Solutions to Assignment 4 of CPSC 368/516 (Spring'23)

February 21, 2023

1 Problem 1

Problem 1.1. *The goal of this problem is to bound bit complexities of certain quantities related to linear programs. Let $A \in \mathbb{Q}^{m\times n}$ be a matrix and $b \in \mathbb{Q}^m$ be a vector and let L be the bit complexity of (A, b). (Thus, in particular, $L > m$ and $L > n$.) We assume that $K = \{x \in \mathbb{R}^n : Ax \le b\}$ is a bounded, full-dimensional polytope in $\mathbb{R}^n$.*

1. *Prove that there is an integer $M \in \mathbb{Z}$ and a matrix $B \in \mathbb{Z}^{m\times n}$ such that $A = \frac{1}{M}B$ and the bit complexities of M and every entry in B are bounded by L.*
2. *Let C be any square and invertible submatrix of A. Consider the matrix norm $\|C\|_2 := \max_{x\neq 0} \frac{\|Cx\|_2}{\|x\|_2}$. Prove that there exists a constant d such that $\|C\|_2 \le 2^{O(L\cdot[\log(nL)]^d)}$ and $\|C^{-1}\|_2 \le 2^{O(nL\cdot[\log(nL)]^d)}$.*
3. *Prove that every vertex of K has coordinates in $\mathbb{Q}$ with bit complexity $O(nL \cdot [\log(nL)]^d)$ for some constant d.*

1.1 Facts

We will use the following facts in our proofs.

Fact 1.2. *For any integers $a, b \in \mathbb{Z}$, $L(ab) \le L(a) + L(b)$.*

Proof.

$$\begin{aligned} L(ab) &= 1 + \lceil \log(|ab| + 1)\rceil \\ &\le 1 + \lceil \log(|a| + 1) + \log(|b| + 1)\rceil \\ &\le 1 + \lceil \log(|a| + 1)\rceil + 1 + \lceil \log(|b| + 1)\rceil \\ &= L(a) + L(b). \end{aligned}$$

$\square$

Fact 1.3. *For any rational $r \in \mathbb{Q}$, $r \le 2^{L(r)}$*

Proof. To see this, suppose $r = \frac{p}{q}$ where p and q are coprime and $q \ge 1$, then $r = \frac{p}{q} \le |p| \le 2^{L(p)} \le 2^{L(r)}$. $\square$

1.2 Part 1

For each $i, j \in [n]$, let $A_{ij} = \frac{p_{ij}}{q_{ij}}$, where p_{ij}, q_{ij} are integers and $q_{ij} \neq 0$. Define M as the LCM of all denominators $\{q_{ij}\}_{ij}$ and B be the matrix whose (i, j)-th entry is $M \cdot \frac{p_{ij}}{q_{ij}}$. Clearly, we have that $A = \frac{1}{M}B$.

We can upper bound the bit complexity of M as follows

$$
\begin{aligned}
L(M) &\leq L\left(\prod_{i=1}^{n}\prod_{j=1}^{n} q_{ij}\right) && \text{(Using that } M \text{ is an integer and } M \leq \textstyle\prod_{i=1}^{n}\prod_{j=1}^{n} q_{ij}\text{)}\\
&\leq \sum_{i=1}^{n}\sum_{j=1}^{n} L\left(q_{ij}\right)\\
&\leq \sum_{i=1}^{n}\sum_{j=1}^{n} L\left(q_{ij}\right) + L\left(p_{ij}\right)\\
&\leq L\left(A\right)\\
&\leq L.
\end{aligned}
$$

Fix any $i, j \in [n]$. The bit complexity of B_{ij} can be upper bounded as follows

$$
\begin{aligned}
L(B_{ij}) &= L\left(M \cdot \frac{p_{ij}}{q_{ij}}\right)\\
&= L\left(p_{ij} \cdot \prod_{u\in[n]\colon u\neq i}\ \prod_{v\in[n]\colon v\neq j} q_{uv}\right)\\
&= L\left(p_{ij}\right) + \sum_{u\in[n]\colon u\neq i}\ \sum_{v\in[n]\colon v\neq j} L\left(q_{uv}\right)\\
&\leq L\left(p_{ij}\right) + \sum_{u\in[n]}\sum_{v\in[n]} L\left(q_{uv}\right)\\
&\leq \sum_{u\in[n]}\sum_{v\in[n]} L\left(p_{uv}\right) + L\left(q_{uv}\right)\\
&= L(A)\\
&\leq A.
\end{aligned}
$$

1.3 Part 2

Consider any $k \times k$ real matrix H. We can bound the matrix norm $\|H\|_2$ as a function of the bit complexity of the entries of H as follows

$$
\begin{aligned}
\|H\|_2 &= \max_{x\neq 0} \frac{\|Hx\|_2}{\|x\|_2}\\
&= \max_{\|x\|_2=1} \|Hx\|_2\\
&= \max_{\|x\|_2=1} \sqrt{\sum_{i=1}^{k}\left(\sum_{j=1}^{k} H_{ij}x_j\right)^2}\\
&\leq \max_{\|x\|_2=1} \sqrt{\sum_{i=1}^{k}\left(\sum_{j=1}^{k} H_{ij}^2\right)\cdot\left(\sum_{j=1}^{k} x_j^2\right)} && \text{(Using the Chauchy–Schwarz inequality)}\\
&\leq \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k} H_{ij}^2} && \text{(Using that } \|x\|_2 = 1\text{)}\\
&\leq \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k} 2^{2L(H_{ij})}}. && \text{(Using that for any rational } x,\ x \leq 2^{L(x)}\text{)} \quad (1)
\end{aligned}
$$

Suppose that C is a $k \times k$ invertible submatrix of A. Since each entry of C is also an entry in A, the bit complexity of any entry of C is bounded by $L(A) \le L$. Next, we bound the bit complexity of the entries of C^{-1}. Toward this, recall that

$$C^{-1} = \frac{\mathrm{adj}(C)}{|C|},$$

where $\mathrm{adj}(C)$ is the adjugate matrix of C. Fix any $i, j \in [n]$.

$$L\left(C_{ij}^{-1}\right) = L\left(\frac{1}{|C|}\left(\mathrm{adj}(C)\right)_{ij}\right) = O\left(L\left(|C|\right)\right) + O\left(L\left(\left(\mathrm{adj}(C)\right)_{ij}\right)\right). \tag{2}$$

Next, we bound both terms in the RHS separately

$$\begin{aligned}
|C| &= \sum_{\sigma \in S_k} \mathrm{sgn}(\sigma) \prod_{i=1}^{k} C_{i,\sigma(i)} \\
&\le \sum_{\sigma \in S_k} \left| \prod_{i=1}^{k} C_{i,\sigma(i)} \right| \\
&\le \sum_{\sigma \in S_k} \left| \prod_{i=1}^{k} 2^{L} \right| && \text{(Using that for all rationals } x\text{, } x \le 2^{L(x)}\text{)} \\
&= (k!) \cdot 2^{kL} \\
&= 2^{kL + O(k \log k)}. && (3)
\end{aligned}$$

Let the (u, v)-th entry of C be $\frac{p_{uv}}{q_{uv}}$. Define $Q \coloneqq \prod_{u,v} q_{uv}$. Note that both Q and $Q \cdot |C|$ are integers, and hence

$$\begin{aligned}
L\left(|C|\right) &\le L\left(Q \cdot |C|\right) + L(Q) \\
&= O(\log(Q \cdot |C|)) + L(Q) && \text{(Using that } Q \cdot |C| \text{ is an integer)} \\
&= O(\log(Q)) + O(\log |C|) + L(Q) \\
&= O(L(Q)) + O(\log |C|) && \text{(Using that } Q \text{ is an integer)} \\
&\overset{(3)}{\le} O(L(Q)) + O(kL + k \log k) \\
&\le O(L(C)) + O(kL + k \log k) \\
&\le O(L(A)) + O(kL + k \log k) \\
&\le O(kL + k \log k). && (4)
\end{aligned}$$

For all $i, j \in [k]$, let M_{ij} be the (i, j)-minor of C. Recall that M_{ij} is the determinant of a $(k-1) \times (k-1)$ submatrix of C. Using an analogous argument to Equation (4), we can bound $L(M_{ij}) \le O\left(kL + k \log k\right)$. Using the bound on $L(M_{ij})$, we get

$$\begin{aligned}
L\left(\left(\mathrm{adj}(C)\right)_{ij}\right) &= L\left((-1)^{i+j} M_{ij}\right) \\
&= O(1) + L\left(M_{ij}\right) \\
&= O\left(kL + k \log k\right). && \text{(Using that } L(M_{ij}) \le O\left(kL + k \log k\right)\text{)} \quad (5)
\end{aligned}$$

Substituting Equations (4) and (5) in Equation (2), we get

$$L\left(C_{ij}^{-1}\right) \le O\left(kL + k \log k\right). \tag{6}$$

Now, we are ready to bound $\|C\|_2$ and $\|C^{-1}\|_2$:

$$
\begin{aligned}
\|C\|_2 &\overset{(1)}{\leq} \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k} 2^{2L(C_{ij})}} \\
&\leq \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k} 2^{2L}} && \text{(Using that } L(C_{ij}) \leq L(A) = L\text{)} \\
&= \sqrt{2^{2L+\log(k^2)}} \\
&= 2^{L+\log k} \\
&= 2^{O(L)}, && \text{(Using that } L > n \geq k\text{)} \\
\left\|C^{-1}\right\|_2 &\overset{(1)}{\leq} \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k} 2^{2L(C_{ij}^{-1})}} \\
&\overset{(6)}{\leq} \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k} 2^{O(kL+k\log k)}} \\
&= \sqrt{2^{O(kL+k\log k+\log(k^2))}} \\
&= 2^{O(kL+k\log k)} \\
&\leq 2^{O(nL)}. && \text{(Using that } L > n \geq k\text{)}
\end{aligned}
$$

1.4 Part 3

We first derive a necessary condition for a point $x \in K$ to be a vertex.

Theorem 1.4. *A point $x \in K$ is a vertex if and only if it is a solution to $A'x = b'$, where A' is a square submatrix of A consisting n linearly independent rows of A and b' is a vector with the corresponding entries of b.*

Proof. Toward a contradiction suppose that there is a vertex $x \in K$ such that it is suitable A'. For all $i \in [m]$, let $a_i^\top$ be the i-th row of A. Let $S \subseteq [m]$ be the set of indices such that

$$
\begin{aligned}
&\text{for all } i \in S, \quad (Ax)_i = \langle a_i, x\rangle = b_i, \\
&\text{for all } i \notin S, \quad (Ax)_i = \langle a_i, x\rangle < b_i.
\end{aligned}
$$

Define $A^=$ to be the submatrix of A consisting of all rows of A whose indices are in S, and $A^<$ to be the submatrix of A consisting of the remaining rows of A. Similarly, let $b^=$ be the vector consisting of all rows b whose indices are in S, and $b^<$ be the vector consisting of the remaining rows of b. By definition of $A^=$ and $A^<$, x satisfies

$$
A^= x = b^= \quad \text{and} \quad A^< x < b^<. \tag{7}
$$

By our assumption either $A^=$ has less than n rows or $A^=$ has linearly dependent rows. In either case, the $\ker(A^=)$ is nonzero, and hence, there exists a nonzero $u \in \mathbb{R}^n$ in $\ker(A^=)$. For variable $\varepsilon \in \mathbb{R}$, consider the family of points $x + \varepsilon u$, we have

$$
\begin{aligned}
A^=(x+\varepsilon u) &= A^= x && \text{(Using that } u \in \ker(A^=)\text{)} \\
&\overset{(7)}{=} b^=. && (8)
\end{aligned}
$$

Further, because of the strict inequality $A^< x < b^<$, one can pick a small enough $\varepsilon > 0$ such that

$$
A^<(x+\varepsilon u) = A^< x + \varepsilon A^< u \leq b^<, \tag{9}
$$
$$
A^<(x-\varepsilon u) = A^< x - \varepsilon A^< u \leq b^<. \tag{10}
$$

Combining Equations (8) to (10) we get that there is a small enough $\varepsilon > 0$ such that

$$A(x+\varepsilon u) = \begin{bmatrix} A^{=} \\ A^{<} \end{bmatrix} (x+\varepsilon u) \leq \begin{bmatrix} b^{=} \\ b^{<} \end{bmatrix} = b,$$
$$A(x-\varepsilon u) = \begin{bmatrix} A^{=} \\ A^{<} \end{bmatrix} (x-\varepsilon u) \leq \begin{bmatrix} b^{=} \\ b^{<} \end{bmatrix} = b.$$

We have found two points $x+\varepsilon u, x-\varepsilon u \in K$, such that, $x = \frac{1}{2}(x+\varepsilon u) + \frac{1}{2}(x-\varepsilon u)$. This is a contradiction to the fact that x is a vertex of K. $\square$

From Theorem 1.4 we know that any vertex of K is a solution to $A'x = b'$, where A' is a square submatrix of A consisting n linearly independent rows of A and b' is a vector with the corresponding entries of b. Since A' is a square matrix with linearly independent rows, it is invertible, and hence, $x = (A')^{-1} b'$. Using Equation (6) we can bound the bit complexity of each entry of $(A')^{-1}$, and hence, also of $\left((A')^{-1} b'\right)_i$:

$$\begin{aligned}
L(x_i) &= L(((A')^{-1} b')_i) \\
&= L\left(\sum_{j=1}^{n} (A')^{-1}_{ij} b'_j\right) \\
&= O\left(\log(n) + \max_{j\in[n]} L\left((A')^{-1}_{ij} b'_j\right)\right) \\
&= O\left(\log(n) + \max_{j\in[n]} L\left((A')^{-1}_{ij}\right) + L\left(b'_j\right)\right) \\
&= O\left(\log(n) + O\left(n\log n + nL\right) + L\right) \\
&\text{(Using Equation (6), the fact that } A' \text{ is an } n\times n \text{ matrix, and that for all } j\in[n],\ L(b_j)\leq L(b)\leq L) \\
&= O(nL). && \text{(Using that } L > n)
\end{aligned}$$

2 Problem 2

Problem 2.1. *Recall that an undirected graph $G = (V, E)$ is said to be bipartite if the vertex set V has two disjoint parts L, R and all edges go between L and R. Consider the case when $n := |L| = |R|$ and $m := |E|$. A perfect matching in such a graph is a set of n edges such that each vertex has exactly one edge incident to it. Let $\mathcal{M}$ denote the set of all perfect matchings in G. Let $1_M \in \{0,1\}^E$ denote the indicator vector of the perfect matching $M \in \mathcal{M}$. Consider the function*

$$f(x) := \ln \sum_{M\in\mathcal{M}} e^{\langle x, 1_M\rangle}.$$

1. *Prove that f is convex.*
2. *Consider the bipartite perfect matching polytope of G defined as*

   $$P := \operatorname{conv}\{1_M : M \in \mathcal{M}\}.$$

   Give a polynomial time separation oracle for this polytope.
3. *Prove that, if there is a polynomial time algorithm to evaluate f given the graph G as input, then one can count the number of perfect matchings in G in polynomial time.*

Since the problem of computing the number of perfect matchings in a bipartite graph is $\#\mathbf{P}$-hard, we have an instance of convex optimization that is $\#\mathbf{P}$-hard.

2.1 Part 1

From Problem 2(b) Assignment 1, we know that the Hessian of f is

$$\nabla^2 f(x) = \frac{\sum_{M\in\mathcal{M}} e^{\langle y,1_M\rangle} 1_M 1_M^\top}{\sum_{M\in\mathcal{M}} e^{\langle y,1_M\rangle}} - \frac{\left(\sum_{M\in\mathcal{M}} e^{\langle y,1_M\rangle} 1_M\right)\cdot\left(\sum_{M\in\mathcal{M}} e^{\langle y,1_M\rangle} 1_M^\top\right)}{\left(\sum_{M\in\mathcal{M}} e^{\langle y,1_M\rangle}\right)^2}.$$

We claim that $\nabla^2 f(x)$ is PSD for all $x \in \mathbb{R}^E$, and hence, f is convex. To see this, fix any $y \in \mathbb{R}^E$ and consider $y^\top \nabla^2 f(x) y$.

$$\begin{aligned}
y^\top \nabla^2 f(x) y &= \frac{\left(\sum_{M\in\mathcal{M}} e^{\langle x,1_M\rangle}\right)\left(\sum_{M\in\mathcal{M}} e^{\langle x,1_M\rangle}\langle y,1_M\rangle^2\right)}{\left(\sum_{M\in\mathcal{M}} e^{\langle x,1_M\rangle}\right)^2} - \frac{\left(\sum_{M\in\mathcal{M}} e^{\langle x,1_M\rangle}\langle y,1_M\rangle\right)\left(\sum_{M\in\mathcal{M}} e^{\langle x,1_M\rangle}\langle y,1_M\rangle\right)}{\left(\sum_{M\in\mathcal{M}} e^{\langle x,1_M\rangle}\right)^2} \\
&= \frac{\sum_{M,N\in\mathcal{M}} e^{\langle x,1_M+1_N\rangle}\langle y,1_M\rangle^2}{\left(\sum_{M\in\mathcal{M}} e^{\langle x,1_M\rangle}\right)^2} - \frac{\left(\sum_{M\in\mathcal{M}} e^{\langle x,1_M\rangle}\langle y,1_M\rangle\right)\left(\sum_{M\in\mathcal{M}} e^{\langle x,1_M\rangle}\langle y,1_M\rangle\right)}{\left(\sum_{M\in\mathcal{M}} e^{\langle x,1_M\rangle}\right)^2} \\
&= \frac{\sum_{M,N\in\mathcal{M}} e^{\langle x,1_M+1_N\rangle}\langle y,1_M\rangle^2}{\left(\sum_{M\in\mathcal{M}} e^{\langle x,1_M\rangle}\right)^2} - \frac{\sum_{M,N\in\mathcal{M}} e^{\langle x,1_M+1_N\rangle}\langle y,1_N\rangle\langle y,1_M\rangle}{\left(\sum_{M\in\mathcal{M}} e^{\langle x,1_M\rangle}\right)^2} \\
&= \frac{\sum_{M,N\in\mathcal{M}} e^{\langle x,1_M+1_N\rangle}\left(\langle y,1_M\rangle^2 - \langle y,1_N\rangle\langle y,1_M\rangle\right)}{\left(\sum_{M\in\mathcal{M}} e^{\langle x,1_M\rangle}\right)^2} \\
&= \frac{\sum_{M,N\in\mathcal{M}} e^{\langle x,1_M+1_N\rangle}\left(\langle y,1_M\rangle^2 - 2\langle y,1_N\rangle\langle y,1_M\rangle + \langle y,1_N\rangle^2\right)}{2\left(\sum_{M\in\mathcal{M}} e^{\langle x,1_M\rangle}\right)^2} \\
&= \frac{\sum_{M,N\in\mathcal{M}} e^{\langle x,1_M+1_N\rangle}\left(\langle y,1_M\rangle - \langle y,1_N\rangle\right)^2}{2\left(\sum_{M\in\mathcal{M}} e^{\langle x,1_M\rangle}\right)^2} \\
&\geq 0.
\end{aligned}$$

2.2 Part 2

A matrix $A \in \mathbb{R}^{n\times n}$ is said to be doubly stochastic if all its entries are nonnegative and each of its rows and columns sum to 1, i.e., if

$$\text{for all } i,j\in[n],\ A_{ij}\geq 0,\ \sum_{\ell=1}^n A_{i\ell} = 1,\ \text{and } \sum_{\ell=1}^n A_{\ell j} = 1.$$

Let $K \subseteq \mathbb{R}^{n\times n}$ be the set of all doubly stochastic matrices

$$K \coloneqq \left\{ A\in\mathbb{R}^{n\times n} \colon \text{for all } i,j\in[n],\ A_{ij}\geq 0,\ \sum_{\ell=1}^n A_{i\ell}=1,\ \text{and } \sum_{\ell=1}^n A_{\ell j}=1 \right\}$$

Given a vector $y \in \mathbb{R}^E$, let $A^{(y)} \in \mathbb{R}^{n\times n}$ be the following matrix

$$(A^{(y)})_{ij} \coloneqq \begin{cases} y_e & \{i,j\} \coloneqq e\in E, \\ 0 & \text{otherwise.} \end{cases}$$

Intuitively, if each edge $e \in E$ is assigned weight y_e, then $(A^{(y)})_{ij}$ denotes the weight of the edge between the i-th vertex in the one bipartition and the j-th vertex in the other bipartition.

Lemma 2.2. *For all $y \in \mathbb{R}^E$, $A^{(y)}$ is a permutation matrix if and only if y is the incidence vector of a perfect matching.*

Proof. In any perfect matching M, each vertex in one bipartition is matched to a unique vertex in the other bipartition and vice versa. Thus, for any perfect matching M, each column of $A^{(1_M)}$ has a single 1 entry and all other 0 entries, and each row of $A^{(1_M)}$ has a single 1 entry and all other 0 entries. This implies that $A^{(1_M)}$ is a permutation matrix. To see the other direction, let M be a set of edges, such that, for all $i, j \in [n]$, M contains the edge $\{i, j\}$ iff $(A^{(y)})_{ij} = 1$. Because $A^{(y)}$ is a permutation matrix, it follows that M is a perfect matching. Further, since the entries of $A^{(y)}$ are zero for all $\{i, j\} \notin E$, it follows that $M \subseteq E$, and hence, M is valid perfect matching for G. Further, by this construction, $y = 1_M$. Thus, y is the incidence vector of a perfect matching if and only if $A^{(y)}$ is a permutation matrix. $\square$

We claim that $y \in P$ if and only if $A^{(y)} \in K$. This claim, along with the definition of $A^{(y)}$ and K, implies the following characterization of P

$$P := \left\{ y \in \mathbb{R}^E \colon \text{for all } e \in E,\ y_e \geq 0,\ \text{and for all } i, j \in [n], \quad \sum_{\ell \colon \{i,\ell\} \in E} y_{\{i,\ell\}} = 1,\ \text{and} \quad \sum_{\ell \colon \{\ell,j\} \in E} y_{\{\ell,j\}} = 1 \right\}.$$

Since in the above form P is a polytope determined by $|E| + 2n = \operatorname{poly}(n)$ inequalities and the bit complexity of G is $\operatorname{poly}(n)$, there is a polynomial time separation oracle for P.

It remains to prove the following lemma

Lemma 2.3. *For all $y \in \mathbb{R}^E$, $y \in P$ if and only if $A^{(y)} \in K$.*

Proof.

If $y \in P$ then $A^{(y)} \in K$. Any vector $y \in P$ can be decomposed as a convex combination of the incidence vectors of perfect matchings. Suppose y decomposes as

$$y = \sum_{M \in \mathcal{M}} \alpha_M 1_M,$$

where $\sum_{M \in \mathcal{M}} \alpha_M = 1$ and for all $M \in \mathcal{M}$, $\alpha_M \geq 0$. Using this decomposition, we can decompose $A^{(y)}$ as a convex combination of $A^{(1_M)}$ as follows: For any $i, j \in [n]$

$$\begin{aligned}
(A^{(y)})_{ij} &:= \begin{cases} y_e & \{i,j\} := e \in E, \\ 0 & \text{otherwise}, \end{cases} \\
&= \sum_{M \in \mathcal{M}} \alpha_M \cdot \begin{cases} (1_M)_e & \{i,j\} := e \in E, \\ 0 & \text{otherwise}, \end{cases} \\
&= \sum_{M \in \mathcal{M}} \alpha_M \left(A^{(1_M)}\right)_{ij}.
\end{aligned}$$

Thus,

$$A^{(y)} = \sum_{M \in \mathcal{M}} \alpha_M A^{(1_M)}.$$

From Lemma 2.2, we know that for all $M \in \mathcal{M}$, $A^{(1_M)} \in K$. This shows that $A^{(y)}$ is a convex combination of elements in K, and hence, by the convexity of K, $A^{(1_M)} \in K$.

If $A^{(y)} \in K$ then $y \in P$. From the Birkhoff–von Neumann theorem, we know that any doubly stochastic matrix can be decomposed as a convex combination of at most n^2 permutation matrices. In particular, since $A^{(y)} \in K$,

$$A^{(y)} = \sum_{\ell=1}^{n^2} \alpha_\ell P_\ell, \tag{11}$$

where $\sum_{\ell=1}^{n^2} \alpha_\ell = 1$ and for all $\ell \in [n^2]$, $\alpha_\ell \geq 0$. Without loss of generality, we assume that for all $\ell \in [n^2]$, $\alpha_\ell > 0$. We claim that for each $\ell \in [n^2]$, there is a perfect matching M such that $A^{(1_M)} = P_\ell$. Fix any $\ell \in [n^2]$. We can prove this as follows:

- First, observe that for all $\{i,j\} \notin E$, $(P_\ell)_{ij} = 0$. Otherwise, we have a contradiction because for all $\{i,j\} \notin E$ $(A^y)_{ij} = 0$ and $(A^y)_{ij} \geq \alpha_\ell (P_\ell)_{ij} > 0$.
- Define M as the set of edges, such that, for all $i,j \in [n]$, M contains the edge $\{i,j\}$ iff $(P_\ell)_{ij} = 1$. Because P_ℓ is a permutation matrix, it follows that M is a perfect matching. Further, since the (i,j)-th entry of P_ℓ is zero for all $\{i,j\} \notin E$, it follows that $M \subseteq E$, and hence, M is valid perfect matching for G.

Since the choice of $\ell \in [n^2]$ was arbitrary, for each P_ℓ we have a perfect matching $M(\ell)$ such that $A^{(1_{M(\ell)})} = P_\ell$. Combining this with Equation (11), we get

$$A^{(y)} = \sum_{\ell=1}^{n^2} \alpha_\ell A^{(1_{M(\ell)})}. \tag{12}$$

Using the above, for any $e = \{i,j\} \in E$,

$$\begin{aligned} y_e &= \left(A^{(y)}\right)_{ij} \\ &= \sum_{\ell=1}^{n^2} \alpha_\ell \left(A^{(1_{M(\ell)})}\right)_{ij} \\ &= \sum_{\ell=1}^{n^2} \alpha_\ell \left(1_{M(\ell)}\right)_{ij}. \end{aligned}$$

This shows that, $y = \sum_{\ell=1}^{n^2} \alpha_\ell 1_{M(\ell)}$, and hence, $y \in P$. $\square$

2.3 Part 3

Observe that $f(0) = \ln(|\mathcal{M}|)$, and hence, $|\mathcal{M}| = e^{f(0)}$. Thus, if we can query the evaluation oracle at 0, read the oracle's output, i.e., $f(0)$, and compute $e^{f(0)}$, in polynomial time, and then compute $|\mathcal{M}|$. However, this might not be possible because $f(0)$ can have a large bit complexity. (In fact, if $|\mathcal{M}| \neq 0$ then $f(0)$ is irrational, and hence, cannot be represented using any finite number of bits). Instead, we show that it suffices to use the approximation $\widehat{f}$ of $f(0)$ with a small bit complexity such that

$$\left| e^{\widehat{f}} - e^{f(0)} \right| \leq \frac{1}{8},$$

and compute an approximation $\widehat{E}$ of $e^{\widehat{f}}$ with a small bit complexity, such that

$$\left| \widehat{E} - e^{\widehat{f}} \right| \leq \frac{1}{8}.$$

Combining these bounds with the triangle inequality implies that

$$\left| \widehat{E} - e^{f(0)} \right| \leq \left| \widehat{E} - e^{\widehat{f}} \right| + \left| e^{\widehat{f}} - e^{f(0)} \right| \leq \frac{1}{4}.$$

Since $e^{f(0)}$ is guaranteed to be an integer and $\left| \widehat{E} - e^{f(0)} \right| \leq \frac{1}{4}$, one can recover $e^{f(0)}$ by rounding the value $\widehat{E}$ to the closest integer. It remains to prove that one can find suitable approximations $\widehat{f}$ and $\widehat{E}$ in polynomial time.

Computing $\widehat{f}$. $\widehat{f}$ can be obtained by reading the first $O(m) = \text{poly}(n)$ bits of $f(0)$ output by the evaluation oracle. This guarantees that

$$\left| \widehat{f} - f(0) \right| \leq 2^{-\Theta(m)}. \tag{13}$$

Let $a := \min(f(0), \widehat{f})$ and $b := \max(f(0), \widehat{f})$. Then, by using the first order Taylor approximation of e^x at $\widehat{f}$, we have that

$$e^{f(0)} = e^{\widehat{f}} + \left(\widehat{f} - f(0)\right) \cdot \max_{a \le z \le b} \frac{de^x}{dx} \Big|_{x=z} .$$

In other words,

$$\begin{aligned}
\left|e^{f(0)} - e^{\widehat{f}}\right| &= \left|\widehat{f} - f(0)\right| \cdot \max_{a \le z \le b} \frac{de^x}{dx} \Big|_{x=z} \\
&= \left|\widehat{f} - f(0)\right| \cdot \frac{de^x}{dx} \Big|_{x=b} \\
&= \left|\widehat{f} - f(0)\right| \cdot e^b \\
&\le \left|\widehat{f} - f(0)\right| \cdot e^{\ln|\mathcal{M}|+1} && \text{(Using that } \widehat{f} - f(0) \le 2^{-\Theta(m)} \le 1.\text{)} \\
&\le \left|\widehat{f} - f(0)\right| \cdot e^{m+1} && \text{(Using that } |\mathcal{M}| \le \tbinom{m}{n} \le 2^m\text{)} \\
&\le 2^{-\Theta(m)} \cdot e^{m+1} && \text{(Using Equation (13))} \\
&\le \frac{1}{8}. && (14)
\end{aligned}$$

Computing $\widehat{E}$. $\widehat{E}$ can be obtained by computing the first $O(m^2) = \mathrm{poly}(n)$ terms in the Taylor expansion of $e^{\widehat{f}}$ at 0. Consider the k-th order Taylor expansion of $e^{\widehat{f}}$ at 0

$$e^{\widehat{f}} = 1 + \widehat{f} + \frac{\widehat{f}^2}{2!} + \cdots + \frac{\widehat{f}^k}{k!} + \left(\frac{de^x}{dx} \Big|_{x=z}\right) \cdot \frac{\widehat{f}^{k+1}}{(k+1)!},$$

where $0 \le z \le \widehat{f}$ is some number. Suppose $\widehat{E}$ is obtained by computing the first k terms in the Taylor approximation of $e^{\widehat{f}}$. Then, we have that

$$\begin{aligned}
\left|e^{\widehat{f}} - \widehat{E}\right| &= \left(\frac{de^x}{dx} \Big|_{x=z}\right) \cdot \frac{\widehat{f}^{k+1}}{(k+1)!} \\
&\le e^{\widehat{f}} \cdot \frac{\widehat{f}^{k+1}}{(k+1)!} && \text{(Using that } 0 \le z \le \widehat{f}\text{)} \\
&\le (|M|+1) \cdot \frac{\widehat{f}^{k+1}}{(k+1)!} && \text{(Using Equation (14) and the fact that } e^{f(0)} = \ln|\mathcal{M}|\text{)} \\
&\le (|M|+1) \cdot \frac{(\ln|\mathcal{M}|+1)^{k+1}}{(k+1)!} && \text{(Using Equation (14) and the fact that } e^{f(0)} = \ln|\mathcal{M}|\text{)} \\
&\le (2^m+1) \cdot \frac{(m+1)^{k+1}}{(k+1)!} && \text{(Using that } |\mathcal{M}| \le \tbinom{m}{n} \le 2^m\text{)} \\
&\le \left(2^{m+1}\right) \cdot \frac{(m+1)^{k+1}}{(k+1)!}.
\end{aligned}$$

Setting $k = O(m^2)$, we get

$$\begin{aligned}
\left|e^{\widehat{f}} - \widehat{E}\right| &\le 2^{m+1} \cdot \frac{(m+1)^{O(m^2)+1}}{O(m^2)!} \\
&\le 2^{m+1} \cdot \frac{(m+1)^{2(m+1)}}{\prod_{\ell=1}^{2m+2} \ell} \cdot \frac{(m+1)^{O(m^2)-2m-1}}{\prod_{\ell=2m+2}^{O(m^2)} \ell} \\
&\le 2^{m+1} \cdot (m+1)^{2(m+1)} \cdot \left(\frac{1}{2}\right)^{O(m^2)-2m-2}
\end{aligned}$$

$$= 2^{3m+3+2(m+1)\log(m+1)-O(m^2)}$$
$$\leq 2^{(5m+4)\log(m+2)-O(m^2)} \qquad \text{(Using that for all } m \geq 0,\ \log(m+2) \geq 1\text{)}$$
$$< \frac{1}{8} \qquad \text{(Using the fact that for all } m \geq 0,\ 2^{(5m+4)\log(m+1)-(m+2)^2} < \tfrac{1}{8}\text{)}$$

3 Problem 3

Problem 3.1. *Let $\mathcal{S}$ be a nonempty family of subsets of $\{1, 2, \ldots, n\}$. For a set $S \in \mathcal{S}$, let $1_S \in \mathbb{R}^n$ be the indicator vector of S, i.e., $1_S(i) = 1$ if $i \in S$ and $1_S(i) = 0$ otherwise. Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ given by*

$$f(x) \coloneqq \ln \sum_{S \in \mathcal{S}} e^{\langle x, 1_S \rangle}.$$

Prove that the gradient of f is L-Lipschitz continuous for some $L > 0$ that depends polynomially on n with respect to the Euclidean norm.

We claim that for all $x \in \mathbb{R}^n$ the maximum eigenvalue of the Hessian $\nabla^2 f(x)$ is at most $2n$, and hence, the maximum eigenvalue of $\left(\nabla^2 f(x)\right)^2$ is at most $4n^2$. This implies that f is $2n$-Lipschitz continuous by the following argument: Consider any $x, y \in \mathbb{R}^n$, $t \in [0, 1]$, and let $z_t \coloneqq x + t(y - x)$. Then

$$\|\nabla f(y) - \nabla f(x)\|_2^2 = \left\| \int_0^1 \nabla^2 f(z_t)(y - x) dt \right\|_2^2 \qquad \text{(Using Lemma 2.6 from the textbook)}$$
$$= \int_0^1 \left\|\nabla^2 f(z_t)(y - x)\right\|_2^2 dt \qquad \text{(Pythagorous theorem for the } \ell_2\text{-norm)}$$
$$= \int_0^1 (y - x)^\top \nabla^2 f(z_t)^\top \nabla^2 f(z_t)(y - x) dt$$
$$= \int_0^1 (y - x)^\top \left(\nabla^2 f(z_t)\right)^2 (y - x) dt \qquad \text{(Using that } \nabla^2 f(x) \text{ is symmetric for all } x \in \mathbb{R}^n\text{)}$$
$$\leq \int_0^1 4n^2 \|(y - x)\|_2^2 dt \qquad \text{(Using Claim 3.2)}$$
$$= 4n^2 \|(y - x)\|_2^2 .$$

It remains to prove the following claim.

Claim 3.2. *For all $x \in \mathbb{R}^n$, the maximum eigenvalue of $\nabla^2 f(x)$ is at most $2n$.*

Proof. From part (d) of Problem 2 in Assignment 1, we know that

$$\nabla^2 f(x) = \frac{\sum_{M \in \mathcal{M}} e^{\langle x, 1_M \rangle} 1_M 1_M^\top}{\sum_{M \in \mathcal{M}} e^{\langle x, 1_M \rangle}} - \frac{\left(\sum_{M \in \mathcal{M}} e^{\langle x, 1_M \rangle} 1_M\right) \cdot \left(\sum_{M \in \mathcal{M}} e^{\langle x, 1_M \rangle} 1_M^\top\right)}{\left(\sum_{M \in \mathcal{M}} e^{\langle x, 1_M \rangle}\right)^2}.$$

To simplify the notation for each $M \in \mathcal{M}$, define

$$\alpha_M \coloneqq \frac{e^{\langle x, 1_M \rangle}}{\sum_{M \in \mathcal{M}} e^{\langle x, 1_M \rangle}}.$$

Note that for all $M \in \mathcal{M}$, $\alpha_M \geq 0$ and $\sum_{M \in \mathcal{M}} \alpha_M = 1$. Consider any vector $z \in \mathbb{R}^n$. It suffices to bound

$z^\top \nabla^2 f(x) z$ by $2n \|z\|_2^2$. A proof of this upper bound is as follows

$$
\begin{aligned}
z^\top \nabla^2 f(x) z &= \sum_{M \in \mathcal{M}} \alpha_M \langle z, 1_M \rangle^2 - \left(\sum_{M \in \mathcal{M}} \alpha_M \langle z, 1_M \rangle \right)^2 \\
&\leq \sum_{M \in \mathcal{M}} \alpha_M \langle z, 1_M \rangle^2 + \left(\sum_{M \in \mathcal{M}} \alpha_M \langle z, 1_M \rangle \right)^2 \\
&\leq \sum_{M \in \mathcal{M}} \alpha_M \|z\|_2^2 \|1_M\|_2^2 + \left(\sum_{M \in \mathcal{M}} \alpha_M \|z\|_2 \|1_M\|_2 \right)^2 && \text{(Using the Cauchy-Shwartz inequality)} \\
&\leq \sum_{M \in \mathcal{M}} \alpha_M \|z\|_2^2 \cdot n + \left(\sum_{M \in \mathcal{M}} \alpha_M \|z\|_2 \cdot \sqrt{n} \right)^2 && \text{(Using that 1_M is a 0/1 vector of length n)} \\
&= n \|z\|_2^2 \left(\sum_{M \in \mathcal{M}} \alpha_M + \left(\sum_{M \in \mathcal{M}} \alpha_M \right)^2 \right) && \text{(Using that $\textstyle\sum_{M \in \mathcal{M}} \alpha_M = 1$)} \\
&= 2n \|z\|_2^2 .
\end{aligned}
$$

□