MATH888: High-dimensional probability and statistics Fall 2021

Lecture 12 — October 4, 2021

Sebastien Roch, UW-Madison *Scribe: Matthew Zurek*

1 Overview

In the last lecture we proved the Hanson-Wright inequality. We also mentioned one corollary:

Theorem 1. *(Theorem 6.3.2 in [1]) For any $m \times n$ matrix B and random vector $X = (X_1, \dots, X_n)$ where the $(X_i)_{i=1}^n$ are independent, mean-zero, unit-variance, and subgaussian, we have*

$$\| \|BX\|_2 - \|B\|_F \|_{\psi_2} \leq C \|B\| \max_{i \in \{1,\dots,n\}} \|X_i\|_{\psi_2}^2$$

In this lecture we began a deeper discussion of random vectors.

2 Isotropic Random Vectors

We saw the definition of isotropic random vectors and some useful facts about them in the slides `https://people.math.wisc.edu/~roch/hdps/roch-hdps-slides12.pdf`.

3 Sub-Gaussian Random Vectors

Definition 2. *A random vector X which takes values in $\mathbb{R}^n$ is sub-gaussian if the quantity*

$$\|X\|_{\psi_2} := \sup_{v \in S^{n-1}} \|\langle X, v \rangle\|_{\psi_2}$$

is finite. $\|X\|_{\psi_2}$ is defined to be the sub-gaussian norm of X.

Recall that S^{n-1} is the set of all unit vectors in $\mathbb{R}^n$. Also note that this is not a circular definition because we have defined the sub-gaussian norm of a *vector* in terms of the sub-gaussian norms of its (*scalar*) marginals. These marginals $\langle X, v \rangle$ can be thought of as the projection of X onto the direction determined by $v \in \mathbb{R}^n$.

First we showed that one way to get a sub-gaussian vector is to let each coordinate be an individually independent sub-gaussian random variable.

Lemma 3. *(Lemma 3.4.2 in [1]) Let $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ be a random vector such that the $(X_i)_{i=1}^n$ are independent, mean-zero, and sub-gaussian. Then*

$$\|X\|_{\psi_2} \lesssim \max_{i \in \{1,\dots,n\}} \|X_i\|_{\psi_2}.$$

Proof. Fix $v \in S^{n-1}$. Then

$$\begin{aligned}
\|\langle X, v\rangle\|_{\psi_2}^2 &= \left\|\sum_{i=1}^n v_i X_i\right\|_{\psi_2}^2 \\
&\precsim \sum_{i=1}^n \|v_i X_i\|_{\psi_2}^2 \\
&= \sum_{i=1}^n v_i^2 \|X_i\|_{\psi_2}^2 \\
&\leq \left(\max_{i\in\{1,\ldots,n\}} \|X\|_{\psi_2}\right) \sum_{i=1}^n v_i^2 \\
&= \max_{i\in\{1,\ldots,n\}} \|X\|_{\psi_2}
\end{aligned}$$

□

The non-trivial step (the first $\precsim$) was shown in class when we discussed Hoeffding's inequality (in fact, owing to the equivalence between subgaussian tail bounds and subgaussian norms, this is equivalent to Hoeffding's inequality).

Next we saw an example of a sub-gaussian random vector which does not have independent coordinates.

Theorem 4. *(Theorem 3.4.6 in [1]) Suppose that*

$$X \sim \mathit{Unif}\left(\sqrt{n} S^{n-1}\right)$$

that is, X is a random vector in $\mathbb{R}^n$ uniformly distributed on the sphere of radius $\sqrt{n}$ centered at the origin. Then X is sub-gaussian and

$$\|X\|_{\psi_2} \leq C.$$

Proof. Due to the rotation invariance of the multivariate standard normal distribution, we can realize X with the desired distribution by letting $g \sim N(0, I_n)$ and then setting

$$X = \sqrt{n}\frac{g}{\|g\|_2}.$$

To bound $\|X\|_{\psi_2}$ we need to bound $\|\langle X, v\rangle\|_{\psi_2}$ for all $v \in S^{n-1}$, but again due to rotational symmetry $\langle X, v\rangle$ has the same distribution for any $v \in S^{n-1}$, so for convenience we can take $v = e_1$ (the unit vector along the positive first axis). Thus $\langle X, e_1\rangle = X_1$.

Before proceeding, we outline the approach. Recall that by the tail bound characterization of subgaussians, it suffices to show that $\mathbb{P}(|X_1| \geq t) \precsim \exp(-C't^2)$ for all $t \geq 0$. Furthermore, since

$$\mathbb{P}\left(|X_1| \geq t\right) = \mathbb{P}\left(\sqrt{n}\frac{|g_1|}{\|g\|_2} \geq t\right)$$

we will use the fact (from a previous lecture and also the Theorem 1 above) that $\|g\|_2$ is close to $\sqrt{n}$ with high probability, and as long as we are in that event, we only need to bound $\mathbb{P}(|g_1| \geq t)$ (which is immediate since g_1 is a gaussian).

To execute this approach, first define the event $\mathcal{E} = \{\|g\|_2 \geq \sqrt{n}/2\}$. Then

$$\begin{aligned}
\mathbb{P}\left(\sqrt{n}\frac{|g_1|}{\|g\|_2} \geq t\right) &= \mathbb{P}\left(\left\{\sqrt{n}\frac{|g_1|}{\|g\|_2} \geq t\right\} \cap \mathcal{E}\right) + \mathbb{P}\left(\left\{\sqrt{n}\frac{|g_1|}{\|g\|_2} \geq t\right\} \cap \mathcal{E}^c\right) \\
&\leq \mathbb{P}\left(|g_1| \geq \frac{t}{2}\right) + \mathbb{P}\left(\left\{\sqrt{n}\frac{|g_1|}{\|g\|_2} \geq t\right\} \cap \mathcal{E}^c\right) \\
&\leq \mathbb{P}\left(|g_1| \geq \frac{t}{2}\right) + \mathbb{P}\left(\mathcal{E}^c\right)
\end{aligned}$$

(both inequalities came from upper-bounding an event by an event which contains it). Now we handle each of these terms individually. Because g_1 is a gaussian, we have that

$$\mathbb{P}\left(|g_1| \geq \frac{t}{2}\right) \leq 2\exp\left(-\frac{t^2}{8}\right).$$

In a previous lecture we saw that $\|\|g\|_2 - \sqrt{n}\|_{\psi_2} \lesssim 1$ (one could also apply Theorem 1 above with $B = I_n$ to obtain this). By the tail bound for sub-gaussian random variables, this implies that (for some absolute constant c)

$$\mathbb{P}\left(|\|g\|_2 - \sqrt{n}| \geq r\right) \leq 2\exp\left(-cr^2\right)$$

and so plugging in $r = \sqrt{n}/2$ we obtain that $\mathbb{P}(\mathcal{E}^c) \leq 2\exp(-cn/4)$.

Using both these bounds we conclude that

$$\mathbb{P}\left(\sqrt{n}\frac{|g_1|}{\|g\|_2} \geq t\right) \leq 2\exp\left(-\frac{t^2}{8}\right) + 2\exp\left(-cn/4\right).$$

Because this second term is constant in t, it looks like we might not have the desired tail bound decay as t gets large. However, this is not an issue because in the case that $t > \sqrt{n}$, we do not even need this bound: trivially $\sqrt{n}|g_1|/\|g\|_2 = |X_1| \leq \|X\|_2$ and $\|X\|_2$ is almost surely equal to $\sqrt{n}$, meaning $P(|X_1| > t) \leq P(|X_1| > \sqrt{n}) = 0$. Therefore we only need to check the tail bound $\mathbb{P}(|X_1| \geq t)$ for $t \leq \sqrt{n}$, and in that case the above bound implies that

$$\begin{aligned}
\mathbb{P}\left(\sqrt{n}\frac{|g_1|}{\|g\|_2} \geq t\right) &\leq 2\exp\left(-\frac{t^2}{8}\right) + 2\exp\left(-c\frac{t^2}{4}\right) \\
&\leq 4\exp(-c't^2)
\end{aligned}$$

as desired (absorbing absolute constants into c'). □

References

[1] R. Vershynin, *High-Dimensional Probability: An introduction with Applications in Data Science*, Cambridge University Press, 2008.