Geometry and Data science - final assignment

Many of the exercises are inspired by the open problems which appear here:
https://people.math.ethz.ch/ abandeira/TenLecturesFortyTwoProblems.pdf

1. The Johnson-Lindenstrauss lemma finds a mapping from $x_1, ..., x_N \in \mathbb{R}^d$ to $y_1, ..., y_N \in \mathbb{R}^k$ that preserves distances in ℓ_2 (the usual Euclidean metric). Suppose that for some $\delta > 0$, $x_1, ..., x_N$ are such that $\|x_i\|_1 \geq \delta d \|x_i\|_\infty$ for all i. In this case, formulate and prove a result of a similar spirit to the Johnson-Lindenstrauss lemma for preserving distances in ℓ_1 (the dimension k should also depend on δ somehow).

2. Use the Johnson-Lindenstrauss lemma to show that for every graph $G = (V, E)$ of maximum degree d, there exists a mapping $f : V \to \mathbb{R}^k$ with $k = 10d^2 \log |V|$ such that for any two vertices $v_1, v_2 \in V$ one has $(v_1, v_2) \in E$ if and only if $\langle f(v_1), f(v_2) \rangle \geq t$ for some threshold $t \in \mathbb{R}$. (Hint: By adding a large enough copy of the identity to the adjacency matrix of the graph, it becomes positive definite, and is therefore the Gramm matrix of some vectors...).

3. The Johnson-Lindenstrauss lemma finds a mapping from points $x_1, ..., x_N \in \mathbb{R}^d$ to $y_1, ..., y_N \in \mathbb{R}^k$ which approximately preserves distances. Can you find a way to map $x_1, ..., x_N$ to $y_1, ..., y_N \in \{-1, 1\}^k$ (hence to the discrete cube instead of Euclidean space), so that for some constant $K > 0$,

$$|\langle x_i, x_j \rangle - K\langle y_i, y_j \rangle| \leq \varepsilon, \ \ \forall i \neq j$$

   with $k = O(\log(n)/\varepsilon^2)$?

4. In class, we saw how to use low-rank matrix recovery for recommendation systems (the Netflix problem) where one has n users and m items, and where the users' rankings of the items are only given for some sparse subset of entries of the $n \times m$ matrix. We assumed that the entries given to us are chosen independently with the same probability. Try to generalize this framework in the following directions:

   (a) What if the entries that we get to see are still chosen independently, but the probability that we see the i, j'th entry depends on the value of this entry? (hence, a user is more likely to watch a movie that she will rate higher).

   (b) What if we don't actually get to see a ranking, but we only get to see which user watched which movies (and assume that the probability of watching a some monotone function of the ranking)?

   (c) What if the entries that we get to see are not chosen independently, but are only chosen independently within every user (hence, some users watch more movies than others, but within each user, different movies are independent)?

5. (Community detection). Let $C > 0$ be some big constant (say, 1000). Let n be a large integer and take $p = C\log(n)^2/n$. Suppose that you're given an n-vertex graph $G = (V, E)$ which was generated as follows: First partition the vertices randomly into two sets of size $n/2$, $V = A \cup B$ (assume that n is even). Then for each pair of vertices i, j such that either $i, j \in A$ or $i, j \in B$, connect i and j with probability $1.1p$. Otherwise, if $i \in A$ and $j \in B$ (or the other way around), connect them with probability $0.9p$. This can be thought of as a social network on two communities where the probability to befriend a person in the same community is slightly larger. Note that you only get to see the graph, you don't know which vertex belongs to A and which belongs to B. Suggest an algorithm that, by looking at the graph, reconstructs the sets A and B. (Hint: use the mechanism that we learnt for low-rank matrix completion. Recall that the mechanism takes a low-rank matrix to which some "noise" was added and reconstructs the matrix).

6. Open problems 5.1 and 5.2: Finding a deterministic matrix satisfying RIP seems hard. However, let's try to reduce the randomness.

   (a) Can you prove an analog of Theorem 5.14 if instead of i.i.d Gaussians one takes i.i.d ± 1 Bernoullis?

   (b) Let $a_1, ..., a_n \in \mathbb{R}^k$ be unit vectors. Suppose that $\max_{i \neq j} |\langle a_i, a_j \rangle| \leq \mu$. Prove that the matrix $[a_1; ...; a_n]$ satisfies $\left(\frac{1}{10\mu}, 1/2\right)$-RIP.

   (c) Use the above to show that, in polynomial time, one can generate a $k \times d$ matrix and certify that it satisfies $(s, 1/2)$-RIP whenever $k \geq 10s^2 \log(N/s)$.

7. Related to Open problem 1.1 (Mallat and Zeitouni).

   (a) Can we expect the conjecture to be true if we replace the Gaussians by Bernoulli random variables? Namely if $(g_i)_{i=1}^p$ are independent and $g_i = \pm\alpha_i$ (with probabilities $\frac{1}{2}/\frac{1}{2}$ for some numbers $(\alpha_i)_{i=1}^p$?

   (b) Try to prove the case $d = 1$.

   (c) A natural attempt at the conjecture would be to try to prove the following: suppose that g is a Gaussian with covariance matrix Σ in $\mathbb{R}^p$. Let $V = [e_1, ..., e_p]$ be the standard basis of $\mathbb{R}_p$, and let U be the 2×2 orthogonal rotation that diagonalizes $(\Sigma_{i,j})_{i,j \in \{1,2\}}$ (hence a rotation under which $e_3, .., e_n$ are invariant). Consider the basis $V' = [e'_1, e'_2, e_3, ...]$ such that $e'_1 = Ue_1$, $e'_2 = Ue_2$. Hopefully, in this case we have that $\mathbb{E}[\Gamma_{V'}] \geq \mathbb{E}[\Gamma_V]$. (a) Can you explain why, were this true, this would imply the conjecture? (b) Can you find a counterexample to this?

8. Open problem 1.2. The monotonicity of the singular values seems like a hard question, but let's try an easier one: consider the quantities $m_k := \frac{1}{p}\mathrm{Tr}\left(\frac{1}{n}(XX^T)^k\right)$. In class, we proved that $\lim_{n\to\infty} \mathrm{Var}[m_k] = 0$. Can you prove that $\mathrm{Var}[m_k]$ is a decreasing sequence? Otherwise, can you find an explicit subsequence for which it is decreasing?

9. Open problem 1.3.

   (a) What if the matrix W is replaced by a non-symmetric matrix of i.i.d Gaussians? Would we still expect the same phenomenon to hold?

   (b) What if the vector $\mathbf{1}\mathbf{1}^T$ is replaced by vv^T for an arbitrary vector v?

   (c) where in the proof is the fact that $r = n$ used?

10. Open problem 2.3 (The planted clique problem).

    (a) Suggest a formulation of this problem as a rank-minimization problem (hence, minimize the rank of a matrix among some convex family of matrices).

    (b) The rank of a matrix is not convex. Try to find a natural convex relaxation of the above (i.e., find a quantity which tends to be small when the rank is small and which can be minimized by convex optimization).

    (c) Can you find any (even very weak) theoretical guarantee for the success of this relaxation? E.g., something of the sort: If the clique is very large and very disconnected from the rest of the graph, and the rest of the graph is very sparse, etc.

11. A question related to Open problem 3.2: Suppose that there is a symmetric $d \times d$ matrix A which is unknown to you. However, it is known that $\|A\|_{OP} \leq 1$ and that one of the two following cases holds: Either (a) A is positive definite or (b) A has an eigenvalue equal to -1. You are given n samples of the value $\langle \Gamma, A\Gamma \rangle$ where $\Gamma \sim N(0, Id)$. How large should n be so that you'll be able to distinguish between the two above cases, with high probability? Try to give both an upper and lower bounds for the dependence of n on d.

12. Somewhat related to open problem 4.2: Fix $\alpha > 0$ and let $X_1, X_2, \dots$ be a sequence of independent Gaussians such that $\mathrm{Var}(X_n) = \frac{1}{\log(n+2)^\alpha}$. Can you give an estimate of $\mathbb{E}[\sup_{n \in \mathbb{N}} X_n]$ in terms of α? For which values of α is it finite?

13. Open problem 4.6: What can you say if for some $p > 1$ it is known that $\mathbb{E}[X_i^p] \leq C$ uniformly for all i? What about the case $p = 2$? What if we assume that the X_i's are Bernoulli random variables (hence they attain only two values)?

14. Open problem 6.1: Show that it is enough to take $M = KN^{0.1}$ random rows.

15. In the max-cut algorithm of Gomans-Williamson that we leant, we rounded the vectors given by the optimization problem according to a hyperplane cut. Denote the vectors given by the SDP by $v_1, ..., v_n \in \mathbb{R}^k$. Let $\Gamma \sim N(0, I_k)$ be a standard Gaussian random vector in $\mathbb{R}^k$ and consider the variables $u_i = \langle \Gamma, v_i \rangle$. The Gomans-Williamson algorithm partitions the vertices according to $\mathrm{sign}(u_i)$. Prove that for any function $f : \mathbb{R} \to \{-1, 1\}$, partitioning the vertices according to $f(u_i)$ gives a worse approximation. Bonus: what if we instead define $u_i = Gv_i$ where G is a matrix of $m \times k$ independent standard Gaussians? (Hint: search for Borell's Gaussian noise stability inequality).

16. Prove a "cheap" version of the geometric result used in the sparsest cut algorithm of Arora-Rao-Vazirani: Let $v_i, i \in [n]$ be a set of unit vectors in $\mathbb{R}^d$ satisfying $\frac{1}{n^2} \sum_{i,j} |v_i - v_j|^2 \geq \frac{1}{2}$. Then there exists two subsets $S, T \subset [n]$ with $|S|, |T| \geq cn$ such that for every $i \in S, j \in T$ one has $|v_i - v_j|^2 \geq c/\log(n)$ (here $c > 0$ is a universal constant). Show that if we replace the expression $c/\log(n)$ by a constant, then the result is not correct.

17. Let $G = (V, E)$ be a graph, and let $f : E \to \mathbb{R}^k$. Try to think of an algorithm that finds an approximation to the c-balanced partition $V = S \cup T$ (balanced means that $|S|, |T| \geq c|V|$) that minimizes that quantity

$$\left| \sum_{e \in E(S,T)} f(e) \right|.$$

What about maximizing the above quantity? (For maximization, assume k is constant, hence the complexity can depend exponentially on k).

18. Open problem 8.3.

    (a) Prove any lower bound on the constant K_G which is strictly bigger than 1.

    (b) Try to give a simple proof that for every matrix $(a_{i,j})_{i,j\in[n]}$ and every n unit vectors $v_1, ..., v_n \in \mathbb{R}^d$ there exist $u_1, ..., u_n \in \{-1, 1\}$ such that

$$\sum_{i,j} a_{i,j}\langle v_i, v_j\rangle \le C\log(n)\sum_{i,j} a_{i,j}u_i u_j.$$

19. Find the VC-dimension of the following families of sets:

    (a) Convex polygons in the plane.

    (b) Triangles in the plane.

    (c) The family of axis-parallel boxes in $\mathbb{R}^d$.

    (d) The family of Euclidean balls (with arbitrary center and radius) in $\mathbb{R}^d$.

    (e) Bonus: Simplices in $\mathbb{R}^d$.

20. Estimate the Gaussian width of the following sets (find the correct order up to constants).

    (a) The set of s-sparse unit vectors.

    (b) The unit ball of the ℓ_p norm in $\mathbb{R}^d$, $p \in [1, \infty]$.

    (c) The set of points $x = (x_1, ..., x_d)$ such that $\|x\|_2 \le 1$ and $x_1 \le x_2 \le ... \le x_d$.

    (d) The set of positive-definite matrices whose operator norm is at most 1.

    (e) The set of positive-definite matrices whose trace is at most 1.

21. In class, we defined the VC-dimension of functions classes for functions whose image is $\{0, 1\}$. Generalize the definition to functions that take values in $\{1, 2, ..., k\}$ for some integer $k \in \mathbb{N}$, and prove an analogue of the Sauer-Shelah lemma.

22. For a set $K \subset \mathbb{R}^d$, denote by $N(K, r)$ the minimum number N such that K is contained in a union of N Euclidean balls of radius r.

    (a) Prove that if K is contained in the unit ball, then

$$\mathrm{GaussianWidth}(K) \le C\inf_{r\ge 0}\left(\sqrt{\log N(K,r)} + r\sqrt{d} + 1\right).$$

    (b) Find an example where the above is not tight up to constants.

23. (a) Let K the ℓ_1 ball in $\mathbb{R}^d$ (hence the convex hull of $\pm e_i$ where e_i are the standard basis vectors). This set has diameter 2. Prove, however, that if L is the span of $d/2$ independent standard Gaussian random vectors (in other words L is a uniformly chosen subspace of dimension $d/2$), then $K \cap L$ typically has a much smaller diameter, namely

$$\mathbb{E}\left[\mathrm{diam}(K\cap L) \le C\sqrt{\frac{\log d}{d}}\right].$$

(b) Try to prove a tail estimate for the above, namely that

$$\mathbb{P}\left[\operatorname{diam}(K \cap L) \geq Ct\sqrt{\frac{\log d}{d}}\right] \leq \exp(-ct), \forall t > 1.$$

24. Let X be an $n \times n$ matrix of rank r and entries in $[-1, 1]$. Suppose we're given a matrix Y such that $Y_{ij} = \delta_{ij} X_{ij}$ with δ_{ij} being independent $0-1$ Bernoulli's with expectation p. Suppose that additionally, you are given the image of the matrix X.

    (a) Prove that for every $\varepsilon > 0$, it suffices to take $p = \frac{C(\varepsilon,r)}{n}$ in order to build a matrix Z such that

    $$\mathbb{E}\frac{1}{n^2}\left[\sum_{i,j} |Z_{ij} - X_{ij}|\right] \leq \varepsilon.$$

    Here $C(\varepsilon, r)$ is any expression that depends only on ε, r. Note that the theorem in class needed an extra $\log n$ factor, but this statement may have a worse dependence on r and moreover the image of the matrix needs to be known.

    (b) How many entries does one need in order to learn the image of a small rank matrix?

25. (a) Prove the following extension of Dudley's bound (for the supremum of Sub-Gaussian processes, which we did in class) that we did in class: Suppose that $(Z(x))_{x\in\mathcal{I}}$ is a sub-Gaussian process with respect to a metric d, namely

    $$\mathbb{P}\left(|Z(x) - Z(y)| > td(x, y)\right) \leq Ce^{-t^2}, \ \ \forall x, y \in \mathcal{I}$$

    and additionally one has that, almost surely $|Z(x) - Z(y)| \leq Kd(x, y)$ for all x, y, for some constant $K > 0$. Then one has

    $$\mathbb{E}\sup_{x\in\mathcal{I}} |Z(x)| \leq C\int_s^{\operatorname{diam}(\mathcal{I})} \sqrt{\log \mathcal{N}(I, \varepsilon, d)}d\varepsilon + sK.$$

    (b) In class, we discussed a way to use Dudley's bound for sub-Gaussian processes in order to find a set of points $x_1, ..., x_n \in [0, 1]$ such that for every 1-Lipschitz function f on $[0, 1]$ one has that

    $$\left|\int f d\nu - \frac{1}{n}\sum_i f(x_i)\right| \leq \frac{c}{\sqrt{n}}.$$

    Can you use the first part of the question to derive a statement of similar nature in higher dimensions? What is the best dependence on n that you can get in dimension d?

26. (Sparsity of RIP-matrices) Let A be a $m \times n$ matrix with entries in $\{0, 1\}$ satisfying $(s, 1/2)-$RIP. Suppose that the average column sparsity of A is d, i.e. A has nd nonzero entries. Show that either $d \gtrsim s$ or $m \gtrsim n$. What if each non-zero $A_{i,j}$ is drawn from $N(0, 1)$?

27. Let $x_1, \ldots, x_n$ be n points in $\mathbb{R}^d$ and let $X = [x_1, \ldots, x_n]$. Suppose we only have estimates for the Euclidean distances between the points: $d_{ij} \approx \|x_i - x_j\|_2^2$. Let Δ be the matrix with entries d_{ij}.

(a) Show that given $d_{ij} = \|x_i - x_j\|_2^2$, there is a choice of the x_i (note that there is no unique choice since any translation, rotation, or reflection of the coordinate system leaves the distances invariant) such that

$$X^T X = -\frac{1}{2} H \Delta H,$$

where $H = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$.

(b) Describe an algorithm to determine the matrix X.

28. (Low Rank Approximation of the Identity) Let I_n denote the $n \times n$ identity matrix. For $0 < \epsilon < 1$, show that there exists a positive semidefinite $n \times n$ matrix $\tilde{I}_n$ such that $|(I_n - \tilde{I}_n)_{ij}| \leq \epsilon$ for all i, j, and rank$(\tilde{I}_n) = O(\log n/\epsilon^2)$. (Hint: Apply the Johnson-Lindenstrauss lemma).