ACO Comprehensive Exam Spring 2025

Jan 5, 2025

1 Algorithms

We are given three arrays $A[0...n-1]$, $B[0...m-1]$, $M[0...m-1]$ with $m \leq n$ and $M[i] \in [0,1]$ for all i. For vectors $v \in \mathbb{R}^m$ we define the *mask-norm*

$$\|v\|_M^2 := \sum_{k=0}^{m-1} M[k] \cdot v_k^2$$

i.e., the classical L2-norm but we weight entries by $M[k] \in [0,1]$. So array M functions as a "mask" and each entry $M[k]$ specifies how important the entry is on a scale from 0 to 1. For any index i, we interpret $A[i...i+m-1], B[0...m-1] \in \mathbb{R}^m$ as m-dimensional vectors. We want to find index i where $A[i...i+m-1]$ is "closest" to $B[0...m-1]$ with respect to the mask-norm. That is, find i that minimizes

$$\|A[i...i+m-1] - B[0...m-1]\|_M^2 = \sum_{k=0}^{m-1} M[k] \cdot (A[i+k] - B[k])^2.$$

(For example, array B may be a picture of some object, array M is a mask that separates the object from the background in picture B, and we try to find the object within picture A.)

1. (2 points) Show that for any two vectors $u, v \in \mathbb{R}^m$, we have

$$\|u - v\|_M^2 = \|u\|_M^2 - 2\left(\sum_{k=0}^{m-1} u_k \cdot v_k \cdot M[k]\right) + \|v\|_M^2.$$

2. (4 points) Design and analyze an algorithm that, given arrays $C[0...n-1]$, $D[0...m-1]$, computes in $O(n \log n)$ time the array $E[0...n-m]$ with

$$E[i] = \sum_{k=0}^{m-1} C[i+k] \cdot D[k]$$

(Hint: Use FFT in $O(n \log n)$ time for polynomial multiplication.)

3. (4 points) Design and analyze an algorithm that, given arrays $A[0...n-1]$, $B[0...m-1]$, $M[0...m-1]$, returns in $O(n \log n)$ time an index i where $\|A[i...i+m-1] - B[0...m-1]\|_M^2$ is as small as possible. (You may use 2. even if you have not solved it.)

2 Graph Theory

Question. Show that for every positive integer k there exists a positive integer c_k with the following property: If G is a c_k-vertex-connected graph for which there exists $S \subseteq V(G)$ with $|S| = k$ such that

(1) $G - S$ is a bipartite graph with partition sets V_1, V_2, and

(2) for every $T \subseteq S$ and for each $i \in [2]$, $|N_G(T) \cap V_i| \geq |T|$,

then G contains k odd cycles that are pairwise vertex disjoint. Give a sufficient bound on c_k as a function of k alone.

3 Linear Inequalities

A sequence of positive integers $\{a_j\}_{j\in[n]}$ is called super-increasing if $\sum_{j=1}^{k} a_j < a_{k+1}$ for all $k \in [n-1]$. For example, the sequence $a_j = 2^{j-1}$ for $j \in [n]$ is super-increasing. Consider a knapsack set with coefficients being super-increasing:

$$S := \left\{ x \in \{0,1\}^n \;\middle|\; \sum_{j=1}^{n} a_j x_j \leq b \right\},$$

where we assume $a_n \leq b$ and there exists $\mathcal{I} \subseteq [n]$ such that $\sum_{j\in\mathcal{I}} a_j = b$. For $j \in [n] \setminus \mathcal{I}$, define the set: $\mathcal{I}_j := \{j\} \cup \{i \in \mathcal{I} \,|\, i > j\}$.

(a.) (3 pt) Show that if $\hat{x} \in S$, then it satisfies the inequalities:

$$\sum_{i\in\mathcal{I}_j} \hat{x}_i \leq |\mathcal{I}_j| - 1, \tag{1}$$

for all $j \in [n] \setminus \mathcal{I}$.

(b.) (3 pt) Show that if $\hat{x} \in \{0,1\}^n \setminus S$, then there exists $j' \in [n] \setminus \mathcal{I}$ such that $\hat{x}$ does not satisfy (1) corresponding to j'.

(c.) (4 pt) Show that $P := \{x \in [0,1]^n \,|\, (1),\ \forall j \in [n] \setminus \mathcal{I}\}$ is an integral polytope, thus establishing that $\mathrm{conv}(S) = P$.

4 Solutions

4.1 Algorithms

Let $M^{1/2}u$ be the entry-wise product, i.e., $(M^{1/2}u)_i = M[i]^{1/2} \cdot u_i$. Then

$$\begin{aligned} \|u-v\|_M^2 = \|M^{1/2}u - M^{1/2}v\|^2 = \|M^{1/2}u\|^2 - 2\langle M^{1/2}u, M^{1/2}v\rangle + \|Mv\|^2 \\ = \|u\|_M^2 - 2\left(\sum_i M[i] \cdot u_i \cdot v_i\right) + \|v\|_M^2. \end{aligned}$$

Algorithm: The algorithm first construct two polynomials $f(x) = \sum_i C[i]x^i$, $g(x) = \sum_i D[m-i-1]x^i$, and computes the product $h(x) = f(x)g(x)$ via FFT. The array E are the coefficients of h, when ignoring the first m terms. That is, $E[i]$ is the coefficient of x^{i+m} in h.

Complexity Constructing the polynomials f, g is $O(n)$ time. The polynomial products via FFT take $O(n \log n)$ time. So overall complexity is $O(n \log n)$.

Correctness The polynomial $h(x)$ has as coefficient of x^i

$$\begin{aligned} \sum_{k=0}^{m-1} f_{i-k} g_k &= \sum_{k=0}^{m-1} C[i-k]D[m-k-1] \\ &= \sum_{k=0}^{m-1} C[i-m+1+k]D[k] \end{aligned}$$

In other words, the coefficient in $h(x)$ of $x^{i+(m-1)}$ is

$$\sum_k C[i+k]D[k]$$

So by ignoring the first m coefficients of h, and only keeping the coefficients of $x^{i+(m-1)}$ for $i = 0...n-m$, we get the desired sums for $E[i]$.

Algorithm We compute $\|B\|_M^2 = \sum_{i=0}^{m-1} B[i]^2 * M[i]$. Then we construct arrays A', B' with

$$A'[i] = A[i]^2$$

$$B'[i] = B[i] \cdot M[i]$$

for $i = 0...m-1$. Then we compute arrays C, D with

$$C[i] = \sum_{k=0}^{m-1} A[i+k] \cdot B'[k]$$

$$D[i] = \sum_{k=0}^{m-1} A'[i+k] \cdot M[k]$$

via the algorithm of Q2. Finally, for $i = 0...n-m-1$ we set $E[i] = \|B\|_M^2 + D[i] - 2C[i]$, and return the index i with largest $E[i]$.

Complexity Computing $\|B\|_M^2$ takes $O(n)$ time. Constructing A', B' takes $O(n)$ time. Computing C,D takes $O(n \log n)$ by Q2, and the last for-loop also takes $O(n)$ time. Thus overall time complexity is $O(n \log n)$.

Correctness Correctness follows directly from Q1 and the values computed in Q2:

$$\begin{aligned}
&\|A[i...i+m-1] - B[0..m-1]\|_M^2 \\
\overset{Q1}{=}\ &\|A[i...i+m-1]\|_M^2 - 2\left(\sum_k A[i+k] \cdot B[k] \cdot M[k]\right) + \|B[0...m-1]\|_M^2 \\
=\ &\left(\sum_{k=0}^{m-1} A[i+k]^2 \cdot M[k]\right) - 2\left(\sum_k A[i+k] \cdot B[k] \cdot M[k]\right) + \|B[0...m-1]\|_M^2 \\
=\ &D[i] - 2 \cdot C[i] + \|B\|_M^2.
\end{aligned}$$

4.2 Graph Theory

Solution. Choose c_k to be sufficiently large, such that $(c_k - k)$-connected graphs are k-linked. Note such c_k exists by a theorem of Larmen and Mani and independently Jung. (Indeed, we may take $c_k = 11k$ by a theorem of Thomas and Wollan.)

For $i \in [2]$, let G_i be the subgraph of G induced by the edges of G from S to V_i. Then condition (2) allows us to apply Hall's theorem to conclude that G_i has a complete matching from S to V_i. For $i \in [2]$, let $\{s_j v_j^i : j \in [k]\}$ denote such a matching, where $S = \{s_1, \ldots, s_k\}$ and $\{v_1^i, \ldots, v_k^i\} \subseteq V_i$.

Since G is c_k-connected, $G - S$ is $(c_k - k)$-connected; hence, $G - S$ is k-linked. Therefore, $G - S$ has k pairwise disjoint paths P_j from v_j^1 to v_j^2, $j \in [k]$. Since $G - S$ is bipartite, P_j has odd length. Let C_j be the cycle obtained from P_j by adding the path $v_j^1 s_j v_j^2$. Clearly, $C_1, \ldots, C_k$ are pairwise disjoint odd cycles.

4.3 Linear Inequalities

(1): Assume by contradiction $\sum_{i \in \mathcal{I}_j} \hat{x}_i = |\mathcal{I}_j|$, that is $\hat{x}_i = 1$ for all $i \in \mathcal{I}_j$ for some $j \in [n] \backslash \mathcal{I}$. In that case

$$\sum_{i \in [n]} a_i \hat{x}_i \geq \sum_{i \in \mathcal{I}_j} a_i \hat{x}_i = a_j + \sum_{i \in \mathcal{I}: i > j} a_i > \sum_{i \in \mathcal{I}: i < j} a_i + \sum_{i \in \mathcal{I}: i > j} a_i = b, \tag{2}$$

where the strict inequality follows from the fact the sequence is super increasing. (2) contradicts $\hat{x} \in S$.

(2): Let $j' := \operatorname{argmax}\{i \in [n] \setminus \mathcal{I} \,|\, \hat{x}_i = 1\}$. Note that if $i' > j'$, $i' \in \mathcal{I}$, then $\hat{x}_{i'} = 1$. Otherwise, if $x_{i'} = 0$, then ...

$$\sum_{i \in [n]} a_i \hat{x}_i \leq \sum_{i \in [j']} a_i + \sum_{i \in \mathcal{I}: i > j', i \neq i'} a_i < \sum_{i \in \mathcal{I}: i > j'} a_i \leq b,$$

contradicting $\hat{x} \notin S$. We used the super increasing property for the second (strict) inequality Thus, $\hat{x}_i = 1$ for all $i \in \mathcal{I}_{j'}$, hence it does not satisfy (1) corresponding to j'.

(3): It is clear that $P \cap \{0,1\}^n = S$. We will show that left-hand-side of the constraint matrix of P is totally unimodular (TU). As the right-hand-side is an integral vector, this will show that P is integral.

In class, we have shown that if A is TU, then so is $\begin{bmatrix} A \\ I \\ -I \end{bmatrix}$ where I is the identity matrix. Thus, it is sufficient to show that the left-hand-side of the inequalities (1) is TU. Note now that by rearranging the columns, for example, by first writing the columns corresponding to variables in $\mathcal{I}$ and then followed by the columns corresponding to variables in $[n] \setminus \mathcal{I}$, the constraint matrix can be seen to be of the form $[C \ I]$ where C is an interval matrix and I is the identity matrix. Thus the left-hand-side of the inequalities (1) is TU, completing the proof.