TOPICS IN PROBABILITY. BETWEEN PARTS I AND II: RANDOM PROJECTIONS, FIRST STEPS TOWARDS UNIVERSALITY

Exercise sheet 6: log-Sobolev inequality and random projections

Exercise 1 (Proof of Gaussian log-Sobolev inequality). *(1) Complete the proof (sketched in class) of Gaussian log-Sobolev inequality*

$$\mathrm{Ent}[f] \leq \frac{1}{2}\mathbb{E}[\|\nabla f\|^2/f]$$

under the assumption on boundedness of derivatives of $f > 0$ up to third order by filling in the missing details.

(2) Extend the result to continuously differentiable functions $f \geq 0$.

Proof. Using tensorization of entropy we can reduce the problem to the one-dimensional case: to show is that for $f : \mathbb{R} \to [0, \infty) \in C^1$, $\mathrm{Ent}[f] \leq \frac{1}{2}\mathbb{E}[(f')^2/f]$, where $\mathbb{E}$ is taken w.r.t. standard Gaussian measure.

We start with the proof under the assumption that $f > 0 \in C^3$ with $|f|, |f'|, |f''|$ and $|f'''|$ bounded, say, by $C > 0$. For $n \in \mathbb{N}$, we define $g : \mathbb{R}^n \to \mathbb{R}_+$ by $g(x_1, \dots, x_n) = f(\sum_{i=1}^n x_i/\sqrt{n})$. Let $X = (X_1, \dots, X_n)$ be a standard Gaussian vector. Since $f(X_1)$ has the same law as $g(X)$, we need to show that $\mathrm{Ent}[g(X)] \leq \frac{1}{2}\mathbb{E}[((f')^2/f)(X_1)]$. By tensorisation of entropy,

$$\mathrm{Ent}[g] \leq \mathbb{E}\left[\sum_{i=1}^n \mathrm{Ent}_i[g]\right],$$

where $\mathrm{Ent}_i[g] = \mathbb{E}_i[g(X)\log\frac{g(X)}{\mathbb{E}_i[g(X)]}]$ and $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot|(X_j)_{j\neq i}]$. Recall that $\mathbb{P}[|X_i| > (\log n)^2] \leq 2e^{-(\log n)^2/2}$ and by assumptions on f, $g\log g$ is uniformly bounded, thus,

$$\mathrm{Ent}[g] \leq \mathbb{E}\left[\sum_{i=1}^n \mathbb{E}_i[g(X)\log\frac{g(X)}{\mathbb{E}_i[g(X)]}\mathbf{1}_{\{|X_i|\leq(\log n)^2\}}]\right] + \mathcal{O}(ne^{-(\log n)^2}).$$

For any $1 \leq i \leq n$ and $x \in \mathbb{R}^n$, set $x^i = (x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n)$. Then, by Taylor's theorem, since $x - x^i = (0, \dots, 0, x_i, 0, \dots, 0)$:

$$g(x) = g(x^i) + \partial_i g(x^i)x_i + \frac{1}{2}\partial_{ii}g(x^i)x_i^2 + R_{(i,i,i)}(x^i)x_i^3$$

with $|R_{(i,i,i)}(y)|$ bounded by maximum of the supremum norm of any third order partial derivative, thus (since $\partial_{ijk}g = n^{-3/2}f'''$) by $C/n^{3/2}$. Furthermore,

$$\begin{aligned}\mathbb{E}_i[g(X)] &= g(X^i) + \partial_i g(X^i)\mathbb{E}[X_i] + \frac{1}{2}\partial_{ii}g(X^i)\mathbb{E}[X_i^2] + \mathcal{O}(n^{-3/2})\\ &= g(X^i) + \frac{1}{2}\partial_{ii}g(X^i) + \mathcal{O}(n^{-3/2}).\end{aligned}$$

Here we have additionally used that X_i is independent of the remaining coordinates $(X_j)_{j\neq i}$, properties of conditional expectation and the fact that $\mathbb{E}[|X_i|^3]$ is a finite constant. Since by assumption $f > 0$ is bounded away from 0 and clearly so is g then, we further get that for n sufficiently large,

$$\frac{g(X^i)}{\mathbb{E}_i[g(X)]} = \left(1 + \frac{1}{2}\frac{\partial_{ii}g(X^i)}{g(X^i)} + \mathcal{O}(n^{-3/2})\right)^{-1} = 1 - \frac{1}{2}\frac{\partial_{ii}g(X^i)}{g(X^i)} + \mathcal{O}(n^{-3/2})$$

where we have used that $\partial_{ii}g = f''/n = \mathcal{O}(1/n)$ uniformly on $\mathbb{R}^n$, and that $(1+x)^{-1} = 1-x+\mathcal{O}(x^2)$ for all $|x|$ sufficiently small. Note further that $\frac{\partial_i g}{g}\frac{\partial_{ii}g}{g} = \mathcal{O}(n^{-3/2})$, $(\frac{\partial_{ii}g}{g})^2 = o(n^{-3/2})$,

$$\frac{g(X)}{\mathbb{E}_i[g(X)]} = 1 + \frac{\partial_i g(X^i)}{g(X^i)}X_i + \frac{1}{2}\frac{\partial_{ii}g(X^i)}{g(X^i)}(X_i^2 - 1) + \mathcal{O}(n^{-3/2}(1 + X_i + X_i^2 + X_i^3)).$$

Here $\mathcal{O}(n^{-3/2}(1 + X_i + X_i^2 + X_i^3))$ means that variables X_i to the respective powers $0, 1, 2, 3$ appear with some pre-coefficients (might be different and random) that are in $\mathcal{O}(n^{-3/2})$ (deterministic bound).

From now on, let us work on the event $\{|X_i| \leq (\log n)^2\}$. We then get that $\frac{g(X)}{\mathbb{E}_i[g(X)]} = 1 + \frac{\partial_i g(X^i)}{g(X^i)}X_i + \frac{1}{2}\frac{\partial_{ii}g(X^i)}{g(X^i)}(X_i^2 - 1) + \mathcal{O}(n^{-3/2}\log^6 n) = 1 + \mathcal{O}(n^{-1/2})$. Using the above expansion of g and the fact that $\log(1+x) = x - \frac{x^2}{2} + \mathcal{O}(x^3)$ for all x close to zero, we obtain

$$\begin{aligned} g(X)\log\frac{g(X)}{\mathbb{E}_i[g(X)]} &= \left(g(X^i) + \partial_i g(X^i)X_i + \frac{1}{2}\partial_{ii}g(X^i)X_i^2 + \mathcal{O}(n^{-3/2}\log^6 n)\right) \times \\ &\quad \times \left(\frac{\partial_i g(X^i)}{g(X^i)}X_i + \frac{1}{2}\frac{\partial_{ii}g(X^i)}{g(X^i)}(X_i^2 - 1) - \frac{1}{2}(\frac{\partial_i g(X^i)}{g(X^i)})^2 X_i^2 + \mathcal{O}(n^{-3/2}\log^6 n)\right) \\ &= \partial_i g(X^i)X_i + \frac{1}{2}\partial_{ii}g(X^i)(X_i^2 - 1) + \frac{1}{2}\frac{(\partial_i g(X^i))^2}{g(X^i)}X_i^2 + \mathcal{O}(n^{-3/2}\log^6 n). \end{aligned}$$

Therefore, by properties of conditional expectation,

$$\begin{aligned} \mathbb{E}_i[g(X)\log\frac{g(X)}{\mathbb{E}_i[g(X)]}\mathbf{1}_{\{|X_i|\leq(\log n)^2\}}] &= \mathcal{O}(n^{-3/2}\log^6 n) + \partial_i g(X^i)\mathbb{E}[X_i\mathbf{1}_{\{|X_i|\leq(\log n)^2\}}] \\ &\quad + \frac{1}{2}\partial_{ii}g(X^i)\mathbb{E}[(X_i^2 - 1)\mathbf{1}_{\{|X_i|\leq(\log n)^2\}}] + \frac{1}{2}\frac{(\partial_i g(X^i))^2}{g(X^i)}\mathbb{E}[X_i^2\mathbf{1}_{\{|X_i|\leq(\log n)^2\}}]. \end{aligned}$$

By symmetry of centered Gaussian law, $\mathbb{E}[X_i\mathbf{1}_{\{|X_i|\leq(\log n)^2\}}] = 0$; and since $\mathbb{E}[X_i^2] = 1$ and $\mathbb{E}[(X^2+1)\mathbf{1}_{\{|X_i|>(\log n)^2\}}] \leq 2(2e^{-(\log n)^2/2} + (\log n)^2 e^{-(\log n)^2/2}) \ll n^{-3/2}$, we further obtain

$$\mathbb{E}_i[g(X)\log\frac{g(X)}{\mathbb{E}_i[g(X)]}\mathbf{1}_{\{|X_i|\leq(\log n)^2\}}] = \frac{1}{2}\frac{(\partial_i g(X^i))^2}{g(X^i)} + \mathcal{O}(n^{-3/2}\log^6 n).$$

Altogether, since X_i's are i.i.d., and $\sum_{i=2}^n X_i/\sqrt{n}$ has the same law as $\sqrt{\frac{n-1}{n}}X_1$, which, in turn, converges to X_1 (a.s. or in law), by continuity and boundedness of $frac(f')^2 f$,

$$\begin{aligned}
\mathrm{Ent}[g] &\le \frac{1}{2}\mathbb{E}\left[\sum_{i=1}^n \frac{(\partial_i g(X^i))^2}{g(X^i)}\right] + \mathcal{O}(n^{-1/2}\log^6 n) \\
&= \frac{1}{2}\mathbb{E}\left[\frac{(f')^2}{f}(\sqrt{\frac{n-1}{n}}X_1)\right] + \mathcal{O}(n^{-1/2}\log^6 n) \xrightarrow{n\to\infty} \frac{1}{2}\mathbb{E}\left[\frac{(f')^2}{f}(X_1)\right].
\end{aligned}$$

Let us now extend the result to strictly positive C^1 functions f. Note that in this case, $g \coloneqq \sqrt{f}$ is continuously differentiable, and the desired inequality rewritten in terms of g takes form $\mathrm{Ent}[g^2] \le 2\mathbb{E}[(g')^2]$. We will show that the latter inequality is true for any $g : \mathbb{R} \to \mathbb{R} \in C^1$. To this end, assume that $\mathbb{E}[(g')^2] < \infty$, otherwise there is nothing to show, and note that by the first part of this exercise we know that if $g \neq 0 \in C^3$ (so that $f = g^2 > 0$) with g, g', g'' and g''' uniformly bounded (say, by $C > 0$ finite), the inequality is verified. Furthermore, since $x \mapsto x\log x$ is uniformly continuous on $[0,(C+1)^2]$, $\mathrm{Ent}[(g+\varepsilon)^2] \to \mathrm{Ent}[g^2]$ as $\varepsilon \in (0,1)$ tends to zero. This observation allows us to extend the result to $g \in C^3(\mathbb{R},\mathbb{R})$ with g, g', g'' and g''' uniformly bounded. We now proceed as in the proof of the extension of Gaussian Poincaré inequality, see Exercise 1 Sheet 3. Namely, let η_ε be a standard mollifier and ψ_n be a smooth cut-off function with uniformly (independent of n) bounded derivatives up to third order and taking values 1 in $[-n,n]$ and 0 in $[-n-1,n+1]^c$. Set $g_\varepsilon = g * \eta_\varepsilon$ and $g_{\varepsilon,n} = \psi_n g_\varepsilon \in C_c^\infty$. By the above,

$$\mathrm{Ent}[g_{\varepsilon,n}^2] \le 2\mathbb{E}[(g'_{\varepsilon,n})^2]$$

We saw in the proof of Exercise 1 Sheet 3 that by first taking limit $\varepsilon \to 0$ and then $n \to \infty$, the r.h.s. converges to $2\mathbb{E}[(g')^2]$ as desired. Note that by uniform continuity of $x \mapsto x\log x$ on any compact subset of $[0,\infty)$ and uniform convergence over compacts of g_ε towards g, we may conclude that $\mathrm{Ent}[g_{\varepsilon,n}^2]$ converges to $\mathrm{Ent}[g_n^2]$ as $\varepsilon \to 0$, and also $\mathbb{E}[g_n^2 \log g_n^2 \mathbf{1}_{g^2\le 1}]$ to $\mathbb{E}[g^2\log g^2 \mathbf{1}_{g^2\le 1}]$ as n tends to infinity. By Monotone convergence theorem, we further deduce that $\mathbb{E}[g_n^2]$ converges to $\mathbb{E}[g^2]$ and $\mathbb{E}[\mathbf{1}_{[-n,n]} g_n^2 \log g_n^2 \mathbf{1}_{g^2>1}] = \mathbb{E}[\mathbf{1}_{[-n,n]} g^2 \log g^2 \mathbf{1}_{g^2>1}]$ towards $\mathbb{E}[g^2\log g^2\mathbf{1}_{g^2>1}]$ (since $x\log x$ is increasing on $[1,\infty)$) as n tends to infinity. It only remains to show that $\mathbb{E}[\mathbf{1}_{(-n-1,-n)\cup(n,n+1)} g^2\psi_n^2 \log(g^2\psi_n^2)\mathbf{1}_{g^2>1}]$ vanishes in the limit $n \to \infty$. We recall that in the proof of Exercise 1 Sheet 3, we further proved that $g^2(x)e^{-x^2/2} \to 0$ as $|x| \to \infty$, hence, $|\mathbb{E}[\mathbf{1}_{(-n-1,-n)\cup(n,n+1)} g^2\psi_n^2 \log(\psi_n^2)\mathbf{1}_{g^2>1}]| \le \sup_{x\in[0,1]}(x\log x)\max_{y\in(-n-1,-n)\cup(n,n+1)} g^2(y)e^{-y^2/2} \to 0$ as $n \to \infty$. Analogously we show that $g^2(x)\log g^2(x) e^{-x^2/2} \to 0$ as $|x| \to \infty$ (on the event $g^2 > 1$) to complete this part of the proof. Suppose by contradiction (wlog) that $\liminf_{x\to\infty} g^2(x)\log g^2(x)e^{-x^2/2} > 0$, this implies that $g^2(x) = \Omega(e^{x^2/2}/x^2)$, and hence, $(g')^2(x) = \Omega(e^{x^2/2}(1-1/x^2)^2) = \Omega(e^{x^2/2})$ as $x \to \infty$. The latter contradicts our assumption that $\mathbb{E}[(g')^2] < \infty$.

Let $f \ge 0 \in C^1$, then by the previous part, for any $\varepsilon > 0$, $\mathrm{Ent}[f+\varepsilon] \le \frac{1}{2}\mathbb{E}[(f')^2/(f+\varepsilon)]$. Note that $\mathbb{E}[(f')^2/(f+\varepsilon)] \le \mathbb{E}[(f')^2/f]$ and $\mathbb{E}[(f+\varepsilon)\log(f+\varepsilon)\mathbf{1}_{f>1/e}] \ge \mathbb{E}[f\log(f)\mathbf{1}_{f>1/e}]$ (as $x\log x$ is increasing on $[1/e,\infty)$). By uniform continuity of $x \mapsto x\log x$ on any compact subset of $[0,\infty)$, $\mathbb{E}[(f+\varepsilon)\log(f+\varepsilon)\mathbf{1}_{f\le 1/e}]$ converges as $\varepsilon \in (0,1/2] \to 0$ towards $\mathbb{E}[f\log(f)\mathbf{1}_{f\le 1/e}]$. The remaining term $\mathbb{E}[(f+\varepsilon)]\log(\mathbb{E}[f+\varepsilon])$ clearly converges to $\mathbb{E}[f]\log(\mathbb{E}[f])$. $\square$

Exercise 2 (log-Sobolev inequality for general Gaussians)**.** *Let $X \in \mathbb{R}^n$ has centered Gaussian distribution with covariance matrix Ξ. Show that for any continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ the following holds,*

$$\mathrm{Ent}[f^2] \le 2\mathbb{E}[\langle \Xi \nabla f(X), \nabla f(X) \rangle].$$

Proof. Let A be the square root of the positive semi-definite matrix Ξ. Then, X is equal in law to AZ, where $Z \in \mathbb{R}^n$ is a standard Gaussian vector. Let us introduce $g(x) \coloneqq f(Ax)$, which is continuously differentiable with $\nabla g(x) = A^T(\nabla f)(Ax)$. By applying log-Sobolev inequality to $g(Z)$, we get

$$\mathrm{Ent}[f^2] = \mathrm{Ent}[g^2(Z)] \le 2\mathbb{E}[\|\nabla g(Z)\|^2] = 2\mathbb{E}[\langle \Xi \nabla f(X), \nabla f(X) \rangle].$$

□

Exercise 3 (log-Sobolev implies Poincaré)**.** *Prove that Gaussian log-Sobolev inequality (for standard Gaussian vector) implies Gaussian Poincaré inequality.*

Hint: Let $\varepsilon > 0$ be small and use the log-Sobolev inequality for $(1 + \varepsilon f)$. Show that $\mathrm{Ent}[(1+\varepsilon f)^2] = 2\varepsilon^2 \mathrm{Var}[f(X)] + \mathcal{O}(\varepsilon^3)$.

Proof. Let $\varepsilon > 0$ be small. By Gaussian log-Sobolev inequality, we obtain

$$\mathrm{Ent}[(1+\varepsilon g)^2] \le 2\varepsilon^2 \mathbb{E}[\|\nabla g\|^2].$$

Use that $\log(1+y) = y - y^2/2 + \mathcal{O}(y^3)$ for $y \to 0$. In particular, you get that

$$\mathrm{Ent}[(1+y)^2] = 2\mathbb{E}\left[(1+2y)\left(y - \mathbb{E}[y] - \frac{1}{2}(y^2 - \mathbb{E}[y]^2)\right) + \mathcal{O}(y^3)\right] = 2\mathrm{Var}[y] + \mathcal{O}(\mathbb{E}[y^3]).$$

By plugging εg, resp., instead of y we get that $\mathrm{Ent}[(1+\varepsilon g)^2] = 2\varepsilon^2 \mathrm{Var}[g] + \mathcal{O}(\varepsilon^3)$. Together with the bounds for the entropy dividing by ε^2 and taking the limit $\varepsilon \to 0$ yields the desired result. □

Exercise 4 (Weak Poincaré lemma)**.** *Let $m \in \mathbb{N}$ be fixed. Consider a random vector X^N uniformly distributed on the unit sphere $S^{N-1} \subset \mathbb{R}^N$. Let $X^{m,N}$ denote the vector consisting of its first m coordinates. Prove that as N tends to infinity the law of $\sqrt{N}X^{m,N}$ converges to standard Gaussian distribution in dimension m.*

Proof. Let Y be a N-dimensional standard Gaussian vector. Then $Y/\|Y\|_2$ is uniformly distributed on the $N-1$ sphere. There are two ways to see this. The first one is based on the fact that the unique law on the unit N-dimensional vectors which is invariant under rotations is the uniform distribution on the $N-1$-sphere. So, since standard Gaussian distribution is invariant under rotations and rotations do not change the value of the norm of the vector, we conclude that $Y/\|Y\|_2$ has to be uniformly distributed on the sphere. Alternatively one can just consider $\mathbb{E}[f(Y/\|Y\|_2)]$ for all measurable non-negative functions f, changing to polar coordinates in the integral (w.r.t. the Gaussian density) yields the result. So, instead of X we will now work with $Y/\|Y\|_2$. With the same notation as for $X^{m,N}$, we further conclude that $X^{m,N}$ has the same law as $Y^{m,N}/\|Y\|_2$.

The idea is to use concentration of $\|Y\|_2$ around $\sqrt{N}$. Let us write f_k for the density of the standard k-dimensional Gaussian vector. For any continuity set A of standard m-dimensional

Gaussian measure and Z_i i.i.d. standard normal random variables,

$$R_{N,\varepsilon}(A) := \mathbb{P}\left[\sqrt{N}\frac{Y^{m,N}}{\|Y\|_2} \in A\right] - \mathbb{P}[\mathcal{N}(0, I_m) \in A]$$
$$= \int \mathrm{d}v f_m(v) \left(\mathbb{P}\left[\frac{\sqrt{N}v}{\sqrt{|v|^2 + \sum_{k=1}^{N-m} Z_k^2}} \in A \right] - \mathbf{1}_{\{v \in A\}} \right).$$

Furthermore, for any $\varepsilon > 0$

$$\mathbb{P}\left[\frac{v}{\sqrt{\frac{|v|^2}{N} + \frac{1}{N}\sum_{k=1}^{N-m} Z_k^2}} \in A \right]$$
$$\leq \mathbb{P}\left[\frac{\sqrt{N}v}{\sqrt{|v|^2 + \sum_{k=1}^{N-m} Z_k^2}} \in A, \left| \sqrt{\frac{|v|^2}{N} + \frac{1}{N}\sum_{k=1}^{N-m} Z_k^2} - 1 \right| \leq \varepsilon \right]$$
$$+ \mathbb{P}\left[\left| \sqrt{\frac{|v|^2}{N} + \frac{1}{N}\sum_{k=1}^{N-m} Z_k^2} - 1 \right| > \varepsilon \right]$$
$$\leq \mathbf{1}_{\{v \in [1-\varepsilon, 1+\varepsilon]A\}} + \mathbb{P}\left[\left| \sqrt{\frac{|v|^2}{N} + \frac{1}{N}\sum_{k=1}^{N-m} Z_k^2} - 1 \right| > \varepsilon \right].$$

The goal is to show that the last term in the above inequality is bounded by $2e^{-CN\varepsilon^2}$ for some suitable absolute constant $C > 0$ and all N sufficiently large (depending on v). Indeed, since then by DCT, we get that

$$\limsup_{N\to\infty} |R_{N,\varepsilon}(A)| \leq \int \mathrm{d}v f_m(v) |\mathbf{1}_{\{v\in[1-\varepsilon,1+\varepsilon]A\}} - \mathbf{1}_{\{v\in A\}}|.$$

By continuity of Gaussian density and DCT, we further conclude that the latter converges to zero as ε tends to zero.

For any fixed v, $\varepsilon \in (0, 1)$ and all N large enough such that $(|v|^2 + M)/N < \varepsilon/4$ and $N/(N - M) \geq 2/3$, we have that

$$\mathbb{P}\left[\left| \sqrt{\frac{|v|^2}{N} + \frac{1}{N}\sum_{k=1}^{N-m} Z_k^2} - 1 \right| > \varepsilon \right] \leq \mathbb{P}\left[\left| \frac{|v|^2}{N} + \frac{1}{N}\sum_{k=1}^{N-m} Z_k^2 - 1 \right| > \varepsilon \vee \varepsilon^2 \right]$$
$$\leq \mathbb{P}\left[\left| \frac{1}{N}\sum_{k=1}^{N-m} Z_k^2 - \frac{N-M}{N} \right| > \frac{3\varepsilon}{4} \right] \leq \mathbb{P}\left[\left| \frac{1}{N-M}\sum_{k=1}^{N-m} Z_k^2 - 1 \right| > \frac{\varepsilon}{2} \right]$$

since for all $z \geq 0$, $|z - 1| \geq \delta$ implies that $|z^2 - 1| \geq \max(\delta, \delta^2)$. The desired result follows from (0.1) that is proved (or rather explained how to) in the solution of the next exercise. $\square$

Exercise 5 (Almost isometric projection of uniformly distributed point on the sphere)**.** *Let N be very large and let $S^{N-1} \subset \mathbb{R}^N$ be a unit $N-1$-sphere. Let X be a point chosen uniformly on S^{N-1} and $T : \mathbb{R}^N \to \mathbb{R}^m$ be a projection on the first m coordinates. Find a suitable normalization of T by some power of $\frac{m}{N}$ so that the following holds $1 - \varepsilon \leq \|c_{norm}TX\|_2 \leq$*

$1+\varepsilon$[1] *with probability at least* $1-4e^{-c\varepsilon^2 m}$ *for some uniform constant* $c>0$. *Here* c_{norm} *is some power of* $\frac{m}{N}$, *which you need to find.*

You might proceed as follows

- *Find the normalization constant by computing* L^2*-norm of* Tx *and observing that for the normalized operator it has to be equal to* 1 *(why?). Check yourself* [2].
- *Let* Y *be a standard* N*-dimensional Gaussian vector. Prove that* X *has the same law as* $Y/\|Y\|_2$. *Conclude that* $\|Y\|_2 X$ *has the law of* N*-dimensional standard Gaussian vector.*
- *To prove that you actually get the "almost isometry" property for the normalized operator* $c_{norm}T$, *compare* $c_{norm}TX$ *and* $\frac{1}{\sqrt{m}}T(\|Y\|_2 X)$ *(what is the law of the latter variable?).*
- *Additionally to the last step: estimate the concentration probability of the norm of* m*-dimensional standard Gaussian around* $\sqrt{m}$. *Chernoff-type bounds might be helpful.*

Proof. Let X be uniformly distributed on S^{N-1}. We first choose a normalizing constant: we want $c_{\text{norm}}TX$ to be of almost unit norm with high probability, therefore, we have to pick a constant in such a way that $c_{\text{norm}}TX$ has a unit norm at least w.r.t. L^2-norm. To this end, $\mathbb{E}[\|TX\|_2^2]=\sum_{i\leq m}\mathbb{E}[X_i^2]=m\mathbb{E}[X_1^2]$. On the other hand, $1=\mathbb{E}[\|X\|_2^2]=\mathbb{E}[\sum_{i\leq N}X_i^2]=N\mathbb{E}[X_1^2]$. Thus, $\|TX\|_{\mathrm{L}^2}=\sqrt{\frac{m}{N}}$, which suggests the normalizing constant $c_{\text{norm}}=\sqrt{\frac{N}{m}}$. We saw in the previous exercise that $\sqrt{N}TX$ as N tends to infinity converges in law to m-dimensional Gaussian. We want to use this fact as an intuition for our choice of the auxiliary variable. Recall from the previous exercise as well that if Y is an N-dimensional standard Gaussian vector, then X and $Y/\|Y\|_2$ have the same law. On top of that one can as well show that $\|Y\|_2 X$ has standard Gaussian law. Indeed, it follows directly from independence of $Y/\|Y\|_2$ and $\|Y\|_2$. The latter, in turn, you can check by considering $\mathbb{P}[Y/\|Y\|_2\in A, \|Y\|_2\in(a,b)]$: write it as an integral over Gaussian density and change to polar coordinates, this way you get two decoupled densities — one of uniform distribution on the sphere and the other of chi-distribution, which is the law of $\|Y\|_2$.

Now since $\|Y\|_2 X$ is an N-dimensional standard Gaussian vector, by projecting on the first m coordinates, we get an m-dimensional standard Gaussian vector. So, $T\|Y\|_2 X=\|Y\|_2 TX$ is an m-dimensional standard normal vector. Let us show that the norm of Gaussian is concentrated around square root of its dimension. For this, let Z be a standard m-dimensional Gaussian. Then, by Bernstein inequality[3].

$$\mathbb{P}\left[\left|\frac{\|Z\|_2^2}{m}-1\right|\geq u\right]\leq 2e^{-cm\min(u,u^2)}. \tag{0.1}$$

Alternatively, one can obtain this inequality by applying exponential Markov inequality to $\frac{\|Z\|_2^2}{m}-1$, using that MGF of Z_i^2 is explicitly given by $(1-2\lambda)^{-1/2}$ for $\lambda<1/2$, Tayloring $\log(1-2\lambda)$ for $|\lambda|$ sufficiently small and optimizing in admissible λ; repeating the same

[1] note that $\|X\|_2=1$

[2] $c_{\text{norm}}=\sqrt{\frac{N}{m}}$

[3] proof is based on Chernoff inequalities/exponential Markov inequality and the fact that Z_i^2 are subexponential. The latter follows directly from one of the equivalent definitions of both subgaussian + subexponential. For the details please check Theorem 2.8.1 in "High-Dimensional Probability", Roman Vershynin)

procedure with $-\frac{\|Z\|_2^2}{m}+1$.
Since for all $z \geq 0$, $|z-1| \geq \delta$ implies that $|z^2-1| \geq \max(\delta, \delta^2)$, we can conclude

$$\mathbb{P}\left[\left|\frac{\|Z\|_2}{\sqrt{m}}-1\right| \geq u\right] \leq 2e^{-cmu^2}.$$

In our original setup we get the following,

$$\begin{aligned}
\mathbb{P}&\left[\left|\sqrt{\frac{N}{m}}\|TX\|_2-1\right| \geq u\right] \\
&\leq \mathbb{P}\left[\left|\sqrt{\frac{N}{m}}\|TX\|_2-\frac{1}{\sqrt{m}}\|\|Y\|_2 TX\|_2\right| \geq \frac{u}{2}\right]+\mathbb{P}\left[\left|\frac{1}{\sqrt{m}}\|\|Y\|_2 TX\|_2-1\right| \geq \frac{u}{2}\right] \\
&\leq \mathbb{P}\left[\frac{1}{\sqrt{m}}\left|\sqrt{N}-\|Y\|_2\right| \geq \frac{u}{2}\right]+2e^{-c'mu^2} \leq 4e^{-c'mu^2}
\end{aligned}$$

where we used that $\|TX\|_2 \leq 1$. This finishes the proof □

Exercise 6 (log-Sobolev for Rademacher random variables)**.** *Let $X_1,\ldots,X_n$ be i.i.d. symmetric Rademacher random variables, $f:\mathbb{R}^n \to \mathbb{R}$. Show that*

$$\operatorname{Ent}[f^2] \leq \sum_{i=1}^n \mathbb{E}[(f-f^{(i)})^2],$$

where $f=f(X_1,\ldots,X_n)$ and $f^{(i)}=f(X_1,\ldots,X_{i-1},X_i',X_{i+1},\ldots,X_n)$ with X_i' being an independent copy of X_i.

You may proceed as follows:

(1) Use tensorization of entropy to reduce to a one-dimensional problem;
(2) Verify the following inequality and prove that it yields the desired result,

$$\forall a,b \in \mathbb{R}: \quad \frac{a^2}{2}\log a^2+\frac{b^2}{2}\log b^2-\frac{a^2+b^2}{2}\log\frac{a^2+b^2}{2} \leq \frac{(a-b)^2}{2}.$$

Proof. Recall that by tensorization of entropy,

$$\operatorname{Ent}[f^2] \leq \mathbb{E}\left[\sum_{i=1}^n \operatorname{Ent}^{(i)}[f^2]\right],$$

where $\operatorname{Ent}^{(i)}[f^2]=\mathbb{E}[f^2\log(f^2)|(X_j)_{j\neq i}]-\mathbb{E}[f^2|(X_j)_{j\neq i}]\log\mathbb{E}[f^2|(X_j)_{j\neq i}]$. Thus, it suffices to prove that $\operatorname{Ent}^{(i)}[f^2] \leq \mathbb{E}[(f-f^{(i)})^2|(X_j)_{j\neq i}]$. Note that given $(X_j)_{j\neq i}$, $f(X)$ can take two different values with equal probability. Let us call them a,b. The desired inequality takes the form,

$$\frac{a^2}{2}\log(a^2)+\frac{b^2}{2}\log(b^2)-\frac{a^2+b^2}{2}\log\left(\frac{a^2+b^2}{2}\right) \leq \frac{(a-b)^2}{2}.$$

Thus, it remains to verify this inequality for any $a,b \in \mathbb{R}$. Note that since $(|a|-|b|)^2 \leq (a-b)^2$ (and the l.h.s. does not depend on the signs of a,b), we may assume wlog that $a,b \geq 0$. By symmetry we may further assume that $a \geq b$. For a fixed value $b \geq 0$, define

$$g(a)=\frac{a^2}{2}\log(a^2)+\frac{b^2}{2}\log(b^2)-\frac{a^2+b^2}{2}\log\left(\frac{a^2+b^2}{2}\right)-\frac{(a-b)^2}{2}.$$

Observe that $g(b) = 0$,

$$g'(a) = a \log \frac{2a^2}{a^2 + b^2} - (a - b);$$

$$g''(a) = \log \frac{2a^2}{a^2 + b^2} + \frac{2b^2}{a^2 + b^2} - 1 = \log \frac{2a^2}{a^2 + b^2} - \frac{2a^2}{a^2 + b^2} + 1.$$

In particular, $g'(b) = 0$ and since $\log x - x + 1 \leq 0$ for any $x \geq 0$, $g''(a) \leq 0$ for any $a \geq b$ with strict inequality for $a \neq b$ (hence, g (strictly) concave on $[b, \infty)$). Altogether this implies that $g(a) \leq 0$ for all $a \geq b$ as desired. $\square$