SUBMATRIX PROBLEM

Consider the following problem : We have an $m \times n$ matrix $\mathbf{W}$ (whose set of rows, we denote by R and set of columns by C) with real entries (positive, negative or zero.) For any $S \subseteq R$ and $T \subseteq C$, we let

$$\mathbf{W}(S,T) = \sum_{i \in S, j \in T} \mathbf{W}_{ij}.$$

We wish to find the submatrix which maximizes the absolute value of the sum of the entries. I.e., we want to find $Y \subseteq R$ and $X \subseteq C$ such that $|\mathbf{W}(Y,X)|$ is maximized. It is easily seen that this problem is NP-hard. So, we would like to approximately maximize. We can divide the problem into two parts - first we find a Y, X such that $\mathbf{W}(Y,X)$ is as high as possible and then secondly, we find a Y, X such that $\mathbf{W}(Y,X)$ is as low as possible; clearly one of these two is the solution. The two problems are similar, so we only look at the first problem - of maximizing the sum of the submatrix entries. It is easy to see that it is enough to know Y; then we would take X to be the set of columns whose entries in the Y rows sum to greater than 0. I.e., letting

$$P(Y) = \{j \in C : \sum_{i \in Y} \mathbf{W}_{ij} > 0\} \forall Y \subseteq R,$$

we see that we may take $X = P(Y)$.

Here is an intuitive idea of the method for finding a good Y : we choose a subset U of rows of $\mathbf{W}$ as follows : each row is put in U with probability p/m, independently, where p will be very small ($O(1)$). So the expected size of U will be p. Then **if we knew** $Y \cap U$, since U was random, $Y \cap U$ is a random subset of Y, so for every column j, we may argue that $\frac{m}{p}\mathbf{W}(Y \cap U, j)$ gives us an estimate of $\mathbf{W}(Y, j)$. [This is just saying that to estimate the sum of $|Y|$ numbers, we pick a sample, where each number is put into the sample with probability p, then the sum of the sample is roughly p/n times the sum.] Thus, we may try $P(Y \cap U)$ instead of $P(Y)$ as the subset of columns. Now there is still as question of finding the subset of rows to use; but once we decide to use $P(U \cap Y)$ as the subset of columns, we may just use the rows in which the sum of entries in the $P(U \cap Y)$ columns is positive. Again defining

$$P(X) = \{i \in R : \sum_{j \in X} \mathbf{W}(i,j) > 0\} \forall X \subseteq C,$$

the submatrix we hope will approximately maximize the sum has $P(Y \cap U)$ as its column set and $P(P(Y \cap U))$ as its row set.

Finally, of course, we have to tackle the fact that we do not know $Y \cap U$; but since U is small, we try every subset of U as a possible candidate for $P \cap U$. Here the "Las Vegas" nature of the algorithm is useful - for every subset U' of U, we get a candidate submatrix on $P(P(U')) \times P(U')$. For each candidate submatrix, we may find the sum and then take the best one. We introduce a useful piece of notation : $||\mathbf{W}||_F$ will denote the square root of the sum of squares of all the entries of the matrix $\mathbf{W}$.

Lemma 1 *Suppose* $\mathbf{W}$ *is an* $m \times n$ *matrix with set of rows* R *and set of columns* C. *Let* $Y \subseteq R$ *be such that* $\mathbf{W}(P(Y), Y)$ *attains the maximum sum of any submatrix of* $\mathbf{W}$. *(*Y *is of course unknown.) Suppose* U *is a random subset of* R *chosen as above with expected cardinality* p. *Then*

$$\mathbf{E}_U(\mathbf{W}(Y, P(U \cap Y))) \geq \mathbf{W}(Y, P(Y)) - \sqrt{\frac{mn}{p}}||\mathbf{W}||_F. \tag{1}$$

Let $Z = P(Y)$ and $Z' = P(U \cap Y)$. We write

$$\mathbf{W}(Y, Z') = \mathbf{W}(Y, Z) - \mathbf{W}(Y, B_1) + \mathbf{W}(Y, B_2), \tag{2}$$

where

$$\begin{aligned} B_1 &= \{z \in C : \mathbf{W}(Y, z) > 0 \text{ and } \mathbf{W}(U \cap Y, z) < 0\}, \\ B_2 &= \{z \in C : \mathbf{W}(Y, z) < 0 \text{ and } \mathbf{W}(U \cap Y, z) > 0\}. \end{aligned}$$

Now let $X_z = \mathbf{W}(U \cap Y, z)$. Then note that

$$X_z = \sum_{u \in Y} \mathbf{W}(u, z) I(u),$$

where the 0-1 valued "indicator" random variable $I(u)$ is 1 if u is picked to be in U and 0 otherwise. Note that the $\{I(u)\}$ are independent random variables, and so

$$\mathbf{E}(X_z) = \frac{p}{m}\mathbf{W}(Y, z) \text{ and } \mathbf{Var}(X_z) = \sum \mathbf{Var}(I(u))\mathbf{W}(u, z)^2 \leq \frac{p}{m} W_2(z),$$

where $W_2(z) = \sum_{u \in Y} \mathbf{W}(u, z)^2$. Hence, for any $\xi > 0$,

$$\mathbf{Pr}\left(\left|X_z - \frac{p}{m}\mathbf{W}(Y, z)\right| \geq \xi\right) \leq \frac{pW_2(z)}{m\xi^2} \tag{3}$$

If $z \in B_1$ then $X_z - (p/m)\mathbf{W}(Y, z) \leq -(p/m)\mathbf{W}(Y, z)$ and so applying (3) with $\xi = p\mathbf{W}(Y, z)/m$ we get that for each fixed z,

$$\mathbf{Pr}(z \in B_1) \leq \frac{mW_2(z)}{p\mathbf{W}(Y, z)^2}.$$

Thus,

$$\begin{aligned} \mathbf{E}\left(\sum_{z \in B_1} \mathbf{W}(Y, z)\right) &\leq \sum_{\{z \in C:\, \mathbf{W}(Y, z) > 0\}} \min\left\{\mathbf{W}(Y, z), \frac{mW_2(z)}{p\mathbf{W}(Y, z)}\right\} \\ &\leq \sum_{\{z \in C:\, \mathbf{W}(Y, z) > 0\}} \sqrt{\frac{mW_2(z)}{p}} \end{aligned} \tag{4}$$

By an identical argument we obtain

$$\mathbf{E}\left(\sum_{z\in B_2}\mathbf{W}(Y,z)\right) \geq -\sum_{\{z\in C:\ \mathbf{W}(Y,z)<0\}}\sqrt{\frac{mW_2(z)}{p}}.$$

Hence, (using the Cauchy-Schwartz inequality),

$$\mathbf{E}(\mathbf{W}(Y,Z')) \geq \mathbf{W}(Y,Z) - \sum_{z\in C}\sqrt{\frac{mW_2(z)}{p}} \geq \mathbf{W}(Y,Z) - \sqrt{\frac{mn}{p}}||\mathbf{W}||_F.$$

¿From Lemma 1 and the Markov inequality applied to the *non-negative* random variable $\mathbf{W}(Y,P(Y)) - \mathbf{W}(Y,P(U\cap Y))$ we see that

$$\mathbf{Pr}\left(\mathbf{W}(Y,Z) - \mathbf{W}(Y,Z') \geq 2\sqrt{\frac{mn}{p}}||\mathbf{W}||_F\right) \quad \leq \quad 1/2. \tag{5}$$

Thus, with probability at least $1/2$, we have

$$\begin{aligned}
\mathbf{W}(Y,P(U\cap Y)) &\geq \mathbf{W}(Y,P(Y)) - 2\sqrt{\frac{mn}{p}}||\mathbf{W}||_F\\
\text{thus, } \mathbf{W}(P(P(U\cap Y)),P(U\cap Y)) &\geq \mathbf{W}(Y,P(Y)) - 2\sqrt{\frac{mn}{p}}||\mathbf{W}||_F\\
\text{thus, } \exists U'\subseteq U : \mathbf{W}(P(P(U')),P(U')) &\geq \mathbf{W}(Y,P(Y)) - 2\sqrt{\frac{mn}{p}}||\mathbf{W}||_F.
\end{aligned}$$

¿From this, it follows that with probability at least $1-\frac{1}{2^l}$, the following algorithm succeeds in finding a $S\subseteq R, T\subseteq C$ such that $\mathbf{W}(S,T)$ is at least the maximum value possible minus $2\sqrt{\frac{mn}{p}}||\mathbf{W}||_F$. If we are given an error parameter $\epsilon > 0$, by choosing $p = 4/\epsilon^2$, we can make the error at most $\epsilon\sqrt{mn}||\mathbf{W}||_F$ which is at most $\epsilon mn \max_{i,j}|\mathbf{W}_{i,j}|$. The algorithm then runs in time a polynomial in m,n times $2^{O(1/\epsilon^2)}$.

Now the algorithm is as follows :

(1) Pick at random a subset U of cardinality p (picked to make error smaller than a prespecified quantity).

(2) For each subset U' of U, find $\mathbf{W}(P(P(U')),P(U'))$.

Repeat the steps l times, then take the largest submatrix ever seen.