TOPICS IN HIGH-DIMENSIONAL PROBABILITY

PROBLEM SETS

(Last updated: April 24, 2024)

- For each set, write up solutions for at least 5 exercises. Your grade will be based on your best 5.
- You are welcome to collaborate with other students. If you do so, you must list your collaborators at the top of your writeup.
- It is okay to use a result established in an earlier exercise even if you aren't writing that one up.
- Problems will be added as we proceed through the course (I may for instance need to relegate a proof I had planned to present in class to the problem sets.)
- "Vershynin Exercise ..." refers to exercises from Vershynin's text (linked in the course syllabus).
- Updates and corrections are in blue.

Notation: As in the lectures, C, c, c' etc. denote positive, finite constants, independent of all parameters unless otherwise noted, and *their value may vary from line to line.* For $a, b \in \mathbb{R}$, $a = O(b)$ and $a \lesssim b$ mean $|a| \le Cb$ for some absolute constant C. For $a, b > 0$, $a \gtrsim b$ means $b \lesssim a$. A random variable $\xi \in \mathbb{R}$ is said to be *standardized* if $\mathbb{E}\xi = 0$ and $\mathbb{E}\xi^2 = 1$. In asymptotic notation, dependence of implicit constants on parameters is indicated with subscripts – for instance, $a = O_q(b)$ and $a \lesssim_q b$ mean that $|a| \le Cb$ for a constant C that may depend on the parameter q.

1. Problem set 1 (due Feb 8th)

Exercise 1.1. Let $X \in \mathbb{R}^n$ be an isotropic random vector – that is, a vector $X = (X_1, \dots, X_n)$ with $\mathbb{E}X_i = 0$ for each i and covariance matrix $\mathbb{E}XX^\mathsf{T} = I_n$ (the $n \times n$ identity matrix).

(a) Show that for any deterministic real matrix M with n columns, we have
$$\mathbb{E}\|MX\|^2 = \|M\|_{\mathrm{HS}}^2 \tag{1.1}$$
where $\|M\|_{\mathrm{HS}} = (\mathrm{Tr}(M^\mathsf{T} M))^{1/2} = (\sum_{i,j} M_{i,j}^2)^{1/2}$ is the Hilbert–Schmidt norm of M (also known as the Frobenius norm).

(b) Use (1.1) to prove the parallelogram law in $\mathbb{R}^n$ – that the sum of squared lengths of the $n2^{n-1}$ edges of an n-dimensional parallelepiped is equal to the sum of squared lengths of the 2^{n-1} diagonals.

(c) Show that for $X \in \mathbb{R}^n$ an isotropic vector and $V \subset \mathbb{R}^n$ a subspace (deterministic, or random and independent of X with $\dim(V)$ deterministic) we have $\mathbb{E}\,\mathrm{dist}(X, V)^2 = n - \dim(V)$.

Exercise 1.2 (Equivalence of concentration about the mean and median)**.** Recall that $m \in \mathbb{R}$ is a median for a real-valued random variable if
$$\mathbb{P}(X \le m) \ge 1/2 \quad \text{and} \quad \mathbb{P}(X \ge m) \ge 1/2.$$

(a) Show that any real random variable X has at least one median, and that the set of all medians of X is a closed interval. Give an example of a random variable having more than one median value.

(b) Let m be any median of X, and suppose there are $a \in \mathbb{R}$ and $K > 0$ such that
$$\mathbb{P}(|X - a| \ge t) \le 2\exp(-t^2/K^2) \quad \forall t \ge 0. \tag{1.2}$$
Show that $|a - m| = O(K)$, and deduce that
$$\mathbb{P}(|X - m| \ge t) \le 2\exp(-ct^2/K^2) \quad \forall t \ge 0 \tag{1.3}$$

for some universal constant $c > 0$. Deduce from this that $|m - \mathbb{E}X| = O(K)$, and that (1.3) holds with m replaced by $\mathbb{E}X$, for a possibly smaller universal constant $c > 0$.

(Hint: Note that the bound (1.3) holds trivially for $t \le K\sqrt{(\log 2)/c}$, so by shrinking c we may assume without loss of generality that $t \ge CK$ for any fixed constant $C > 0$ as large as we please.)

(c) Show that if (1.2) holds then $\mathrm{Var}(X) = O(K^2)$. Deduce that if $X \ge 0$ almost surely, then (1.3) holds with m replaced by $(\mathbb{E}X^2)^{1/2}$, for a possibly smaller universal constant $c > 0$.

Exercise 1.3 (Concentration of degrees in random graphs)**.** Vershynin Exercises 2.4.2, 2.4.3, 2.4.4, 2.4.5

Exercise 1.4 (Equivalent characterizations of sub-exponential tails)**.** Vershynin Exercises 2.7.2, 2.7.4

Exercise 1.5 (Sub-Gaussian vectors have large support)**.** Vershynin Exercise 3.4.5

Exercise 1.6 (Gaussian concentration from isoperimetry)**.** The isoperimetric theorem for n-dimensional Gauss space (i.e. $\mathbb{R}^n$ equipped with the Euclidean distance $d_2(x, y) = \|x - y\|_2$ and the standard Gaussian measure γ_n) states that for any $m \in (0, 1)$ and $r > 0$, among all Borel sets $A \subset \mathbb{R}^n$ of measure $\gamma_n(A) = m$, the ones that minimize $\gamma_n(A_r)$ are half-spaces, i.e. sets of the form $H_{u,a} = \{x \in \mathbb{R}^n : \langle x, u\rangle \le a\}$ for some $u \in S^{n-1}$ and $a \in \mathbb{R}$ (recall the notation $A_r := \{x \in \mathbb{R}^n : d_2(x, A) \le r\}$ for the r-widening of A). This is to be compared with the case of Lebesgue rather than Gaussian measure, where the minimizers are balls. Use this fact to show that for any 1-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$, $f(G) - \mathbb{E}f(G)$ is K-sub-Gaussian with $K = O(1)$, where $G \sim N(0, I_n)$ is a standard Gaussian vector. (You may find a result from Exercise 1.2 helpful for this.) *Bonus: find the optimal value of K.*

Exercise 1.7 (Concentration for permutations and random regular graphs)**.** Let S_n be the symmetric group of permutations on $[n] = \{1, \dots, n\}$ (i.e. bijections $\sigma : [n] \to [n]$ with group multiplication given by composition of functions). A label $i \in [n]$ is a *fixed point* of $\sigma \in S_n$ if $\sigma(i) = i$. Recall that $\tau \in S_n$ is a *transposition* if all but two elements $i \ne j$ of $[n]$ are fixed points (thus $\sigma(i) = j$ and $\sigma(j) = i$). Write τ_{ij} for the transposition that exchanges i and j.

(a) Suppose $F : S_n \to \mathbb{R}$ has the property that

$$|F(\sigma \circ \tau_{ij}) - F(\sigma)| \le b \qquad \forall \sigma \in S_n\,,\ 1 \le i < j \le n \tag{1.4}$$

for some $b > 0$. Show that if π is a uniform random element of S_n, then $F(\pi) - \mathbb{E}F(\pi)$ is $O(b\sqrt{n})$-sub-Gaussian.

(b) For $n, d \in \mathbb{N}$, the *permutation model* for a random $2d$-regular multigraph is a graph with labeled vertex set $[n]$ and with (undirected) edges determined by d independent uniform random permutations $\pi_1, \dots, \pi_d \in S_n$ as follows. Letting A be the $n \times n$ symmetric matrix with entry A_{ij} equal to the number of edges connecting i and j, we have

$$A_{ij} = \sum_{k=1}^{d} 1_{\pi_k(i)=j} + 1_{\pi_k(j)=i}.$$

Or, in terms of the permutation matrices $P^{(k)}_{ij} = 1_{\pi_k(i)=j}$ we have $A = \sum_{k=1}^{d} P^{(k)} + (P^{(k)})^\mathsf{T}$. Thus every vertex has $2d$ neighbors, counting multiplicity. (This is a multigraph since it allows for multiple edges connecting a fixed pair of vertices, and also allows self-loops, though when $d = o(n)$ most edges will not occur in this way with high probability (optional exercise!).) For fixed disjoint sets $U, V \subset [n]$ of vertices, let $e(U, V)$ be the number of edges with one endpoint in U and the other in V (counting multiplicity). Show that $e(U, V) - \frac{2d}{n}|U||V|$ is $O(\sqrt{d|U|})$-sub-Gaussian.

(c) (Optional). Prove a Bernstein-type tail

$$\mathbb{P}\big(\big|e(U, V) - \tfrac{2d}{n}|U||V|\big| \ge t\big) \le 2\exp\Big(-\frac{ct^2}{\frac{d|U||V|}{n} + t}\Big) \qquad \forall t \ge 0. \tag{1.5}$$

For what ranges of t and $|V|$ does this improve on the result of part (b)?

Remark: One should compare this random graph with the Erdős–Rényi random graph $G_{n,p}$ from Exercise 1.3, where each pair $\{i,j\} \subset [n]$ is included as an edge independently with probability p. Taking $p = 2d/(n-1)$, the degree of each vertex is binomially distributed with expectation $2d$. You can show the same results from (b,c) hold for this model by easier arguments!

Exercise 1.8. Recall the following important concentration result of Talagrand:

Theorem 1.1 ([Tal96])**.** *Let $X = (\xi_1, \dots, \xi_n)$ be a vector of independent random variables with $|\xi_i| \le 1$ a.s. for each $i \in [n]$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function that is 1-Lipschitz under the Euclidean norm. Then for any median m of $f(X)$ and any $t \ge 0$, we have*

$$\mathbb{P}(|f(X) - m| \ge t) \le 4\exp(-t^2/16).$$

In particular, $f(X) - m$ is $O(1)$-sub-Gaussian (and so is $f(X) - \mathbb{E}f(X)$ by the result of Exercise 1.2).

(a) Use Theorem 1.1 to prove the generalized statement that if the components of X are almost-surely bounded by some $B < \infty$, and f is convex and L-Lischitz, then for any median a of $f(X)$,

$$\mathbb{P}(|f(X) - a| \ge t) \le 4\exp(-\frac{t^2}{16B^2L^2}) \qquad \forall t \ge 0. \tag{1.6}$$

(b) Show that if ξ is K-sub-Gaussian, then for any $\beta > 0$,

$$\mathbb{E}\exp(\frac{1}{2K^2}\xi^2\, \mathrm{I}(|\xi| > \beta K)) \le 1 + 2\exp(-\beta^2/2). \tag{1.7}$$

(c) Show that if we relax the boundedness assumption in Theorem 1.1 to the assumption that the variables ξ_i are all K-sub-Gaussian, then $f(X) - m$ is $O(K\sqrt{\log n})$-sub-Gaussian. (From Exercise 1.2 the same holds for $f(X) - \mathbb{E}f(X)$.)

(Hint: split $X = X_> + X_\le$, with $X_\le$ having entries $\xi_j\, \mathrm{I}(|\xi_j| \le CK\sqrt{\log n})$ for some constant $C > 0$, and use part (b) to control the event that $\|X_>\|_2$ exceeds $t/4$, say.)

Exercise 1.9 (Applying Talagrand's inequality)**.** .

(a) Let $X = (\xi_1, \dots, \xi_n) \in \mathbb{R}^n$ be a random vector with independent standardized components ξ_i satisfying $|\xi_i| \le B$ almost surely, and let V be a subspace of $\mathbb{R}^n$ (deterministic or random and independent of X) of dimension d. Show that $\mathrm{dist}(X, V) - \sqrt{n-d}$ is $O(B)$-sub-Gaussian.

(Hint: A result from Exercise 1.2 may be helpful for this.)

(b) Let X be an $n \times n$ matrix with entries ξ_{ij} that are iid copies of a standardized random variable ξ with $|\xi| \le B$ almost surely for some fixed $B < \infty$. Use the Bai–Yin law (that $\frac{1}{\sqrt{n}}\|X\|_{\mathrm{op}} \to 2$ in probability, where $\|X\|_{\mathrm{op}} := \sup_{u \in \mathbb{S}^{n-1}} \|Xu\|_2$) and Talagrand's inequality to show that for any fixed $\varepsilon > 0$,

$$\mathbb{P}(\|X\|_{\mathrm{op}} > (2+\varepsilon)\sqrt{n}) \le 2\exp(-c\varepsilon^2 n/B^2)$$

for all n sufficiently large.

(c) *(Bonus)* More generally, show that if X is as in part (b), then for each $1 \le k \le n$ we have that for any fixed $\varepsilon > 0$,

$$\mathbb{P}(\sigma_k(X) > (2+\varepsilon)\sqrt{n}) \le 2\exp(-c\varepsilon^2 kn/B^2)$$

for all n sufficiently large (recall $\sigma_k(X)$ is the kth *largest* singular value of X).

Exercise 1.10 (Maurey's empirical method for constructing nets)**.** For $p \in [1, \infty]$ we denote the unit ℓ_p-ball in $\mathbb{R}^n$ by $\mathbb{B}_p^n$. That is, $\mathbb{B}_p^n = \{v \in \mathbb{R}^n : \sum_{i=1}^n |v_i|^p \le 1\}$. For $p = 2$ we generally drop the subscript. The set of r-sparse unit vectors in $\mathbb{R}^n$ is denoted

$$S_{n,r} = \{u \in \mathbb{S}^{n-1} : |\mathrm{supp}(u)| \le r\}. \tag{1.8}$$

(a) Let $w_1, \dots, w_m \in \mathbb{R}^n$ be m points in the cube $\mathbb{B}_\infty^n$, i.e. $\|w_i\|_\infty \le 1$ for each i, and let T be their convex hull. For a given $y = \sum_{k=1}^m \alpha_k w_k \in T$, let $Y_1, \dots, Y_N$ be iid vectors in $\{w_1, \dots, w_m\}$ with distribution $\sum_{k=1}^m \alpha_k \delta_{w_k}$ (so $\mathbb{P}(Y_i = w_k) = \alpha_k$ for each i, k). With $\overline{Y}_N = \frac{1}{N}\sum_{i=1}^N Y_i$ the sample mean, show that for any $\varepsilon > 0$,

$$\mathbb{P}(\|y - \overline{Y}_N\|_\infty > \varepsilon) \le 2n\exp(-c\varepsilon^2 N).$$

(Hoeffding's inequality will be useful for this.)

(b) Deduce that T can be covered by $\exp(O(\varepsilon^{-2}(\log n)(\log m)))$ translates of $\varepsilon \cdot \mathbb{B}_\infty^n$ with centers in T (i.e. T has an ε-net under the ℓ_∞ metric of size $\exp(O(\varepsilon^{-2}(\log n)(\log m))))$.

(c) Let H be an $n \times n$ matrix with entries bounded by 1. With $S_{n,r}$ as in (1.8), show that $S_{n,r} \subset \sqrt{r}\mathbb{B}_1^n$, and use this to construct an ε-net for $HS_{n,r} = \{Hu : u \in S_{n,r}\}$ under the ℓ_∞ metric of size $\exp(O(\varepsilon^{-2}r(\log n)^2))$. *(Hint: Note that $\mathbb{B}_1^n$ is the convex hull of the $2n$ signed standard basis vectors $\pm e_1, \dots, \pm e_n$.)*

Commentary: This construction of a net for sparse vectors was an important part of proofs in [RV08, Bou14] that m row vectors sampled independently and uniformly from a matrix H having orthogonal rows with bounded entries (such as discrete Fourier matrices or Hadamard matrices) form an RIP matrix with high probability if $m \geq Cr\varepsilon^{-C}(\log n)^C$.

2. Problem set 2 (due Mar 7th)

Exercise 2.1 (Covering and packing numbers)**.** Vershynin Exercises 4.2.5, 4.2.9, 4.2.10.

Exercise 2.2 (Bounding the norm of symmetric matrices)**.** Vershynin Exercises 4.4.3, 4.4.4; then read Sections 4.4.2 and 4.6 (in class we followed the approach sketched in Exercise 4.6.4).

Exercise 2.3 (Spectral clustering)**.** Vershynin Exercise 4.7.6.

Exercise 2.4. The *Lévy concentration function* of a random variable $X \in \mathbb{R}$ is defined as

$$\mathcal{L}(X,t) := \sup_{a\in\mathbb{R}} \mathbb{P}(|X-a| \leq t)\,, \quad t \geq 0. \tag{2.1}$$

Bounds on $\mathcal{L}(X,t)$ are known as *anticoncentration* or *small-ball* estimates.

(a) Show that if X is standardized (i.e. has mean 0 and variance 1), then

$$\mathcal{L}(X, \tfrac{1}{4}\mathbb{E}|X|) \leq 1 - c_0(\mathbb{E}|X|)^2 \tag{2.2}$$

for some absolute constant $c_0 > 0$. Show this bound is sharp in the sense that for arbitrarily small $\varepsilon > 0$ there is a standardized random variable X with $\mathbb{E}|X| \leq \varepsilon$ and for which the reverse of the above inequality holds (for some possibly modified value of c_0 – you don't need to find the sharp constant).

(b) Show that if we further assume $\mathbb{E}|X|^q \leq A$ for some $q > 2$ and $A < \infty$ then

$$\mathcal{L}(X, 0.99) \leq 1 - c_1 \tag{2.3}$$

for some $c_1 > 0$ depending only on q and A. (Thus, a mild concentration assumption – namely, the moment bound $\mathbb{E}|X|^q \leq A$ – is enough to guarantee some amount of anticoncentration for a standardized variable X.)

(Hint: use (or adapt the proof of) the Paley–Zygmund inequality.)

Exercise 2.5 (Anti-concentration from Berry–Esseen)**.** Recall the following:

Theorem 2.1 (Berry–Esseen theorem for non-identically distributed summands)**.** *Let $\zeta_1, \dots, \zeta_n$ be independent centered random variables with $\mathbb{E}|\zeta_i|^3 < \infty$ for each $i \in [n]$, set $S = (\sum_{i=1}^n \zeta_i)/\sqrt{\sum_{i=1}^n \mathbb{E}\zeta_i^2}$. and let g be a standard Gaussian variable. For any $t \in \mathbb{R}$ we have*

$$|\mathbb{P}(S < t) - \mathbb{P}(g < t)| \lesssim \frac{\sum_{i=1}^n \mathbb{E}|\zeta_i|^3}{\left(\sum_{i=1}^n \mathbb{E}\zeta_i^2\right)^{3/2}}.$$

Using the above result, show that if $X = (\xi_1, \dots, \xi_n)$ is uniform in $\{-1,1\}^n$ and $u \in \mathbb{S}^{n-1}$ is a fixed unit vector satisfying

$$\sum_{i=1}^n u_i^2 1_{|u_i| \leq b/\sqrt{n}} \geq a^2$$

then

$$\mathcal{L}(\langle X, u\rangle, t) \lesssim t/a \qquad \forall t \ge \frac{b}{\sqrt{n}}$$

where the Lévy concentration function was defined in (2.1). Thus, although $\langle X, u\rangle$ is a discrete random variable, it effectively has bounded density at scales $\gg n^{-1/2}$ if u has a constant proportion of its ℓ_2 mass on coordinates of size $O(1/\sqrt{n})$ (a property that holds for generic $u \in \mathbb{S}^{n-1}$).

(Hint: condition on variables ξ_i for which u_i is large.)

Exercise 2.6 (Tensorization of Anticoncentration))**.** The Lévy concentration function for a random vector $X \in \mathbb{R}^d$ is defined

$$\mathcal{L}(X, t) := \sup_{x_0 \in \mathbb{R}^d} \mathbb{P}(\|X - x_0\|_2 \le t)\,, \qquad t \ge 0 \tag{2.4}$$

generalizing (2.1) for the case $d = 1$. Suppose $X = (\xi_1, \dots, \xi_d)$ has independent components.

(a) Show that if $\mathcal{L}(\xi_i, a) \le b$ for some $a > 0$ and $b \in (0,1)$ and all $i \in [n]$, then $\mathcal{L}(X, c\sqrt{d}) \le \exp(-cd)$ for some $c > 0$ depending only on a, b.

(b) Show that if $\mathcal{L}(\xi_i, \varepsilon) \le L\varepsilon$ for all $\varepsilon \ge \varepsilon_0$ and $i \in [n]$, then $\mathcal{L}(X, \varepsilon\sqrt{d}) \le O(L\varepsilon)^d$ for all $\varepsilon \ge \varepsilon_0$.

(Hint: after fixing x_0, you can control the event that a sum S of independent random variables is small by bounding an inverse exponential moment $\mathbb{E}\exp(-\lambda S)$ for some $\lambda > 0$.)

Exercise 2.7 (Lower bounds for most singular values)**.**

(a) Let M be an $n \times n$ invertible square matrix with real or complex entries. Denote its rows and columns by $\mathrm{row}_i(M), \mathrm{col}_j(M)$. For each $i \in [n]$ let $V_{(i)}$ be the span of $\{\mathrm{col}_j(M) : j \in [n]\setminus\{i\}\}$ (that is, all but the ith column). Prove that $\|\mathrm{row}_i(M^{-1})\|_2 = \mathrm{dist}(\mathrm{col}_i(M), V_{(i)})^{-1}$. *(Hint: consider the ith row of the equation $M^{-1}M = I$.)*

(b) Deduce the *inverse second moment identity*: if M is an $m \times n$ matrix with complex entries and $m \ge n$, then

$$\sum_{i=1}^n \frac{1}{\sigma_i(M)^2} = \sum_{i=1}^n \frac{1}{\mathrm{dist}(\mathrm{col}_i(M), V_{(i)})^2}\,.$$

(Hint: you can argue to reduce to the square case $m = n$ by projecting the columns to their span.)

(c) Show that if X is an $n \times (n-k)$ matrix with independent standardized entries ξ_{ij} almost-surely bounded by $B < \infty$, then

$$\mathbb{P}\bigg(\bigcap_{1 \le i \le n-k} \Big\{ \mathrm{dist}(X_i, V_{(i)}) \ge \tfrac{1}{2}\sqrt{k} \Big\} \bigg) \ge 1 - ne^{-ck/B^2} \tag{2.5}$$

where X_i is the ith column of X and $V_{(i)}$ is the span of the remaining $n-k-1$ columns.

(d) Use the Courant–Fisher min-max formula to prove that if M is an $m \times n$ matrix with complex entries, and M' is obtained by removing ℓ columns from M, then $\sigma_i(M') \le \sigma_i(M)$ for every $1 \le i \le n - \ell$ (where we label singular values in non-increasing order).

(e) Show that if X is an $n \times n$ matrix with independent standardized entries ξ_{ij} almost-surely bounded by $B < \infty$, then for any $k \le n$,

$$\mathbb{P}\bigg(\bigcap_{k \le i \le n} \Big\{ \sigma_{n-i+1}(\tfrac{1}{\sqrt{n}}X) \ge c\tfrac{i}{n} \Big\} \bigg) \ge 1 - ne^{-ck/B^2} \tag{2.6}$$

(where the constant $c > 0$ may differ from the one in part (c)).

Note this is consistent with the asymptotic Marchenko–Pastur law (also known the quarter-circular law in the square case), which says that the proportion of singular values of $\frac{1}{\sqrt{n}}X$ lying in a fixed interval $I \subset \mathbb{R}_+$ converges to $\frac{1}{\pi}\int_I \sqrt{4 - x^2}\,dx$, so we would expect the ith smallest singular value of $\frac{1}{\sqrt{n}}X$ to be of size $\Theta(i/n)$. The above shows this holds as a lower bound with high probability for all but the $O(B^2 \log n)$ smallest singular values.

Exercise 2.8. *(In this and the following exercise we explore arguments bounding the minimal singular value of square matrices with entries of bounded density that completely avoid the use of nets; in particular they require no* a priori *control on the norm.)*

(a) Show that for any $m \times n$ random matrix M with $n \le m$ and any $t > 0$,

$$\mathbb{P}(\sigma_n(M) < t) \le \sum_{j=1}^{n} \mathbb{P}(\mathrm{dist}(\mathrm{col}_j(M), V_{(j)}) < t\sqrt{n})$$

where $\mathrm{col}_j(M)$ is the jth column of M and $V_{(j)}$ is the span of the $n-1$ columns of M with the jth column left out.

(b) Let $X \in \mathbb{R}^m$ have independent components ξ_j. Show that for any nonzero $v \in \mathbb{R}^m$ and any $s > 0$,

$$\mathcal{L}(\langle X, v\rangle, s) \le \min_{i \in [m]: v_i \neq 0} \mathcal{L}(\xi_i, s/|v_i|)\,.$$

(c) Let $n \le m$. Show that If X is an $m \times n$ matrix with independent entries having density bounded by L, then

$$\mathbb{P}(\sigma_n(X) < t) \le 2Ltm^{1/2}n^{3/2}$$

for all $t > 0$. (While this bound loses a factor $n^{3/2}$ over the optimal bound when X is a standardized $n \times n$ Gaussian matrix, say, (see the exercise below) the upshot is that we have made no assumptions on the means or variances of X, let alone whether these exist. A bound of this form was used in the study of the spectrum of non-Hermitian heavy-tailed matrices in [BCC11]; see also [BC12].)

(Hint for parts (a) and (c): given a unit vector $u \in \mathbb{S}^{n-1}$, we can always find a component of size at least $1/\sqrt{n}$ (why?).)

Exercise 2.9 (Optimal bound on σ_n for square Gaussian matrices)**.** Let G be an $n \times n$ matrix with independent standardized Gaussian entries.

(a) Show that for any fixed $v \in \mathbb{S}^{n-1}$ and any $t > 0$, $\mathbb{P}(\|G^{-1}v\|_2 > 1/t) \lesssim t$. *(Hint: use the fact that the distribution of G is invariant under multiplication on the left or right by orthogonal matrices.)*

(b) Show that for any invertible $n \times n$ matrix M and any $v \in \mathbb{S}^{n-1}$, letting $u_1 \in \mathbb{S}^{n-1}$ be such that $\|M^{-1}u_1\|_2 = \|M^{-1}\|_{\mathrm{op}}$ (thus u_1 is the first right-singular vector of M^{-1}) we have

$$\|M^{-1}v\|_2 \ge \|M^{-1}\|_{\mathrm{op}}|\langle v, u_1\rangle|.$$

Deduce from this and part (a) that for $v \in \mathbb{S}^{n-1}$ uniform random and independent of G,

$$\mathbb{P}\Big(\|G^{-1}\|_{\mathrm{op}} > 1/t\,,\ |\langle v, u_1\rangle| \ge \frac{1}{\sqrt{n}}\Big) \lesssim t\sqrt{n}$$

for any $t > 0$, where u_1 is the first right-singular vector of G^{-1}.

(c) Conclude that $\mathbb{P}(\sigma_n(G) < t) = \mathbb{P}(\|G^{-1}\|_{\mathrm{op}} > 1/t) \lesssim t\sqrt{n}$ for all $t > 0$. (Note it follows from part (c) of the previous exercise that $\mathbb{P}(\sigma_n(G) = 0) = 0$, i.e. G is invertible almost surely, though there are easier ways to see this!)

Commentary: A variant of this argument shows the same holds for $M+G$ for any fixed $n \times n$ matrix M – it could even have norm n^{100}; see the original paper [SST06], where it was used to analyze the "smoothed complexity" of algorithms involving small noisy perturbations of a fixed matrix. (Contrast with our need to have $\|X\|_{\mathrm{op}} = O(\sqrt{n})$ with high probability in our treatment for general entry distributions.) A similar result for symmetric Gaussian matrices was used to study the localization for eigenvectors of random band matrices in [APS+17, PSSS19]. This has been conjectured to extend to random matrices with independent entries of bounded density [Vu]. The conjecture was established in the affirmative under some additional tail conditions on the entry distribution in [Tik20].

Exercise 2.10 (Symmetrization for suprema)**.** Vershynin Exercise 7.1.9

Exercise 2.11 (Multivariate Gaussian integration by parts)**.** Vershynin Exercise 7.2.6

3. Problem set 3 (due April 8th)

Exercise 3.1 (Sudakov–Fernique inequality)**.** Vershynin 7.2.12

Exercise 3.2 (Gordon's inequality and the smallest singular value)**.** Vershynin 7.2.14 and 7.3.4

Exercise 3.3 (Properties of the Gaussian width)**.** Vershynin 7.5.3, 7.5.4, 7.5.11

Exercise 3.4 (Tail bounds from chaining)**.** Vershynin 8.1.7 and 8.5.6

Exercise 3.5 (Non-sharpness/sharpness of Dudley's bound and γ_2-functional)**.** Vershynin 8.1.12, 8.5.2

Exercise 3.6 (Symmetrization for empirical processes)**.** Exercise 17.3 in the class notes.

Exercise 3.7 (Supremum of Boolean empirical processes)**.** Exercises 17.4 and 17.5 in the class notes.

4. Problem set 4 (due May 2nd)

Exercise 4.1 (L^q-Poincaré inequalities)**.** [vH, Problem 8.5]

Exercise 4.2 (Concentration for non-Lipschitz functions)**.** [vH, Problem 8.6]

Exercise 4.3 (Escape from small sets)**.** Let $p \in (0, \frac{1}{2})$, and equip $\mathcal{X} = \{0,1\}^n$ with the product Bernoulli(p) measure μ_p. Consider the continuous-time Markov process $(X(t))_{t\geq 0}$ on $\mathcal{X}$ where to each coordinate $i \in [n]$ we associate an independent Poisson clock of rate 1 (that is, a Poisson process $(N_i(t))_{t\geq 0}$), and every time a clock "ticks" (i.e. every time t when $N_i(t)$ jumps in value) we replace the associated coordinate $X_i(t)$ with an independent Bernoulli(p) variable. This process has stationary distribution μ_p and is reversible with respect to this measure. See Example 8.14 and Section 2.3.2 in [vH]. We consider the process $(X(t))_{t\geq 0}$ with $X(0) \sim \mu_p$; thus, $X(t) \sim \mu_p$ for all $t \geq 0$.

(a) Show that for any set $A \subset \mathcal{X}$, we have

$$\mathbb{P}(X(t) \in A | X(0) \in A) \leq \mu_p(A)^{0.49} \tag{4.1}$$

for all $t \geq C\log(1/p)$, where $C > 0$ is a universal constant (the same holds with 0.49 replaced by any other fixed constant in $(0, \frac{1}{2})$ after appropriate modification of C).

(Hint: Express $\mathbb{P}(X(0) \in A, X(t) \in A)$ in terms of the indicator function 1_A and the semigroup $(P_t)_{t\geq 0}$ associated to the Markov process, and apply Cauchy–Schwarz and hypercontractivity. You may use the fact from [vH, Exercise 8.3] that the semigroup is hypercontractive with constant $c_p^* = \frac{1}{2}(1-2p)^{-1}\log\frac{1-p}{p}$.)

(b) Show that (4.1) cannot be improved in general; that is, find a set A such that for any $p \in (0, \frac{1}{2})$ the inequality is reversed if $t < c\log(1/p)$ for an appropriate constant $c > 0$.

Exercise 4.4 (Tribes function)**.** [vH, Problem 8.9].

Exercise 4.5 (Superconcentration for the BRW [Cha14])**.** Recall the branching random walk model defined in the lectures, which we may formulate as follows: Consider a rooted binary tree of depth n; thus, including the root vertex there are $2^{n+1}-1$ vertices, 2^n of which are leaves at the nth level. To each edge attach an independent standard Gaussian, and for each leaf v let X_v be the sum of the n Gaussians along the path leading from v back to the root. Let M_n be the maximum of X_v over all 2^n leaves v.

(a) Show that the Gaussian Poincaré inequality implies $\mathrm{Var}(M_n) = O(n)$.
(b) Prove using Talagrand's $L^1 - L^2$ bound for the Gaussian measure that M_n has variance $O(\log n)$.

Exercise 4.6. Exercises 3 and 4 from this blog post.

Exercise 4.7. Exercises 6 and 7 from this blog post.

References

[APS+17] Michael Aizenman, Ron Peled, Jeffrey Schenker, Mira Shamis, and Sasha Sodin. Matrix regularizing effects of Gaussian perturbations. *Commun. Contemp. Math.*, 19(3):1750028, 22, 2017.

[BC12] Charles Bordenave and Djalil Chafaï. Around the circular law. *Probab. Surv.*, 9:1–89, 2012.

[BCC11] Charles Bordenave, Pietro Caputo, and Djalil Chafaï. Spectrum of non-Hermitian heavy tailed random matrices. *Comm. Math. Phys.*, 307(2):513–560, 2011.

[Bou14] Jean Bourgain. An improved estimate in the restricted isometry problem. In *Geometric aspects of functional analysis*, volume 2116 of *Lecture Notes in Math.*, pages 65–70. Springer, Cham, 2014.

[Cha14] Sourav Chatterjee. *Superconcentration and related topics.* Springer Monographs in Mathematics. Springer, Cham, 2014.

[PSSS19] Ron Peled, Jeffrey Schenker, Mira Shamis, and Sasha Sodin. On the Wegner orbital model. *Int. Math. Res. Not. IMRN*, (4):1030–1058, 2019.

[RV08] Mark Rudelson and Roman Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61(8):1025–1045, 2008.

[SST06] Arvind Sankar, Daniel A. Spielman, and Shang-Hua Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM J. Matrix Anal. Appl.*, 28(2):446–476 (electronic), 2006.

[Tal96] Michel Talagrand. A new look at independence. *Ann. Probab.*, 24(1):1–34, 1996.

[Tik20] Konstantin Tikhomirov. Invertibility via distance for noncentered random matrices with continuous distributions. *Random Structures Algorithms*, 57(2):526–562, 2020.

[vH] Ramon van Handel. Probability in high dimensions. Lectures notes. https://web.math.princeton.edu/~rvan/APC550.pdf.

[Vu] V. H. Vu. Private communication.