Ayush Sahni

Roll no 21111019

CS 698C End sem

~~Ans 1~~

~~(a) One possi~~

Ans 1

(a) One possible way to preserve the norm of

$$\min_{x} \|Ax - b\|^2 + \|Cx - d\|^2$$

can be to first convert the ~~summ~~ term in a closed form solution and then multiply with S accordingly to preserve norms. It can be done by using gradient descent to find an x such that $\|Ax-b\|^2 + \|Cx-d\|^2$ is minimum. Let it be x^*. Now we have to preserve this x^* in lower dimension which is possible only if we preserve column space of A and C. ~~~~

Also, we should preserve b and d.

~~So~~ By Gamma-net, no. of vectors in $\text{Colspace}(A) = O(5^d)$

$\text{Colspace}(C) = O(5^d)$

Since $b, d \in \mathbb{R}^n$ are only vectors so we can easily preserve colspace of $[A \ b]$ and $[C \ d]$

So, $k = O\left(\frac{1}{\varepsilon^2} \log \frac{5^{d+1} + 5^{d+1}}{\delta}\right) = O\left(\frac{d}{\varepsilon^2} \log \frac{1}{\delta}\right)$

(b)

$$\|Ax - b\|^2 + \|Cx - d\|^2$$

$$= (Ax-b)^T(Ax-b) + (Cx-d)^T(Cx-d)$$

$$= (x^TA^T - b^T)(Ax-b) + (x^TC^T - d^T)(Cx-d)$$

$$= x^TA^TAx - x^TA^Tb - b^TAx + b^Tb + x^TC^TCx - x^TC^Td - d^TCx + d^Td$$

- To find the x which minimises it, I can differentiate wrt x & equate the result to 0. The x obtained as result will be optimum.

$$\frac{d}{dx}\left[x^TA^TAx - x^TA^Tb - b^TAx + b^Tb + x^TC^TCx - x^TC^Td - d^TCx + d^Td\right] = 0$$

$$\frac{d}{dx}\left[x^TA^TAx\right] - A^Tb - b^TA + 0 + \frac{d}{dx}\left[x^TC^TCx\right] - C^Td - d^TC + 0 = 0$$

We know that

$$\frac{d}{dx}\left[x^TVx\right] = 2x^TV \quad \text{symmetric}$$

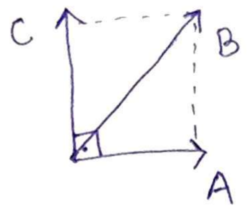
$$\Rightarrow 2x^TA^TA - A^Tb - b^TA + 2x^TC^TC - C^Td - d^TC = 0$$

$$\Rightarrow x^T(2A^TA + C^TC) = (A^Tb)^T + (C^Td)^T$$

(B)

Ans 2

(a) $\min\limits_{rank(X) \leq K} \|AX - B\|_F$



C is a hyperplane perpendicular to A. B can be written as sum of its projection on A and its projection on C.

AA^- is an orthogonal projection matrix on the column space of A. $\cancel{AA^- = U\Sigma V^T (U\Sigma V^T)}$

So, $I - AA^-$ is projection matrix on a space perpendicular to column space of A.

$$B = \underbrace{(AA^-)B} + \underbrace{(I - AA^-)B}$$ They are perpendicular

$$\min_{rank(X) \leq K} \|AX - B\|_F$$

$$= \min_{rank(X) \leq K} \| \underbrace{\overbrace{AX}^{\text{in colsp(A)}} - \overbrace{AA^-B}^{\text{in colsp(A)}}}_{\substack{\text{in column space} \\ \text{of A}}} - \underbrace{(I - AA^-)B}_{\substack{\text{in left nullspace} \\ \text{of } A \perp \text{colsp}(A)}} \|_F$$

$$= \min_{rank(X) \leq K} \|(AX - AA^-B) - (I - AA^-)B\|_F^2$$

$$= \min_{rank(X) \leq K} \left(\underbrace{\|AX - AA^-B\|_F^2}_{\substack{\text{we need to minimise} \\ \text{only this}}} + \|(I - AA^-)B\|_F^2 \right)$$ {because of orthogonality}

$$= \min_{\text{rank}(X) \le K} \| AX - AA^{-}B \|_f^2$$

If X^* is the optimal solution then

$$AX^* = AA^{-}B \Rightarrow \boxed{X^* = A^{-}[AA^{-}B]}$$

(b) Given, $S_{ij} \sim$ Normal distribution & S is $r \times m$ matrix

$AA^{-}[AA^{-}B]_K X$ lies in the column space of A. X is not known. So to preserve this, the whole column space should be preserved.

The no. of ~~vectors~~ column vectors in matrix A is n.

By Gamma-net, we know that total vectors in the column space of A are bound by 5^n.

Also, all the norms of vectors in B should be preserved

The no. of column vectors in B is p.

For single vector, the condition on r is

$$r = O\left(\frac{1}{\varepsilon^2} \log \frac{1}{\delta}\right)$$, where δ = error probability, ε = error

If we preserve colspace(A) and vectors in B, then,

$$r = O\left(\frac{1}{\varepsilon^2} \log \frac{1}{\delta/(5^n+p)}\right) = O\left(\frac{1}{\varepsilon^2} \log \frac{5^n+p}{\delta}\right)$$

$$\Rightarrow \boxed{r = O\left(\frac{n}{\varepsilon^2} \log \frac{1}{\delta} + \frac{1}{\varepsilon^2} \log \frac{p}{\delta}\right)}$$

(c) $\|SAX - SB\|_F \in (1 \pm \epsilon) \|AX - B\|_F \quad \forall X \in \mathbb{R}^{n \times p}$

$$\min_X \|SAX - SB\|_F \leq (1+\epsilon) \min_X \|AX - B\|_F = (1+\epsilon)\|A - B\|$$

$$\hat{X} = (SA)^{-}(SB)$$

By affine subspace embedding norm preservation,

$$(1-\epsilon)\|A\hat{X} - B\|_F \leq \|SA\hat{X} - SB\|_F \leq (1+\epsilon)\|A - B\|_F$$

$$\|A\hat{X} - B\|_F \leq (1+\epsilon) \min_{\text{rank}(X) \leq K} \|AX - B\|_F$$

Yes, $\text{rank}(\hat{X}) \leq K$.

(e) $\|YSAA^{-}BR - BR\|_F^2$

adding & subtracting $BR(SAA^{-}BR)^{-}(SAA^{-}BR)$

$$= \|\underbrace{YSAA^{-}BR - BR(SAA^{-}BR)^{-}(SAA^{-}BR)} + \underbrace{BR(SAA^{-}BR)^{-}(SAA^{-}BR) - BR}\|_F^2$$

Orthogonal

$$= \|YSAA^{-}BR - BR(SAA^{-}BR)^{-}(SAA^{-}BR)\|_F^2 + \|BR(SAA^{-}BR)^{-}(SAA^{-}BR) - BR\|_F^2$$

Ans (3)

(a) Let $(Sx)_i$ denote the rows of Sx, $i = 1, 2, \ldots, k$.

$$(Sx)_i = \sum_{j=1}^{n} b_{ij} \epsilon_{ij} x_j \quad , \quad i = 1, 2, \ldots, k.$$

$$(Sx)_i^2 = \left(\sum_{j=1}^{n} b_{ij} \epsilon_{ij} x_j \right)^2$$

$$= \sum_{j=1}^{n} b_{ij}^2 \epsilon_{ij}^2 x_j^2 + 2 \sum_{j<j'} b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}$$

$\epsilon_{ij}^2 = 1$ {because ϵ_{ij} takes value -1 or 1}

$b_{ij}^2 = b_{ij}$ {because b_{ij} takes value 0 or 1}

$$(Sx)_i^2 = \sum_{j=1}^{n} b_{ij} x_j^2 + 2 \sum_{1 \le j < j' \le n} b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}$$

(b) $$\|Sx\|^2 = \sum_{i=1}^{k} (Sx)_i^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} b_{ij} x_j^2 + 2 \sum_{i=1}^{k} \sum_{1 \le j < j' \le n} b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}$$

In first term, interchanging $\sum_i$ and $\sum_j$,

$$\sum_{j=1}^{n} \left(\sum_{i=1}^{k} b_{ij} \right) x_j^2 = \sum_{j=1}^{n} x_j^2 = \|x\|^2 \qquad \left\{ \sum_{i=1}^{k} b_{ij} = 1 \right\}$$

So, $$\|Sx\|^2 = \|x\|^2 + 2 \sum_{i=1}^{k} \sum_{1 \le j \le j' \le n} b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}$$

Since $\|x\|^2 = 1$ because x is unit vector,

$$\boxed{\|Sx\|^2 - 1 = \sum_{i=1}^{k} \sum_{1 \le j \le j' \le n} 2 b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}}$$

(c) Taking expectation of both side

$$E[\|Sx\|^2 - 1] = \sum_{i=1}^{k} \sum_{1\le j<j'\le n} 2\, E[b_{ij} b_{ij'}]\, E[\epsilon_{ij}\epsilon_{ij'}]\, x_j x_{j'}$$

Binomial RV b_{ij} are independent from Rademacher RV ϵ_{ij}

Now, ϵ_{ij} & $\epsilon_{ij'}$ are in same row & different columns.
(Same goes for b_{ij} & $b_{ij'}$)
ϵ_{ij} & $\epsilon_{ij'}$ are independent Rademache RVs.
So, $E[\epsilon_{ij}\epsilon_{ij'}] = E[\epsilon_{ij}]E[\epsilon_{ij'}] = 0.0 = 0$ $\quad \{\because E[\epsilon_{ij}] = 0\}$

So, $E[\|Sx\|^2 - 1] = 0$

Ans(4) (a) $i \neq i'$

We have to show $E[\epsilon_{ij}\epsilon_{ij'}\epsilon_{i'l}\epsilon_{i'l'}] = 0$

Rademacher R.Vs are ~~~~ used in the element of random matrix. They are all independent irrespective of the values of (j, j') and (l, l')

So, $E[\epsilon_{ij}\epsilon_{ij'}\epsilon_{i'l}\epsilon_{i'l'}] = E[\epsilon_{ij}]E[\epsilon_{ij'}]E[\epsilon_{i'l}]E[\epsilon_{i'l'}]$
$= 0$ $\quad \{\because E[\epsilon_{ij}] = 0\}$

(b) $E[T^{i}_{j,j'}\, T^{i'}_{l,l'}] = E[(2 b_{ij} b_{ij'} \epsilon_{ij}\epsilon_{ij'} x_j x_{j'})(2 b_{i'l} b_{i'l'} \epsilon_{i'l}\epsilon_{i'l'} x_l x_{l'})]$

$= 4\, E[b_{ij} b_{ij'} b_{i'l} b_{i'l'}]\, \underbrace{E[\epsilon_{ij}\epsilon_{ij'}\epsilon_{i'l}\epsilon_{i'l'}]}\, x_j x_{j'} x_l x_{l'}$

→ If $i \neq i'$, clearly the value of expectation will come out to be zero because of independence of rademacher R.Vs.
So, for $T^{i}_{j,j'} \cdot T^{i'}_{l,l'}$ to have non-zero expectation, $i = i'$.

(c) The elements of S_0 are defined as

$$S_{ij} = b_{ij} \epsilon_{ij} \qquad 1 \le i \le k \qquad 1 \le j \le n$$

In a column, there can be only one row element which can be non zero, all other elements of the column are zero.

$i = i'$ is a necessary but not sufficient condition for the product expectation to be non zero. because :-

$$E[\epsilon_{ij} \epsilon_{ij'} \epsilon_{i'l} \epsilon_{i'l'}] = E[\epsilon_{ij} \epsilon_{ij'} \epsilon_{il} \epsilon_{il'}] \qquad \{\text{since } i = i'\}$$

$$= E[\epsilon_{ij}] E[\epsilon_{ij'}] E[\epsilon_{il}] E[\epsilon_{il'}]$$

$$= 0$$

It will be nonzero only if pairs of edge are also matched ie., $(j, j') = (l, l')$ and $i = i'$ because

$$E[\epsilon_{ij} \epsilon_{ij'} \epsilon_{i'l} \epsilon_{i'l'}] = E[\epsilon_{ij} \epsilon_{ij'} \epsilon_{ij} \epsilon_{ij'}] = E[\epsilon_{ij}^2 \epsilon_{ij'}^2]$$

$$= E[\epsilon_{ij}^2] E[\epsilon_{ij'}^2] = 1 \cdot 1.$$

$\{$because $E[\epsilon_{ij}^2] = 1\}$

(d) We have shown that $i = i'$ for product expectation to be non zero. So $K^i = K^{i'}$, that means our complete graph is same. We also showed that $\{j, j'\} = \{l, l'\}$. So this means we are talking about same pair of vertices ~~These~~ So, we should consider pair of edges (j, j') from same graph $(K^i = K^{i'})$ twice to form cycle of length 2.

~~(e)~~

(e) ~~I have derived in part (b)~~ :- Since $i = i'$ & $\{j, j'\} = \{l, l'\}$

$$T^{i}_{j,j'} T^{i'}_{l,l'} = T^{i}_{j,j'} T^{i}_{j,j'} = (T^{i}_{j,j'})^2$$

$$= (2\, b_{ij}\, b_{ij'}\, \epsilon_{ij}\, \epsilon_{ij'}\, x_j\, x_{j'})^2$$

$$= 4\, b_{ij}^2\, b_{ij'}^2\, \epsilon_{ij}^2\, \epsilon_{ij'}^2\, x_j^2\, x_{j'}^2$$

~~Its expectation is~~

~~$4\,E[b_{ij}]^2$~~

b_{ij} takes values 0 or 1, so $b_{ij}^2 = b_{ij}$

ϵ_{ij} takes values -1 or 1, so $\epsilon_{ij}^2 = 1$

$$\boxed{T^{i}_{j,j'} T^{i}_{j,j'} = 4\, b_{ij}\, b_{ij'}\, 1 \cdot 1 \cdot x_j^2 x_{j'}^2}$$

(f) Taking expectation of part (e)

$$E[(T^{i}_{j,j'})^2] = 4\, E[b_{ij}\, b_{ij'}\, x_j^2\, x_{j'}^2] \quad -(1)$$

$$= 4\, E[b_{ij}\, b_{ij'}]\, x_j^2\, x_{j'}^2$$

It is given that the family $\{b_{ij}\}$, accross distinct j's is 4 wise independent (and hence 3-wise and 2-wise independent also). Hash functions maps $j \in \{1, 2, \ldots n\} \rightarrow i \in \{1, 2, \ldots k\}$

For each j, it can map to one of the k possible i's.

So, $P[h(j) = i] = \frac{1}{k}$ or $P[b_{ij} = 1] = \frac{1}{k}$

from equation (1) :-

$$E[(T_{j,j'}^{i})^2] = 4\, E[b_{ij}\, b_{ij'}]\, x_j^2\, x_{j'}^2 = 4\, E[b_{ij}]\, E[b_{ij'}]\, x_j^2\, x_{j'}^2$$

$$= 4\,(0 * P[b_{ij}=0] + 1 * P[b_{ij}=1])(0 * P[b_{ij'}=0]$$
$$+ 1 * P[b_{ij'}=1])\, x_j^2 x_{j'}^2$$

$$= 4\left(1 * \frac{1}{k}\right)\left(1 * \frac{1}{k}\right) x_j^2 x_{j'}^2$$

$$\boxed{= \frac{4}{k^2} x_j^2 x_{j'}^2}$$

(g) $$E[(\|Sx\|^2 - 1)^2] = E\left[\left(\sum_{i=1}^{k} \sum_{j<j'} 2 b_{ij} b_{ij'} \epsilon_{ij} \epsilon_{ij'} x_j x_{j'}\right)^2\right]$$

{from ques 3 (b)}

$$= E\left[\left(\sum_{i=1}^{k} \sum_{j<j'} T^i_{jj'}\right)^2\right]$$

$$= E\left[\left(\sum_{i=1}^{k} \sum_{j<j'} T^i_{jj'}\right)\left(\sum_{i=1}^{k} \sum_{j<j'} T^i_{jj'}\right)\right]$$

$$= E\Big[\underbrace{\sum_{i} \sum_{i'} \sum_{j<j'} \sum_{l<l'} T^i_{jj'} T^{i'}_{ll'}}\Big]$$

In this, there are all the product terms having $i = i'$, $i \neq i'$, $\{j, j'\} = \{l, l'\}$ & $\{j, j'\} \neq \{l, l'\}$. In earlier parts of question (part (b) and (c)) we have shown that expectation of product terms which have either $i \neq i'$ or $\{j, j'\} \neq \{l, l'\}$ is zero. By graph analogy, these are the product terms that do not correspond to a 2-cycle. So, I am removing terms in which $i \neq i'$ or $\{j, j'\} \neq \{l, l'\}$

$$= \sum_{i} \sum_{i'} \sum_{j<j'} \sum_{l<l'} E[T^i_{jj'} T^{i'}_{ll'}] = \sum_{i} \sum_{j<j'} E[(T^i_{jj'})^2]$$

~~As shown in~~

As shown in ques 4(f), $E[(T_{jj'}^{i})^2] = \frac{4}{K^2} x_j^2 x_{j'}^2$

So, $E[(\|Sx\|^2 - 1)^2]$

$$= \sum_i \sum_{j<j'} E[(T_{jj'}^{i})^2]$$

$$= \sum_{i=1}^{k} \sum_{j<j'} \frac{4}{K^2} x_j^2 x_{j'}^2$$

$$= \left(\sum_{i=1}^{k} \frac{4}{K^2}\right)\left(\sum_{j<j'} x_j^2 x_{j'}^2\right)$$

$$= \left(\frac{4}{K^2}\right) K \left(\sum_{j<j'} x_j^2 x_{j'}^2\right)$$

$$= \frac{4}{K} \sum_{j<j'} x_j^2 x_{j'}^2$$

$$= \frac{4}{K}\Big(x_1^2 x_2^2 + x_1^2 x_3^2 + \dots + x_1^2 x_n^2 + x_2^2 x_3^2 + \dots + x_2^2 x_n^2 + \dots + x_{n-1}^2 x_n^2 \Big)$$

$$\boxed{\le \frac{4}{K}\left[\left(\sum_j x_j^2\right)^2 - \sum_i x_i^4\right]}$$

This is because:- for example:

$$(x_1^2 + x_2^2 + x_3^2)^2 = x_1^4 + x_2^4 + x_3^4 + 2x_1^2 x_2^2 + 2x_1^2 x_3^2 + 2x_2^2 x_3^2$$

$$\Rightarrow \left(\sum_{j=1}^{3} x_j^2\right)^2 - \left(\sum_{i=1}^{3} x_i^4\right) = 2 \sum_{j<j'} x_j^2 x_{j'}^2$$

$$\Rightarrow \left(\sum_{j=1}^{3} x_j^2\right)^2 - \left(\sum_{i=1}^{3} x_i^4\right) > \sum_{j<j'} x_j^2 x_{j'}^2$$

$$\Rightarrow \frac{4}{K}\left[\left(\sum_{j=1}^{3} x_j^2\right)^2 - \left(\sum_{i=1}^{3} x_i^4\right)\right] > \frac{4}{K} \sum_{j<j'} x_j^2 x_{j'}^2$$ {because k is positive}

So,

$$E[(\|Sx\|^2-1)^2] \leq \frac{4}{K}\left[\left(\sum_j x_j^2\right)^2 - \left(\sum_i x_i^4\right)\right]$$

Now, $\left(\sum_j x_j^2\right)^2 - \left(\sum_i x_i^4\right) \geq \underbrace{\sum_{j<j'} x_j^2 x_{j'}^2}_{\text{positive quantity}}$

Since x is a unit vector, $\left(\sum_{j=1}^{n} x_j^2\right) = 1$.

$$\boxed{\begin{aligned} \Rightarrow E[(\|Sx\|^2-1)^2] &\leq \frac{4}{K}\Big[1 - \underbrace{\left(\sum_i x_i^4\right)}_{\text{lies between 0 to 1}}\Big] \\ &\leq \frac{4}{K} \qquad \left\{\text{because, } 0<\left(1-\sum_i x_i^4\right)<1\right\} \end{aligned}}$$

(h) By chebychev's inequality,

$$P[\,|\|Sx\|^2-1| \geq \epsilon\,] \leq \frac{E[(\|Sx\|^2-1)^2]}{\epsilon^2}$$

$$\leq \frac{4}{K\epsilon^2} \qquad \{\text{using result of part (g)}\}$$

— eqn 2

(i) if $k \geq \frac{4}{\epsilon^2 \delta}$

$\Rightarrow \delta \geq \frac{4}{K\epsilon^2}$

From eqn 2 :-

$$\boxed{P[\,|\|Sx\|^2-1| \geq \epsilon\,] \leq \frac{4}{K\epsilon^2} \leq \delta}$$