You can work with other students, but you must write-up your solutions by yourself and indicate at the top who you worked with!

Recall that rank$^+(M)$ is the smallest $r$ such that there are entry-wise nonnegative matrices $A$ and $W$ with inner-dimension $r$, satisfying $M = AW$.

**Problem 1**

Which of the following are equivalent definitions of nonnegative rank? For each, give a proof or a counter-example.

(a) the smallest $r$ such that $M$ can be written as the sum of $r$ rank one, nonnegative matrices

(b) the smallest $r$ such that there are $r$ nonnegative vectors $v_1, v_2, ..., v_r$ such that the cone generated by them contains all the columns of $M$

(c) the largest $r$ such that there are $r$ columns of $M$, $M_1, M_2, ..., M_r$ such that no column in set is contained in the cone generated by the remaining $r - 1$ columns

**Problem 2**

Let $M \in \mathbb{R}^{n \times n}$ where $M_{i,j} = (i - j)^2$. Prove that rank$(M) = 3$ and that rank$^+(M) \geq \log_2 n$. *Hint:* To prove a lower bound on rank$^+(M)$ it suffices to consider just where it is zero and where it is non-zero.

**Problem 3**

(a) [Papadimitriou et al. ’97] considered the following document model: $M = AW$ and each column of $W$ has only one non-zero and the support of each column of $A$ is disjoint. Prove that the left singular vectors of $M$ are the columns of $A$ (after rescaling). You may assume that all the non-zero singular values of $M$ are distinct. *Hint:* $MM^T$ is block diagonal, after applying a permutation $\pi$ to its rows and columns.
(b) Let $M$ be $n \times m$ with rows corresponding to terms and columns corresponding to documents. For each document $j$, let column $j$ of $M$ sum to 1, representing a probability distribution $\pi_j$ over terms. Let document $j$ consist of $N$ terms (not necessarily distinct), each drawn independently from $\pi_j$. Let
\[
\hat{M}_{ij} = \frac{1}{N} \# \text{ occurrences of term } i \text{ in document } j
\]
be the matrix of observed term frequencies. Give a bound $t = t(n, m, N, \delta)$ such that with probability $\geq 1 - \delta$ we have that every entry $i, j$ satisfies $|M_{ij} - \hat{M}_{ij}| \leq t$. Use Hoeffding’s inequality:

**Theorem** (Hoeffding). If $X_1, \ldots, X_n$ are independent random variables with $X_i \in [0, 1]$ then $\mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^{n} (X_i - \mathbb{E}[X_i]) \right| \geq t \right) \leq 2 \exp(-2nt^2)$.

(c) We have the following perturbation bound for singular subspaces:

**Theorem** (Papadimitriou et al. ’97). Let $M$ be an $n \times m$ matrix with singular value decomposition $M = U\Sigma V^\top$. Suppose that, for some $k$, the singular values satisfy $21/20 \geq \sigma_1 \geq \cdots \geq \sigma_k \geq 19/20$ and $1/20 \geq \sigma_{k+1} \geq \cdots$. Let $E$ be an arbitrary $n \times m$ matrix with Frobenius norm $\|E\|_F \leq \epsilon \leq 1/20$. Let $M' = M + E$ and let $U'\Sigma'V'^\top$ be its singular value decomposition. Let $U_k$ and $U'_k$ be $n \times k$ matrices consisting of the first $k$ columns of $U$ and $U'$ respectively. Then, $U'_k = U_k R + G$ for some $k \times k$ orthogonal matrix $R$ and some $n \times k$ matrix $G$ with $\|G\|_F \leq 9\epsilon$.

Let $M = AW$ as in part (a), and let $\hat{M}$ be the observed word frequencies from part (b). Suppose the non-zero singular values of $M$ satisfy $19c/20 \leq \sigma_i \leq 21c/20$ for some $c$. State a precise bound (in terms of $n, m, N, \delta, c$) showing that given $\hat{M}$, we can approximately recover the span of the columns of $A$. (Your bound should be an upper bound on $\|G\|_F$ as above.)

**Problem 4**

**Greedy Anchorwords**

1. Set $S = \emptyset$
2. Add the row of $M$ with the largest $\ell_2$ norm to $S$
3. For $i = 2$ to $r$
4. Project the rows of $M$ orthogonal to the span of vectors in $S$
5. Add the row with the largest $\ell_2$ norm to $S$
6. End
Let $M = AW$ where $A$ is separable and the rows of $M$, $A$ and $W$ are normalized to sum to one. Also assume $W$ has full row rank. Prove that Greedy Anchorwords finds all the anchor words and nothing else. Hint: the $\ell_2$ norm is strictly convex — i.e. for any $x \neq y$ and $t \in (0,1)$, $\|tx + (1-t)y\|_2 < t\|x\|_2 + (1-t)\|y\|_2$.

**Problem 5**

In the *multi-reference alignment* problem (considered by e.g. [Perry et al. '17]) we observe many noisy copies of the same unknown signal $x \in \mathbb{R}^d$, but each copy has been circularly shifted by a random offset (as pictured).

![Figure: Three shifted copies of the true signal $x$ are shown in gray. Noisy samples $y_i$ are shown in red. (Figure credit: [Bandeira et al. '17])](image)

Formally, for $i = 1, 2, \ldots, n$ we observe

$$y_i = R_{\ell_i}x + \xi_i$$

where: the $\ell_i$ are drawn uniformly and independently from $\{0,1,\ldots,d-1\}$; $R_{\ell}$ is the operator that circularly shifts a vector by $\ell$ indices; $\xi_i \sim \mathcal{N}(0, \sigma^2 I_{d \times d})$ with $\{\xi_i\}_i$ independent; and $\sigma > 0$ is a known constant. Think of $d$, $x$ and $\sigma$ as fixed while $n \to \infty$. The goal is to recover $x$ (or a circular shift of $x$).

(a) Consider the tensor $T(x) = \frac{1}{d} \sum_{\ell=0}^{d-1} (R_{\ell}x) \otimes (R_{\ell}x) \otimes (R_{\ell}x)$. Show how to use the samples $y_i$ to estimate $T$ (with error tending to zero as $n \to \infty$). Take extra care with the entries that have repeated indices (e.g. $T_{aab}, T_{aaa}$).

(b) Given $T(x)$, prove that Jennrich’s algorithm can be used to recover $x$ (up to circular shift). Assume that $x$ is *generic* in the following sense: let $x' \in \mathbb{R}^d$ be arbitrary and let $x$ be obtained from $x'$ by adding a small perturbation $\delta \sim \mathcal{N}(0,\epsilon)$ to the first entry. *Hint*: form a matrix with rows $\{R_{\ell}x\}_{0 \leq \ell < d}$, arranged so that the diagonal entries are all $x_1$. 

---

**Algorithmic Aspects of Machine Learning**

**PS # 1**