EE378B Homework 1 Solution

Due to: Chen Cheng

1 Problem 1

1.1 Part (a)

First we show

$$\max\{\|\boldsymbol{M}\boldsymbol{x}\|_2 : \|\boldsymbol{x}\|_2 = 1\} = \max\{\langle \boldsymbol{x}, \boldsymbol{M}\boldsymbol{y}\rangle : \|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2 = 1\}. \tag{1}$$

In fact the two sides are well defined since the space of $\|\boldsymbol{x}\| = 1, \|\boldsymbol{y}\| = 1$ is compact, the supremum can be attained by a maximal point. On that space by Cauchy-Schwarz

$$\langle \boldsymbol{x}, \boldsymbol{M}\boldsymbol{y}\rangle \le \|\boldsymbol{x}\|_2 \|\boldsymbol{M}\boldsymbol{y}\|_2 = \|\boldsymbol{M}\boldsymbol{y}\|_2, \tag{2}$$

one has LHS $\ge$ RHS. On the other hand

$$\|\boldsymbol{M}\boldsymbol{y}\|_2 = \langle \frac{My}{\|\boldsymbol{M}\boldsymbol{y}\|_2}, My\rangle \le \max\{\langle \boldsymbol{x}, \boldsymbol{M}\boldsymbol{y}\rangle : \|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2 = 1\}, \tag{3}$$

which implies LHS $\le$ RHS. Putting together completes the proof for the first part. Next we only have to show

$$\sigma_1(\boldsymbol{M}) = \max\{\langle \boldsymbol{x}, \boldsymbol{M}\boldsymbol{y}\rangle : \|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2 = 1\}. \tag{4}$$

Recall the SVD of $\boldsymbol{M} \in \mathbb{R}^{m\times n}$,

$$\boldsymbol{M} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top, \qquad r \le m \wedge n, \quad \sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r > 0, \tag{5}$$

where $\boldsymbol{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r] \in \mathbb{R}^{m\times r}$ and $\boldsymbol{V} = [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r] \in \mathbb{R}^{n\times r}$ are orthonormal. On the one hand, we have

$$\sigma_1(\boldsymbol{M}) = \boldsymbol{u}_1^\top \boldsymbol{M}\boldsymbol{v}_1 = \langle \boldsymbol{u}_1, \boldsymbol{M}\boldsymbol{v}_1\rangle \le \max\{\langle \boldsymbol{x}, \boldsymbol{M}\boldsymbol{y}\rangle : \|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2 = 1\}. \tag{6}$$

On the other hand, for any two normalized vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, we can deduce by Cauchy-Schwarz that

$$\langle \boldsymbol{x}, \boldsymbol{M}\boldsymbol{y}\rangle = \sum_{i=1}^{r} \sigma_i \langle \boldsymbol{x}, \boldsymbol{u}_i\rangle\langle \boldsymbol{y}, \boldsymbol{v}_i\rangle \le \sigma_1 \sqrt{\sum_{i=1}^{r}\langle \boldsymbol{x}, \boldsymbol{u}_i\rangle^2 \sum_{i=1}^{r}\langle \boldsymbol{y}, \boldsymbol{v}_i\rangle^2} = \sigma_1 \|P_{\boldsymbol{U}}\boldsymbol{x}\|_2 \|P_{\boldsymbol{V}}\boldsymbol{y}\|_2 \le \sigma_1 \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2 \le \sigma_1(\boldsymbol{M}). \tag{7}$$

where $P_{\boldsymbol{U}}, P_{\boldsymbol{V}}$ are projections onto subspaces $\boldsymbol{U}$ and $\boldsymbol{V}$. Taken collectively, the proof is complete.

1.2 Part (b)

Given the SVD of $\boldsymbol{M}$, we can directly write out the SVD of $\boldsymbol{M}^\top$ as

$$\boldsymbol{M}^\top = \sum_{i=1}^{r} \sigma_i \boldsymbol{v}_i \boldsymbol{u}_i^\top, \tag{8}$$

and by the result shown in part (a), we see that $\|\boldsymbol{M}\|_{\text{op}} = \sigma_1(\boldsymbol{M}) = \sigma_1 = \sigma_1(\boldsymbol{M}^\top) = \|\boldsymbol{M}^\top\|_{\text{op}}$. Secondly,

$$\begin{aligned}
\|\boldsymbol{AB}\|_{\text{op}} &= \max\{\|\boldsymbol{ABx}\|_2 : \|\boldsymbol{x}\|_2 = 1\} \\
&= 0 \vee \max\{\|\boldsymbol{ABx}\|_2 : \|\boldsymbol{x}\|_2 = 1, \|\boldsymbol{Bx}\|_2 \neq 0\} \\
&= 0 \vee \max\left\{\frac{\|\boldsymbol{ABx}\|_2}{\|\boldsymbol{Bx}\|_2} \cdot \|\boldsymbol{Bx}\|_2 : \|\boldsymbol{x}\|_2 = 1, \|\boldsymbol{Bx}\|_2 \neq 0\right\} \\
&\leq 0 \vee \max\left\{\frac{\|\boldsymbol{ABy}\|_2}{\|\boldsymbol{By}\|_2} \cdot \|\boldsymbol{Bx}\|_2 : \|\boldsymbol{x}\|_2 = 1, \|\boldsymbol{By}\|_2 \neq 0\right\} \\
&\leq 0 \vee (\max\{\|\boldsymbol{Az}\|_2 : \|\boldsymbol{z}\|_2 = 1, \boldsymbol{z} = \boldsymbol{By}\} \cdot \max\{\|\boldsymbol{Bx}\|_2 : \|\boldsymbol{x}\|_2 = 1\}) \\
&\leq \|\boldsymbol{A}\|_{\text{op}}\|\boldsymbol{B}\|_{\text{op}},
\end{aligned} \tag{9}$$

where the convention $\max\emptyset = -\infty$ is used.

1.3 Part (c)

(i) Note that for any $\boldsymbol{x}$, $\|a\boldsymbol{Mx}\|_2 = |a|\|\boldsymbol{Mx}\|_2$. By the first definition of operator norm, we get

$$\|a\boldsymbol{M}\|_{\text{op}} = |a|\|\boldsymbol{M}\|_{\text{op}}. \tag{10}$$

(ii) Again we use the first definition, and deduce that

$$\begin{aligned}
\|\boldsymbol{A}+\boldsymbol{B}\|_{\text{op}} &= \max\{\|(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}\|_2 : \|\boldsymbol{x}\|_2 = 1\} \\
&= \max\{\|\boldsymbol{Ax}\|_2 + \|\boldsymbol{Bx}\|_2 : \|\boldsymbol{x}\|_2 = 1\} \\
&\leq \max\{\|\boldsymbol{Ax}\|_2 + \|\boldsymbol{Bx}\|_2 : \|\boldsymbol{x}\|_2 = 1\} \\
&\leq \max\{\|\boldsymbol{Ax}\|_2 + \|\boldsymbol{By}\|_2 : \|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2 = 1\} \\
&= \max\{\|\boldsymbol{Ax}\|_2 : \|\boldsymbol{x}\|_2 = 1\} + \max\{\|\boldsymbol{By}\|_2 : \|\boldsymbol{y}\|_2 = 1\} \\
&\leq \|\boldsymbol{A}\|_{\text{op}} + \|\boldsymbol{B}\|_{\text{op}}.
\end{aligned} \tag{11}$$

(iii) Finally, we invoke the maximum singular value definition. Since $\|\boldsymbol{M}\|_{\text{op}} = \sigma_1(\boldsymbol{M}) = 0$, one must have

$$\boldsymbol{M} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \boldsymbol{0}. \tag{12}$$

2 Problem 2

2.1 Part (a)

Let $\boldsymbol{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k] \in \mathbb{R}^{n\times k}$. For any other orthonormal basis of W, we can denote it by $\boldsymbol{U}' = \boldsymbol{U}\boldsymbol{O}$, where $\boldsymbol{O} \in \mathbb{R}^{k\times k}$ is an orthogonal matrix. Then

$$
\begin{aligned}
\sum_{j=1}^{k} \langle \boldsymbol{u}_j, \boldsymbol{A}\boldsymbol{u}_j \rangle &= \operatorname{tr}\left[\boldsymbol{u}_i^\top \boldsymbol{A}\boldsymbol{u}_j\right]_{1\le i,j\le k} = \operatorname{tr}\left(\boldsymbol{U}^\top \boldsymbol{A}\boldsymbol{U}\right) = \operatorname{tr}\left(\boldsymbol{U}^\top \boldsymbol{A}\boldsymbol{U}\boldsymbol{O}\boldsymbol{O}^\top\right) \\
&= \operatorname{tr}\left(\boldsymbol{O}^\top \boldsymbol{U}^\top \boldsymbol{A}\boldsymbol{U}\boldsymbol{O}\right) \\
&= \operatorname{tr}\left(\boldsymbol{U}'^\top \boldsymbol{A}\boldsymbol{U}'\right) \\
&= \sum_{j=1}^{k} \langle \boldsymbol{u}_j', \boldsymbol{A}\boldsymbol{u}_j' \rangle,
\end{aligned} \tag{13}
$$

where we use the fact that $\boldsymbol{O}\boldsymbol{O}^\top = \boldsymbol{I}$ and $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{B}\boldsymbol{A})$ for all $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{B} \in \mathbb{R}^{n\times m}$. Thus the definition of $\operatorname{tr}(\boldsymbol{A}|_W)$ doesn't depend on the choice of orthonormal basis.

2.2 Part (b)

Since

$$
\dim\left(W \cap V_1\right) = 1 = \left\{\alpha \boldsymbol{u}_1 \middle| \|\boldsymbol{u}_1\|_2 = 1, \alpha \in \mathbb{R}, \boldsymbol{u}_1 \in V_1\right\}, \tag{14}
$$

we can choose a orthonormal basis of W with the first vector exactly being $\boldsymbol{u}_1$. Let the complete basis be $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k\}$. Define

$$
W' = \boldsymbol{u}_2 \oplus \cdots \oplus \boldsymbol{u}_k, \tag{15}
$$

we show that $W' \in \mathcal{S}_{k-1}\left(V_2, \cdots, V_k\right)$. In fact since $\boldsymbol{u}_1 \perp W'$, for all $1 \le j \le k$,

$$
\begin{aligned}
\dim(W' \cap V_j) &= \dim(\boldsymbol{u}_1 \oplus (W' \cap V_j)) - 1 \\
&= \dim((\boldsymbol{u}_1 \oplus W') \cap V_j) - 1 \\
&= \dim(W \cap V_j) - 1 \\
&= j - 1,
\end{aligned} \tag{16}
$$

where in the second line we use the fact $\boldsymbol{u}_1 \in V_j$. Next we show the desired inequality, now it's clear that

$$
\begin{aligned}
\operatorname{tr}(\boldsymbol{A}|_W) &= \sum_{j=1}^{k} \langle \boldsymbol{u}_j, \boldsymbol{A}\boldsymbol{u}_j \rangle \\
&= \langle \boldsymbol{u}_1, \boldsymbol{A}\boldsymbol{u}_1 \rangle + \sum_{j=2}^{k} \langle \boldsymbol{u}_j, \boldsymbol{A}\boldsymbol{u}_j \rangle \\
&= \langle \boldsymbol{u}_1, \boldsymbol{A}\boldsymbol{u}_1 \rangle + \operatorname{tr}(\boldsymbol{A}|_{W'}) \\
&\ge \lambda_{i_1}(\boldsymbol{A}) + \operatorname{tr}(\boldsymbol{A}|_{W'}),
\end{aligned} \tag{17}
$$

while the final line follows from $\boldsymbol{u}_1 \in V_1 \Rightarrow \boldsymbol{u}_1 = \sum_{l=1}^{i_1} \langle \boldsymbol{u}_1, \boldsymbol{v}_l \rangle \boldsymbol{v}_l$, and thus

$$
\langle \boldsymbol{u}_1, \boldsymbol{A}\boldsymbol{u}_1 \rangle = \sum_{l=1}^{i_1} \langle \boldsymbol{u}_1, \boldsymbol{v}_l \rangle^2 \lambda_l(\boldsymbol{A}) \ge \lambda_{i_1}(\boldsymbol{A}) \sum_{l=1}^{i_1} \langle \boldsymbol{u}_1, \boldsymbol{v}_l \rangle^2 = \lambda_{i_1}(\boldsymbol{A}). \tag{18}
$$

2.3 Part (c)

Let $W_0 := W$, we prove by induction that there exists $W_j \in \mathcal{S}_{k-j}(V_{j+1}, \cdots, V_k)$ (in particular, $W_k = \emptyset$) such that

$$\mathrm{tr}(\boldsymbol{A}|_{W_0}) \geq \lambda_{i_1}(\boldsymbol{A}) + \cdots + \lambda_{i_j}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{A}|_{W_j}), \qquad \forall 1 \leq j \leq k. \tag{19}$$

By part (b) we see that the induction hypothesis holds for $j = 1$. Suppose it's true for some $j < k$. We invoke part (b) for $W_j \in \mathcal{S}_{k-j}(V_{j+1}, \cdots, V_k)$. Then there must be some $W_{j+1} \in \mathcal{S}_{k-j-1}(V_{j+2}, \cdots, V_k)$ such that

$$\mathrm{tr}(\boldsymbol{A}|_{W_j}) \geq \lambda_{i_{j+1}}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{A}|_{W_{j+1}}). \tag{20}$$

The conclusion also holds for $j + 1$. The proof is complete by induction and using $W_k = \emptyset$. Finally by definition, we thus have for this special choice of $(V_1, \cdots, V_k)$,

$$\mathcal{R}\left(\boldsymbol{A}; i_1, \cdots, i_k\right) \geq \inf_{W \in \mathcal{S}_k(V_1, \cdots, V_k)} \mathrm{tr}(\boldsymbol{A}|_W) \geq \lambda_{i_1}(\boldsymbol{A}) + \cdots + \lambda_{i_k}(\boldsymbol{A}). \tag{21}$$

2.4 Part (d)

We use the same notations $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ in part (b) for $\boldsymbol{A}$'s eigenvectors. We construct W in the following way. Let $U_1 := \mathrm{span}(\boldsymbol{v}_{i_1}, \cdots, \boldsymbol{v}_n)$, then $\dim(U_1) = n - i_1 + 1$, while $\dim(V_1) = i_1$, thus

$$\dim(V_1 \cap U_1) \geq 1, \tag{22}$$

we can choose some $\boldsymbol{u}_1 \in V_1 \cap U_1$ of unit norm. Then we can simply choose normalized vectors $\boldsymbol{u}_j \in V_j \backslash V_{j-1}, \boldsymbol{u}_j \perp V_{j-1}$ for $j = 2, \cdots, k$ by Gram-Schmidt orthogonalization since $V_j \backslash V_{j-1} \neq \emptyset$. Let $W = \mathrm{span}(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k)$ and $W' = \mathrm{span}(\boldsymbol{u}_2, \cdots, \boldsymbol{u}_k)$. Clearly since $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k$ are orthonormal, and for any $1 \leq j \leq k$, $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_j \in V_j$, $\boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_k \perp V_j$, one has

$$\dim(W \cap V_j) = \dim(\mathrm{span}(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_j)) = j. \tag{23}$$

Following the same argument, one can easily check $\dim(W' \cap V_j) = j - 1$ for all $1 \leq j \leq k$. Therefore

$$\mathrm{tr}(\boldsymbol{A}|_W) = \langle \boldsymbol{u}_1, \boldsymbol{A}\boldsymbol{u}_1 \rangle + \mathrm{tr}(\boldsymbol{A}|_{W'}) \leq \lambda_{i_1}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{A}|_{W'}), \tag{24}$$

where the inequality follows from $\boldsymbol{u}_1 \in U_1$, and

$$\langle \boldsymbol{u}_1, \boldsymbol{A}\boldsymbol{u}_1 \rangle = \sum_{l=i_1}^{n} \langle \boldsymbol{u}_1, \boldsymbol{v}_l \rangle^2 \lambda_l(\boldsymbol{A}) \leq \lambda_{i_1}(\boldsymbol{A}) \sum_{l=i_1}^{n} \langle \boldsymbol{u}_1, \boldsymbol{v}_l \rangle^2 = \lambda_{i_1}(\boldsymbol{A}). \tag{25}$$

2.5 Part (e) - Optional

We can construct the chain of $W =: W_0 \supset W_1 \cdots \supset W_k = \emptyset$ from below by induction. First of all, we want to construct a perturbed orthonormal basis of $\mathbb{R}^n$, such that

$$[\boldsymbol{g}_1, \cdots, \boldsymbol{g}_n] =: \boldsymbol{G} = \boldsymbol{V}\boldsymbol{O}, \qquad \|\boldsymbol{O} - \boldsymbol{I}\|_{\max} \leq \epsilon,$$

where $\boldsymbol{V}$ is the eigenvector matrix of $\boldsymbol{A}$ and $\boldsymbol{O}$ is orthogonal. We define $U_j := \mathrm{span}(\boldsymbol{g}_{i_j}, \cdots, \boldsymbol{g}_n)$ for $1 \leq j \leq k$. We claim there exists a $\boldsymbol{G}$ for any $\epsilon > 0$ such that

$$U_j \cap V_{j-1} = \emptyset, \qquad \forall j = 2, 3, \cdots, k. \tag{26}$$

Lemma 1. *For $1 \leq j \leq k$ and any $W_j \in \mathcal{S}_{k-j}(V_{j+1}, \cdots, V_k)$ that takes the form $W_j = \mathrm{span}(\boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n)$ where the n orthonormal basis $\boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_k$ satisfy $\boldsymbol{u}_l \in U_l$ for $j + 1 \leq l \leq k$, there exists a $W_{j-1} \in \mathcal{S}_{k-j+1}(V_j, \cdots, V_k)$ such that $W_{j-1} = \mathrm{span}(\boldsymbol{u}_j, \cdots, \boldsymbol{u}_n) = \mathrm{span}(\boldsymbol{u}_j, W_j)$ where $\boldsymbol{u}_j \in U_j$.*

Proof. Since $\dim(U_j) = n - i_j + 1$ and $\dim(V_1) = i_j$, thus

$$\dim(V_j \cap U_j) \geq 1 \tag{27}$$

and we can choose some $\tilde{\boldsymbol{u}}_j \in V_j \cap U_j$ of unit norm. For any $j+1 \leq l \leq k$, since $\boldsymbol{u}_l \in W_j$ and $W_j \cap V_j \subset U_{j+1} \cap V_j = \emptyset$, we know $\tilde{\boldsymbol{u}}_j$ is linearly independent of $\boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_k$. Therefore, by setting $W_{j-1} = \operatorname{span}(\tilde{\boldsymbol{u}}_j, W_j)$, we have

$$\dim(W_{j-1}) = \dim(W_j) + 1 = j, \tag{28}$$

$$\dim(W_{j-1} \cap V_l) = \dim((W_j + \tilde{\boldsymbol{u}}_j) \cap V_l) = \dim(W_j \cap V_l) + 1 = j, \qquad \forall j \leq l \leq k, \tag{29}$$

where the second line comes form $\tilde{\boldsymbol{u}}_j \notin W_j$ and $\tilde{\boldsymbol{u}}_j \in V_j \subset V_l$. Therefore $W_{j-1} \in \mathcal{S}_{k-j+1}(V_j, \cdots, V_k)$. Finally, we take $\boldsymbol{u}_j$ by Gram-Schmidt orthogonalization

$$\boldsymbol{u}_j = \frac{\tilde{\boldsymbol{u}}_j - \sum_{l=j+1}^{n} \langle \tilde{\boldsymbol{u}}_j, \boldsymbol{u}_l \rangle \boldsymbol{u}_l}{\left\| \tilde{\boldsymbol{u}}_j - \sum_{l=j+1}^{n} \langle \tilde{\boldsymbol{u}}_j, \boldsymbol{u}_l \rangle \boldsymbol{u}_l \right\|_2}. \tag{30}$$

Note that $\tilde{\boldsymbol{u}}_j \in U_j, \boldsymbol{u}_l \in U_l \subset U_j$ for all $j+1 \leq l \leq n$, one must have $\boldsymbol{u}_j \in U_j$. The proof of the lemma is complete. □

Therefore, by iteratively invoking the lemma starting from $W_k := \emptyset$, we can construct a sequence of subspace $W_j \in \mathcal{S}_{k-j}(V_{j+1}, \cdots, V_k)$ where $W := W_0$ has an orthonormal basis $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k$ such that $\boldsymbol{u}_j \in U_j$. Note that

$$\langle \boldsymbol{u}_j, \boldsymbol{A}\boldsymbol{u}_j \rangle \leq \max_{\|\boldsymbol{x}\|_2 = 1, \boldsymbol{x} \in U_j} \langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{x} \rangle. \tag{31}$$

Clearly for any $\delta > 0$, we can always choose $\epsilon > 0$ small enough such that

$$\langle \boldsymbol{u}_j, \boldsymbol{A}\boldsymbol{u}_j \rangle \leq \max_{\|\boldsymbol{x}\|_2 = 1, \boldsymbol{x} \in \operatorname{span}(\boldsymbol{v}_{i_j}, \cdots, \boldsymbol{v}_n)} \langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{x} \rangle + \delta \leq \lambda_{i_j}(\boldsymbol{A}) + \delta, \qquad \forall 1 \leq j \leq k, \tag{32}$$

since

$$\lim_{\epsilon \to 0} [\boldsymbol{g}_{i_j}, \cdots, \boldsymbol{g}_n] \to [\boldsymbol{v}_{i_j}, \cdots, \boldsymbol{v}_n], \qquad \forall 1 \leq j \leq k. \tag{33}$$

Therefore

$$\operatorname{tr}(\boldsymbol{A}|_W) = \sum_{j=1}^{k} \langle \boldsymbol{u}_j, \boldsymbol{A}\boldsymbol{u}_j \rangle \leq \sum_{j=1}^{k} \lambda_{i_j}(\boldsymbol{A}) + k\delta. \tag{34}$$

Note that δ can be arbitrarily small, we thus can conclude

$$\inf_{W \in \mathcal{S}_k(V_1, \cdots, V_k)} \operatorname{tr}(\boldsymbol{A}|_W) \leq \sum_{j=1}^{k} \lambda_{i_j}(\boldsymbol{A}). \tag{35}$$

We only need to prove for any $\epsilon > 0$, $\boldsymbol{G}$ exists such that

$$U_j \cap V_{j-1} = \emptyset, \qquad \forall j = 2, 3, \cdots, k. \tag{36}$$

In fact for any $2 \leq j \leq k$, $\dim(V_{j-1}) \leq n - 1$. Under the basis $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$, V_{j-1} is a lower dimensional hyperplane inside $\mathbb{R}^n$. While U_j has an induced Haar measure from the orthogonal transform $\boldsymbol{O}$. The Haar measure is uniform and therefore the measure of its intersection with a fixed lower dimensional hyperplane being nonempty is 0. Hence such $\boldsymbol{G}$ must exists since $\|\boldsymbol{O} - \boldsymbol{I}\|_{\max} \leq \epsilon$ has positive Haar measure.

3 Problem 3

We first show for any subspace W of dimension k,

$$\mathrm{tr}(\boldsymbol{B}|_W) \leq \lambda_1(\boldsymbol{B}) + \cdots + \lambda_k(\boldsymbol{B}). \tag{37}$$

Suppose the i-th normalized eigenvector of $\boldsymbol{B}$ is $\boldsymbol{z}_i$, let $\boldsymbol{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k] \in \mathbb{R}^{n\times k}$ be an orthonormal basis of W. Let $\boldsymbol{Z} = [\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n] \in \mathbb{R}^{n\times n}$. Note that

$$\boldsymbol{B} = \sum_{i=1}^{n} \lambda_i(\boldsymbol{B})\boldsymbol{z}_i\boldsymbol{z}_i^\top. \tag{38}$$

Then

$$\mathrm{tr}(\boldsymbol{B}|_W) = \sum_{j=1}^{k}\langle \boldsymbol{u}_j, \boldsymbol{B}\boldsymbol{u}_j\rangle = \sum_{j=1}^{k}\sum_{i=1}^{n}\lambda_i(\boldsymbol{B})\langle \boldsymbol{u}_j, \boldsymbol{z}_i\rangle^2 := \sum_{i=1}^{n}\lambda_i(\boldsymbol{B})\sum_{j=1}^{k} c_{ij}, \tag{39}$$

where $c_{ij} = \langle \boldsymbol{z}_i, \boldsymbol{u}_j\rangle^2$ and

$$0 \leq p_i := \sum_{j=1}^{k} c_{ij} \leq 1, \qquad \sum_{i=1}^{n} p_i = \sum_{j=1}^{k}\sum_{i=1}^{n} c_{ij} = \sum_{j=1}^{k} 1 = k. \tag{40}$$

Thus given $\lambda_1(\boldsymbol{B}) \geq \cdots \geq \lambda_n(\boldsymbol{B})$ and at most k steps of greedy algorithm, we can show

$$\mathrm{tr}(\boldsymbol{B}|_W) = \sum_{i=1}^{n}\lambda_i(\boldsymbol{B})p_i \leq \sum_{i=1}^{n}\lambda_i(\boldsymbol{B})\mathbb{1}_{i\leq k} = \lambda_1(\boldsymbol{B}) + \cdots + \lambda_k(\boldsymbol{B}). \tag{41}$$

Similarly

$$\mathrm{tr}(\boldsymbol{B}|_W) = \sum_{i=1}^{n}\lambda_i(\boldsymbol{B})p_i \geq \sum_{i=1}^{n}\lambda_i(\boldsymbol{B})\mathbb{1}_{i\geq n-k+1} = \lambda_{n-k+1}(\boldsymbol{B}) + \cdots + \lambda_n(\boldsymbol{B}). \tag{42}$$

Finally we can finish the proof using Problem 2, which tells us

$$\begin{aligned}
\lambda_{i_1}(\boldsymbol{A}+\boldsymbol{B}) + \cdots + \lambda_{i_k}(\boldsymbol{A}+\boldsymbol{B}) &= \mathcal{R}\left(\boldsymbol{A}+\boldsymbol{B}; i_1, \cdots, i_k\right) \\
&= \sup_{(V_1,\cdots,V_k)\in\mathcal{F}(i_1,\cdots,i_k)} \inf_{W\in\mathcal{S}_k(V_1,\cdots,V_k)} \mathrm{tr}(\boldsymbol{A}+\boldsymbol{B}|_W) \\
&= \sup_{(V_1,\cdots,V_k)\in\mathcal{F}(i_1,\cdots,i_k)} \inf_{W\in\mathcal{S}_k(V_1,\cdots,V_k)} \left(\mathrm{tr}(\boldsymbol{A}|_W) + \mathrm{tr}(\boldsymbol{B}|_W)\right) \\
&\leq \sup_{(V_1,\cdots,V_k)\in\mathcal{F}(i_1,\cdots,i_k)} \inf_{W\in\mathcal{S}_k(V_1,\cdots,V_k)} \left(\mathrm{tr}(\boldsymbol{A}|_W) + \lambda_1(\boldsymbol{B}) + \cdots + \lambda_k(\boldsymbol{B})\right) \\
&= \left(\sup_{(V_1,\cdots,V_k)\in\mathcal{F}(i_1,\cdots,i_k)} \inf_{W\in\mathcal{S}_k(V_1,\cdots,V_k)} \mathrm{tr}(\boldsymbol{A}|_W)\right) + \lambda_1(\boldsymbol{B}) + \cdots + \lambda_k(\boldsymbol{B}) \\
&= \mathcal{R}\left(\boldsymbol{A}; i_1, \cdots, i_k\right) + \lambda_1(\boldsymbol{B}) + \cdots + \lambda_k(\boldsymbol{B}) \\
&= \lambda_{i_1}(\boldsymbol{A}) + \cdots + \lambda_{i_k}(\boldsymbol{A}) + \lambda_1(\boldsymbol{B}) + \cdots + \lambda_k(\boldsymbol{B}).
\end{aligned} \tag{43}$$

Similarly, we can get the inequality in the other direction,

$$\lambda_{i_1}(\boldsymbol{A}+\boldsymbol{B}) + \cdots + \lambda_{i_k}(\boldsymbol{A}+\boldsymbol{B}) \geq \lambda_{i_1}(\boldsymbol{A}) + \cdots + \lambda_{i_k}(\boldsymbol{A}) + \lambda_{n-k+1}(\boldsymbol{B}) + \cdots + \lambda_n(\boldsymbol{B}). \tag{44}$$