A note on the improved sparse Hanson-Wright inequalities

Guozheng Dai[‡], Yiyun He[*], Ke Wang[†] and Yizhe Zhu[§]

Abstract

In this paper, we establish sparse Hanson-Wright inequalities for quadratic forms of sparse α-sub-exponential random vectors where $\alpha \in (0, 2]$. When only considering the regime $0 < \alpha \leq 1$, we derive a sharper sparse Hanson-Wright inequality that achieves optimality in certain special cases. These results generalize some classical Hanson-Wright inequalities without sparse structure.

1 Introduction

Let $\xi = (\xi_1, \cdots, \xi_n)^\top$ be a random vector with independent centered entries and $A = (a_{ij})_{n\times n}$ be a fixed matrix. Exploring the concentration properties of the following quadratic random variables

$$S_A(\xi) := \xi^\top A\xi = \sum_{i,j} a_{ij}\xi_i\xi_j$$

is a classic topic in probability. A well-known result, proved by Hanson and Wright [7], claims that if ξ_i are independent centered 2-sub-exponential (sub-gaussian) random variables such that $\max_i \|\xi_i\|_{\Psi_2} \leq L$, then for $t \geq 0$ (the following version was provided in [15])

$$\mathbb{P}\Big\{|S_A(\xi) - \mathbb{E}S_A(\xi)| \geq t\Big\} \leq 2\exp\Big(-c\min\Big\{\frac{t^2}{L^4\|A\|_F^2}, \frac{t}{L^2\|A\|_{l_2\to l_2}}\Big\}\Big), \tag{1.1}$$

where $\|\cdot\|_F$ and $\|\cdot\|_{l_2\to l_2}$ are the Frobenius norm and the spectral norm of a matrix respectively. Recall that a random variable η_1 is α-sub-exponential if satisfying

$$\mathbb{P}\{|\eta_1| \geq t\} \leq 2\exp\big(-\frac{t^\alpha}{K^\alpha}\big), \quad t \geq 0.$$

[‡]Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, guozhengdai1997@gmail.com.

[*]Department of Mathematics, University of California Irvine, Irvine, CA 92697, USA, yiyunh4@uci.edu.

[†]Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, kewang@ust.hk.

[§]Department of Mathematics, University of Southern California, 3620 Vermont Avenue Los Angeles, CA 90089, USA, yizhezhu@usc.edu.

Its α-sub-exponential norm is defined as

$$\|\eta_1\|_{\Psi_\alpha} := \inf\big\{t > 0 : \mathbb{E}\exp\big(\frac{|\eta_1|^\alpha}{t^\alpha}\big) \le 2\big\}.$$

There is a lot of work dedicated to extending the classical Hanson-Wright inequality (1.1) to more general cases. For example, assume that $\xi_1, \cdots, \xi_n$ are independent centered α-sub-exponential variables and satisfy $\max_i \|\xi_i\|_{\Psi_\alpha} \le L(\alpha)$. When $1 \le \alpha \le 2$, Adamczak and Latała [1] proved for $t \ge 0$ (see also [2])

$$\mathbb{P}\{|S_A(\xi) - \mathbb{E}S_A(\xi)| \ge L(\alpha)^2 t\} \le 2\exp(-c(\alpha) f_1(t)), \tag{1.2}$$

where

$$\begin{aligned} f_1(t) = \min\Big\{\Big(\frac{t}{\|A\|_F}\Big)^2, \frac{t}{\|A\|_{l_2\to l_2}}, \Big(\frac{t}{\|A\|_{l_{\alpha^*}(l_2)}}\Big)^\alpha, \\ \Big(\frac{t}{\|A\|_{l_2\to l_{\alpha^*}}}\Big)^{\frac{2\alpha}{2+\alpha}}, \Big(\frac{t}{\|A\|_{l_\alpha\to l_{\alpha^*}}}\Big)^{\frac{\alpha}{2}}\Big\}. \end{aligned} \tag{1.3}$$

One can refer to (1.9) below for the definitions of the matrix's norms and $\alpha^* = \alpha/(\alpha-1)$. When $0 < \alpha \le 1$, Kolesko and Latała [11] proved for $t \ge 0$ (see also [6])

$$\mathbb{P}\{|S_A(\xi) - \mathbb{E}S_A(\xi)| \ge L(\alpha)^2 t\} \le 2\exp(-c(\alpha) f_2(t)), \tag{1.4}$$

where

$$f_2(t) = \min\Big\{\Big(\frac{t}{\|A\|_F}\Big)^2, \frac{t}{\|A\|_{l_2\to l_2}}, \Big(\frac{t}{\|A\|_{l_2\to l_\infty}}\Big)^{\frac{2\alpha}{2+\alpha}}, \Big(\frac{t}{\|A\|_\infty}\Big)^{\frac{\alpha}{2}}\Big\}. \tag{1.5}$$

As we have seen above, the Hanson-Wright inequalities (1.2) and (1.4) are quite intricate. By evaluating the family of norms used therein, Sambale [16] obtained a simplified version. This simplified version is not only easily calculable but also sufficient for many applications (see [16] for details). In particular, Sambale proved for $0 < \alpha \le 2$ and $t \ge 0$

$$\begin{aligned} &\mathbb{P}\big\{|S_A(\xi) - \mathbb{E}S_A(\xi)| \ge t\big\} \\ \le& 2\exp\Big(-c(\alpha)\min\Big\{\frac{t^2}{L(\alpha)^4\|A\|_F^2}, \Big(\frac{t}{L(\alpha)^2\|A\|_{l_2\to l_2}}\Big)^{\alpha/2}\Big\}\Big). \end{aligned} \tag{1.6}$$

Motivated by the covariance estimation problem in the matrix variate model, Zhou [20] showed a sparse Hanson-Wright type inequality for sub-gaussian random variables. In particular, assume that $\{\xi_i = \delta_i \cdot \zeta_i, 1 \le i \le n\}$ is a sequence of independent random variables, where $\delta_i \sim \text{Bernoulli}(p_i)$ (that is, taking values $0, 1$ with probability $1 - p_i$ and p_i respectively) and ζ_i is a centered sub-gaussian variable independent of δ_i. Zhou proved for $t \ge 0$

$$\mathbb{P}\big\{|S_A(\xi) - \mathbb{E}S_A(\xi)| \ge t\big\} \le 2\exp\Big(-c\min\Big\{\frac{t^2}{L^4\gamma_1(A)}, \frac{t}{L^2\|A\|_{l_2\to l_2}}\Big\}\Big), \tag{1.7}$$

where $L = \max_i \|\zeta_i\|_{\Psi_2}$ and $\gamma_1(A) = \sum_k a_{kk}^2 p_k + \sum_{i\ne j} a_{ij}^2 p_i p_j$.

Recently, He, Wang and Zhu [8] established sparse Hanson-Wright type inequalities in a more general setting. Specifically, let $\{\xi_i = \delta_i \cdot \zeta_i, 1 \le i \le n\}$ be a sequence of independent random variables, where $\delta_i \sim$ Bernoulli(p_i) and ζ_i is a centered α-sub-exponential variable independent of δ_i, $1 \le i \le n$. [8, Corollary 1] showed that for $\alpha > 0$ and $t \ge 0$

$$\begin{aligned}&\mathbb{P}\big\{|S_A(\xi) - \mathbb{E}S_A(\xi)| \ge t\big\}\\ \le& 2\exp\Big(-c(\alpha)\min\Big\{\frac{t^2}{L(\alpha)^4\gamma_1(A)}, \frac{t}{L(\alpha)^2\gamma_2(A)}, \Big(\frac{t}{L(\alpha)^2\|A\|_\infty}\Big)^{\min\{\frac{\alpha}{2},\frac{1}{2}\}}\Big\}\Big),\end{aligned} \tag{1.8}$$

where $L(\alpha) = \max_i \|\zeta_i\|_{\Psi_\alpha}$, $\|A\|_\infty = \max_{ij}|a_{ij}|$, and

$$\gamma_2(A) = \max_i \Big\{\sum_{j:j\neq i}|a_{ij}|p_j, \sum_{j:j\neq i}|a_{ji}|p_j, |a_{ii}|\Big\}.$$

The work in [8] further extends to sparse versions of Bernstein and Bennett type inequalities.

In addition to the works mentioned above, there are numerous other papers that delve into the Hanson-Wright inequality with sparse structure. We refer interested readers to [3, 5, 9, 14, 17, 18, 19, 21] and the references therein for further exploration.

Notation Given a fixed vector $x = (x_1, \cdots, x_n)^\top \in \mathbb{R}^n$, denote by $\|x\|_r = (\sum_i |x_i|^r)^{1/r}$ the l_r norm. Use $\|\xi\|_{L_r} = (\mathbb{E}|\xi|^r)^{1/r}$ as the L_r norm of a random variable ξ. Let $A = (a_{ij})$ be an $m \times n$ matrix. We use the following notations of the matrix norms:

$$\begin{aligned}\|A\|_F =& \sqrt{\sum_{i,j}|a_{ij}|^2}, \quad \|A\|_\infty = \max_{i,j}|a_{ij}|,\\ \|A\|_{l_r(l_2)} =& (\sum_{i\le m}(\sum_{j\le n}|a_{ij}|^2)^{r/2})^{1/r},\\ \|A\|_{l_{r_1}\to l_{r_2}} =& \sup\{|\sum_{i,j}a_{ij}x_jy_i| : \|x\|_{r_1} \le 1, \|y\|_{r_2^*} \le 1\},\end{aligned} \tag{1.9}$$

where $1 \le r_1, r_2 < \infty$ and $r_2^* = r_2/(r_2 - 1)$. Indeed, $\|A\|_{l_2\to l_2}$ is the spectral norm of A and $\|A\|_{l_1\to l_\infty} = \|A\|_\infty$.

Let $\xi_1, \cdots, \xi_n$ be a sequence of random variables and $f : \mathbb{R}^n \to \mathbb{R}$ be a measurable function. Then, $\mathbb{E}_{\xi_{i_1},\cdots,\xi_{i_s}} f(\xi_1, \cdots, \xi_n)$ means only taking expectation with respect to random variables $\xi_{i_1}, \cdots, \xi_{i_s}$, where $\{i_1, \cdots, i_s\} \subset \{1, \cdots, n\}$. Denote $\xi \sim \mathcal{W}_s(\alpha)$ when ξ is a symmetric Weibull variable with the scale parameter 1 and the shape parameter α. Specifically, $-\log\mathbb{P}\{|\xi| > x\} = x^\alpha, x \ge 0$.

Unless otherwise stated, denote by $C, C_1, c, c_1, \cdots$ the universal constants independent of any parameters and the dimension n. Besides, let $C(\delta), c(\delta)\cdots$ be the constants depending only on the parameter δ. Their values can change from line to line. For convenience, we say $f \lesssim g$ if $f \le Cg$ for some universal constant C. We write $f \lesssim_\delta g$ if $f \le C(\delta)g$ for some constant $C(\delta)$. Besides, we say $f \asymp g$ if $f \lesssim g$ and $g \lesssim f$, so does $f \asymp_\delta g$.

Organization of the paper The rest of the paper is organized as follows. In the remaining part of Section 1, we will first present our main results. Then, we will draw comparisons between our results and existing ones to highlight the novelties and improvements. To conclude Section 1, we will pose two open questions that warrant further exploration. Section 2 is dedicated to introducing several key lemmas. These lemmas cover aspects such as the relationship between tails and moments, the decoupling inequality, the contraction principle, and concentration inequalities. They serve as fundamental tools for the subsequent theoretical derivations. In Section 4, we shall prove our main results, and in Appendix A, we offer proofs for Proposition 1.1, Corollary 1, and Lemma 2.8.

1.1 Main results

This paper concentrates on the sparse Hanson-Wright inequalities in α-sub-exponential random variables, $0 < \alpha \leq 2$. Our first main result in this regime reads as follows:

THEOREM 1. *Assume that $\{\delta_i, 1 \leq i \leq n\}$ and $\{\zeta_i, 1 \leq i \leq n\}$ are two independent sequences of random variables, where $\delta_i \sim \mathrm{Bernoulli}(p_i)$ are independent and ζ_i are independent centered α-sub-exponential random variables. Consider a random vector $\xi = (\xi_1, \cdots, \xi_n)$ with $\xi_i = \delta_i \cdot \zeta_i$. Let $A = (a_{ij})_{n\times n}$ be a symmetric fixed matrix. Then for $0 < \alpha \leq 2$ and $t \geq 0$*

$$\begin{aligned}&\mathbb{P}\{|\xi^\top A\xi - \mathbb{E}\xi^\top A\xi| \geq L(\alpha)^2 t\}\\ \leq &2\exp\left(-c(\alpha)\min\left\{\frac{t^2}{\sum_k a_{kk}^2 p_k + \sum_{i\neq j} a_{ij}^2 p_i p_j}, \left(\frac{t}{\|A\|_{l_2\to l_2}}\right)^{\alpha/2}\right\}\right),\end{aligned}$$

where $L(\alpha) = \max_i \|\zeta_i\|_{\Psi_\alpha}$ and $c(\alpha)$ is a positive constant depending only on α.

REMARK 1.1 (Comparison of Theorem 1 with Hanson-Wright inequalities in [16, 20])**.** *(i). When $p_1 = \cdots = p_n = 1$, i.e., $\delta_i = 1$ for $1 \leq i \leq n$, Theorem 1 recovers Sambale's result [16], see* (1.6).

(ii). When $\alpha = 2$, Theorem 1 recovers the sparse Hanson-Wright inequality for sparse sub-Gaussian random vectors obtained by Zhou in [20], see (1.7).

Moreover, when only considering the case $0 < \alpha \leq 1$, we have the following more refined result.

THEOREM 2. *In the same setting of Theorem 1, we have for $0 < \alpha \leq 1$ and $t \geq 0$*

$$\mathbb{P}\{|\xi^\top A\xi - \mathbb{E}\xi^\top A\xi| \geq L(\alpha)^2 t\} \leq 2\exp(-c(\alpha)f(t)),$$

where $L(\alpha) = \max_i \|\zeta_i\|_{\Psi_\alpha}$, $c(\alpha)$ is a positive constant depending only on α, and

$$\begin{aligned}f(t) = \min\Big\{&\frac{t^2}{\sum_k a_{kk}^2 p_k + \sum_{i\neq j} a_{ij}^2 p_i p_j}, \frac{t}{\|(a_{ij}\sqrt{p_i p_j})_{n\times n}\|_{l_2\to l_2}},\\ &\Big(\frac{t}{\max_i\big(\sum_j a_{ij}^2 p_j\big)^{1/2}}\Big)^{\frac{2\alpha}{2+\alpha}}, \Big(\frac{t}{\|A\|_\infty}\Big)^{\frac{\alpha}{2}}\Big\}.\end{aligned}$$

REMARK 1.2 (Theorem 2 improves Theorem 1)**.** *Theorem 1 for the case $0 < \alpha \le 1$ can be derived from Theorem 2. Indeed, a direct integration of the tail probability in Theorem 2 yields that*

$$\begin{aligned}\left\|\xi^\top A\xi - \mathbb{E}\xi^\top A\xi\right\|_{L_r} \lesssim_\alpha & r^{2/\alpha}\|A\|_\infty + r^{1/2+1/\alpha}\max_i\Big(\sum_j a_{ij}^2 p_j\Big)^{1/2} \\ & + r\|(a_{ij}\sqrt{p_i p_j})\|_{l_2\to l_2} + \sqrt{r}\Big(\sum_k a_{kk}^2 p_k + \sum_{i\neq j} a_{ij}^2 p_i p_j\Big)^{1/2} \\ \lesssim_\alpha & r^{2/\alpha}\|A\|_{l_2\to l_2} + \sqrt{r}\Big(\sum_k a_{kk}^2 p_k + \sum_{i\neq j} a_{ij}^2 p_i p_j\Big)^{1/2},\end{aligned}$$

where we use the following fact

$$\begin{aligned}&\max_i\Big(\sum_j a_{ij}^2 p_j\Big)^{1/2} \le \|A\|_{l_2\to l_\infty} \le \|A\|_{l_2\to l_2}, \\ &\|(a_{ij}\sqrt{p_i p_j})\|_{l_2\to l_2} \le \|A\|_{l_2\to l_2}, \quad \|A\|_\infty \le \|A\|_{l_2\to l_2}.\end{aligned}$$

Then, one can obtain Theorem 1 by Lemma 2.1 below.

REMARK 1.3 (Comparison of Theorem 2 with Hanson-Wright inequalities in [8, 11])**.** *(i). When $p_1 = \cdots = p_2 = 1$, Theorem 2 recovers the Hanson-Wright inequality* (1.4) *due to Kolesko and Latała in [11].*

(ii). Note that for $r \ge 1$,

$$\begin{aligned}r^{\frac{1}{2}+\frac{1}{\alpha}}\max_i(\sum_j a_{ij}^2 p_j)^{1/2} &\le (r^{1/\alpha}\|A\|_\infty^{1/2})\cdot(r^{1/2}\max_i(\sum_j |a_{ij}|p_j)^{1/2}) \\ &\le \frac{1}{2}r^{2/\alpha}\|A\|_\infty + \frac{1}{2}r\max_i\sum_j |a_{ij}|p_j\end{aligned}$$

and

$$\|(a_{ij}\sqrt{p_i p_j})\|_{l_2\to l_2} \le \max_i\sum_j |a_{ij}|\sqrt{p_i p_j} \le K\max_i\sum_j |a_{ij}|p_j,$$

where $K = K(p_1,\cdots,p_n) = \max p_i/\min p_j$. Hence, Theorem 2 yields that

$$\begin{aligned}\left\|\xi^\top A\xi - \mathbb{E}\xi^\top A\xi\right\|_{L_r} \lesssim_\alpha & r^{2/\alpha}\|A\|_\infty + r^{1/2+1/\alpha}\max_i\Big(\sum_j a_{ij}^2 p_j\Big)^{1/2} \\ & + r\|(a_{ij}\sqrt{p_i p_j})\|_{l_2\to l_2} + \sqrt{r}\Big(\sum_k a_{kk}^2 p_k + \sum_{i\neq j} a_{ij}^2 p_i p_j\Big)^{1/2} \\ \lesssim_\alpha & r^{2/\alpha}\|A\|_\infty + Kr\max_i\sum_i |a_{ij}|p_j + \sqrt{r}\Big(\sum_k a_{kk}^2 p_k + \sum_{i\neq j} a_{ij}^2 p_i p_j\Big)^{1/2}.\end{aligned}$$

Then, by Lemma 2.1 below, we deduce (1.8) *from Theorem 2 but with an extra factor K.*

Note that, when A is a diagonal-free matrix, the decay rate of the Hanson-Wright inequality (1.4) is optimal in the sense that, there exist independent centered α-sub-exponential random variables $\eta_1, \cdots, \eta_n$ such that a matching lower bound holds. See Proposition 1.1 for details and its proof is given in Appendix A.

PROPOSITION 1.1 (Optimality of Theorem 2 when $p_1 = \cdots = p_2 = 1$)**.** *Let $\eta_1, \cdots, \eta_n \overset{i.i.d.}{\sim} \mathcal{W}_s(\alpha)$ and A be a symmetric diagonal-free matrix. Then for $0 < \alpha \le 1$ and $t \ge 0$*

$$\mathbb{P}\{|\eta^\top A\eta - \mathbb{E}\eta^\top A\eta| \ge t\} \ge C(\alpha)\exp(-c(\alpha)f_2(t)),$$

where $\eta = (\eta_1, \cdots, \eta_n)^\top$, $C(\alpha), c(\alpha)$ are positive constants depending only on α, and $f_2(t)$ is defined in (1.5)*.*

REMARK 1.4 (Generalization to general A)**.** *Theorems 1 and 2 assume that the matrix A is symmetric. This assumption was made primarily for the convenience of presentation. Indeed, the above results can be extended to general square matrices. The only modification required is that in many places, A should be replaced by $\frac{1}{2}(A + A^\top)$.*

1.2 Disscussion

In this section, we assume that $\{\delta_i, 1 \le i \le n\}$ is a sequence of independent random variables with $\delta_i \sim$ Bernoulli(p_i) and $\{\zeta_i, 1 \le i \le n\}$ is a sequence of independent centered α-sub-exponential random variables. Assume δ_i's are independent of ζ_i's. Denote by $\xi = (\xi_1, \cdots, \xi_n)$ the random vector with $\xi_i = \delta_i \cdot \zeta_i$. Let $A = (a_{ij})_{n\times n}$ be a symmetric fixed matrix and $a = (a_1, \cdots, a_n)^\top$ be a fixed vector. Set $L(\alpha) = \max_i \|\zeta_i\|_{\Psi_\alpha}$. We mention the following two open questions:

1. For $1 \le \alpha \le 2$ and $t > 0$, it holds (see Theorem 6.1 in [2])

$$\mathbb{P}\Big\{\big|\sum_{i=1}^n a_i\zeta_i\big| \ge t\Big\} \le 2\exp\Big(-c(\alpha)\min\Big\{\frac{t^2}{L(\alpha)^2\|a\|_2^2}, \Big(\frac{t}{L(\alpha)\|a\|_{\alpha^*}}\Big)^\alpha\Big\}\Big)$$

   and

$$\mathbb{P}\{|\zeta^\top A\zeta - \mathbb{E}\zeta^\top A\zeta| \ge L(\alpha)^2 t\} \le 2\exp(-c(\alpha)f_2(t)),$$

   where $\alpha^* = \alpha/(\alpha-1)$ and $f_1(t)$ is defined in (1.3). How to obtain upper bounds for

$$\mathbb{P}\Big\{\big|\sum_{i=1}^n a_i\xi_i\big| \ge t\Big\}, \quad \mathbb{P}\{|\xi^\top A\xi - \mathbb{E}\xi^\top A\xi| \ge t\}$$

   which recover the above results when $\delta_1 = \cdots \delta_n = 1$ is an interesting question.

2. In this paper, we give upper bounds for the linear and quadratic forms of sparse α-sub-exponential random vectors:

$$\big\|\sum_{i=1}^n a_i\xi_i\big\|_{L_r}, \quad \|\xi^\top A\xi - \mathbb{E}\xi^\top A\xi\|_{L_r}.$$

   Investigating their lower bounds is an interesting question, which may help us establish the optimal bounds for the above quantities.

2 Preliminaries

2.1 Tails and Moments

In this subsection, we first introduce the following lemma, which provides the link between L_r estimates and tail probability inequalities. Although such type results are by now common, we give a brief proof for the sake of reading.

LEMMA 2.1. *Let ξ be a random variable such that for $r \geq r_0$*

$$\|\xi\|_{L_r} \leq \sum_{k=1}^{m} C_k r^{\beta_k} + C_{m+1},$$

where $C_1, \cdots, C_{m+1} > 0$ and $\beta_1, \cdots, \beta_m > 0$. Then for $t > 0$,

$$\mathbb{P}\big\{|\xi| > e(mt + C_{m+1})\big\} \leq e^{r_0} \exp\Big(-\min\big\{\big(\frac{t}{C_1}\big)^{1/\beta_1}, \cdots, \big(\frac{t}{C_m}\big)^{1/\beta_m}\big\}\Big).$$

Proof. Define the following function:

$$f(t) := \min\big\{\big(\frac{t}{C_1}\big)^{1/\beta_1}, \cdots, \big(\frac{t}{C_m}\big)^{1/\beta_m}\big\}.$$

If $f(t) \geq r_0$, then

$$\|\xi\|_{L_{f(t)}} \leq \sum_{k=1}^{m} C_k f(t)^{\beta_k} + C_{m+1} \leq mt + C_{m+1}.$$

Hence, Markov's inequality yields

$$\mathbb{P}\{|\xi| > e(mt + C_{m+1})\} \leq \mathbb{P}\{|\xi| > e\|\xi\|_{L_{f(t)}}\} \leq e^{-f(t)}.$$

As for $f(t) < r_0$, we have the following trivial bound

$$\mathbb{P}\{|\xi| > e(mt + C_{m+1})\} \leq 1 \leq e^{r_0} e^{-f(t)}.$$

Hence, we have for $t \geq 0$

$$\mathbb{P}\{|\xi| > e(mt + C_{m+1})\} \leq e^{r_0} e^{-f(t)}.$$

□

The next lemma estimates the r-th moments of random linear and bilinear sums.

LEMMA 2.2 (Theorem 1.1 in [10]). *Let $\xi_1, \cdots, \xi_n$ be independent symmetric random variables with log-convex tails, i.e., $\log \mathbb{P}\{|\xi_i| \geq t\}$ is a convex function for $t \geq 0$. Then for $r \geq 2$*

$$\Big\|\sum_i \xi_i\Big\|_{L_r} \asymp \Big(\sum_i \mathbb{E}|\xi_i|^r\Big)^{1/r} + \sqrt{r}\Big(\sum_i \mathbb{E}\xi_i^2\Big)^{1/2}.$$

In the following lemmas of this subsection, we assume that $\xi_1, \cdots, \xi_n \overset{i.i.d.}{\sim} \mathcal{W}_s(\alpha)$ and $A = (a_{ij})_{n\times n}$ is a fixed symmetric diagonal-free matrix. Assume further that $\tilde{\xi}_i$ is an independent copy of ξ_i, $1 \le i \le n$. For any $1 \le \alpha \le 2$, let α^* be its conjugate exponent, i.e., $1/\alpha + 1/\alpha^* = 1$.

LEMMA 2.3 (Theorem 6.1 in [2]). *In the case $1 \le \alpha \le 2$, we have for $r \ge 2$*

$$\Big\| \sum_{i,j} a_{ij}\xi_i\tilde{\xi}_j \Big\|_{L_r} \asymp_\alpha r^{1/2}\|A\|_F + r\|A\|_{l_2\to l_2} + r^{1/\alpha}\|A\|_{l_{\alpha^*}(l_2)} + r^{(\alpha+2)/2\alpha}\|A\|_{l_2\to l_{\alpha^*}} + r^{2/\alpha}\|A\|_{l_\alpha \to l_{\alpha^*}},$$

where $\|A\|_{l_{\alpha*}(l_2)}$ is defined in (1.9).

LEMMA 2.4 (Example 3 in [11]). *In the case $0 < \alpha \le 1$, we have for $r \ge 2$*

$$\Big\| \sum_{i,j} a_{ij}\xi_i\tilde{\xi}_j \Big\|_{L_r} \asymp_\alpha r^{1/2}\|A\|_F + r\|A\|_{l_2\to l_2} + r^{(\alpha+2)/2\alpha}\|A\|_{l_2\to l_\infty} + r^{2/\alpha}\|A\|_\infty.$$

By employing a similar argument as in [16], we can derive the following simplified result from Lemmas 2.3 and 2.4. For ease of reference, the proof is provided in Appendix B.

COROLLARY 1. *In the case $0 < \alpha \le 2$, we have for $r \ge 2$,*

$$\Big\| \sum_{i,j} a_{ij}\xi_i\tilde{\xi}_j \Big\|_{L_r} \lesssim_\alpha r^{1/2}\|A\|_F + r^{2/\alpha}\|A\|_{l_2\to l_2}.$$

2.2 Decoupling inequality

Decoupling is a technique of replacing quadratic forms of random variables by bilinear forms. The monograph [4] systematically studies decoupling and its applications. In this subsection, we introduce a classic decoupling inequality as follows:

LEMMA 2.5 (Theorem 3.1.1 in [4]). *Let $F : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex function and $A = (a_{ij})_{n\times n}$ be a diagnal-free matrix. If $\{\xi_i, i \le n\}$ is a sequence of independent centered random variables and $\tilde{\xi}_i$ is an independent copy of ξ_i, then there exists a universal constant C such that*

$$\mathbb{E}F\Big(\big|\sum_{i,j} a_{ij}\xi_i\xi_j\big|\Big) \le \mathbb{E}F\Big(C\big|\sum_{i,j} a_{ij}\xi_i\tilde{\xi}_j\big|\Big). \tag{2.1}$$

REMARK 2.1. *If, moreover, $A = (a_{ij})$ is a symmetric diagnal-free matrix, then* (2.1) *can be reversed, that is,*

$$\mathbb{E}F\Big(\frac{1}{C}\big|\sum_{i,j} a_{ij}\xi_i\tilde{\xi}_j\big|\Big) \le \mathbb{E}F\Big(\big|\sum_{i,j} a_{ij}\xi_i\xi_j\big|\Big).$$

2.3 Contraction principle

In this subsection, we present a well-known contraction principle in Banach space. This principle enables the extension of results from Weibull random variables to α-sub-exponential random variables.

LEMMA 2.6 (Lemma 4.6 in [13]). *Let $F : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex non-decreasing function. Assume that $\{\eta_i, i \le n\}$ and $\{\xi_i, i \le n\}$ are two sequences of independent symmetric random variables such that for some constant $K \ge 1$*

$$\mathbb{P}\{|\eta_i| > t\} \le K\mathbb{P}\{|\xi_i| > t\}, \quad i \le n, \quad t > 0. \tag{2.2}$$

Then, for any sequence $\{a_i, i \le n\}$ in a Banach space $(\mathcal{A}, \|\cdot\|)$,

$$\mathbb{E}F\Big(\big\|\sum_{i=1}^{n} a_i\eta_i\big\|\Big) \le \mathbb{E}F\Big(K\big\|\sum_{i=1}^{n} a_i\xi_i\big\|\Big).$$

REMARK 2.2. *If $\{\eta_i, i \le n\}$ and $\{\xi_i, i \le n\}$ are two sequences of independent centered random variables satisfying* (2.2)*, the result in Lemma 2.6 is still valid by a symmetrization argument. Indeed, let $\tilde{\eta}_i$ be independent copy of η_i, $1 \le i \le n$. Then, Jensen inequality yields*

$$\begin{aligned}\mathbb{E}F\Big(\big\|\sum_{i=1}^{n} a_i\eta_i\big\|\Big) &= \mathbb{E}F\Big(\big\|\sum_{i=1}^{n} a_i(\eta_i - \mathbb{E}_{\tilde{\eta}}\tilde{\eta}_i)\big\|\Big)\\ &\le \mathbb{E}F\Big(\big\|\sum_{i=1}^{n} a_i(\eta_i - \tilde{\eta}_i)\big\|\Big) \le \mathbb{E}F\Big(2\big\|\sum_{i=1}^{n} a_i\varepsilon_i\eta_i\big\|\Big),\end{aligned}$$

where $\{\varepsilon_i, i \le n\}$ are independent Rademacher variables. On the other hand, by the convexity of F

$$\mathbb{E}F\Big(2\big\|\sum_{i=1}^{n} a_i\varepsilon_i\eta_i\big\|\Big) \le \mathbb{E}F\Big(2\big\|\sum_{i=1}^{n} a_i\varepsilon_i(\eta_i - \tilde{\eta}_i)\big\|\Big) \le \mathbb{E}F\Big(4\big\|\sum_{i=1}^{n} a_i\eta_i\big\|\Big).$$

Hence, we have

$$\mathbb{E}F\Big(\big\|\sum_{i=1}^{n} a_i\eta_i\big\|\Big) \le \mathbb{E}F\Big(2\big\|\sum_{i=1}^{n} a_i\varepsilon_i\eta_i\big\|\Big) \le \mathbb{E}F\Big(4\big\|\sum_{i=1}^{n} a_i\eta_i\big\|\Big).$$

2.4 Concentration inequality

In this subsection, we shall introduce Talagrand's concentration inequality for convex Lipschitz functions.

LEMMA 2.7 (Corollary 4.10 in [12]). *Let $\xi_1, \cdots, \xi_n$ be independent random variables such that $|\xi_i| \le K$ for all $1 \le i \le n$. Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is a convex and 1-Lipschitz function. Then for $t > 0$*

$$\mathbb{P}\big\{\big|f(\xi_1, \cdots, \xi_n) - \mathbb{E}f(\xi_1, \cdots, \xi_n)\big| > Kt\big\} \le 4e^{-t^2/4}.$$

2.5 The sparse Bernstein inequality

In this subsection, we introduce the following sparse Bernstein inequality in α-sub-exponential random variables, $0 < \alpha \leq 1$. With this property, one can estimate the diagonal sum of the quadratic forms.

LEMMA 2.8. *Assume that $\{\delta_i, 1 \leq i \leq n\}$ and $\{\zeta_i, 1 \leq i \leq n\}$ are two independent sequences of random variables, where $\delta_i \sim \text{Bernoulli}(p_i)$ are independent and ζ_i are independent centered α-sub-exponential random variables. Let $a = (a_1, \cdots, a_n)$ be a fixed vector. Then for $0 < \alpha \leq 1$ and $t \geq 0$*

$$\mathbb{P}\Big\{\big|\sum_{i=1}^{n} a_i\delta_i\zeta_i\big| \geq t\Big\} \leq 2\exp\Big(-c(\alpha)\min\Big\{\frac{t^2}{L(\alpha)^2(\sum_i a_i^2 p_i)}, \Big(\frac{t}{L(\alpha)\|a\|_\infty}\Big)^\alpha\Big\}\Big),$$

where $L(\alpha) = \max_i \|\zeta_i\|_{\Psi_\alpha}$ and $c(\alpha)$ is a positive constant depending only on α.

REMARK 2.3. *(i). This property, though likely known in the literature, was explicitly stated and proved using a combinatorial approach in He et al. (see Theorem 3 in [8]). In Appendix C we present an alternative, more concise proof.*

(ii). For $0 < \alpha \leq 1$ and $t \geq 0$, it holds (see Corollary 1.4 in [6])

$$\mathbb{P}\Big\{\big|\sum_{i=1}^{n} a_i\zeta_i\big| \geq t\Big\} \leq 2\exp\Big(-c(\alpha)\min\Big\{\frac{t^2}{L(\alpha)^2\|a\|_2^2}, \Big(\frac{t}{L(\alpha)\|a\|_\infty}\Big)^\alpha\Big\}\Big).$$

Hence, Lemma 2.8 recovers this classical result when $p_1 = \cdots = p_n = 1$.

3 Proofs of main results

3.1 The sparse Hanson-Wright inequality for $0 < \alpha \leq 2$

We begin by proving Theorem 1 for the special case where matrix A has zero diagonal elements.

PROPOSITION 3.1. *Assume that $\{\delta_i, 1 \leq i \leq n\}$ and $\{\zeta_i, 1 \leq i \leq n\}$ are two independent sequences of random variables, where $\delta_i \sim \text{Bernoulli}(p_i)$ are independent and ζ_i are independent centered α-sub-exponential random variables. Consider a random vector $\xi = (\xi_1, \cdots, \xi_n)$ with $\xi_i = \delta_i \cdot \zeta_i$. Let $A = (a_{ij})_{n\times n}$ be a symmetric diagonal-free matrix. Then for $0 < \alpha \leq 2$ and $t \geq 0$,*

$$\mathbb{P}\{|\xi^\top A\xi| \geq t\} \leq 2\exp\Big(-c(\alpha)\min\Big\{\frac{t^2}{L(\alpha)^4\sum_{i,j} a_{ij}^2 p_i p_j}, \big(\frac{t}{L(\alpha)^2\|A\|_{l_2\to l_2}}\big)^{\alpha/2}\Big\}\Big),$$

where $L(\alpha) = \max_i \|\zeta_i\|_{\Psi_\alpha}$ and $c(\alpha)$ is a positive constant depending only on α.

Proof. Let $\tilde{\xi}_i = \tilde{\delta}_i \cdot \tilde{\zeta}_i$ for $1 \leq i \leq n$, where $\{\tilde{\delta}_i, 1 \leq i \leq n\}$ and $\{\tilde{\zeta}_i, 1 \leq i \leq n\}$ are independent copies of $\{\delta_i, 1 \leq i \leq n\}$ and $\{\zeta_i, 1 \leq i \leq n\}$, respectively. Lemma 2.5 yields for $r \geq 1$

$$\|\xi^\top A\xi\|_{L_r} \lesssim \|\xi^\top A\tilde{\xi}\|_{L_r},$$

where $\tilde{\xi} = (\tilde{\xi}_1, \cdots, \tilde{\xi}_n)$. Let $\eta_1, \cdots, \eta_n \overset{i.i.d.}{\sim} \mathcal{W}_s(\alpha), 0 < \alpha \le 2$, and denote by $\tilde{\eta}_i$ the independent copy of η_i, $1 \le i \le n$. Using the conditional probability and Lemma 2.6 twice, we have for $r \ge 1$

$$\begin{aligned}\|\xi^\top A\tilde{\xi}/L(\alpha)^2\|_{L_r} &= \Big(\mathbb{E}_\xi\mathbb{E}_{\tilde{\xi}}\Big|\frac{1}{L(\alpha)^2}\sum_j \tilde{\xi}_j \sum_i a_{ij}\xi_i\Big|^r\Big)^{1/r} \\ &\lesssim \Big(\mathbb{E}_\xi\mathbb{E}_{\tilde{\delta},\tilde{\eta}}\Big|\frac{1}{L(\alpha)}\sum_j \tilde{\delta}_j\tilde{\eta}_j \sum_i a_{ij}\xi_i\Big|^r\Big)^{1/r} \lesssim \Big\|\sum_{i,j} a_{ij}\delta_i\eta_i\tilde{\delta}_j\tilde{\eta}_j\Big\|_{L_r}.\end{aligned}$$

Corollary 1 yields for $r \ge 2$

$$\begin{aligned}\mathbb{E}_{\eta,\tilde{\eta}}\big|\sum_{i,j} a_{ij}\delta_i\eta_i\tilde{\delta}_j\tilde{\eta}_j\big|^r \le& C_1(\alpha)^r r^{r/2}\Big(\sum_{i,j} a_{ij}^2\delta_i\tilde{\delta}_j\Big)^{r/2} \\ &+ C_2(\alpha)^r r^{2r/\alpha}\|(a_{ij}\delta_i\tilde{\delta}_j)_{n\times n}\|_{l_2\to l_2}^r,\end{aligned} \tag{3.1}$$

where $\mathbb{E}_{\eta,\tilde{\eta}}$ means taking expectation with respect to $\eta_i, \tilde{\eta}_i, 1 \le i \le n$.

Let $\Lambda = \text{Diag}(\delta_1, \cdots, \delta_n)$ be a diagonal matrix with entries $\delta_1, \cdots, \delta_n$, and $\tilde{\Lambda} = \text{Diag}(\tilde{\delta}_1, \cdots, \tilde{\delta}_n)$. As for the second term on the right side of (3.1), we have

$$\|(a_{ij}\delta_i\tilde{\delta}_j)_{n\times n}\|_{l_2\to l_2} = \|\Lambda A\tilde{\Lambda}\|_{l_2\to l_2} \le \|\Lambda\|_{l_2\to l_2}\|A\|_{l_2\to l_2}\|\tilde{\Lambda}\|_{l_2\to l_2} \le \|A\|_{l_2\to l_2}.$$

We next turn to bounding the first term of (3.1). Note that, $f(x_1, \cdots, x_n) = \sqrt{\sum_i a_i^2 x_i^2}$ is a convex and Lipschitz function with $\|f\|_{\text{Lip}} = \max_i |a_i|$. Hence, Lemma 2.7 yields for $t > 0$

$$\mathbb{P}_\delta\Big\{\Big|\big(\sum_i \delta_i \sum_j a_{ij}^2\tilde{\delta}_j\big)^{1/2} - \mathbb{E}_\delta\big(\sum_i \delta_i \sum_j a_{ij}^2\tilde{\delta}_j\big)^{1/2}\Big| > \max_i(\sum_j a_{ij}^2\tilde{\delta}_j)^{1/2}t\Big\} \le 4e^{-t^2/4},$$

where $\mathbb{P}_\delta$ and $\mathbb{E}_\delta$ mean taking probability and expectation with respect to $\delta_1, \cdots, \delta_n$. Then, a direct integration yields for $r \ge 1$

$$\begin{aligned}&\Big(\mathbb{E}_\delta\big|\big(\sum_i \delta_i \sum_j a_{ij}^2\tilde{\delta}_j\big)^{1/2}\big|^r\Big)^{1/r} \\ \lesssim& \mathbb{E}_\delta\big(\sum_i \delta_i \sum_j a_{ij}^2\tilde{\delta}_j\big)^{1/2} + \sqrt{r}\max_i(\sum_j a_{ij}^2\tilde{\delta}_j)^{1/2}. \\ \le& \big(\sum_{i,j} a_{ij}^2 p_i\tilde{\delta}_j\big)^{1/2} + \sqrt{r}\max_i(\sum_j a_{ij}^2)^{1/2}.\end{aligned} \tag{3.2}$$

It is obvious to see that

$$\max_i(\sum_j a_{ij}^2)^{1/2} = \|A\|_{l_2\to l_\infty} \le \|A\|_{l_2\to l_2}.$$

Using Lemma 2.7 again, we have for $t > 0$

$$\mathbb{P}\Big\{\Big|\big(\sum_{i,j} a_{ij}^2 p_i\tilde{\delta}_j\big)^{1/2} - \mathbb{E}\big(\sum_{i,j} a_{ij}^2 p_i\tilde{\delta}_j\big)^{1/2}\Big| > \max_j(\sum_i a_{ij}^2 p_i)^{1/2}t\Big\} \le 4e^{-t^2/4}.$$

Then for $r \geq 1$

$$\big\|\big(\sum_{i,j} a_{ij}^2 p_i \tilde{\delta}_j\big)^{1/2}\big\|_{L_r} \lesssim \mathbb{E}\big(\sum_{i,j} a_{ij}^2 p_i \tilde{\delta}_j\big)^{1/2} + \sqrt{r}\max_j\big(\sum_i a_{ij}^2 p_i\big)^{1/2}$$
$$\leq \big(\sum_{i,j} a_{ij}^2 p_i p_j\big)^{1/2} + \sqrt{r}\|A\|_{l_2\to l_2}. \tag{3.3}$$

By virtue of (3.2) and (3.3), we have for $r \geq 1$

$$\sqrt{r}\big\|\big(\sum_{i,j} a_{ij}^2 \delta_i \tilde{\delta}_j\big)^{1/2}\big\|_{L_r} \lesssim \sqrt{r}\big(\sum_{i,j} a_{ij}^2 p_i p_j\big)^{1/2} + r\|A\|_{l_2\to l_2}. \tag{3.4}$$

As $0 < \alpha \leq 2$, we can absorb the last term above with $r^{2/\alpha}\|A\|_{l_2\to l_2}$. Combining (3.1) with (3.4), we have for $r \geq 2$

$$\big\|\sum_{ij} a_{ij}\delta_i\eta_i\tilde{\delta}_j\tilde{\eta}_j\big\|_{L_r} \lesssim_\alpha \sqrt{r}\big(\sum_{ij} a_{ij}^2 p_i p_j\big)^{1/2} + r^{2/\alpha}\|A\|_{l_2\to l_2}.$$

Hence, we finish the proof by Lemma 2.1 and adjusting the universal constant. $\square$

Now, we are prepared to prove Theorem 1.

Proof of Theorem 1. By the triangle inequality, we have for $r \geq 1$

$$\big\|\xi^\top A\xi - \mathbb{E}\xi^\top A\xi\big\|_{L_r} \leq \big\|\sum_{i\neq j} a_{ij}\xi_i\xi_j\big\|_{L_r} + \big\|\sum_i a_{ii}(\xi_i^2 - \mathbb{E}\xi_i^2)\big\|_{L_r}.$$

Proposition 3.1 yields that

$$\big\|\sum_{i\neq j} a_{ij}\xi_i\xi_j/L(\alpha)^2\big\|_{L_r} \lesssim_\alpha \sqrt{r}\big(\sum_{i\neq j} a_{ij}^2 p_i p_j\big)^{1/2} + r^{2/\alpha}\|A\|_{l_2\to l_2}. \tag{3.5}$$

Let $\tilde{\xi}_i$ be an independent copy of ξ_i, $1 \leq i \leq n$. Note that

$$\big\|\sum_i a_{ii}(\xi_i^2 - \mathbb{E}\xi_i^2)\big\|_{L_r} = \big\|\sum_i a_{ii}(\xi_i^2 - \mathbb{E}_{\tilde{\xi}_i}\tilde{\xi}_i^2)\big\|_{L_r} \leq 2\big\|\sum_i a_{ii}\varepsilon_i\delta_i\zeta_i^2\big\|_{L_r},$$

where $\varepsilon_1, \cdots, \varepsilon_n$ is a sequence of i.i.d. Rademacher random variables. Due to that

$$\mathbb{P}\{|\varepsilon_i\zeta_i^2| > L(\alpha)^2 t\} \leq \mathbb{P}\{|\zeta_i| > L(\alpha)\sqrt{t}\} \leq 2e^{-ct^{\alpha/2}},$$

Lemma 2.8 yields that

$$\big\|\sum_i a_{ii}(\xi_i^2 - \mathbb{E}\xi_i^2)/L(\alpha)^2\big\|_{L_r} \lesssim_\alpha \sqrt{r}\big(\sum_i a_{ii}^2 p_i\big) + r^{2/\alpha}\max_i |a_{ii}|. \tag{3.6}$$

By virtue of (3.5) and (3.6), we have

$$\begin{aligned}&\big\|(\xi^\top A\xi - \mathbb{E}\xi^\top A\xi)/L(\alpha)^2\big\|_{L_r}\\ \lesssim_\alpha &\sqrt{r}\big(\sum_{i\neq j} a_{ij}^2 p_i p_j + \sum_i a_{ii}^2 p_i\big)^{1/2} + r^{2/\alpha}\|A\|_{l_2\to l_2},\end{aligned}$$

where we use the fact $\max_i |a_{ii}| \leq \|A\|_{l_2\to l_2}$. Hence, the desired result follows from Lemma 2.1. $\square$

3.2 An improved sparse Hanson-Wright inequality for $0 < \alpha \le 1$

We complete the proof of Theorem 2 in this subsection.

Proof of Theorem 2. We first assume $a_{ii} = 0, 1 \le i \le n$. By Lemmas 2.5 and 2.6, we only need to bound the quantity $\| \sum_{i,j} a_{ij}\delta_i\eta_i\tilde{\delta}_j\tilde{\eta}_j\|_{L_r}$. Here, $\eta_1, \cdots, \eta_n \overset{i.i.d.}{\sim} \mathcal{W}_s(\alpha)$, and for $1 \le i \le n$, $\tilde{\delta}_i, \tilde{\eta}_i$ are independent copies of δ_i, η_i, respectively.

Note that, for $1 \le i \le n$,

$$-\log \mathbb{P}\{|\delta_i\eta_i| \ge t\} = \begin{cases} 0, & t = 0 \\ t^\alpha - \log p_i, & t > 0 \end{cases}$$

is a concave function for $t \ge 0$. Hence, by Lemma 2.2, we have for $r \ge 2$

$$\begin{aligned}\Big(\mathbb{E}_{\delta,\eta}\big|\sum_i \delta_i\eta_i\sum_j a_{ij}\tilde{\delta}_j\tilde{\eta}_j\big|^r\Big)^{1/r} \lesssim_\alpha r^{1/\alpha}\Big(\sum_i p_i\big|\sum_j a_{ij}\tilde{\delta}_j\tilde{\eta}_j\big|^r\Big)^{1/r} \\ + \sqrt{r}\Big(\sum_i p_i(\sum_j a_{ij}\tilde{\delta}_j\tilde{\eta}_j)^2\Big)^{1/2},\end{aligned} \tag{3.7}$$

where $\mathbb{E}_{\delta,\eta}$ means taking expectation with respect to $\{\delta_i, \eta_i, 1 \le i \le n\}$.

Combining Fubini's theorem with (3.7), we have for $r \ge 2$

$$\begin{aligned}\Big\|\sum_{i,j} a_{ij}\delta_i\eta_i\tilde{\delta}_j\tilde{\eta}_j\Big\|_{L_r} &= \Big(\mathbb{E}_{\tilde{\delta},\tilde{\eta}}\mathbb{E}_{\delta,\eta}\big|\sum_i \delta_i\eta_i\sum_j a_{ij}\tilde{\delta}_j\tilde{\eta}_j\big|^r\Big)^{1/r} \\ &\lesssim_\alpha r^{1/\alpha}\Big(\mathbb{E}\sum_i p_i\big|\sum_j a_{ij}\tilde{\delta}_j\tilde{\eta}_j\big|^r\Big)^{1/r} \\ &\quad + \sqrt{r}\Big\|\Big(\sum_i p_i(\sum_j a_{ij}\tilde{\delta}_j\tilde{\eta}_j)^2\Big)^{1/2}\Big\|_{L_r},\end{aligned} \tag{3.8}$$

where the last inequality we use the fact $(a+b)^r \le 2^r(|a|^r + |b|^r)$.

Next, we turn to the terms on the right side of (3.8). For the first term, we have

$$\begin{aligned}\Big(\mathbb{E}\sum_i p_i\big|\sum_j a_{ij}\tilde{\delta}_j\tilde{\eta}_j\big|^r\Big)^{1/r} &= \Big(\sum_i p_i\Big\|\sum_j a_{ij}\tilde{\delta}_j\tilde{\eta}_j\Big\|_{L_r}^r\Big)^{1/r} \\ &\lesssim \Big(\sum_i p_i\Big((\sum_j \mathbb{E}|a_{ij}\tilde{\delta}_j\tilde{\eta}_j|^r)^{1/r} + \sqrt{r}(\sum_j \mathbb{E}|a_{ij}\tilde{\delta}_j\tilde{\eta}_j|^2)^{1/2}\Big)^r\Big)^{1/r} \\ &\lesssim_\alpha r^{1/\alpha}\Big(\sum_{i,j}|a_{ij}|^r p_ip_j\Big)^{1/r} + \sqrt{r}\Big(\sum_i p_i(\sum_j a_{ij}^2p_j)^{r/2}\Big)^{1/r},\end{aligned}$$

where we use Lemma 2.2 in the second step and use the triangle inequality in the third step.

For the second term on the right side of (3.8), note that for $g_1, \cdots, g_n \overset{i.i.d.}{\sim} N(0,1)$ and $a_1, \cdots, a_n \in \mathbb{R}$

$$\Big\|\sum_i a_ig_i\Big\|_{L_r} \asymp \sqrt{r}\big(\sum_i a_i^2\big)^{1/2}.$$

Hence, we have by Lemma 2.2

$$
\begin{aligned}
&\mathbb{E}\Big(r\sum_i p_i(\sum_j a_{ij}\tilde{\delta}_j\tilde{\eta}_j)^2\Big)^{r/2} \\
\le& C^r\mathbb{E}\mathbb{E}_g\Big(\sum_i\sqrt{p_i}g_i\sum_j a_{ij}\tilde{\delta}_j\tilde{\eta}_j\Big)^r \\
=& C^r\mathbb{E}\mathbb{E}_{\tilde{\delta},\tilde{\eta}}\Big(\sum_j\tilde{\delta}_j\tilde{\eta}_j\sum_i a_{ij}\sqrt{p_i}g_i\Big)^r \\
\le& C_1(\alpha)^r\mathbb{E}\Big(r^{r/\alpha}\sum_j p_j\big|\sum_i a_{ij}\sqrt{p_i}g_i\big|^r + r^{r/2}\big(\sum_j p_j\big(\sum_i a_{ij}\sqrt{p_i}g_i\big)^2\big)^{r/2}\Big) \\
\le& C_2(\alpha)^r r^{r/\alpha+r/2}\sum_j p_j\big(\sum_i a_{ij}^2p_i\big)^{r/2} + C_3(\alpha)^r\mathbb{E}\Big(\sum_{i,j}a_{ij}\sqrt{p_ip_j}g_i\tilde{g}_j\Big)^r,
\end{aligned}
$$

where $\tilde{g}_j$ is an independent copy of g_j, $1\le j\le n$. Note that g_i are also sub-gaussian random variables. Hence, we have by Corollary 1 and Lemma 2.6

$$
\Big\|\sum_{i,j}a_{ij}\sqrt{p_ip_j}g_i\tilde{g}_j\Big\|_{L_r}\lesssim\sqrt{r}\Big(\sum_{i,j}a_{ij}^2p_ip_j\Big)^{1/2}+r\|(a_{ij}\sqrt{p_ip_j})_{n\times n}\|_{l_2\to l_2}.
$$

Up to now, we have proved for diagonal free symmetric matrix A

$$
\begin{aligned}
\Big\|\sum_{i,j}a_{ij}\delta_i\eta_i\tilde{\delta}_j\tilde{\eta}_j\Big\|_{L_r}\lesssim_\alpha& r^{2/\alpha}\Big(\sum_{ij}|a_{ij}|^rp_ip_j\Big)^{1/r}+r^{1/2+1/\alpha}\Big(\sum_i p_i\big(\sum_j a_{ij}^2p_j\big)^{r/2}\Big)^{1/r} \\
&+r\|(a_{ij}\sqrt{p_ip_j})_{n\times n}\|_{l_2\to l_2}+\sqrt{r}\Big(\sum_{i,j}a_{ij}^2p_ip_j\Big)^{1/2}.
\end{aligned}
$$

Note that for $r\ge 3$, by Young's inequality, we have

$$
\begin{aligned}
\Big(\sum_{i,j}|a_{ij}|^rp_ip_j\Big)^{1/r}&\le\max_{i,j}|a_{ij}|^{(r-2)/r}\cdot\Big(\sum_{i,j}a_{ij}^2p_ip_j\Big)^{1/r} \\
&=\Big(e^{2r/(r-2)}\max_{i,j}|a_{ij}|\Big)^{(r-2)/r}\cdot\Big(\sum_{i,j}a_{ij}^2p_ip_je^{-2r}\Big)^{1/r} \\
&\le\frac{r-2}{r}\Big(e^{2r/(r-2)}\max_{i,j}|a_{ij}|\Big)+\frac{2}{r}\Big(\sum_{i,j}a_{ij}^2p_ip_je^{-2r}\Big)^{1/2} \\
&\le e^6\|A\|_\infty+e^{-r}\Big(\sum_{i,j}a_{ij}^2p_ip_j\Big)^{1/2}.
\end{aligned}
$$

Similarly

$$
\Big(\sum_i p_i\big(\sum_j a_{ij}^2p_j\big)^{r/2}\Big)^{1/r}\le e^6\max_i\Big(\sum_j a_{ij}^2p_j\Big)^{1/2}+e^{-r}\Big(\sum_{i,j}a_{ij}^2p_ip_j\Big)^{1/2}.
$$

Note that the e^{-r} factors we introduced will eliminate the $r^{2/\alpha}$, $r^{1/2+1/\alpha}$ factors in the corresponding terms. Hence, we have for $r \geq 3$

$$\begin{aligned}\Big\| \sum_{i,j} a_{ij}\delta_i\eta_i\tilde{\delta}_j\tilde{\eta}_j \Big\|_{L_r} \lesssim_\alpha & r^{2/\alpha}\|A\|_\infty + r^{1/2+1/\alpha} \max_i \Big(\sum_j a_{ij}^2 p_j \Big)^{1/2} \\ & + r\|(a_{ij}\sqrt{p_i p_j})_{n\times n}\|_{l_2\to l_2} + \sqrt{r}\Big(\sum_{i,j} a_{ij}^2 p_i p_j \Big)^{1/2}. \end{aligned} \tag{3.9}$$

For a general symmetric matrix A, we only need to estimate the diagonal case. Following the same line as in the proof of Theorem 1, we have

$$\big\| \sum_i a_{ii}(\xi_i^2 - \mathbb{E}\xi_i^2)/L(\alpha)^2 \big\|_{L_r} \lesssim_\alpha \sqrt{r}\big(\sum_i a_{ii}^2 p_i \big)^{1/2} + r^{2/\alpha} \max_i |a_{ii}|. \tag{3.10}$$

Combining (3.9) with (3.10), the desired result follows from Lemma 2.1. □

Acknowledgment K. Wang was partially supported by Hong Kong RGC grant GRF 16304222. Y. Zhu was partially supported by an AMS-Simons Travel Grant.

References

[1] Radosław Adamczak and Rafał Latała. Tail and moment estimates for chaoses generated by symmetric random variables with logarithmically concave tails. In *Annales de l'IHP Probabilités et statistiques*, volume 48, pages 1103–1136, 2012.

[2] Radosław Adamczak and Paweł Wolff. Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. *Probability Theory and Related Fields*, 162:531–586, 2015.

[3] Abhishek Chakrabortty and Arun K Kuchibhotla. Tail bounds for canonical U-statistics and U-processes with unbounded kernels. *arXiv preprint arXiv:2504.01318*, 2025.

[4] Victor De la Pena and Evarist Giné. *Decoupling: from dependence to independence*. Springer Science & Business Media, 2012.

[5] Evarist Giné, Rafał Latała, and Joel Zinn. Exponential and moment inequalities for U-statistics. In *High Dimensional Probability II*, pages 13–38. Springer, 2000.

[6] Friedrich Götze, Holger Sambale, and Arthur Sinulis. Concentration inequalities for polynomials in α-sub-exponential random variables. *Electronic Journal of Probability*, 26, 2021.

[7] David Lee Hanson and Farroll Tim Wright. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics*, 42(3):1079–1083, 1971.

[8] Yiyun He, Ke Wang, and Yizhe Zhu. Sparse Hanson-Wright inequalities with applications. *arXiv preprint arXiv:2410.15652*, 2024.

[9] Yukun He, Antti Knowles, and Matteo Marcozzi. Local law and complete eigenvector delocalization for supercritical Erdős–Rényi graphs. *The Annals of Probability*, 47(5):3278–3302, 2019.

[10] Paweł Hitczenko, Stephen J Montgomery-Smith, and Krzysztof Oleszkiewicz. Moment inequalities for sums of certain independent symmetric random variables. *Studia Math*, 123(1):15–42, 1997.

[11] Konrad Kolesko and Rafał Latała. Moment estimates for chaoses generated by symmetric random variables with logarithmically convex tails. *Statistics & Probability Letters*, 107:210–214, 2015.

[12] Michel Ledoux. *The concentration of measure phenomenon*. Number 89. American Mathematical Soc., 2001.

[13] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media, 2013.

[14] Seongoh Park, Xinlei Wang, and Johan Lim. Sparse Hanson–Wright inequality for a bilinear form of sub-Gaussian variables. *Stat*, 12(1):e539, 2023.

[15] Mark Rudelson and Roman Vershynin. Hanson-Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18:1–9, 2013.

[16] Holger Sambale. Some notes on concentration for α-subexponential random variables. In *High Dimensional Probability IX: The Ethereal Volume*, pages 167–192. Springer, 2023.

[17] Warren Schudy and Maxim Sviridenko. Bernstein-like concentration and moment inequalities for polynomials of independent random variables: multilinear case. *arXiv preprint arXiv:1109.5193*, 2011.

[18] Warren Schudy and Maxim Sviridenko. Concentration and moment inequalities for polynomials of independent random variables. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 437–446. SIAM, 2012.

[19] David Wu and Anant Sahai. Precise asymptotic generalization for multiclass classification with overparameterized linear models. *Advances in Neural Information Processing Systems*, 36:42576–42622, 2023.

[20] Shuheng Zhou. Sparse Hanson–Wright inequalities for subgaussian quadratic forms. *Bernoulli*, 25(3):1603–1639, 2019.

[21] Shuheng Zhou. Concentration of measure bounds for matrix-variate data with missing values. *Bernoulli*, 30(1):198–226, 2024.

A Proof of Proposition 1.1

Proof. Let $\tilde{\eta} = (\tilde{\eta}_1, \cdots, \tilde{\eta}_n)^\top$ be an independent copy of η. Lemma 2.6 and Remark 2.2 yield for $r \geq 1$

$$\|\eta^\top A\eta - \mathbb{E}\eta^\top A\eta\|_{L_r} \asymp \|\eta^\top A\tilde{\eta}\|_{L_r}.$$

By virtue of Lemma 2.4, we have for $r \geq 2$

$$\begin{aligned}\|\eta^\top A\tilde{\eta}\|_{L_r} \asymp_\alpha & \sqrt{r}\|A\|_F + r\|A\|_{l_2\to l_2} \\ & + r^{1/\alpha+1/2} \max_i \big(\sum_j a_{ij}^2\big)^{1/2} + r^{2/\alpha}\|A\|_\infty := f_1(A, r).\end{aligned}$$

Hence, for $r \geq 2$, we have by the Paley-Zygmund inequality

$$\begin{aligned}\mathbb{P}\Big\{|\eta^\top A\eta - \mathbb{E}\eta^\top A\eta| \geq \frac{C_1(\alpha)}{2} f_1(A, r)\Big\} &\geq \mathbb{P}\Big\{|\eta^\top A\tilde{\eta}| \geq \frac{1}{2}\|\eta^\top A\tilde{\eta}\|_{L_r}\Big\} \\ &\geq (1 - 2^{-r})^2 \Big(\frac{\|\eta^\top A\tilde{\eta}\|_{L_r}}{\|\eta^\top A\tilde{\eta}\|_{L_{2r}}}\Big)^{2r} \geq \frac{1}{2} e^{-c_1(\alpha)r},\end{aligned}$$

where in the last inequality, we use Lemma 2.4. If we lower bound $e^{-c_1(\alpha)r}$ by $e^{-c_1(\alpha)(r+2)}$, this inequality is valid for all $r \geq 0$. Hence, we have for $t \geq 0$

$$\mathbb{P}\Big\{|\eta^\top A\eta - \mathbb{E}\eta^\top A\eta| \geq \frac{C_1(\alpha)}{2} f_1(A, t)\Big\} \geq \frac{1}{2} e^{-2c_1(\alpha)} e^{-c_1(\alpha)t},$$

which immediately yields the desired result. □

B Proof of Corollary 1

Proof. We first consider the case $0 < \alpha \leq 1$. Note that

$$\|A\|_\infty \leq \|A\|_{l_2\to l_\infty} \leq \|A\|_{l_2\to l_2}.$$

Therefore, in this case, Corollary 1 can be immediately deduced from Lemma 2.4.

We next turn to the case $1 \leq \alpha \leq 2$. In this case, note that

$$\|A\|_{l_\alpha\to l_{\alpha^*}} \leq \|A\|_{l_2\to l_{\alpha^*}} \leq \|A\|_{l_2\to l_2}.$$

Hence, to finish the proof, it is enough to prove for $1 \leq \alpha \leq 2$

$$r^{1/\alpha}\|A\|_{l_{\alpha^*}(l_2)} \leq r^{1/2}\|A\|_F + r\|A\|_{l_2\to l_\infty}.$$

Define the following set for $1 \leq \alpha \leq 2$

$$I(r) := \{(x_{ij}) = (z_i y_{ij}) \in \mathbb{R}^{n\times n} : \sum_{i=1}^n |z_i|^\alpha \leq r, \max_{i=1,\cdots,n} \sum_{j=1}^n y_{ij}^2 \leq 1\}.$$

We have the following relation

$$r^{1/\alpha}\|A\|_{l_{\alpha^*}(l_2)} = \sup\big\{\sum_{i,j} a_{ij}x_{ij} : (x_{ij}) \in I(r)\big\}.$$

Indeed, on the one hand, we have

$$\sup_{(x_{ij})\in I(r)} \sum_{i,j} a_{ij}z_i y_{ij} \le \sup_{(z_i)} \sum_i z_i \sup_{(y_{ij})} \sum_j a_{ij}y_{ij} \le r^{1/\alpha}\|A\|_{l_{\alpha^*}(l_2)}.$$

On the other hand, letting

$$y_{ij} = \frac{a_{ij}}{(\sum_{j=1}^n a_{ij}^2)^{1/2}}, \quad z_i = \frac{r^{1/\alpha}(\sum_{j=1}^n a_{ij}^2)^{(\alpha^*-1)/2}}{(\sum_{i=1}^n(\sum_{j=1}^n a_{ij}^2)^{\alpha^*/2})^{1/\alpha}}$$

yields the inverse inequality. Define another subset of $\mathbb{R}^{n\times n}$

$$I_1(r) := \{(x_{ij}) = (z_i y_{ij}) \in \mathbb{R}^{n\times n} : \sum_{i=1}^n \min\{|z_i|^\alpha, z_i^2\} \le r, \max_{i=1,\cdots,n} \sum_{j=1}^n y_{ij}^2 \le 1\}.$$

Obviously, $I(r) \subset I_1(r)$. Given any (z_i) and (y_{ij}) satisfying the conditions of $I_1(r)$, we have

$$\begin{aligned}
\big|\sum_{i,j} a_{ij}z_i y_{ij}\big| \le& \sum_{i=1}^n |z_i|\big(\sum_{j=1}^n a_{ij}^2\big)^{1/2}\big(\sum_{j=1}^n y_{ij}^2\big)^{1/2} \le \sum_{i=1}^n |z_i|\big(\sum_{j=1}^n a_{ij}^2\big)^{1/2} \\
\le& \sum_{i=1}^n |z_i|\mathbb{I}_{\{|z_i|\le 1\}}\big(\sum_{j=1}^n a_{ij}^2\big)^{1/2} + \sum_{i=1}^n |z_i|\mathbb{I}_{\{|z_i|>1\}}\big(\sum_{j=1}^n a_{ij}^2\big)^{1/2} \\
\le& \|A\|_F\big(\sum_{i=1}^n z_i^2 \mathbb{I}_{\{|z_i|\le 1\}}\big)^{1/2} + \|A\|_{l_2\to l_\infty}\sum_{i=1}^n |z_i|\mathbb{I}_{\{|z_i|>1\}} \\
\le& r^{1/2}\|A\|_F + r\|A\|_{l_2\to l_\infty}.
\end{aligned}$$

This completes the proof. □

C Proof of Lemma 2.8

Proof. Let $\eta_1, \cdots, \eta_n \overset{i.i.d.}{\sim} \mathcal{W}_s(\alpha)$, $0 < \alpha \le 1$. Note that, for $1 \le i \le n$,

$$-\log \mathbb{P}\{|\delta_i \eta_i| \ge t\} = \begin{cases} 0, & t = 0 \\ t^\alpha - \log p_i, & t > 0 \end{cases}$$

is a concave function for $t \ge 0$. Hence, Lemma 2.2 yields for $r \ge 2$

$$\big\|\sum_i a_i\delta_i\eta_i\big\|_{L_r} \asymp_\alpha r^{1/\alpha}\big(\sum_i p_i|a_i|^r\big)^{1/r} + \sqrt{r}(\sum_i p_i a_i^2)^{1/2}. \tag{C.1}$$

Here, we use the fact $\|\eta_1\|_{L_r} \asymp r^{1/\alpha}$. As for the first term on the right side of (C.1), we have for $r \geq 3$

$$\begin{aligned}
\big(\sum_i p_i |a_i|^r\big)^{1/r} &\leq \max_i |a_i|^{(r-2)/r} \cdot \big(\sum_i p_i a_i^2\big)^{1/r} \\
&= \big(e^{2r/(r-2)} \max_i |a_i|\big)^{(r-2)/r} \cdot \big(\sum_i p_i a_i^2 e^{-2r}\big)^{1/r} \\
&\leq \frac{r-2}{r}\big(e^{2r/(r-2)} \max_i |a_i|\big) + \frac{2}{r}\big(\sum_i p_i a_i^2 e^{-2r}\big)^{1/2} \\
&\leq e^6 \|a\|_\infty + e^{-r}\big(\sum_i p_i a_i^2\big)^{1/2}.
\end{aligned}$$

This result with (C.1) yields for $r \geq 3$,

$$\big\| \sum_i a_i \delta_i \eta_i \big\|_{L_r} \lesssim_\alpha \sqrt{r} (\sum_i p_i a_i^2)^{1/2} + r^{1/\alpha} \|a\|_\infty.$$

For $t \geq 0$ and $1 \leq i \leq n$, we have

$$\mathbb{P}\{|\delta_i \zeta_i| \geq L(\alpha) t\} \leq 2\mathbb{P}\{|\tilde{\delta}_i \eta_i| \geq t\},$$

where $\tilde{\delta}_i$ is an independent copy of δ_i, $1 \leq i \leq n$. Hence, Lemma 2.6 yields for $r \geq 3$

$$\big\| \sum_i a_i \delta_i \zeta_i / L(\alpha) \big\|_{L_r} \lesssim \big\| \sum_i a_i \tilde{\delta}_i \eta_i \big\|_{L_r} \lesssim_\alpha \sqrt{r} (\sum_i p_i a_i^2)^{1/2} + r^{1/\alpha} \|a\|_\infty.$$

Then, using Lemma 2.1 and adjusting the universal constant yield the desired result. $\square$