Online Supplement for "Robust Convex Optimization: A New Perspective That Unifies And Extends"

Dimitris Bertsimas Dick den Hertog Jean Pauphilet Jianzhe Zhen

Generalization of the Hanson-Wright inequality

The goal of this online supplement is to derive large deviation bounds on quadratic transforms of sub-Gaussian random variables. In the Gaussian case, the chi-square distribution with n degrees of freedom is defined as the distribution of $\sum_{i=1}^n \tilde{z_i}^2$ where the $\tilde{z_i}$'s are n independent Gaussian random variable with mean 0 and variance 1. In this section we provide analogous results about the distribution of $\tilde{z_i}^2$ and $\sum_{i=1}^n \tilde{z_i}^2$ when $\tilde{z_i}$'s are sub-Gaussian random variables.

We start by a technical lemma regarding the moment generating function of $\tilde{z}^2$.

Lemma OS.1. *Let $\tilde{z}$ be a sub-Gaussian random variable with variance proxy 1, and $\lambda \geq 0$ be a non-negative scalar. Then, for any $\theta \in \mathbb{R}$ and $\beta > 2$ such that $\beta\theta\lambda \leq 1$, we have*

$$\mathbb{E}\left[\exp\left(\theta\lambda\tilde{z}^2\right)\right] \leq \exp\left(\frac{1}{2}\beta\theta\lambda \ln\left(\tfrac{\beta}{\beta-2}\right)\right).$$

Proof. We follow the same proof technique as Bertsimas and Sim [2006]. Introducing $\theta, \beta \in \mathbb{R}$,

$$\mathbb{E}\left[\exp\left(\theta\lambda\tilde{z}^2\right)\right] = \mathbb{E}\left[\exp\left(\beta\theta\lambda\tfrac{\tilde{z}^2}{\beta}\right)\right] = \mathbb{E}\left[\left(e^{\tilde{z}^2/\beta}\right)^{\beta\theta\lambda}\right] \leq \mathbb{E}\left[e^{\tilde{z}^2/\beta}\right]^{\beta\theta\lambda},$$

where the last inequality follows from Jensen's inequality under the condition that $\beta\theta\lambda \leq 1$. Since $\tilde{z}$ is sub-Gaussian with variance proxy 1, we have $\mathbb{E}\left[e^{s\tilde{z}^2/2}\right] \leq \frac{1}{\sqrt{1-s}}$ for any $s \in [0, 1)$ [Wainwright, 2019, Theorem 2.6]. Hence, if $\beta > 2$,

$$\mathbb{E}\left[\exp\left(\theta\lambda\tilde{z}^2\right)\right] \leq \exp\left(\frac{1}{2}\beta\theta\lambda \ln\left(\tfrac{\beta}{\beta-2}\right)\right).$$

□ □

We now state a generalization of the Hanson-Wright inequality [Hanson and Wright, 1971] to the case of sub-Gaussian random variables.

Theorem OS.2. *Let $\tilde{\boldsymbol{z}} = (\tilde{z_1}, \dots, \tilde{z_n}) \in \mathbb{R}^n$ be a random vector with independent sub-Gaussian coordinates $\tilde{z_j}$ with variance proxy σ^2. Let $\boldsymbol{A}$ be an $m \times n$ matrix. Then, denoting $\boldsymbol{\lambda}$ the vector of eigenvalues of $\boldsymbol{A}^\top\boldsymbol{A}$, for every $t > \sigma^2\|\boldsymbol{\lambda}\|_1$,*

$$\mathbb{P}\left(\|\boldsymbol{A}\tilde{\boldsymbol{z}}\|_2^2 > t\right) \leq \left[\sqrt{\frac{et}{\sigma^2\|\boldsymbol{\lambda}\|_1}} \exp\left(-\frac{t}{2\sigma^2\|\boldsymbol{\lambda}\|_1}\right)\right]^{\|\boldsymbol{\lambda}\|_1/\|\boldsymbol{\lambda}\|_\infty}.$$

Remark OS.3. *Hsu et al. [2012] proved a similar tail bound for quadratic forms of sub-Gaussian vectors. Under the assumptions and notations of Theorem OS.2, for a sub-Gaussian random vector with variance proxy 1, they proved that (Theorem 2.1)*

$$\forall s > 0, \quad \mathbb{P}\left(\|\boldsymbol{A}\tilde{\boldsymbol{z}}\|_2^2 > \|\boldsymbol{\lambda}\|_1 + 2\|\boldsymbol{\lambda}\|_2^2\sqrt{s} + 2\|\boldsymbol{\lambda}\|_\infty s\right) \leq e^{-s}.$$

Setting $t = \|\boldsymbol{\lambda}\|_1 + 2\|\boldsymbol{\lambda}\|_2\sqrt{s} + 2\|\boldsymbol{\lambda}\|_\infty s > \|\boldsymbol{\lambda}\|_1$, we would similarly obtain a bound which exponentially decreases in t as $\exp\left(-t/2\|\boldsymbol{\lambda}\|_\infty\right)$.

Proof. /Without loss of generality, we can prove the result for $\sigma^2 = 1$. The general result will follow after observing that

$$\mathbb{P}\left(\|\boldsymbol{A}\tilde{\boldsymbol{z}}\|_2^2 > t\right) = \mathbb{P}\left(\|\boldsymbol{A}\tilde{\boldsymbol{y}}\|_2^2 > t/\sigma^2\right),$$

where $\tilde{\boldsymbol{y}} := \tilde{\boldsymbol{z}}/\sigma^2$ is a sub-Gaussian vector with variance proxy 1. The rest of the proof is divided in three steps.

Step 1: Decoupling

Our proof technique relies on a decoupling mechanism similar to Hsu et al. [2012]. Let us introduce a m-dimensional Gaussian vector $\tilde{\boldsymbol{g}} = (\tilde{g_1}, \dots, \tilde{g_m}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_m)$. We fix some $t > 0$ and $\theta \in \mathbb{R}$. On the one side we have,

$$\mathbb{E}\left[\exp(\theta\, \tilde{\boldsymbol{g}}^\top \boldsymbol{A}\tilde{\boldsymbol{z}})\right] = \mathbb{E}\left[\mathbb{E}\left[\exp(\theta\, \tilde{\boldsymbol{g}}^\top \boldsymbol{A}\tilde{\boldsymbol{z}})\,|\tilde{\boldsymbol{g}}\right]\right] \le \mathbb{E}\left[\exp\left(\frac{\theta^2}{2}\ \|\boldsymbol{A}^\top\tilde{\boldsymbol{g}}\|_2^2\right)\right],$$

where the last inequality follows from the fact that $\tilde{\boldsymbol{z}}$ is sub-Gaussian with variance proxy 1. On the other side,

$$\begin{aligned}
\mathbb{E}\left[\exp(\theta\, \tilde{\boldsymbol{g}}^\top \boldsymbol{A}\tilde{\boldsymbol{z}})\right] &\ge \mathbb{E}\left[\exp(\theta\, \tilde{\boldsymbol{g}}^\top \boldsymbol{A}\tilde{\boldsymbol{z}})\mathbb{I}(\|\boldsymbol{A}\tilde{\boldsymbol{z}}\|_2^2 > t)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\exp(\theta\, \tilde{\boldsymbol{g}}^\top \boldsymbol{A}\tilde{\boldsymbol{z}})\mathbb{I}(\|\boldsymbol{A}\tilde{\boldsymbol{z}}\|_2^2 > t)\,|\tilde{\boldsymbol{z}}\right]\right] \\
&= \mathbb{E}\left[\exp\left(\frac{1}{2}\theta^2\|\boldsymbol{A}\tilde{\boldsymbol{z}}\|_2^2\right)\mathbb{I}(\|\boldsymbol{A}\tilde{\boldsymbol{z}}\|_2^2 > t)\right] \\
&\ge \exp\left(\frac{\theta^2 t}{2}\right)\mathbb{P}\left(\|\boldsymbol{A}\tilde{\boldsymbol{z}}\|_2^2 > t\right).
\end{aligned}$$

All in all,

$$\mathbb{P}\left(\|\boldsymbol{A}\tilde{\boldsymbol{z}}\|_2^2 > t\right) \le \exp\left(-\frac{\theta^2 t}{2}\right)\mathbb{E}\left[\exp\left(\frac{1}{2}\theta^2\ \|\boldsymbol{A}^\top\tilde{\boldsymbol{g}}\|_2^2\right)\right].$$

Step 2: Bounding the moment generating function of the Gaussian chaos

Consider the singular value decomposition of $\boldsymbol{A}$, $\boldsymbol{A} = \boldsymbol{U}^\top\boldsymbol{\Lambda}\boldsymbol{V}$ where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthonormal matrices and $\boldsymbol{\Lambda} \in \mathbb{R}^{m\times n}$ is a diagonal matrix whose diagonal coefficients are $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_{\min(m,n)}}$. By rotational invariance, $\tilde{\boldsymbol{y}} := \boldsymbol{U}\tilde{\boldsymbol{g}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_m)$ and $\|\boldsymbol{A}^\top\tilde{\boldsymbol{g}}\|_2 = \|\boldsymbol{\Lambda}\tilde{\boldsymbol{y}}\|_2$. Hence,

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\frac{1}{2}\theta^2\ \|\boldsymbol{A}^\top\tilde{\boldsymbol{g}}\|_2^2\right)\right] &= \mathbb{E}\left[\exp\left(\frac{1}{2}\theta^2\ \|\boldsymbol{\Lambda}^\top\tilde{\boldsymbol{y}}\|_2^2\right)\right] \\
&= \prod_{i=1}^{\min(m,n)}\mathbb{E}\left[\exp\left(\frac{1}{2}\theta^2\lambda_i\tilde{y_i}^2\right)\right].
\end{aligned}$$

We now apply Lemma OS.1 to bound each moment generating function: Given $\beta > 2$ such that $\frac{1}{2}\beta\lambda_i\sigma^2 \le 1$, we have

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\theta^2\ \|\boldsymbol{A}^\top\tilde{\boldsymbol{g}}\|_2^2\right)\right] \le \exp\left(\frac{1}{4}\beta\theta^2\|\boldsymbol{\lambda}\|_1\ln\left(\tfrac{\beta}{\beta-2}\right)\right),$$

where $\|\boldsymbol{\lambda}\|_1 = \sum_{i=1}^{\min(m,n)}\lambda_i$.

Step 3: Conclusion

Putting pieces together, we finally obtain

$$\mathbb{P}\left(\|\boldsymbol{A}\tilde{\boldsymbol{z}}\|_2^2 > t\right) \leq \exp\left(-\frac{1}{2}\theta^2 t + \frac{1}{4}\beta\theta^2\|\boldsymbol{\lambda}\|_1 \ln\left(\tfrac{\beta}{\beta-2}\right)\right).$$

We fix the value of θ such that $\frac{1}{2}\beta\|\boldsymbol{\lambda}\|_\infty\theta^2 = 1$. As a result the condition $\frac{1}{2}\beta\lambda_i\theta^2 \leq 1$ is naturally satisfied for all $i = 1, \ldots, n$ and

$$\mathbb{P}\left(\|\boldsymbol{A}\tilde{\boldsymbol{z}}\|_2^2 > t\right) \leq \exp\left(-\frac{t}{\beta\|\boldsymbol{\lambda}\|_\infty} + \frac{\|\boldsymbol{\lambda}\|_1}{2\|\boldsymbol{\lambda}\|_\infty} \ln\left(\tfrac{\beta}{\beta-2}\right)\right).$$

Taking derivatives and choosing the best β , we have

$$\beta = \frac{2t}{t - \|\boldsymbol{\lambda}\|_1} > 2 \text{ as long as } t > \|\boldsymbol{\lambda}\|_1.$$

Substituting and simplifying, we finally obtain

$$\mathbb{P}\left(\|\boldsymbol{A}\tilde{\boldsymbol{z}}\|_2^2 > t\right) \leq \exp\left(\frac{\|\boldsymbol{\lambda}\|_1}{2\|\boldsymbol{\lambda}\|_\infty}\right) \exp\left(\frac{1}{2}\frac{\|\boldsymbol{\lambda}\|_1}{\|\boldsymbol{\lambda}\|_\infty} \ln\left(\frac{t}{\|\boldsymbol{\lambda}\|_1}\right)\right) \exp\left(-\frac{t}{2\|\boldsymbol{\lambda}\|_\infty}\right),$$

for $t > \|\boldsymbol{\lambda}\|_1\sigma^2$. □ □

We conclude this section with a weaker, yet simpler, version of the Hanson-Wright inequality of Theorem OS.2.

Corollary OS.4. *Under the same assumptions as Theorem OS.2, for every* $t > \sigma^2\|\boldsymbol{\lambda}\|_1$*,*

$$\mathbb{P}\left(\|\boldsymbol{A}\tilde{\boldsymbol{z}}\|_2^2 > t\right) \leq \sqrt{\frac{et}{\sigma^2\|\boldsymbol{\lambda}\|_1}} \exp\left(-\frac{t}{2\sigma^2\|\boldsymbol{\lambda}\|_1}\right).$$

Proof. It follows from Theorem OS.2 after observing that $\|\boldsymbol{\lambda}\|_1/\|\boldsymbol{\lambda}\|_\infty \geq 1$. □ □

Remark OS.5. *Notice that the bound in Corollary OS.4 only involves* $\|\boldsymbol{\lambda}\|_1$*, which can be computed directly from* $\boldsymbol{A}$ *as* $\|\boldsymbol{\lambda}\|_1 = tr(\boldsymbol{A}^\top\boldsymbol{A})$*.*

References

Dimitris Bertsimas and Melvyn Sim. Tractable approximations to robust conic optimization problems. *Mathematical Programming*, 107(1-2):5–36, 2006.

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.

Daniel Hsu, Sham Kakade, Tong Zhang, et al. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17(52):1–6, 2012.

Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, UK, 2019.