MATHEMATICAL TOOLS IN COMPUTER SCIENCE
FINAL EXAM 5769

ANONYMOUS

Anonymous
Questions answered:

CONTENTS

1. Probability - Question 2

Notation. $G = (V, E)$ is a 3-uniform hypergraph with vertex set V and edge set $E \subset \binom{V}{3}$. $x \in V$ will be called *isolated* if it is not part of any edge, and formally, if $\forall e \in E : x \notin e$.

I assume $|V| \geq 3$ because otherwise the graph creation algorithm would fail.

The probability space. We create random graphs G over a vertex set V as above by randomly choosing edges. The probability space for choosing a single edge will be $\mathcal{D} = (\Omega, P)$, where Ω is the set of possible edges, $\Omega = \binom{V}{3}$, and the probability function uniform: $\forall \omega \in \Omega : P(\omega) = \frac{1}{|\Omega|} = \frac{1}{\binom{|V|}{3}}$.

Unless specifically noted, all random variables used will be indicator functions, and I will define them by the event they indicate. The notation $P(X)$ and $P(\neg X)$ will mean the same as $P(X = 1)$ and $P(X = 0)$ respectively.

Single edge events. The basic single edge event we will want to look at is if a specific vertex $v \in V$ is included in an edge e chosen with distribution $\mathcal{D}$. Due to future use, we will actually look at the complement, the event that $v \in V$ is not in the chosen edge. Let this random (indicator) variable be called X_v or X_v^1. Obviously, due to symmetry and linearity of expectation $E(X_v) = P(v \notin e) = 1 - \frac{3}{|V|}$. More directly, we can count the edges which do not include v, and divide by $|\Omega|$ since the probability is uniform. This set is the choice of 3 vertices from $V \setminus \{v\}$, which gives:

$E(X) = P(X_v) = \frac{\binom{|V|-1}{3}}{\binom{|V|}{3}} = \frac{(|V|-1)! \cdot 3! \cdot (|V|-3)!}{|V|! \cdot 3! \cdot (|V|-4)!} = \frac{|V|-3}{|V|}$ as required.

Repeated I.I.D. If we randomly choose k edges for G I.I.D. from $\mathcal{D}$ (which allows repetition), we want to find the probability that vertex v is isolated. If $k = 1$, the answer is clearly $E(X_v)$, which is the expectation of a single edge "missing" v.

Let the random variables $X_{v,1}, X_{v_2}, \cdots, X_{v,k}$ be chosen I.I.D from distribution $\mathcal{D}$, (identical to X_v above). The requirement that v be isolated after k edges are chosen is equivalent to the requirement that each of the k chosen edges "miss" v. The probability of the j_{th} edge missing v is $E(X_{v,j})$, edges are chosen independently, so the probability of all edges missing is: (We will call this event X_v^k in a little abuse of notation.)

$$\begin{aligned}
E(X_v^k) &= P(X_v^k) && \text{Indicator, mixed notation} \\
&= P(\textstyle\bigwedge_{i=1}^k X_{v,i}) && \text{Splitting the event} \\
&= \textstyle\prod_{i=1}^k P(X_{v,i}) && \text{Independence} \\
&= \textstyle\prod_{i=1}^k E(X_{v,i}) && \text{Indicator} \\
&= \textstyle\prod_{i=1}^k E(X_{v,1}) && \text{Identical} \\
&= E(X_{v,1})^k = E(X_v)^k && \\
&= \left(1 - \tfrac{3}{|V|}\right)^k && \text{Above}
\end{aligned}$$

Note that event idicated by X_v^k is actually in the probability space $\mathcal{D}^k = (\Omega, P)^k$.

Note also that the derivation used here is consistent with the trivial k, and as expected $X_v^0 = 1$.

Expected isolated vertices. Now we know the probability that in a random graph with $|V|$ vertices and k edges, a specific vertex v will be isolated. Due to the symmetry of the probability space, there is nothing special about v, and we will get the same result for each of the vertices.

Let Y be a random variable counting the number of isolated vertices in a k-edged graph. From its definition, $Y = \sum_{v \in V} X_v^k$, since each X_v^k tells us if a single vertex is isolated. The various random variables X_v^k are definitely not independent. (A trivial example is $|V| = 4, k = 1$) However expectation is linear, and all we want is the expected number of isolated vertices. Therefore, using this, symmetry, and the previous results:

$$\begin{aligned}
E(Y) &= E(\textstyle\sum_{v \in V} X_v^k) && \text{Definition} \\
&= \textstyle\sum_{v \in V} E(X_v^k) && \text{Linearity of expectation} \\
&= |V| \cdot E(X_{\hat{v}}^k) && \text{Identical} \\
&= |V| \cdot \left(1 - \tfrac{3}{|V|}\right)^k && \text{Above}
\end{aligned}$$

The threshold. We can find the threshold by setting $E(Y) = 1$ and showing k as a function of $|V|$.

$$1 = E(Y) = |V| \cdot \left(1 - \tfrac{3}{|V|}\right)^k \iff \tfrac{1}{|V|} = \left(1 - \tfrac{3}{|V|}\right)^k$$

Note that now k can be shown as a function of $|V|$. For now I will call the threshold I'll find as $\hat{k}(|V|)$, using the previous equation. Since the bound we have to show is only an asymptotic bound, using o, ω notation, it will be enough to show linear constants which bound from above and below.

Above: 2. Assume we use $k = 2 \cdot \hat{k}(|V|)$. Then from above we have:

$$\begin{aligned}
E(Y) = \ & |V| \cdot \left(1 - \tfrac{3}{|V|}\right)^k && \text{Above} \\
&= |V| \cdot \left(1 - \tfrac{3}{|V|}\right)^{2\cdot\hat{k}(|V|)} && \text{Inserting threshold} \\
&= |V| \cdot \left(\left(1 - \tfrac{3}{|V|}\right)^{\hat{k}(|V|)}\right)^2 \\
&= |V| \cdot \left(\tfrac{1}{|V|}\right)^2 \\
&= \tfrac{1}{|V|} \xrightarrow{|V|\to\infty} 0
\end{aligned}$$

Y is a nonnegative random variable which accepts integer values, so $P(Y \neq 0) = P(Y > 0) = P(Y \geq 1)$, and we can use the Markov inequality, (with $a = 1$) giving us:

$P(Y \neq 0) = P(Y \geq 1) \leq \frac{E(Y)}{1} = \frac{E(Y)}{1} \xrightarrow{|V|\to\infty} 0$ as required.

Below: $\frac{1}{4}$. Assume we use $k = \frac{\hat{k}(|V|)}{4}$. Then from above we have:

$$\begin{aligned}
E(Y) &= |V| \cdot \left(1 - \tfrac{3}{|V|}\right)^k \\
&= |V| \cdot \left(1 - \tfrac{3}{|V|}\right)^{\frac{\hat{k}(|V|)}{4}} \\
&= |V| \cdot \sqrt[4]{\left(1 - \tfrac{3}{|V|}\right)^{\hat{k}(|V|)}} \\
&= |V| \cdot \sqrt[4]{\tfrac{1}{|V|}} \\
&= \sqrt[4]{|V|}^3 \xrightarrow{|V|\to\infty} \infty
\end{aligned}$$

This time we will use the Chebyshev inequality.

What is $Var(Y)$? Unfortunately, Y is a sum of correlated random variables. Therefore:

$Var(Y) = Var(\sum_{v\in V} X_v^k) = \sum_{v\in V}\sum_{w\in V} Cov(X_v^k, X_w^k)$.

However, not all is lost, since the various vertices are correlated negatively with each other.

$\forall v \in V, w \in V, v \neq w :$

$$\begin{aligned}
Cov(X_v^k, X_w^k) &= E(X_v^k \cdot X_w^k) - E(X_v^k) \cdot E(X_w^k) && \text{Definition} \\
&= P(X_v^k, X_w^k) - P(X_v^k) \cdot P(X_w^k) && \text{Indicator random variables} \\
&= P(X_v^k | X_w^k) \cdot P(X_w^k) - P(X_v^k) \cdot P(X_w^k) && \text{Chain rule} \\
&= (P(X_v^k | X_w^k) - P(X_v^k)) \cdot P(X_w^k) && \text{Associativity} \\
&= \left(\left(1 - \tfrac{3}{|V|-1}\right)^k - \left(1 - \tfrac{3}{|V|}\right)^k\right) \cdot P(X_w^k) && \text{Combinatorics} \\
&\leq 0 && \text{Probability is nonnegative}
\end{aligned}$$

Therefore, for the parameters of the Chebyshev inequality, we have $\mu = E(Y) = \sqrt[4]{|V|}^3$ for the expectation. For the variance we have combining the above results:

$$\begin{aligned}
Var(Y) &= Var(\textstyle\sum_{v\in V} X_v^k) && \text{Definition of } Y \\
&= \textstyle\sum_{v\in V}\sum_{w\in V} Cov(X_v^k, X_w^k) && \text{Variance of sum} \\
&= \textstyle\sum_{v\in V} Cov(X_v^k, X_v^k) + \sum_{v\neq w\in V} Cov(X_v^k, X_w^k) && \text{Split index} \\
&\leq \textstyle\sum_{v\in V} Cov(X_v^k, X_v^k) && \text{Negative covariance} \\
&= \textstyle\sum_{v\in V} Var(X_v^k) && \text{Definition of variance} \\
&= |V| \cdot P(X_v^k) \cdot (1 - P(X_v^k)) && \text{Identical} \\
&\leq |V| && \text{Probability less than 1}
\end{aligned}$$

And finally we get to use Chebyshev's inequality:

$$\begin{aligned} P(Y=0) &\leq P(|Y-E(Y)| \geq E(Y)) && \text{Subevent} \\ &\leq \frac{Var(Y)}{(E(Y)-\varepsilon)^2} && \text{Chebyshev's inequality} \\ &\leq \frac{|V|}{\left(\sqrt[4]{|V|}^3\right)^2} && \text{Previous results} \\ &\leq \frac{2\cdot|V|}{\left(\sqrt[4]{|V|}^3\right)^2} \\ &\leq \frac{2}{\sqrt{|V|}} \xrightarrow{|V|\to\infty} 0 \end{aligned}$$

as required in both directions.

Cleaning up. We still need to specify $\hat{k}$. We have from before that: $1 = |V| \cdot \left(1 - \frac{3}{|V|}\right)^k$ and therefore:

$$0 = \log(1) = \log\left(|V| \cdot \left(1 - \frac{3}{|V|}\right)^k\right) = \log|V| + k \cdot \log\left(1 - \frac{3}{|V|}\right) = \log|V| + k \cdot \log\left(\frac{|V|-3}{|V|}\right)$$

And therefore:

$k = \frac{-\log|V|}{\log\left(\frac{|V|-3}{|V|}\right)} = \frac{-\log|V|}{\log(|V|-3) - log|V|} = \frac{\log|V|}{\log|V| - log(|V|-3)}$

And $\hat{k}(|V|) = \frac{\log(|V|)}{\log|V| - log(|V|-3)}$.

And looking at the Taylor expansion of the logarithm function near $x = 1$ shows that:

$\hat{k}(|V|) = \frac{-\log|V|}{\log\left(\frac{|V|-3}{|V|}\right)} = \frac{-\log|V|}{\log\left(1-\frac{3}{|V|}\right)} = \frac{-\log|V|}{-\sum_{n=1}^{\infty}\frac{(\frac{3}{|V|})^n}{n}} \xrightarrow{|V|\to\infty} \frac{|V|\cdot\log|V|}{3} \in \Theta(|V| \cdot \log|V|)$.

Writing a limit in this form is an abuse of notation, but the result is correct. More rigorous would be to prove $lim_{|V|\to\infty}\frac{\hat{K}(|V|)}{|V|\cdot\log|V|} = \frac{1}{3}$, which is of course also correct.

2. Linear Algebra - Question 3

Notation. Let $G = (V, E)$ be a graph with vertex set V and edge set E. G's line graph will be $L(G) = (\bar{V}, \bar{E})$. The adjacency matrix of these graphs will be A_G and $A_{L(G)}$. The incidence matrix of G will be called M, with the rows referring to vertices, and the columns to edges.

Relationship between the matrices. Multiplying the adjacency matrix of graph G, M, with its transpose, M^t, gives a matrix which shows in each cell the common vertices (or edges) of two edges (or vertices) of the multiplication (depending if we have $M^t \cdot M$ or $M \cdot M^t$). Note that for different edges (or vertices), the common vertices (or edges) will be 1 if they share a common vertex (or edge), otherwise 0. For the same vertex or edge (the diagonal), we get the rank of the vertex or 2 for both sides of the edge.

To recover the adjacency matrices, we just subtract the diagonals, giving: $A_{L(G)} = M^t \cdot M - 2 \cdot I$, and $A_G = M \cdot M^t - diag(M \cdot M^t)$. (Where I is the identity matrix, and $diag(A)$ the diagonal mask of A.)

Minimal eigenvalues of $L(G)$. We mentioned that the adjecency matrix of $L(G)$, $A_{L(G)} = M^t \cdot M - 2 \cdot I$. M is a real valued matrix, so $M^t \cdot M$ is a positive semidefinite matrix and diagonalizable. Since all vectors are eigenvectors of $2 \cdot I$, any eigenvector for $M^t \cdot M$ will also be an eigenvector for $A_{L(G)}$. However, since $M^t \cdot M$ is positive semidefinite, the minimal eigenvalue for it is at least 0. Therefore, for any eigenvector of $M^t \cdot M$, (and also $A_{L(G)}$), we find that:
$M^t \cdot M \cdot x = \lambda \cdot x \geq 0 \cdot x$, and therefore:
$A_{L(G)} \cdot x = (M^t \cdot M - 2 \cdot I) \cdot x = M^t \cdot M \cdot x - 2 \cdot I \cdot x = \lambda \cdot x - 2 \cdot x \geq 0 \cdot x - 2 \cdot x = -2 \cdot x$,
and all the eigenvectors are at least -2, as required.

Eigenvalues -2. In general, we have $M^t \cdot M \in \mathbb{R}^{|E| \times |E|}$, and $M \cdot M^t \in \mathbb{R}^{|V| \times |V|}$. Therefore $rank(M^t \cdot M) = rank(M \cdot M^t) = rank(M) \leq \min(|V|, |E|)$. If $|E| > |V|$, then the vectors in $M^t \cdot M$ cannot be independent, $|E|$ vectors spanning a dimension of $rank(M) \leq |V| < |E|$, which means it has an eigenvalue of 0, and a matching eigenvector.

As before, let x be such an eigenvector: $A_{L(G)} \cdot x = (M^t \cdot M - 2 \cdot I) \cdot x = M^t \cdot M \cdot x - 2 \cdot I \cdot x = 0 \cdot x - 2 \cdot x = -2 \cdot x$, and -2 is an eigenvalue, as required.

The rest of the eigenvalues. If G is d-regular ($d \geq 1$), then each vertex shares exactly d edges with itself, and we can view A_G as above as: $A_G = M \cdot M^t - diag(M \cdot M^t) = M \cdot M^t - d \cdot I$.
We already saw that: $A_{L(G)} = M^t \cdot M - 2 \cdot I$.
Let B, C be block matrices as defined below, M is a general $m \times n$ block, I is the identity matrix of the necessary dimensions.
$B = \begin{pmatrix} I & M \\ M^t & xI \end{pmatrix}$, and $C = \begin{pmatrix} xI & -M \\ 0 & I \end{pmatrix}$. Then $BC = \begin{pmatrix} xI & 0 \\ * & xI - M^tM \end{pmatrix}$, and $CB = \begin{pmatrix} xI - MM^t & 0 \\ * & xI \end{pmatrix}$.
Since $det(BC) = det(CB)$, and they are triangular block matrices, we find that:
$det(xI - M^tM) = x^{m-n} \cdot det(xI - MM^t)$.
This proves that the first $min(|V|, |E|)$ eigenvalues of M^tM and MM^t are the same, with the rest completed with zeros, due to the matching of the characteristic polynomial.

We can now repeat the -2 proof above and show that for any eigenvalue $\lambda \neq 0$ in (both) M^tM and MM^t, with eigenvectors x, y (of the correct dimension) we get:
$A_{L(G)} \cdot x = (M^t \cdot M - 2 \cdot I) \cdot x = M^t \cdot M \cdot x - 2 \cdot I \cdot x = \lambda \cdot x - 2 \cdot x = (\lambda - 2) \cdot x$.
$A_G \cdot y = (M \cdot M^t - d \cdot I) \cdot y = M \cdot M^t \cdot y - d \cdot I \cdot y = \lambda \cdot y - d \cdot y = (\lambda - d) \cdot y$.

Therefore, combining the above, we find that for every eigenvalue λ of A_G, there is an eigenvalue $\lambda + d - 2$ for $A_{L(G)}$, with the same multiplicity. All the excess eigenvalues ($|E| - |V|$ of them) will be additional -2 eigenvalues, if necessary.

When $d \geq 3$ this is necessary, and -2 is added. The above formula works also for $d = 2$, except that additional -2 eigenvalues are not necessary. In this case $\lambda + d - 2 = \lambda$ as required in the question. It is also easy to see that in this case $L(G) = G$ which obviously have the same spectrum.

In the case of $d = 1$, we need to make corrections to the formula, since $|V| > |E|$, but it is easy to see that the the line graph in this case is empty, since no 2 edges share a column. Obviously, The spectrum of the empty graph is only 0. Note that this also fits the formula with connections, since the eigenvalues of A_G are $1, -1$, with 1 changing to 0 according to the formula, and -1 lost in the rank loss of the multiplication.

The real matrix A**.** Let $A \in \mathbb{R}^{\binom{n}{2}\times\binom{n}{3}}$ be defined with: $A_{S,T}\begin{cases}1 & S\subset T\\ 0 & S\not\subset T\end{cases}, |S|=2, |T|=3$

The goal is to find the nonzero eigenvalues of $\sqrt{A\cdot A^t}$, or $\sqrt{A^t\cdot A}$, which are of course the same.

We will use $A\cdot A^t$ which will give a matrix which is easier to work with. This matrix has a dimension of $\binom{n}{2}\times\binom{n}{2}$, with each cell counting the common 3-sets to both 2-sets. Each 2-set is part of $n-2$ 3-sets, thus the diagonal (common with itself) will be a constant $n-2$. Any other pair of 2-sets will equal 1, if there is a common point between them (two pairs with a common point define a triple), and 0 otherwise (4 points must be in the 3-set). Therefore:

$$[A\cdot A^t]_{S_i,S_j} = |S_i\cap S_j| = \begin{cases}n-2 & i=j\\ 1 & i\neq j, S_i\cap S_j\neq\emptyset\\ 0 & S_i\cap S_j=\emptyset\end{cases}, |S_i|=|S_j|=2$$

But if we remove the diagonal, we find that:

$$[A\cdot A^t-(n-2)\cdot I]_{S_i,S_j} = |S_i\cap S_j| = \begin{cases}1 & |S_i\cap S_j|=1\\ 0 & \text{else}\end{cases}, |S_i|=|S_j|=2.$$

But this is the adjacency matrix of the line grah of the complete graph K_n. Its adjacency matrix is of dimension $n\times n$ with

$$A(K_n)_{ij}=\begin{cases}0 & i=j\\ 1 & i\neq j\end{cases}.$$

The edges set of K_n is isomorphic to $\binom{n}{2}$, with edges sharing a vertex being 2-sets sharing a point. This describes $[A\cdot A^t-(n-2)\cdot I]$, which is $A_{L(K_n)}$.

The eigenvalues of a complete graph K_n are $n-1$ with multiplicity 1 (with eigenvector $(1,1,\cdots,1)$), and -1 with multiplicity $n-1$, with eigenvectors any such that $\sum_i v_i=0$. (This can also be seen trivially by looking at $A_{K_n}=\vec{1}-I$, with $\vec{1}$ being the ones matrix.)

K_n is of course $(n-1)$-regular, and therefore, according to the previous section, the eigenvalues of the line graph of K_n will be $\lambda_i+n-1-2$ for each eigenvalue λ_i of K_n, with -2 filling out the rest of the $\binom{n}{2}-n$ values. This gives the eigenvalues $2n-4$ with multiplicity 1, $n-4$ with multiplicity $n-1$, and -2 with multiplicity $\binom{n}{2}-n$.

Finally, $A\cdot A^t=A_{L(K_n)}+(n-2)\cdot I$, so following the proof above, the eigenvalues of $A\cdot A^t$ will be the the eigenvalues of $A_{L(K_n)}$, adding a constant $n-2$. This gives $3n-6$ with multiplicity 1, $2n-6$ with multiplicity $n-1$, and $n-4$ with multiplicity $\binom{n}{2}-n$.

Then the singular values are the roots of these, giving: $\{\sqrt{3n-6}(\times 1), \sqrt{2n-6}(\times n-1), \sqrt{n-4}(\times\binom{n}{2}-n)\}$.

This is all assuming $\binom{n}{3}\geq\binom{n}{2}$. $n=3$ happens to fit in, since $2n-6=0$, $\binom{n}{2}-n=0$, and if $n=2$, you just need to skip the additional singular values of $\sqrt{n-4}$.

3. Linear Programming - Question 5

$b(n,k,p)$ **in LP form.** Since $b(n,k,p)$ is simpler than $a(n,k,p)$, I will turn it into LP form first.

From the definition of expectency and the bounds of B: $E(f(B)) = \sum_{i=0}^{n} P(B=i) \cdot f(i)$.

Calling f's coefficients β_j gives: $E(f(B)) = \sum_{i=1}^{n} P(B=i) \cdot f(i) = \sum_{i=0}^{n} \sum_{j=0}^{k} P(B=i) \cdot \beta_j \cdot i^j$.

The only unknowns at this point are the β_js, and we finally get the linear form: $\sum_{j=0}^{k} \beta_j \cdot \left[\sum_{i=0}^{n} P(B=i) \cdot i^j\right]$.

This can also be converted to the much nicer: $\sum_{j=0}^{k} \beta_j \cdot \left[\sum_{i=0}^{n} P(B=i) \cdot i^j\right] = \sum_{j=0}^{k} \beta_j \cdot E(B^j)$.

The constraints are only on the values f gets at specific points, namely: $\forall i \in [n-1] : f(i) \geq 0$ and $f(n) \geq 1$.

And as above, $f(n)$ can be easily converted to the linear form, giving constraints $f(i) = \sum_{j=0}^{k} \beta_j \cdot i \geq 0/1$.

$a(n,k,p)$ **in LP form.** Next, we convert $a(n,k,p) = \max[P(X=n)|(X_1, X_2, \cdots, X_n)]$ into a linear form. The above result for $b(n,k,p)$ can give us a hint where we want to go.

This time the target is trivial, maximizing a single variable. The variable vector I will use will be the probability vector for X. I will also show that the condition $P(X_i) = p$ is not necessary.

Given a probability vector for X, $E(X^j)$ is completely determined and is not dependent on the X_is themselves. Furthermore, if $E(X) = E(B) = n \cdot p$, there exists a joint distribution of $(X_1, X_2, \cdots, X_n)$ such that $X = \sum_{i=1}^{n} X_i$, and each X_i is a boolean indicator function with $E(X_i) = p$. The proof is simple: use independence of X_is, given X.

Define $S \subseteq (1, 2, \cdots, n)$. We now define a distribution using $P(X_i = 1 \iff i \in S) = P(X = |S|) \cdot \frac{1}{\binom{n}{|S|}}$. This is a distribution, since:

$$\begin{aligned}
& \textstyle\sum_{S \subseteq (1,2,\cdots,n)} P(X_i = 1 \iff i \in S) && \\
= \;& \textstyle\sum_S P(X = |S|) \cdot \frac{1}{\binom{n}{|S|}} && \text{Above} \\
= \;& \textstyle\sum_{j=0}^{n} \sum_{|S|=j} P(X = j) \cdot \frac{1}{\binom{n}{j}} = && \text{Splitting the sum} \\
= \;& \textstyle\sum_{j=0}^{n} \binom{n}{j} \cdot P(X = j) \cdot \frac{1}{\binom{n}{j}} && \text{Symmetry,combinatorics} \\
= \;& \textstyle\sum_{j=0}^{n} P(X = j) = 1 && \text{Distribution on } X
\end{aligned}$$

This distribution is consistent with the probability vector, because for any specific j, $\sum_{|S|=j} P(X=j) \cdot \frac{1}{\binom{n}{j}} = \binom{n}{j} \cdot P(X=j) \cdot \frac{1}{\binom{n}{j}} = P(X=j)$, as required.

For every X_i, $E(X_i) = \frac{E(X)}{n} = p$ due to linearity of expectation and symmetry. Or directly, specifying $X_i(S)$ as the value of random variable X_i indicating if the event $X_i = 1 \iff i \in S$ takes place:

$$\begin{aligned}
E(X_i) \;& = P(X_i = 1) = && \\
& = \textstyle\sum_S X_i(S) \cdot P(X_i = 1 \iff i \in S) && \text{Chain rule} \\
& = \textstyle\sum_{i \in S} P(X_i = 1 \iff i \in S) && \text{sum only when } X_i(S) = 1 \\
& = \textstyle\sum_{j=0}^{n} \sum_{i \in S, |S|=j} P(X_i = 1 \iff i \in S) && \text{split by size} \\
& = \textstyle\sum_{j=0}^{n} \sum_{i \in S, |S|=j} P(X = j) \cdot \frac{1}{\binom{n}{j}} && \text{As above} \\
& = \textstyle\sum_{j=0}^{n} \binom{n-1}{j-1} \cdot P(X = j) \cdot \frac{1}{\binom{n}{j}} && \text{combinatorics, since } i \text{ is specified} \\
& = \textstyle\sum_{j=0}^{n} P(X = j) \cdot \frac{(n-1)! \cdot j! \cdot (n-j)!}{n! \cdot (j-1)! \cdot (n-j)!} && \\
& = \textstyle\sum_{j=0}^{n} P(X = j) \cdot \frac{j}{n} = E(\frac{X}{n}) = \frac{E(X)}{n} = \frac{n \cdot p}{n} = p && \text{as required.}
\end{aligned}$$

Therefore we can ignore the conditions of $E(X_i) = p$, and $X = \sum_i X_i$.

Finally we can write the conditions of $a(n,k,p)$ as $E(X^j) = E(B^j)$ for $j \in (0, 1, \cdots, k)$. Note that we don't really need to include $j = 0$, since it is a constant. We include it to use the same dimensions as for $b(n,k,p)$.

Calling the probability vector for X, α, we have:

$E(X^j) = \sum_{i=0}^{n} \alpha_i \cdot i^j$

$E(B^j) = \sum_{i=0}^{n} P(B=i) \cdot i^j$

And therefore constraints: $\sum_{i=0}^{n} \alpha_i \cdot i^j = \sum_{i=0}^{n} P(B=i) \cdot i^j$.

In this case we also have the constraint $\alpha \geq 0$, since α is a probability vector.

Comparing the two. To summarize the results until now: Let $A \in \mathbb{R}^{(n+1)\times(k+1)}$ be the matrix for the constraints. (the indexes will start from 0 here.) A will be defined by $A_{ij} = i^j$. Using this notation we have:

$b(n,k,p) = \min_\beta \langle \{E(X^j)\}_{j=0}^{k}, \beta \rangle$ s.t. $A \cdot \beta \geq (0, 0, \cdots, 0, 1)^t$, and:

$a(n,k,p) = \max_\alpha \langle \alpha, (0, 0, \cdots, 0, 1)^t \rangle$ s.t. $\alpha^t \cdot A = [\{E(X^j)\}_{j=0}^{k}]^t, \alpha \geq 0$.

This is one of the standard statements of the linear optimization duality problem. To convert this to the regular standard form with nonnegative variables and inequality, we replace each β_j with a pair of nonnegative variables such that $\beta_j = \beta_{j+} - \beta_{j-}$. This conversion of $b(n, k, p)$ would combine with the splitting of the equalities in the $a(n, k, p)$ constraints into pairs of inequalities, leaving both problems in the regular standard form.

Duality theorems. Let us name two functions, $a'(n, \alpha) = P(X = n)$, and $b'(n, \beta) = E(f(B))$. That is, they are placements which attempt to solve a, b. Note they might not be optimal and might not follow the constraints.

The weak duality theorem states that for any feasible solutions α, β which are compatible with the constraints, the $a'(\alpha) < b'(\beta)$. This means that if one is unbounded (in the correct direction), the other is not feasible. This does not state that at least one will be feasible. This also means that if both are feasible, they are both bound (and have an optimal solution due to compactness).

The strong duality theorem states that in a linear programming problem, if the primal (or the dual) has an optimal solution, then the dual (or primal) also has an optimal solution, and they both reach the same optimal target.

Combining this with the above, we only need to prove that both problems are feasible, and then the two theorems tell us that the optimal targets are the same, and we get $a(n, k, p) = b(n, k, p)$ as required.

$a(n, k, p)$ is feasible since there is always a distribution which is “similar” to distribution B, B itself. Clearly for any j, $E(B^j) = E(B^j)$.

$b(n, k, p)$ is feasible since there is always a polynomial which follows the constraints for any set of (n, k, p). This polynomial is the constant $f(x) = 1$. It is always positive and more than 1 for any n. It is of degree 0, thus fits for any k. p has nothing to do with the constraints.

Since both problems are feasible, $a(n, k, p) = b(n, k, p)$ for any appropriate (n, k, p) (in terms of type, if $0 \leq p \leq 1, n \in \mathbb{N}, k \in \mathbb{N}$), as required.

Additional notes. We get for free that $a(n, 1, p) = b(n, 1, p)$.

We can calculate $P(B = i)$ for any i, but we don’t need to. for the same reason we could ignore the composition of X, we can ignore the composition of B. The proof would have stayed the same even if we were given an arbitrary probability vector for B. The optimal result might change, but the proof would not.

4. Harmonic Alalysis - Question 7

Some notation. In the course of this question I will consider 3 bases. The first basis is the standard basis: $\delta_a(x) = \begin{cases} 1 & a = x \\ 0 & a \neq x \end{cases}$

The second is the fourier basis: $\chi_a(x) = (-1)^{\langle a,x \rangle}$.

The third basis (which will be proven later) is what is defined in the question, $P_S(x) = \prod_{i \in S} x_i$, which is clearly $P_S(x) = \begin{cases} 1 & S \subseteq x \\ 0 & S \nsubseteq x \end{cases}$

I will mix the notations of sets (S) and vectors (a) and numbers ($num(x_i) = 2^i$), using the natural isomorphism.

Since everything in this question is in $\mathbb{Z}_2^n$ and therefore with real numbers, it will always hold that $\langle x, y \rangle = \langle y, x \rangle$, and $\chi_a(x) = \chi_x(a)$, and $\delta_a(x) = \delta_x(a)$. I will freely switch these indexes.

$\{P_S\}_{S \subseteq [n]}$ is a basis. First of all the size of the set $\{P_S\}_{S \subseteq [n]}$ is the dimension of the space (2^n). This means that all that has to be shown is that the set covers the function space. If it does, then the dimensions make sure they are a basis.

We will prove coverage using induction. The basis of the induction will be $\delta(2^n - 1)$, or all S, where $|S| = n$. (There is only one.)

Clearly $\delta_{[n]}(2^n - 1) = P_{[n]}(2^n - 1)$. The only $S \subseteq [n]$ such that $[n] \subseteq S$ is obviously $[n]$ itself, which means that $\forall x \neq 2^n - 1 : \delta_{[n]}(x) = P_{[n]}(x) = 0$, and $\delta_{[n]}(2^n - 1) = P_{[n]}(2^n - 1) = 1$, as required.

Now assume we covered all S with $|S| \geq k + 1$. Let a be a set with $|a| = k$.

From the definition in the notations above, we find $P_a(x) = \sum_{a \subseteq S \subseteq [n]} \delta_S(x) = \delta_a(x) + \sum_{a \subset S \subseteq [n]} \delta_S(x)$.

$\delta_a(x) = P_a(x) - \sum_{a \subset S \subseteq [n]} \delta_S(x)$.

However, all the sets $S \supset a$ have cardinality of at least $k + 1$, thus were already covered previously acccording to the assumption. Therefore, since all elements of the standard basis can be covered by functions P_S, the set $\{P_S\}_{S \subseteq [n]}$ is a basis, as required.

Note that I did not say anything about what the coefficient vectors look like. However, if we want, using the same induction as before can also tell us about the coefficient vectors, following combinatorial rules.

Matching the bases. Remember that the Fourier coefficients on functions $\{0,1\}^n \to \mathbb{C}$ are defined by $\chi_a(x) = (-1)^{\langle x,a \rangle}$.

Let G be the set of all functions $\{0,1\}^n \to \mathbb{C}$. Let M_k be the set of functions with expansion in the given basis with $degM \leq k$. Let F_k be the set of functions with degree of at most k in the Fourier basis. The sets M_k, F_k are closed to addition, (if $m_1, m_2 \in M_k$, then for all $|S| > k$, $\breve{m}_1(s) = \breve{m}_1(s) = 0$, and scalar multiplication, (same reason) and therefore $m_1 \breve{+} m_2(s) = \breve{m}_1(s) + \breve{m}_1(s) = 0$). Therefore M_k is a linear subspace of G. Identical reasoning shows that F_k is also a linear subspace. Counting the number of base members with a degree of at most k shows that $dim(M_0) = dim(F_0) = 0 = \binom{n}{0}$. The dimensions added for each degree is simply $dim(M_{k+1}) - dim(M_k) = dim(F_{k+1}) - dim(F_k) = \binom{n}{k+1}$.

Since the dimensions match, to show the subspaces match too, we only need to show one direction of containment. The other direction comes for free. To do this we will show that the basis of M_k is covered by the basis of F_k.

The only elements of the two bases which are of degree 0, are χ_0 and $P_\emptyset$. Since $\forall x \in \{0,1\}^n : \breve{g}(0) = \hat{g}(0) = 1$, these are the same function, and clearly span the same subspace. Let us assume the subspaces $M_{k-1} = F_{k-1}$, we will prove that also $M_k = F_k$. Let P_a be a character of degree k. Since all the bits are symmetric, we can assume without loss of generality that $a = 2^k - 1$. That means that $P_a(x) = 1$ iff the k final bits are 1.

But this function can be viewed as a standard basis function ($\delta^k_{2^k - 1}$) from $\{0,1\}^k \to \mathbb{C}$ which clearly has a Fourier transform of degree k. This is $f(x) \propto \sum_{a=0}^{2^k-1} (-1)^{\langle 2^k - 1, a \rangle} \chi_a(x)$ which is clearly of degree k.

What if we add more irrelevant bits? It does not matter. This gives:

$P_{2^k-1}(x) \propto \sum_{a=0}^{2^n-1} (-1)^{\langle 2^k-1,a \rangle} \chi_a(x) = \sum_{j=0}^{2^{n-k}} \sum_{a=j \cdot 2^k}^{(j+1) \cdot 2^k - 1} (-1)^{\langle 2^k-1,a \rangle} \chi_a(x) \propto \sum_{a=0}^{2^k-1} (-1)^{\langle 2^k-1,a \rangle} \chi_a(x)$

Due to symmetry between bits, this is true for any set of k bits chosen.

We find that any basis member of the given basis of degree k can be shown in the Fourier character basis with the same degree. This means $M_k \subseteq F_k$, and since the dimensions are compatible, $M_k = F_k$ for every k, and $\forall g \in G : deg_M(g) = deg_F(g)$, as required.

Back to the examples. Since $\forall g \in G : deg_M(g) = deg_F(g)$, we only need to show the degree in one of the two bases.

The constant function $g(x) = 5$ which was mentioned above is represented by $g(x) = 5 \cdot \chi_\emptyset(x) = 5 \cdot P_\emptyset(x)$, which are clearly of degree 0: $deg_M(g) = deg_F(g) = 0$.

The dictatorship of x_7 is very simple to show over the given basis: $h(x) = x_7 = P_{\{x_7\}}$. The degree for the Fourier basis must be the same, and therefore: $deg_M(h) = deg_F(h) = 1$. It is still not difficult to find the representation $h(x) = \frac{1}{2} \cdot (\chi_\emptyset(x) - \chi_{\{x_7\}}(x))$.

5. Bonus Question - Linear Programming - Question 6

General notation. The notation in the question is a little confusing, so I will try to be more clear here.

$V = (v^1, v^2, \cdots, v^n)$ will be a set of vectors in $\mathbb{R}^d$ for some d. (I need the subscript later) Since d can always be raised, adding zeros to all the vectors, we can view these vectors as being in $\mathbb{R}^{\mathbb{N}}$ with the constraint that they have a finite number of nonzero elements.

Matrix $A \in \mathbb{R}^{n \times n}$ will be a distance matrix created by defining $A_{ij} = ||v^i - v^j||_1$. Since all the vectors have finite nonzero elements, the difference norm is also finite.

$\mathcal{L}_n$ will be the collection of all matrices which can be created this way.

First we define an offset operator on vectors. Define $o : (\mathbb{R}^{\mathbb{N}} \times \mathbb{N}) \to \mathbb{R}^{\mathbb{N}}$. If $v = (v_0, v_1, \cdots, v_{k-1}, 0, \cdots)$, then $o(v,d)_i = \begin{cases} 0 & i < d \\ v_{i-d} & i \geq d \end{cases}$

I will call the d-offset of vector v, $o(v,d) = {}^{d}v$.

Closure to addition. Assume $A, B \in \mathcal{L}_n$, where A is created by a set of vectors $V = (v^1, v^2, \cdots, v^n)$ all bounded in $\mathbb{R}^{d_v}$, and B from a set $W = (w^1, w^2, \cdots, w^n)$ all bounded in $\mathbb{R}^{d_w}$.

I claim the $A + B\mathcal{L}$ based on the set of vectors $\{v^i + {}^{d_v}w^i\}_{i=1}^n$. This set is of course bounded by $d_v + d_w$, the offset we pushed the second set, plus its own dimension.

Proof: We will call the matrix created by the above set C.

$$\begin{aligned}
C_{ij} &= ||(v^i + {}^{d^v}w^i) - (v^j + {}^{d_v}w^j)||_1 && \text{Definition} \\
&= \textstyle\sum_{k=0}^{d_v+d_w-1} |((v^i + {}^{d_v}w^i) - (v^j + {}^{d_v}w^j))_k| && \text{Norm-1 definition} \\
&= \textstyle\sum_{k=0}^{d_v-1} |((v^i + {}^{d_v}w^i) - (v^j + {}^{d_v}w^j))_k| + \sum_{k=d_v}^{d_v+d_w-1} |((v^i + {}^{d_v}w^i) - (v^j + {}^{d_v}w^j))_k| && \text{Splitting the sum} \\
&= \textstyle\sum_{k=0}^{d_v-1} |(v^i - v^j)_k| + \sum_{k=d_v}^{d_v+d_w-1} |((v^i + {}^{d_v}w^i) - (v^j + {}^{d_v}w^j))_k| && \text{Offset} \\
&= \textstyle\sum_{k=0}^{d_v-1} |(v^i - v^j)_k| + \sum_{k=d_v}^{d_v+d_w-1} |({}^{d_v}w^i - {}^{d_v}w^j)_k| && v^i \in \mathbb{R}^{d_v} \\
&= A_{ij} + B_{ij} && \text{Creation of } A, B
\end{aligned}$$

Therefore $C = A + B$, $C \in \mathcal{L}$ and therefore also $A + B \in \mathcal{L}$, and $\mathcal{L}$ is closed to addition, as required.

The set Z. Z is defined to be all $(0,1)$-matrices in $\mathcal{L}$. Let $V = (v^1, v^2, \cdots, v^n)$ be a set of vectors which create a matrix $A \in Z \subseteq \mathcal{L}_n$.

A is defined by norms, so $A_{ij} = 0 \iff ||v^i - v^j||_1 = 0 \iff v^i = v^j$. Since equality is an equivalence relation, an $(0,1)$-matrix B is in Z iff there is a partition of V into sets of equal vectors. That means there is a partition $J : [n] \to [n]$ of where: $B_{ij} = \begin{cases} 0 & J(i) = J(j) \\ 1 & J(i) \neq J(j) \end{cases}$

One direction of containment (no partition then not exists) was already shown. For the other direction we first mention the vectors of the standard basis, e^i, where $e^i_j = ind_i(j)$, where ind_i is the indicator function for i.

Finally we can show how to build vector set, $V = (v^1, v^2, \cdots, v^n)$, for every matrix with a partition:

$v^i = \frac{1}{2} \cdot e_{J(i)}$. Obviously, any pair of vectors in the same part of the partition will be the same, otherwise, assume i, j, are not in the same partition, then $||v^i - v^j||_1 = ||\frac{1}{2} \cdot e_{J(i)} - \frac{1}{2} \cdot e_{J(j)}||_1 = \frac{1}{2} + \frac{1}{2} = 1$, therefore the existence of a partition shows existence in $\mathcal{L}_n$.

Proving $cone(Z) \subseteq \mathcal{L}_n$. This direction is the easy direction. $\mathcal{L}_n$ is closed to multiplication by a positive scalar. If a set of vectors $V = (v^1, v^2, \cdots, v^n)$ prove $A \in \mathcal{L}_n$, then for any $\lambda \geq 0$, $\lambda \cdot V = (\lambda \cdot v^1, \lambda \cdot v^2, \cdots, \lambda \cdot v^n)$ proves that $\lambda \cdot A \in \mathcal{L}_n$, due to the positive scalar linearity of norms. Above I have proven that $\mathcal{L}_n$ is closed with regards to addition. Therefore, any positive linear combination of matrices in Z are in $\mathcal{L}_n$, proving $cone(Z) \subseteq \mathcal{L}_n$

Proving $cone(Z) \supseteq \mathcal{L}_n$. Due to the independence of the dimensions of the vectors, we only need to prove this for scalars.

Let $A\mathcal{L}$ be a matrix proven by $V = (v^1, v^2, \cdots, v^n)$ of dimension d.

According to the definitions, $A_{ij} = ||v^i - v^j||_1 = \sum_{k=0}^{d-1} |v^i_k - v^j_k|$.

Let us look at the kth element of the vectors of V as vectors in $\mathbb{R}^1$, giving us a set $V_k = (v^1_k, v^2_k, \cdots, v^n_k)$. We call the induced matrices A^k, and immediately get:

$$\begin{aligned} \sum_{k=0}^{d} A_{ij}^{k} &= \sum_{k=0}^{d} ||v_k^i - v_k^j||_1 && \text{Definition} \\ &= \sum_{k=0}^{d} |v_k^i - v_k^j| && \text{Scalars} \\ &= A_{ij} && \text{Definition} \end{aligned}$$

Proving $A = \sum_{k=0}^{d} A^k$.

Therefore, to complete the solution, we only need to show it is correct for matrices induced by scalar vectors. Since membership in both Z, and $\mathcal{L}$ is indifferent to permutations (does not change $(0-1)$ status, and you can change the vector order), we can reorder the proving vector set however we want. Since $\mathbb{R}^1$ has a natural ordering, we will order the set $V = \{\lambda^i\}_{i=1}^n$ from smallest to largest: $\lambda^1 \le \lambda^2 \le \cdots \le \lambda^n$.

We define the binary partitions (as used above) $J_i(k) = \begin{cases} 0 & i < k \\ 1 & i \ge k \end{cases}$.

We define matching matrices in Z, $z^1, z^2, \cdots, z^{n-1}$, where z^i is the matching matrix for partition J_i.

Let A be the induced matrix from $V = \{\lambda^i\}_{i=1}^n$ as defined above.

Then I intend to prove that $A = \sum_{i=1}^{n-1} (\lambda^{i+1} - \lambda^i) \cdot z^i$.

$$\begin{aligned} A_{ij} &= |\lambda^j - \lambda^i| && \text{Definition in } \mathbb{R}^1 \\ &= \sum_{k=min(i,j)}^{max(i,j)-1} \lambda^{k+1} - \lambda^k && \text{Values ordered} \\ &= \sum_{k=min(i,j)}^{max(i,j)-1} (\lambda^{k+1} - \lambda^k) \cdot 1 && \text{1 is netral} \\ &= \sum_{k=1}^{n-1} (\lambda^{k+1} - \lambda^k) \cdot \alpha_k(i,j) && \alpha \text{ defined below} \end{aligned}$$

Defining α as follows:

$$\begin{aligned} \alpha_k(i,j) &= \begin{cases} 1 & i < k \le j \\ 1 & j < k \le i \\ 0 & \text{else} \end{cases} && \text{The direct route} \\ &= \begin{cases} 1 & J_i(k) \ne J_j(k) \\ 0 & J_i(k) = J_j(k) \end{cases} && \text{From definition of } J_k \\ &= z_{ij}^k && z^k \text{ is induced from partition } J_k \end{aligned}$$

Therefore, we have reached that:

$$A_{ij} = \sum_{k=1}^{n-1} (\lambda^{k+1} - \lambda^k) \cdot z^k$$

To summarize the complete proof:

$$\begin{aligned} & A_{ij} = \sum_{k=1}^{n-1} (\lambda^{k+1} - \lambda^k) \cdot z^k && \text{Matrices induced from } \textbf{ordered} \text{ 1-dimensional vector sets can be produced.} \\ & Z \text{ indifferent to permutations} && \text{Matrices induced from } \textbf{all} \text{ 1-dimensional vector sets can be produced.} \\ & A = \sum_{k=0}^{d-1} A^k && \text{Matrices induced from all } \textbf{d}\text{-dimensional vector sets can be produced.} \\ & d \text{ is not bounded} && \text{proof complete} \end{aligned}$$

Finally, this means the positive linear combinations of Z, $cone(Z)$, covers $\mathcal{L}_n$, which means $cone(Z) \supseteq \mathcal{L}_n$.

Since in the previous section I have proven that also $cone(Z) \subseteq \mathcal{L}$, We have reached the required $cone(Z) = \mathcal{L}_n$.