertion can be improved to the second assertion with a singularity as p tends to 2. Thus for $p \ge 2.5$ the assertion holds with an absolute constant. □

REMARK 3.1. The improved Rosenthal inequality allows us to extend Lust-Piquard's noncommutative Khintchine inequality [29, 30] in a twisted setting. We refer to [9] for unexplained notion on the Gaussian measure space construction. The starting point is a discrete group acting on a real Hilbert space H. This means we fix an isometry $b: H \to L_2(\Omega, \Sigma, \mu)$ such that b is linear, and $b(h)$ is a centered Gaussian random variable with variance $\|h\|^2$. For example, for $H = L_2(0, \infty)$ and $B_t = b(1_{[0,t]})$ we recover a well-known method to construct Brownian motion. We may assume that Σ is the minimal sigma algebra generated by the random variables $b(H)$. Then the action of G extends to a family of measure preserving automorphism $\alpha : G \to \mathrm{Aut}(L_\infty(\Omega, \Sigma, \mu))$ such that

$$\alpha_g\big(b(h)\big) = b(g.h).$$

This allows us to form the crossed product $M = L_\infty(\Sigma) \rtimes G$. The crossed product is spanned by random variables of the form

$$x = \sum_g f_g \lambda(g).$$

Here $\lambda(g)$ refers to the regular representation of group. The algebraic structure is determined by $\lambda(g) f \lambda(g^{-1}) = \alpha_g(f)$. The twisted Gaussian random variables are of the form

$$B = \sum_g b(h_g)\lambda(g), \qquad h_g \in H.$$

In order to formulate the Khintchine inequality, we have to recall that there exists trace preserving conditional expectation $E : M \to L(G)$. Here $L(G)$ is the von Neumann subalgebra generated by the image $\lambda(G)$ and the trace is given by

$$\tau\Big(\sum_g f_g \lambda(g)\Big) = \int f_1 \, d\mu.$$

Then we can deduce from Theorem 0.4 that for $p \ge 2$,

$$\|B\|_p \le c\sqrt{p}\big\|E\big(B^*B + BB^*\big)^{1/2}\big\|_p. \tag{3.4}$$

Moreover, the span of the generalized Gaussian random variables is complemented, and the inequality remains true with additional vector valued coefficients. This is a key fact in proving noncommutative Riesz transforms. To illustrate (3.4) let us assume that the action is trivial. Let (e_k) be a basis and

$$B = \sum_{k,g} a(k,g) b(e_k) \otimes \lambda(g) = \sum_k b(e_k) \otimes a_k.$$

Then we find

$$E\big(BB^*\big) = \sum_k a_k a_k^*, \qquad E\big(B^*B\big) = \sum_k a_k^* a_k.$$

Thus the right-hand side gives exactly the square function we expect for Gaussian variables. However, with nontrivial additional group action BB^* and B^*B look quite different, and the group action interferes significantly.

Using (0.8), we can prove Corollary 0.5 which will play a central role in the application to compressed sensing in the next section.

PROOF OF COROLLARY 0.5. By Jensen's inequality, we have

$$
\begin{aligned}
&\left(\mathbb{E}_\delta\left\|\frac{1}{k}\sum_{i=1}^m \delta_i x_i - 1\right\|_{L_p(\mathcal{N},\tau)}^p\right)^{1/p} \\
&\quad = \left(\mathbb{E}_\delta\left\|\frac{1}{k}\sum_{i=1}^m \delta_i x_i - \frac{1}{k}\mathbb{E}_{\delta'}\left(\sum_{i=1}^m \delta_i' x_i\right)\right\|_{L_p(\mathcal{N},\tau)}^p\right)^{1/p} \\
&\quad \le \left(\mathbb{E}_\delta\left(\mathbb{E}_{\delta'}\left\|\frac{1}{k}\sum_{i=1}^m (\delta_i - \delta_i')x_i\right\|_{L_p(\mathcal{N},\tau)}\right)^p\right)^{1/p} \\
&\quad \le \left(\mathbb{E}_{\delta,\delta'}\left\|\frac{1}{k}\sum_{i=1}^m (\delta_i - \delta_i')x_i\right\|_{L_p(\mathcal{N},\tau)}^p\right)^{1/p},
\end{aligned}
$$

where (δ_i') is a sequence of independent selectors with the same distribution as δ_i's. In order to apply Theorem 0.4, it is crucial to choose appropriate probability space. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space generated by (δ_i, δ_i'). We consider the noncommutative probability space as the algebra $\mathcal{M} = L_\infty(\mathbb{P}) \otimes \mathcal{N}$. Then we have a normalized trace $\tilde{\tau} = \mathbb{E} \otimes \tau$ on $\mathcal{M}$. We identify $\mathbb{E}$ as the conditional expectation $\mathbb{E}: \mathcal{M} \to \mathcal{N}$. Clearly, $((\delta_i - \delta_i')x_i)_{i=1}^n$ are fully independent over $\mathcal{N}$. Note that

$$
\mathbb{E}(\delta_i - \delta_i')^2 = \frac{2k}{m}\left(1 - \frac{k}{m}\right) \le \frac{2k}{m} \quad \text{and} \quad \sup_{i=1,\dots,m} |\delta_i - \delta_i'| \le 1.
$$

Since x_i is positive, $x_i^* x_i = x_i^2$. Using (0.8), we obtain

$$
\begin{aligned}
\left(\mathbb{E}\left\|\sum_{i=1}^m (\delta_i - \delta_i')x_i\right\|_{L_p(\mathcal{N},\tau)}^p\right)^{1/p} &= \left\|\sum_{i=1}^m (\delta_i - \delta_i')x_i\right\|_{L_p(\mathcal{M},\tilde{\tau})} \\
&\le C \max\left\{\sqrt{p}\left\|\sum_{i=1}^m \mathbb{E}\big((\delta_i - \delta_i')^2 x_i^2\big)\right\|_{L_{p/2}(\mathcal{N},\tau)}^{1/2}, \right. \\
&\qquad \left. p\left\|\sup_{i=1,\dots,m} |\delta_i - \delta_i'| x_i\right\|_{L_p(\mathcal{M},\tilde{\tau})}\right\}.
\end{aligned}
$$

Since $\tau(1) = 1$ and $x_i \le r$, we obtain $\| |\delta_i - \delta_i'| x_i \|_{L_p(\mathcal{M},\tilde{\tau};\ell_\infty)} \le r$, and

$$
\left\|\sum_{i=1}^m \mathbb{E}(\delta_i - \delta_i')^2 x_i^2\right\|_{L_{p/2}(\mathcal{N},\tau)} \le 2kr\left\|\frac{1}{m}\sum_{i=1}^m x_i\right\|_{L_{p/2}(\mathcal{N},\tau)} = 2kr.
$$

Therefore, we find

$$
\left(\mathbb{E}\left\|\frac{1}{k}\sum_{i=1}^m (\delta_i - \delta_i')x_i\right\|_{L_p(\mathcal{N},\tau)}^p\right)^{1/p} \le C \max\left\{\sqrt{\frac{2pr}{k}}, \frac{pr}{k}\right\}.
$$

We have completed the proof of (0.9) with constant $\sqrt{2}C$. For the "moreover" part, we use the additional norm assumption and obtain

$$\left\|\frac{1}{k}\sum_{i=1}^{m}\delta_i x_i - 1\right\|_{L_\infty(\mathrm{tr})} \le \left\|\frac{1}{k}\sum_{i=1}^{m}\delta_i x_i - 1\right\|_{L_p(\mathrm{tr})}.$$

Then by Chebyshev's inequality and (0.9) for trace $\tau(x) = \mathrm{tr}(x)/\mathrm{tr}(1)$, we have

$$\mathbb{P}\left(\left\|\frac{1}{k}\sum_{i=1}^{m}\delta_i x_i - 1\right\|_{L_\infty(\mathrm{tr})} \ge t\varepsilon\right) \le (t\varepsilon)^{-p}\mathbb{E}\left\|\frac{1}{k}\sum_{i=1}^{m}(\delta_i - \delta_i')x_i\right\|_{L_p(\mathrm{tr})}^p$$
$$\le \mathrm{tr}(1)\max\left\{\sqrt{\frac{C^2pr}{kt^2\varepsilon^2}}, \frac{Cpr}{kt\varepsilon}\right\}^p.$$

Let us first assume $t\varepsilon \le C$. Optimize the first term in p and find $p = t^2\varepsilon^2k/(C^2re)$. Recall that $k = r\varepsilon^{-2}$. Then the first term becomes $e^{-t^2/(2C^2e)}$. Using $t\varepsilon \le C$, this choice of p gives an upper bound of $e^{-t^2/(C^2e)}$ for the second term. Now assume $t\varepsilon \ge C$. The optimal choice for the second term is obtained for $p = kt\varepsilon/(Cre)$. Then the second term becomes $e^{-t/(Ce\varepsilon)}$ and, thanks to $t\varepsilon \ge C$, the first term is less than $e^{-t/(2Ce\varepsilon)}$. The additional assumption on t guarantees that $p \ge 2.5$ in both cases. Therefore,

$$\mathbb{P}\left(\left\|\frac{1}{k}\sum_{i=1}^{m}\delta_i x_i - 1\right\|_{L_\infty(\mathrm{tr})} \ge t\varepsilon\right) \le \mathrm{tr}(1)\begin{cases} e^{-t^2/(2C^2e)}, & \text{if } t\varepsilon \le C,\\ e^{-t/(2Ce\varepsilon)}, & \text{if } t\varepsilon \ge C.\end{cases}$$

The constant C is the same as the constant in the first assertion. □

REMARK 3.2. In this context it is useful to compare our different generalizations of Rosenthal's inequality. We observe that with Corollary 0.3, we can only obtain

$$\left(\mathbb{E}\left\|\frac{1}{k}\sum_{i=1}^{m}\delta_i x_i - 1\right\|_{L_p(\tau)}^p\right)^{1/p} \le C\left(\sqrt{\frac{pr^2}{k}} + \frac{pr}{k}\right),$$

and with inequality (0.7) we obtain

$$\left(\mathbb{E}\left\|\frac{1}{k}\sum_{i=1}^{m}\delta_i x_i - 1\right\|_{L_p(\tau)}^p\right)^{1/p} \le C\left(\sqrt{\frac{pr}{k}} + \frac{pr}{k^{1-1/p}}\right).$$

Both estimates are worse than inequality (0.9).

The following two examples are meant to justify the optimality of $\sqrt{p}$ and p. We refer the reader to [35] for a more detailed discussion on this topic in the framework of classical probability. We will use the standard notation for comparing orders of functions as $p \to \infty$. Recall that $f(p) = O(g(p))$ if there exists a constant C such that $f(p) \le Cg(p)$ asymptotically, $f(p) = \Omega(g(p))$ if there exists a constant c such that $f(p) \ge cg(p)$ asymptotically, $f(p) = \Theta(g(p))$ if there exist constants c and C such that $cg(p) \le f(p) \le Cg(p)$ asymptotically, and $f(p) \sim g(p)$ if $\lim_{p\to\infty} f(p)/g(p) = 1$.

EXAMPLE 3.3 (The optimality of $\sqrt{p}$ in Theorem 0.4). Let us assume that

$$\left\|\sum_{i=1}^n x_i\right\|_p \le A(p)\left(\sum_{i=1}^n \|x_i\|^2\right)^{1/2} + B(p)\left(\sum_{i=1}^n \|x_i\|^p\right)^{1/p} \tag{3.5}$$

for some functions $A(p)$ and $B(p)$. We use $x_i = g_i$. Here (g_i) is a sequence of i.i.d. normal random variables with mean 0 and variance 1. We know $\mathbb{E}|g_1|^p = \frac{2^{p/2}}{\sqrt{\pi}}\Gamma(\frac{p+1}{2})$. By Stirling's formula, we obtain for large p,

$$\|g_1\|_p \sim \sqrt{\frac{p}{e}}.$$

This yields that there exist absolute constants c and C such that $c\sqrt{p} \le \|g_1\| \le C\sqrt{p}$ for all $p \ge 2$. Hence, we obtain

$$c\sqrt{p} \le \|g_1\|_p = \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i\right\|_p \le A(p) + CB(p)\sqrt{p}n^{1/p-1/2}.$$

Sending $n \to \infty$, we have

$$A(p) \ge c\sqrt{p} \qquad \text{for } p > 2.$$

This shows that one cannot reduce the order of $A(p)$, even at the expense of increasing the order of $B(p)$.

EXAMPLE 3.4 (The optimality of p in Theorem 0.4). Following Corollary 0.5, we do a random selector on $\Omega = \{1\}$, that is, $x_i = 1$ and $\mathbb{E}\delta_i = \lambda = k/m$, and then we shall assume that

$$\left(\mathbb{E}\left|\frac{1}{k}\sum_{i=1}^m \delta_i - 1\right|^p\right)^{1/p} \le C\sqrt{\frac{p}{k}} + \frac{f(p)}{k}$$

for some function $f(p)$. Here we choose $m = p$ and $k = ap$ for some very small a. Then we find that for every $1 \le j \le m$,

$$\left|\frac{j}{k} - 1\right|\binom{m}{j}^{1/m}\lambda^{j/m}(1-\lambda)^{1-j/m} \le C\sqrt{\frac{m}{k}} + \frac{f(m)}{k}.$$

Let us first fix $j = \lceil \gamma m \rceil$ and assume that $\gamma \ge 1/4$ and $1/2^m < a \le 1/8$. This gives $\frac{j}{k} \ge \frac{\gamma}{a} \ge \frac{1}{4a} \ge 2$ and hence

$$\left|\frac{j}{k} - 1\right| \ge \frac{1}{8a}.$$

Note that $1 \le \binom{m}{j}^{1/m} \le 2$ so that we cannot expect any help here. Thus we find

$$\frac{1}{16}a^{\gamma-1}(1-a)^{1-\gamma} \le \frac{1}{8}a^{\gamma-1+1/m}(1-a)^{1-\gamma} \le Ca^{-1/2} + \frac{f(p)}{ap}.$$

Let us now fix $\gamma = 1/4$ and choose a such that

$$2Ca^{-1/2} \le \frac{1}{16}\left(\frac{1-a}{a}\right)^{3/4}$$

or equivalently,

$$32Ca^{1/4} \le (1-a)^{3/4}.$$

However, $a \le 1/8$ implies $1 - a \ge 7/8$. Thus

$$a \le \left(\frac{7}{8}\right)^3 \frac{1}{(32C)^4}$$

will do. Then we find

$$\left(a^{1/4}\frac{(7/8)^{3/4}}{32}\right)p \le f(p).$$

Choose $a = (7/8)^3/(32C)^4$. Then we have

$$\frac{c_0}{C}p \le f(p) \tag{3.6}$$

for an absolute constant $c_0 = (7/8)^{3/2}/32^2$. This shows that one cannot reduce the order of $f(p)$, as long as we keep $A(p) \le C\sqrt{p}$ in (3.5).

REMARK 3.5. In fact, Example 3.4 provides more information. Instead of fixing γ, by sending $\gamma \to 0$ and choosing $a \le \gamma/2$ appropriately, we can find a different behavior. Indeed, then we have $|j/k - 1| \ge \gamma/(2a)$ and

$$\frac{\gamma}{4}a^{\gamma-1}(1-a)^{1-\gamma} \le Ca^{-1/2} + \frac{f(p)}{ap},$$

and since $a < \gamma$ and $(1-\gamma)^{1-\gamma} \ge e^{-1}$, we need $8eCa^{-1/2} \le \gamma a^{\gamma-1}$ or

$$a^{1/2-\gamma} \le \frac{\gamma}{8eC}.$$

Note that $(\frac{\gamma}{8eC})^{2/(1-2\gamma)} \le \gamma/2$ for $\gamma \le 1$. Hence with

$$a \le \left(\frac{\gamma}{8eC}\right)^{2/(1-2\gamma)},$$

we have

$$\frac{\gamma a^{\gamma}}{8e}p \le f(p).$$

Put $a = (\frac{\gamma}{8eC})^{2/(1-2\gamma)}$. Then we obtain

$$\left(\frac{\gamma}{8eC}\right)^{1/(1-2\gamma)} Cp \le f(p).$$

Optimizing the left-hand side in γ, we obtain $2\gamma \log(8e^2C) - 2\gamma \log(\gamma) = 1$ and

$$\left(16eC \log \frac{8e^2C}{\gamma}\right)^{-1-1/\log(8eC/\gamma)} Cp \le f(p).$$

Since $\gamma \log \gamma \to 0$ as $\gamma \to 0$, we choose

$$\gamma = \frac{1}{2\log(8e^2C)}.$$

In order to obtain a lower bound for $f(p)$, we need to assume $8C \ge 1$ so that $\gamma \le 1/4$. This yields for $C \ge 1.5$,

$$f(p) \ge \frac{1}{32\sqrt{2}e^{3/2+2/e}\log(8e^2C)} p \ge \frac{p}{c_1 \log C} \tag{3.7}$$

for some absolute constant c_1. Compare (3.7) with (3.6). Estimate (3.7) is better for large C. Let us now fix p and put $C = p^\alpha$. Example 3.3 shows that α has to be nonnegative. (3.7) implies that for $\alpha > 0$,

$$f(p) \ge \frac{p}{c_1 \alpha \log p}.$$

In particular, for $C = \sqrt{p}/\log p$, we obtain $f(p) \ge 2c_1^{-1} p/\log p$, which recovers the best constants obtained in [20].

Example 3.3 and Remark 3.5 yield the following result.

THEOREM 3.6. *Under the hypotheses of Theorem* 0.4, *assume that*

$$\left\|\sum_{i=1}^n x_i\right\|_p \le A(p)\left\|\left(\sum_{j=1}^n E_{\mathcal{N}}\left(x_j x_j^* + x_j^* x_j\right)\right)^{1/2}\right\|_p + B(p)\left(\sum_{j=1}^n \|x_j\|_p^p\right)^{1/p}$$

for some functions $A(p)$ and $B(p)$. Then we have:

(i) *The best possible order of the lower bound for $A(p)$ is $\sqrt{p}$, which cannot be improved, even if the order of $B(p)$ is increased.*

(ii) *If $\Omega(p/\log p) = A(p) = O(p^\beta)$ where $\beta \ge 1$, then the best possible order of $B(p)$ is $p/\log p$.*

The point here is that the random selector model attains the worst case in the noncommutative Rosenthal inequality. In the commutative case, (i) was proved by Pinelis and Utev in [35]. Later, Pinelis proved much stronger results which give different combinations of best constants in the martingale version of Rosenthal inequality in the context of Banach spaces. We refer the interested reader to [34] for more details. We thank Pinelis for pointing this out to us.

4. Illustration in compressed sensing. At the time of this writing there is a large body of work relating tools originating from noncommutative probability to estimates from compressed sensing; see [31, 47] for more details. Since our improvement of the Rosenthal inequality was motivated by problems in compressed sensing, we want to describe this relation toward compressed sensing. Let us briefly recall the background here following [6, 8, 43]. We want to reconstruct an unknown signal $f \in \mathbb{C}^n$ from linear measurements $\Phi f \in \mathbb{C}^k$, where Φ is some known $k \times n$ matrix called the measurement matrix. The reconstruction problem is stated as

$$\min \|f^*\|_0 \quad \text{subject to} \quad \Phi f^* = \Phi f, \tag{4.1}$$

where $\|f\|_0 = |\mathrm{supp}\, f|$ is the number of nonzero element of f. Since this problem is computationally expensive, we consider its convex relaxation instead.

$$\min \|f^*\|_1 \quad \text{subject to} \quad \Phi f^* = \Phi f, \tag{4.2}$$

where $\|f\|_p = (\sum_{j=1}^n |f_j|^p)^{1/p}$ denotes ℓ_p norm throughout this section. Exact reconstruction means that the solutions to (4.1) and (4.2) are both equal to f. f is assumed to be s-sparse, that is, $|\mathrm{supp}\, f| \le s$. We refer to [6, 43] for why (4.2) is a good substitute of (4.1). However, the restricted isometry property (RIP) on Φ is an extremely important tool for exact reconstruction due to Candes and Tao [7]; see also [5]. Let Φ_T denote the $k \times |T|$ matrix consisting of the columns of Φ indexed by T. The RIP constant Δ_s is defined to be the smallest positive number such that the inequality

$$C(1-\Delta_s)\|x\|_2^2 \le \|\Phi_T x\|_2^2 \le C(1+\Delta_s)\|x\|_2^2$$

holds for some number $C > 0$ and for all $x \in \ell_2$ and all subsets $T \subset \{1, \dots, n\}$ of size $|T| \le s$. Candes and Tao proved the following theorem [5, 7]:

THEOREM 4.1. *Let f be an s-sparse signal and Φ be a measurement matrix whose RIP constant satisfies*

$$\Delta_{3s} + 3\Delta_{4s} \le 2.$$

Then f can be recovered exactly.

Since Δ_s is nondecreasing in s, in order to verify RIP, it suffices to show that

$$\Delta_{4s} \le \tfrac{1}{2}$$

or simply $\Delta_s \le \frac{1}{2}$ by adjusting constant if necessary. In this section, we apply Corollary 0.5 to study the problem of reconstruction from Fourier measurements. Two cases will be considered. In the first case, we fix the support T of f. In the second case we allow it to vary. In the following, C will always denote the constant in Corollary 0.5, and $\mathbb{C}^m$ will always denote the m-dimensional complex Euclidean space equipped with ℓ_2 norm.

EXAMPLE 4.2 (Fourier measurements). We consider the discrete Fourier transform $\hat{f} = \Psi f$ where Ψ is a matrix with entries

$$\Psi_{\omega,t} = \frac{1}{\sqrt{n}} e^{-i2\pi\omega t/n}, \qquad \omega, t \in \{0, \dots, n-1\}.$$

We want to reconstruct an s-sparse signal $f \in \mathbb{C}^n$ from linear measurements $\Phi f \in \mathbb{C}^{\Omega}$, where $\Omega \subset \{0, \dots, n-1\}$ is a uniformly random subset with average cardinality k and the measurement matrix Φ is a submatrix of Ψ consisting of random rows with indices in Ω. This is the Fourier measurement matrix considered in [6, 8, 43]. We can formulate this random subset precisely using the Bernoulli model. Let $(\delta_i)_{i=0}^{n-1}$ be a sequence of independent selectors with $\mathbb{E}\delta_i = k/n$, for $i = 0, \dots, n-1$. Then

$$\Omega = \{j : \delta_j = 1\}$$

and $k = \mathbb{E}|\Omega|$.

Let y_i be the ith row of Ψ and T the support of f. Write y_i^T for the restriction of y_i on the coordinate in the set T. For $x, y, z \in \mathbb{C}^n$, we define the tensor $x \otimes y$ as the rank-one linear operator given by $(x \otimes y)(z) = \langle x, z \rangle y$. Then

$$\Phi^*\Phi = \sum_{i\in\Omega} y_i^T \otimes y_i^T = \sum_{i=0}^{n-1} \delta_i y_i^T \otimes y_i^T.$$

Let $x_j = n y_j^T \otimes y_j^T$. Then

$$\frac{1}{n}\sum_{i=0}^{n-1} x_i = id_{\mathbb{C}^T} = I_T \quad \text{and} \quad \|x_j\| = n\|y_j^T \otimes y_j^T\| = n\|y_j^T\|_2^2 \le s.$$

The next proposition follows easily from Corollary 0.5.

PROPOSITION 4.3. *Assume that the average cardinality of a random set Ω is $k = \varepsilon^{-2}s$. Then for $t\varepsilon \le C$,*

$$\mathbb{P}\left(\left\|\frac{n}{k}\sum_{i=0}^{n-1} \delta_i y_i^T \otimes y_i^T - id_{\mathbb{C}^T}\right\| \ge t\varepsilon\right) \le s e^{-t^2/(2C^2 e)}, \tag{4.3}$$

where $\|\cdot\|$ is the operator norm.

Define

$$H = id_{\mathbb{C}^T} - \frac{n}{|\Omega|}\sum_{i=0}^{n-1} \delta_i y_i^T \otimes y_i^T.$$

Then $\Phi^*\Phi = \frac{|\Omega|}{n}(I_T - H)$. By the classical Bernstein inequality, $k/2 \le |\Omega| \le 3k/2$ with high probability; see [8], Lemma 6.6. Therefore, by choosing $t\varepsilon < 1$, we find that the matrix $I_T - H$ is invertible with high probability. The precise meaning of "high probability" will become clear in a moment. This proposition is an analog of [6], Theorem 3.1, and [43], Theorem 3.3, with a *single* set T. We compare our results with previous results in the following remark. It is easy to show that $\mathbb{P}(k/2 \le |\Omega| \le 3k/2)$ given by Bernstein's inequality dominates $1 - se^{-t^2/(2C^2e)}$ for the value of k given below. Hence we only need to consider (4.3) for the probability of success.

REMARK 4.4. (i) For a single set T our result is more general than previous results on the invertibility of $\Phi^*\Phi$ obtained by Candes, Romberg and Tao in the breakthrough paper [6]. In particular, if we put $t\varepsilon = 1/2$ and $\varepsilon^{-2} = 8C^2e(M\log n + \log s)$ for some $M > 0$, then we obtain $k = c_M s \log n$ for some constant c_M, and $I_T - H$ is invertible with probability at least $1 - O(n^{-M})$. This gives [6], Theorem 3.1. Together with [6], Lemma 2.3, or following verbatim the end of the proof of Theorem 4.2 ([41], Section 7.3), we recover the main results of [6].

(ii) Allowing arbitrary choices of k and p, we recover [41], Theorem 7.3, and we would like to thank H. Rauhut for bringing this to our attention. His proof requires considerably more technology. Both proofs are based on the optimal constant in the noncommutative Khintchine inequality (used in Rudelson's lemma) which was discovered independently by the first named author and Pisier; see [37] for more historic comments. We believe that our proof is more direct. Moreover, Rauhut established the exact reconstruction results based on his version of (4.3) cited above, which shows that an estimate like (4.3) is the key to the exact reconstruction problem.

We now investigate the case with multiple choices of T. First, it is clear that (4.3) remains valid for polynomially many sets T. In general, we have

$$\mathbb{P}\left(\sup_{|T|\le s}\left\|\frac{n}{k}\sum_{i=0}^{n-1}\delta_i y_i^T \otimes y_i^T - id_{\mathbb{C}^T}\right\| \ge t\varepsilon\right) \le |S|se^{-t^2/(2C^2e)}, \tag{4.4}$$

where $|S|$ denotes the number of set T with $|T| \le s$. Note that

$$\Delta_s = \inf_{\alpha>0}\sup_{|T|\le s}\left\|\alpha\sum_{i\in\Omega} y_i^T \otimes y_i^T - id_{\mathbb{C}^T}\right\|.$$

It follows that

$$\mathbb{P}(\Delta_s \ge t\varepsilon) \le \mathbb{P}\left(\sup_{|T|\le s}\left\|\frac{n}{k}\sum_{i=0}^{n-1}\delta_i y_i^T \otimes y_i^T - id_{\mathbb{C}^T}\right\| \ge t\varepsilon\right).$$

Assume $s \le n/2$. Since $|S| \le s\binom{n}{s} + 1 \le s(ne/s)^s$, if

$$2\log s + s\log\frac{ne}{s} < \frac{t^2}{2C^2e}, \tag{4.5}$$

then with probability at least $1 - s^2(ne/s)^s e^{-t^2/(2C^2e)}$, we can recover *all* s-sparse signal f from its Fourier measurements Φf. From here we are able to obtain different bounds for k and the corresponding probabilities of success. As an illustration, we have the following result.

PROPOSITION 4.5. *Assume $s \le n/2$. Let $M > 0$ be a precision constant and n be a large integer such that*

$$2\log s + s\log\frac{ne}{s} < (M+1)s\log\frac{n}{s}.$$

Then a random subset Ω of average cardinality

$$k = 8C^2e(M+1)s^2\log\frac{n}{s} = c_M s^2\log\frac{n}{s} \tag{4.6}$$

satisfies RIP with probability at least $1 - s^2e^s(n/s)^{-Ms}$.

PROOF. Put $t\varepsilon = 1/2$ in (4.4). Since $k = s\varepsilon^{-2}$, we obtain $t^2 = 2e(M+1)s\log(n/s)$. Thanks to the assumption on n, (4.5) is true. Then

$$\mathbb{P}\left(\Delta_s \ge \frac{1}{2}\right) \le s^2e^s\left(\frac{n}{s}\right)^{-Ms}.$$

We have proved the assertion. $\square$

REMARK 4.6. We can relax the bound for k a little to obtain polynomial probability of success. Indeed, the same argument as Proposition 4.5 yields that a random subset Ω of average cardinality

$$k = 8C^2e(M+1)s^2\log n = c_M s^2\log n \tag{4.7}$$

satisfies RIP with probability $1 - s^{2-s}e^s n^{-Ms}$.

The good aspect of Proposition 4.5 is that k is linear in $\log n$. Unfortunately, this is weaker than Rudelson and Vershynin's results in [43] $k = O(s\log n\log(s\log n)\log^2 s)$ for fixed probability $1-\varepsilon$ of success, which was strengthened to super-polynomially probability of success by Rauhut following their ideas; see [41]. These results are obtained by using deep Banach spaces techniques. We added our results just for comparison. Of course, simple applications of Khintchine's inequality are not expected to replace either majorizing measure techniques or the iterative methods of [43] for the uniform estimates required for RIP. It seems known in the compressed sensing community that the tails bounds alone are not good enough. To conclude this section, we restate a conjecture on the best bound of k; see [43] (and [41] for further background).

CONJECTURE 4.7. *A random subset $\Omega \subset \{0, 1, \dots, n-1\}$ of average cardinality $k = O(s\log n)$ satisfies RIP with high probability.*

Acknowledgments. We would like to thank W. B. Johnson for bringing [35] to our attention. After our work was completed, we learned from S. Dirksen that he also essentially obtained (0.7) in his Ph.D. thesis [13] using a different method in the UIUC analysis seminar on November 3, 2011. We thank him for helpful conversations. We are also grateful to the warm response from compressed sensing community. Especially, we thank K. Lee, H. Rauhut and J. A. Tropp for their detailed comments on the compressed sensing part of our paper, whose opinions on credits and earlier results have been incorporated in the current version.

We thank the anonymous referee for suggestions on improving the exposition and for bringing Prohorov's inequality to our attention.

Right before this paper is in print, S. Dirksen pointed out that (0.7) can also be deduced from [14], Theorem 6.3. Besides, there is new development on the RIP constant, for which we refer the interested reader to [26].

REFERENCES

[1] AHLSWEDE, R. and WINTER, A. (2002). Strong converse for identification via quantum channels. *IEEE Trans. Inform. Theory* **48** 569–579. MR1889969

[2] ARAKI, H. (1973). Golden–Thompson and Peierls–Bogolubov inequalities for a general von Neumann algebra. *Comm. Math. Phys.* **34** 167–178. MR0341114

[3] BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

[4] BURKHOLDER, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probab.* **1** 19–42. MR0365692

[5] CANDES, E., RUDELSON, M., TAO, T. and VERSHYNIN, R. (2005). Error correction via linear programming. In 46*th Annual IEEE Symposium on Foundations of Computer Science*, 2005. *FOCS* 2005 668–681.

[6] CANDÈS, E. J., ROMBERG, J. and TAO, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory* **52** 489–509. MR2236170

[7] CANDES, E. J. and TAO, T. (2005). Decoding by linear programming. *IEEE Trans. Inform. Theory* **51** 4203–4215. MR2243152

[8] CANDES, E. J. and TAO, T. (2006). Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory* **52** 5406–5425. MR2300700

[9] CHERIX, P.-A., COWLING, M., JOLISSAINT, P., JULG, P. and VALETTE, A. (2001). *Groups with the Haagerup Property: Gromov's a-T-Menability. Progress in Mathematics* **197**. Birkhäuser, Basel. MR1852148

[10] COLLINS, B. and JUNGE, M. (2011). What is a noncommutative Brownian motion? Preprint.

[11] CONWAY, J. B. (1990). *A Course in Functional Analysis*, 2nd ed. *Graduate Texts in Mathematics* **96**. Springer, New York. MR1070713

[12] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics (New York)* **38**. Springer, New York. MR1619036

[13] DIRKSEN, S. (2011). Noncommutative and vector-valued Rosenthal inequalities. Ph.D. thesis, Delft Univ. Technology. Available at http://repository.tudelft.nl/view/ir/uuid:466dbf51-4482-4421-b837-36fdda2e9df4/.

[14] DIRKSEN, S., DE PAGTER, B., POTAPOV, D. and SUKOCHEV, F. (2011). Rosenthal inequalities in noncommutative symmetric spaces. *J. Funct. Anal.* **261** 2890–2925. MR2832586

[15] FACK, T. and KOSAKI, H. (1986). Generalized s-numbers of τ-measurable operators. *Pacific J. Math.* **123** 269–300. MR0840845
[16] GROSS, D. (2011). Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theory* **57** 1548–1566. MR2815834
[17] GUŢĂ, M. and MAASSEN, H. (2002). Generalised Brownian motion and second quantisation. *J. Funct. Anal.* **191** 241–275. MR1911186
[18] GUŢĂ, M. and MAASSEN, H. (2002). Symmetric Hilbert spaces arising from species of structures. *Math. Z.* **239** 477–513. MR1893849
[19] HITCZENKO, P. (1990). Best constants in martingale version of Rosenthal's inequality. *Ann. Probab.* **18** 1656–1668. MR1071816
[20] JOHNSON, W. B., SCHECHTMAN, G. and ZINN, J. (1985). Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Probab.* **13** 234–253. MR0770640
[21] JUNGE, M. (2002). Doob's inequality for non-commutative martingales. *J. Reine Angew. Math.* **549** 149–190. MR1916654
[22] JUNGE, M. (2006). Operator spaces and Araki–Woods factors: A quantum probabilistic approach. *IMRP Int. Math. Res. Pap.* Art. ID 76978, 87. MR2268491
[23] JUNGE, M. and XU, Q. (2003). Noncommutative Burkholder/Rosenthal inequalities. *Ann. Probab.* **31** 948–995. MR1964955
[24] JUNGE, M. and XU, Q. (2005). On the best constants in some non-commutative martingale inequalities. *Bull. Lond. Math. Soc.* **37** 243–253. MR2119024
[25] JUNGE, M. and XU, Q. (2008). Noncommutative Burkholder/Rosenthal inequalities. II. Applications. *Israel J. Math.* **167** 227–282. MR2448025
[26] KRAHMER, F., MENDELSON, S. and RAUHUT, H. (2012). Suprema of chaos processes and the restricted isometry property. Available at arXiv:1207.0235.
[27] KWAPIEŃ, S. and WOYCZYŃSKI, W. A. (1989). Tangent sequences of random variables: Basic inequalities and their applications. In *Almost Everywhere Convergence* (*Columbus*, *OH*, 1988) 237–265. Academic Press, Boston, MA. MR1035249
[28] LIEB, E. H. (1973). Convex trace functions and the Wigner–Yanase–Dyson conjecture. *Adv. Math.* **11** 267–288. MR0332080
[29] LUST-PIQUARD, F. (1986). Inégalités de Khintchine dans C_p $(1 < p < \infty)$. *C. R. Acad. Sci. Paris Sér. I Math.* **303** 289–292. MR0859804
[30] LUST-PIQUARD, F. and PISIER, G. (1991). Noncommutative Khintchine and Paley inequalities. *Ark. Mat.* **29** 241–260. MR1150376
[31] MACKEY, L., JORDAN, M. I., CHEN, R. Y., FARRELL, B. and TROPP, J. A. (2012). Matrix concentration inequalities via the method of exchangeable pairs. Available at arXiv:1201.6002.
[32] NAGAEV, S. V. and PINELIS, I. F. (1978). Some inequalities for the distribution of sums of independent random variables. *Theory Probab. Appl.* **22** 248–256.
[33] OLIVEIRA, R. I. (2009). Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. Available at arXiv:0911.0600.
[34] PINELIS, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.* **22** 1679–1706. MR1331198
[35] PINELIS, I. F. and UTEV, S. A. (1985). Estimates of moments of sums of independent random variables. *Theory Probab. Appl.* **29** 574–577.
[36] PISIER, G. (1998). Non-commutative vector valued L_p-spaces and completely p-summing maps. *Astérisque* **247** vi+131. MR1648908
[37] PISIER, G. (2003). *Introduction to Operator Space Theory*. *London Mathematical Society Lecture Note Series* **294**. Cambridge Univ. Press, Cambridge. MR2006539
[38] PISIER, G. and XU, Q. (1997). Non-commutative martingale inequalities. *Comm. Math. Phys.* **189** 667–698. MR1482934

[39] PROHOROV, Y. V. (1959). An extremal problem in probability theory. *Theory Probab. Appl.* **4** 201–203.

[40] RANDRIANANTOANINA, N. (2007). Conditioned square functions for noncommutative martingales. *Ann. Probab.* **35** 1039–1070. MR2319715

[41] RAUHUT, H. (2010). Compressive sensing and structured random matrices. In *Theoretical Foundations and Numerical Methods for Sparse Recovery*. *Radon Ser. Comput. Appl. Math.* **9** 1–92. de Gruyter, Berlin. MR2731597

[42] ROSENTHAL, H. P. (1970). On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables. *Israel J. Math.* **8** 273–303. MR0271721

[43] RUDELSON, M. and VERSHYNIN, R. (2008). On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.* **61** 1025–1045. MR2417886

[44] RUSKAI, M. B. (1972). Inequalities for traces on von Neumann algebras. *Comm. Math. Phys.* **26** 280–289. MR0312284

[45] SIMON, B. (2005). *Trace Ideals and Their Applications*, 2nd ed. *Mathematical Surveys and Monographs* **120**. Amer. Math. Soc., Providence, RI. MR2154153

[46] TEMME, N. M. (1996). *Special Functions*: *An Introduction to the Classical Functions of Mathematical Physics*. Wiley, New York. MR1376370

[47] TROPP, J. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12** 389–434.

[48] VOICULESCU, D. V., DYKEMA, K. J. and NICA, A. (1992). *Free Random Variables*: *A Noncommutative Probability Approach to Free Products with Applications to Random Matrices*, *Operator Algebras and Harmonic Analysis on Free Groups*. *CRM Monograph Series* **1**. Amer. Math. Soc., Providence, RI. MR1217253

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS 61801
USA
E-MAIL: junge@math.uiuc.edu
zeng8@illinois.edu