An Alternative Proof, via Decoupling, of Hanson-Wright Inequality

Heyuan Yao[1] **Supervisor: Victor H. de la Peña**[2]

[1] Department of Mathematics, Columbia University [2] Department of Statistics, Columbia University

Columbia University Department of Statistics

Sub-Gaussian Random Variables

The tail probability of a random variable is usually considered in both theoretical and empirical studies. Among short-tailed distributions, sub-Gaussian family distribution is widely assumed. The definition is given below (Vershynin [5]).

A random variable X is said to be ***sub-Gaussian*** *if it satisfies that, there exists $K_1 > 0$ such that for all $t \geq 0$,*

$$\mathbb{P}\{|X| \geq t\} \leq 2\exp(-t^2/K_1^2).$$

The ***sub-Gaussian norm*** *of X, denoted by $\|X\|_{\psi_2}$, is defined by any one of the following two equivalent expressions:*

$$\|X\|_{\psi_2} = \inf\{t > 0 : \mathbb{E}\,\exp(X^2/t^2) \leq 2\} = \sup_{p \geq 1} p^{-1/2}\|X\|_p.$$

One important theorem, due to Hanson & Wright (1971, [3]), provides the tail probability of the quadratic form of independent Sub-Gaussian variables.

Consider $X = (X_1, ..., X_n)$, the entries of which are n independent centered sub-Gaussian random variables such that $\max_{1 \leq i \leq n} \|X_i\|_{\psi_2} \leq K$ for some $K > 0$. Then for $X^T AX$, a quadratic form such that $A = (a_{ij})_{n \times n}$ is a symmetric matrix, we have

$$\mathbb{P}\{|X^T AX - \mathbb{E}X^T AX| > t\} \leq 2\exp(-c\min\{\frac{t^2}{K^4\|A\|_{HS}^2}, \frac{t}{K^2\|A\|_2}\}). \quad (1)$$

Tangent Decoupling

Suppose $(\Omega, \mathcal{F}, \{\mathcal{F}_i\}, \mathbb{P},)$ is a filtered probability space with $\{\mathcal{F}_i\}_{i=1}^{\infty}$ a non-decreasing sequence of σ-fields and $\mathcal{F}_i \subset \mathcal{F}$ for all i. Let $\{e_i\}$ and $\{d_i\}$ be two sequences of random variables adapted to the σ-fields $\{\mathcal{F}_i\}$. Then $\{e_i\}$ and $\{d_i\}$ are said to be ***tangent*** *with respect to $\{\mathcal{F}_i\}$ if, for all $i \geq 1$,*

$$\mathscr{L}(d_i|\mathcal{F}_{i-1}) = \mathscr{L}(e_i|\mathcal{F}_{i-1}), \quad (2)$$

where $\mathscr{L}(d_i|\mathcal{F}_{i-1})$ denotes the conditional law of d_i given $\mathcal{F}_{i-1}$. A sequence $\{e_i\}$ adapted to $\{\mathcal{F}_i\}$ is said to satisfy the ***conditional independence (CI) condition*** *if there exists a σ-field $\mathcal{G} \subset \mathcal{F}$, such that $\{e_i\}$ is conditionally independent given $\mathcal{G}$ and $\mathscr{L}(e_i|\mathcal{F}_{i-1}) = \mathscr{L}(e_i|\mathcal{G})$. A sequence $\{e_i\}$ which satisfies the CI condition and which is also tangent to $\{d_i\}$ is said to be a* ***a decoupled tangent sequence*** *to $\{d_i\}$.*

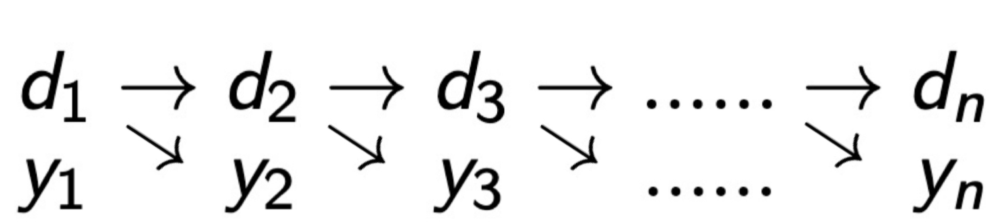

Fig. 1: Construction of Tangent Sequences

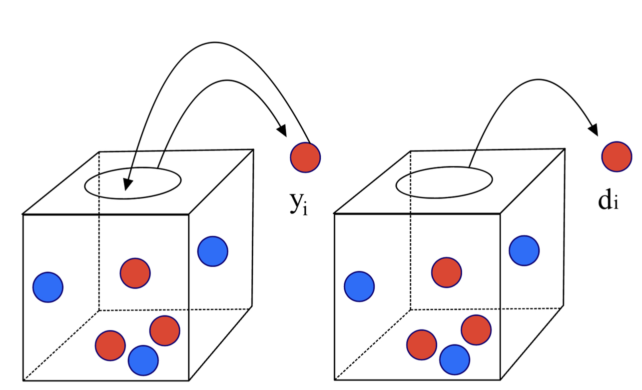

Fig. 2: Example in Random Sampling

In 1999, de la Peña and Giné [2] provide the following lemma: *Let $\{X_i\}$ be a sequence of independent random variables, $\{\tilde{X}_i\}$ and independent copy of $\{X_i\}$ and N be a stopping time adapted to $\sigma(\{X_1, ..., X_j\})$. Let $f_j : \mathbb{R}^j \to \mathbb{R}$, $j = 1, 2, ...,$ be a sequence of measurable functions. Then the sequence*

$$\{f_j(X_1, ..., X_{j-1}; X_j)\mathcal{I}_{\{N \geq j\}}$$

is tangent to

$$\{f_j(X_1, ..., X_{j-1}; \tilde{X}_j)\mathcal{I}_{\{N \geq j\}}$$

with respect to $\mathcal{F}_n := \sigma(\{X_1, ..., X_n; \tilde{X}_1, ..., \tilde{X}_n\})$.

Decoupling Inequality for MGF

The following lemma, introduced by de la Peña in 1994 [1], serves as the central tool for the modern proof of Hanson-Wright inequality. *Let $\{d_i\}$ be a sequence of random variables adapted to an increasing sequence of σ-fields $\{\mathcal{F}_i\} \subset \mathcal{F}$. Let $\{e_i\}$ be any $\{\mathcal{F}_i\}$-tangent sequence to $\{d_i\}$ with $\{e_i\}$ satisfying conditional independence condition given $\mathcal{G}$. Then, for all $\mathcal{G}$-measurable random variables $g \geq 0$ and all finite λ,*

$$\mathbb{E}\left[g\ \exp(\lambda \sum_{i=1}^{n} d_i)\right] \leq \mathbb{E}\left[g^2\ \exp(2\lambda \sum_{i=1}^{n} e_i)\right]. \quad (3)$$

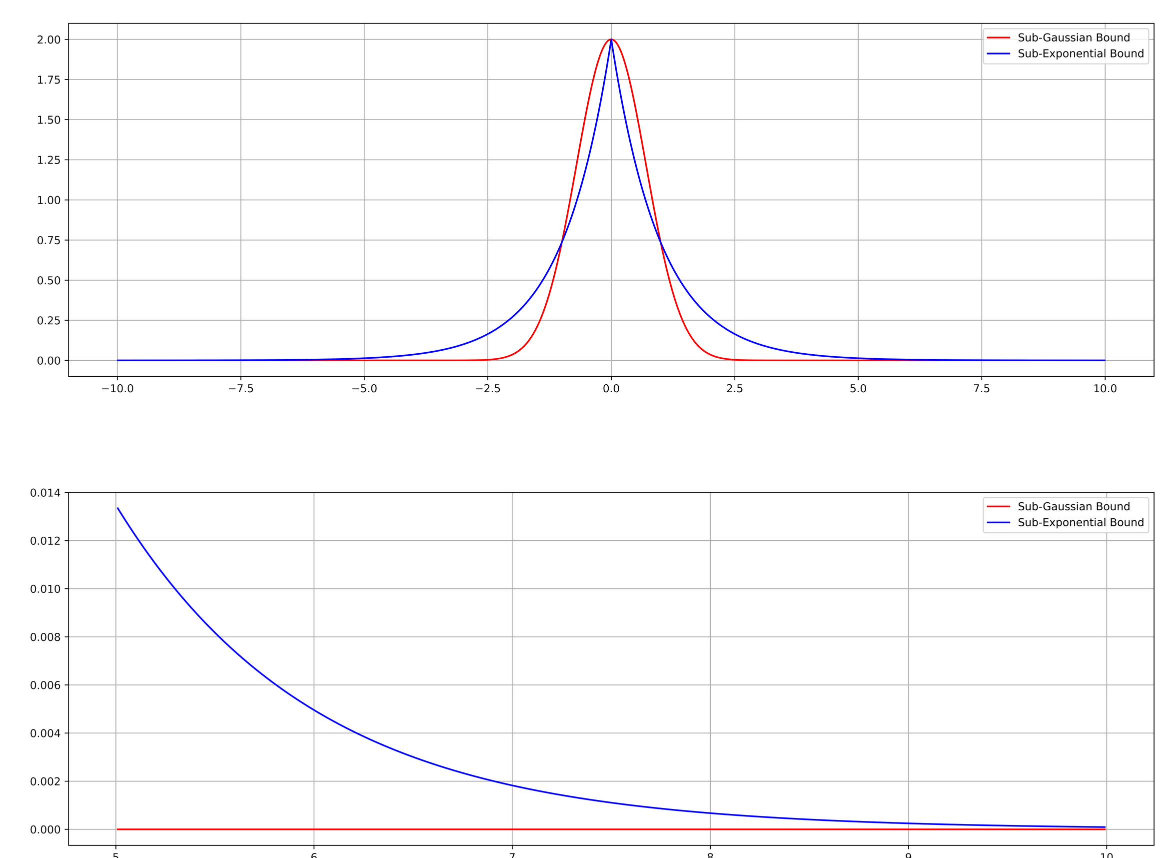

Fig. 3: Sub-Gaussian Bound v.s. Sub-Exponential Bound

Main Theorem

Given $X = (X_1, ..., X_n)$, the entries of which are n independent centered sub-Gaussian random variables such that $\max_{1 \leq i \leq n} \|X_i\|_{\psi_2} \leq K$ for some $K > 0$, and $Y = X^T AX$, a quadratic form such that $A = (a_{ij})_{n \times n}$, with $a_{ii} = 0$ for all i = 1,..., n, the tail probability of Y satisfies

$$\mathbb{P}\{|Y| > t\} \leq 2\exp(-c\min\{\frac{t^2}{K^4\|A\|_{HS}^2}, \frac{t}{K^2\|A\|_2}\}). \quad (4)$$

We provide the pivotal part of the proof, where tangent decoupling appears. Replacing X_i by $\frac{X_i}{K}$ for all i, we can degenerate this proof such that $\max_{1 \leq i \leq n} \|X_i\|_{\psi_2} \leq 1$. We notice that $\mathbb{P}\{|Y| > t\} = \mathbb{P}\{Y > t\} + \mathbb{P}\{Y < -t\}$, and we start to bound $\mathbb{P}\{Y > t\}$. By Markov's inequality, for any $\lambda > 0$,

$$\mathbb{P}(Y > t) = \mathbb{P}(\exp(\lambda Y) > \exp(\lambda t)) \leq \exp(-\lambda t)\mathbb{E}\ \exp(\lambda Y).$$

Then, according to the aforementioned lemma, setting N = n a.s. and for $2 \leq j \leq n$, we have that $d_j = f_j(X_1, ..., X_{j-1}; X_j) = \Sigma_{i=1}^{j-1} 2a_{ij}X_iX_j \in \mathcal{F}_j$ is tangent to $e_j = f_j(X_1, ..., X_{j-1}; \tilde{X}_j) = \Sigma_{i=1}^{j-1} 2a_{ij}X_i\tilde{X}_j \in \tilde{\mathcal{F}}_j$, where $\mathcal{F}_j = \sigma(\{X_1, ..., X_j\})$ and $\tilde{\mathcal{F}}_j = \sigma(\{X_1, ..., X_j; \tilde{X}_i, ..., \tilde{X}_j\}$. And by the decoupling inequality with g = 1 a.s., we have

$$\begin{aligned} \mathbb{E}\ \exp(\lambda Y) &= \mathbb{E}\ \exp(\lambda \sum_{j=2}^{n}\sum_{i=1}^{j-1} 2a_{ij}X_iX_j) = \mathbb{E}\ \exp(\lambda \sum_{j=2}^{n} d_j) \\ &\leq \sqrt{\mathbb{E}\ \exp(2\lambda \sum_{j=2}^{n} e_j)} = \sqrt{\mathbb{E}\ \exp(2\lambda \sum_{j=2}^{n} \tilde{X}_j \sum_{i=1}^{j-1} 2a_{ij}X_i)} \\ &= \left(\mathbb{E}\left[\underbrace{\mathbb{E}\left(\exp(4\lambda \sum_{j=2}^{n} \tilde{X}_j \sum_{i=1}^{j-1} a_{ij}X_i)|\mathcal{F}_n\right)}_{=:E_{\mathcal{F}_n}}\right]\right)^{\frac{1}{2}} \end{aligned}$$

When conditioning $X_1, ..., X_n$, we can to bound $E_{\mathcal{F}_n}$ by the sub-Gaussian properties of $\tilde{X}_1, ..., \tilde{X}_n$, as $\tilde{X}_i$'s are i.i.d. copy of X_i's.

Remarks

- The actual bound we can minimize, with some additional inequalities, is

$$\mathbb{P}(|Y| > t) \leq \begin{cases} 2\exp\left(\frac{-t^2}{256e^2(n-1)\|A\|_{HS}^2}\}\right), & \text{if t} \leq 16\sqrt{2}(n-1)e\|A\|_{HS}, \\ 2\exp\left(\frac{n-1}{2} - \frac{t}{8\sqrt{2}e\|A\|_{HS}}\right), & \text{otherwise}, \end{cases} \quad (5)$$

- To obtain Hanson-Wright inequality, we need to use Bernstein-type inequality (see [4]). The inequality (1) allows the Y to be the quadratic form, where the diagonal entry of A can be nonzero.

Applications and Consequences

- **Concentration of random vectors:** Let $X = (X_1, ..., X_n) \in \mathbb{R}^n$ be a random vector with independent, centered, unit variance, sub-Gaussian coordinates. Let B be a fixed $m \times n$ matrix, then with $K = \max_i \|X_i\|_{\psi_2}$,

$$\mathbb{P}\{|\|BX\|_2^2 - \|B\|_{HS}^2| \geq t\} \leq 2\exp\left(-c\min(t, t^2)\frac{1}{K^4\|B\|^2}\right).$$

  And with the following implication:

$$|\|BX\|_2^2 - \|B\|_{HS}| \geq \delta\|B\|_{HS} \implies |\|BX\|_2^2 - \|B\|_{HS}^2| \geq \epsilon\|_{HS}^2$$

  for some $\epsilon > 0$, we obtain

$$\mathbb{P}\{|\|BX\|_2 - \|B\|_{HS}| \geq t\} \leq 2\exp\left(-\frac{ct^2}{K^4\|B\|^2}\right).$$

- **Distance to a subspace:** Consider the same X aforementioned, and E a subspace of $\mathbb{R}^n$ with dimension d. Then for any $t \geq 0$

$$\mathbb{P}\|dist(X, E)| - \sqrt{n-d}| > t\} \leq 2\exp(-ct^2/K^4).$$

Open Questions

We invite enthusiastic individuals to either substantiate or disprove that the inequality (5) is "sharp". Should this inequality be found not to be sharp, the ensuing question pertains to methods for improving this inequality.

Acknowledgements

The presenter would like to thank Prof. Victor H. de la Peña's supervision and valuable suggestions. The presenter also appreciates the support from Department of Statistics, especially the organizers, Dr. Ronald Neath and Ms. Dood Kalicharan.

References

[1] V. H. de la Peña. "A bound on the moment generating function of a sum of dependent variables with an application to simple random sampling without replacement". In: *Annales de l'IHP Probabilités et statistiques.* Vol. 30. 2. 1994, pp. 197–211.

[2] V. H. de la Peña and E. Giné. *Decoupling: from dependence to independence.* Springer Science & Business Media, 2012.

[3] F. T. Hanson D.L.and Wright. "A bound on tail probabilities for quadratic forms in independent random variables". In: *The Annals of Mathematical Statistics* 42.3 (1971), pp. 1079–1083.

[4] M. Rudelson and R. Vershynin. "Hanson-wright inequality and sub-gaussian concentration". In: (2013).

[5] R. Vershynin. *High-dimensional probability: An introduction with applications in data science.* Vol. 47. Cambridge university press, 2018.