Homework 2 Solutions

Please do not distribute.

1. Mutual coherence (*40 points*)

For an arbitrary pair of orthonormal bases $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \cdots, \boldsymbol{\psi}_n] \in \mathbb{R}^{n\times n}$ and $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_n] \in \mathbb{R}^{n\times n}$, the mutual coherence $\mu(\boldsymbol{\Psi}, \boldsymbol{\Phi})$ of these two bases is defined by

$$\mu(\boldsymbol{\Psi}, \boldsymbol{\Phi}) = \max_{1\le i,j\le n} \left|\boldsymbol{\psi}_i^\top \boldsymbol{\phi}_j\right| \tag{1}$$

(a) Show that

$$\frac{1}{\sqrt{n}} \le \mu(\boldsymbol{\Psi}, \boldsymbol{\Phi}) \le 1.$$

Solution: *By definition,*

$$\mu(\boldsymbol{\Psi}, \boldsymbol{\Phi}) = \max_{1\le i,j\le n} \left|\boldsymbol{\psi}_i^\top \boldsymbol{\phi}_j\right| \ge \sqrt{\frac{1}{n^2}\sum_{1\le i,j\le n} \left|\boldsymbol{\psi}_i^\top \boldsymbol{\phi}_j\right|^2} \tag{2}$$

$$= \sqrt{\frac{1}{n^2}\|\boldsymbol{\Psi}^\top \boldsymbol{\Phi}\|_{\mathrm{F}}^2}. \tag{3}$$

Recognizing that $\|\boldsymbol{\Psi}^\top \boldsymbol{\Phi}\|_{\mathrm{F}}^2 = \mathrm{tr}(\boldsymbol{\Psi}^\top \boldsymbol{\Phi}\boldsymbol{\Phi}^\top\boldsymbol{\Psi}) = n$, *we obtain* $\mu(\boldsymbol{\Psi}, \boldsymbol{\Phi}) \ge \frac{1}{\sqrt{n}}$.

(b) Let $\boldsymbol{\Psi} = \boldsymbol{I}$, and suppose that $\boldsymbol{\Phi} = [\phi_{i,j}]_{1\le i,j\le n}$ is a Gaussian random matrix such that the $\phi_{i,j}$'s are i.i.d. random variables drawn from $\phi_{i,j} \sim \mathcal{N}(0, 1/n)$. Can you provide an upper estimate on $\mu(\boldsymbol{\Psi}, \boldsymbol{\Phi})$ as defined in (1)? Since $\boldsymbol{\Phi}$ is a random matrix, we expect your answer to be a function $f(n)$ such that $\mathbb{P}\{\mu(\boldsymbol{\Psi}, \boldsymbol{\Phi}) > f(n)\} \to 0$ as n scales.

Hint: to simplify analysis, you are allowed to use the crude approximation $\mathbb{P}\{|z| > \tau\} \approx \exp(-\tau^2/2)$ for large $\tau > 0$, where $z \sim \mathcal{N}(0, 1)$.

Solution: *Note that* $\boldsymbol{\Psi} = [\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n]$ *with* $\boldsymbol{e}_i$ *the* i*th standard basis vector. One has*

$$\mu(\boldsymbol{\Psi}, \boldsymbol{\Phi}) = \max_{1\le i,j\le n} |\langle \boldsymbol{e}_i, \boldsymbol{\phi}_j| \le \max_{1\le i,j\le n} |\phi_{i,j}|.$$

This boils down to understanding the maximum magnitude of a set of i.i.d. Gaussian variables. For any sufficiently large $\tau > 0$,

$$\mathbb{P}\left\{\max_{1\le i,j\le n} |\phi_{i,j}| > \frac{\tau}{\sqrt{n}}\right\} \le \sum_{1\le i,j\le n} \mathbb{P}\left\{\sqrt{n}|\phi_{i,j}| > \tau\right\} \tag{4}$$

$$\approx n^2 \exp\left(-\frac{\tau^2}{2}\right) \tag{5}$$

$$= \exp\left(2\log n - \frac{\tau^2}{2}\right), \tag{6}$$

where (4) follows from the union bound, and (5) is a consequence of the (crude) approximation $\mathbb{P}\{|z| > \tau\} \approx \exp\left(-\tau^2/2\right)$. *As a result, taking*

$$\tau = (1+\epsilon)2\sqrt{\log n}$$

for any constant $0 < \epsilon < 1$, *we obtain*

$$\begin{aligned} \mathbb{P}\left\{\max_{1\leq i,j\leq n} |\phi_{i,j}| > \frac{\tau}{\sqrt{n}}\right\} \quad &\leq \quad \exp\left(2\log n - 2\left(1+\epsilon\right)^2 \log n\right) \\ &= \quad \exp\left(-\left(2\epsilon+\epsilon^2\right) 2\log n\right) \\ &\overset{n\to\infty}{\to} \quad 0. \end{aligned}$$

This reveals that with probability approaching one,

$$\mu\left(\boldsymbol{\Psi}, \boldsymbol{\Phi}\right) \leq \frac{(1+\epsilon)2\sqrt{\log n}}{\sqrt{n}}$$

for any small constant $\epsilon > 0$.

(c) Set $n = 100$. Generate a random matrix $\boldsymbol{\Phi}$ as in Part (b), and compute $\mu(\boldsymbol{I}, \boldsymbol{\Phi})$. Report the empirical distribution (i.e. histogram) of $\mu(\boldsymbol{I}, \boldsymbol{\Phi})$ out of 1000 simulations. How does your simulation result compare to your estimate in Part (b)?

Solution: *For this case, the mutual coherence can be simply computed as*

$$\mu(\boldsymbol{I}, \boldsymbol{\Phi}) = \max_{i,j} |\langle \boldsymbol{e}_i, \boldsymbol{\phi}_j\rangle| = \max_{i,j} |\Phi_{i,j}|.$$

Fig. 1 shows the empirical distribution of $\mu(\boldsymbol{I}, \boldsymbol{\Phi})$ *out of 1000 simulations for* $n = 100$. *The estimate given in part (b) is*

$$\frac{2\sqrt{\log n}}{\sqrt{n}} \approx 0.4292,$$

which is depicted in black line. It is shown that for $n = 100$, *84.4% of 1000 simulations are lower than this estimate.*

(d) We now generalize the mutual coherence measure to accommodate a more general set of vectors beyond two bases. Specifically, for any given matrix $\boldsymbol{A} = [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_p] \in \mathbb{R}^{n\times p}$ obeying $n \leq p$, define the mutual coherence of $\boldsymbol{A}$ as

$$\mu(\boldsymbol{A}) = \max_{1\leq i,j\leq p,\ i\neq j} \left| \frac{\boldsymbol{a}_i^\top \boldsymbol{a}_j}{\|\boldsymbol{a}_i\|\|\boldsymbol{a}_j\|} \right|.$$

Show that

$$\mu(\boldsymbol{A}) \geq \sqrt{\frac{p-n}{p-1}\cdot\frac{1}{n}}.$$

This is a special case of the Welch bound.

Hint: you may want to use the following inequality: for any positive semidefinite $\boldsymbol{M} \in \mathbb{R}^{n\times n}$, $\|\boldsymbol{M}\|_{\mathrm{F}}^2 \geq \frac{1}{n}\left(\sum_{i=1}^n \lambda_i(\boldsymbol{M})\right)^2$.

Solution: *Without loss of generality, it is assumed that* $\|\boldsymbol{a}_i\| = 1$ *for all* $1 \leq i \leq p$. *To begin with, we find it convenient to work with the Gram matrix* $\boldsymbol{A}^\top\boldsymbol{A}$, *since the* (i, j) *entry of* $\boldsymbol{A}^\top\boldsymbol{A}$ *is exactly* $\boldsymbol{a}_i^\top\boldsymbol{a}_j$. *It is*

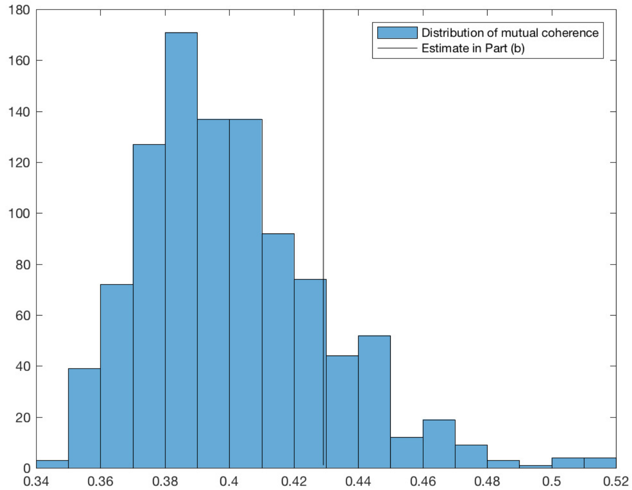

Figure 1: Empirical distribution of the mutual coherence.

seen that

$$\begin{aligned}
\sum_{1\le i,j\le p}\left(\boldsymbol{a}_i^\top \boldsymbol{a}_j\right)^2 &= \left\|\boldsymbol{A}^\top\boldsymbol{A}\right\|_{\mathrm{F}}^2 = \sum_{i=1}^n \lambda_i^2\left(\boldsymbol{A}^\top\boldsymbol{A}\right) && (7)\\
&\ge \frac{1}{n}\left(\sum_{i=1}^n \lambda_i\left(\boldsymbol{A}^\top\boldsymbol{A}\right)\right)^2 && (8)\\
&= \frac{1}{n}\left(\mathrm{Tr}\left(\boldsymbol{A}^\top\boldsymbol{A}\right)\right)^2 = \frac{p^2}{n}, && (9)
\end{aligned}$$

where (8) follows from the elementary inequality $\frac{1}{n}\sum_{i=1}^n x_i \le \sqrt{\frac{1}{n}\sum_{i=1}^n x_i^2}$ *(an immediate consequence of the Cauchy-Schwarz inequality).*

On the other hand, we can connect $\sum_{1\le i,j\le p}\left(\boldsymbol{a}_i^\top\boldsymbol{a}_j\right)^2$ *with the mutual coherence measure as follows*

$$\begin{aligned}
\sum_{1\le i,j\le p}\left(\boldsymbol{a}_i^\top \boldsymbol{a}_j\right)^2 &= \sum_{1\le i\le p}\left(\boldsymbol{a}_i^\top \boldsymbol{a}_i\right)^2 + \sum_{1\le i,j\le p,\ i\ne j}\left(\boldsymbol{a}_i^\top \boldsymbol{a}_j\right)^2\\
&= p + \sum_{1\le i,j\le p,\ i\ne j}\left(\boldsymbol{a}_i^\top \boldsymbol{a}_j\right)^2\\
&\le p + p(p-1)\mu^2(\boldsymbol{A}).
\end{aligned}$$

Putting the above two bounds together yields

$$p + p(p-1)\mu^2(\boldsymbol{A}) \ge \frac{p^2}{n},$$

which immediately establishes the claim.

2. ℓ_1 minimization (*30 points*)

Suppose that $\boldsymbol{A}$ is an $n \times 2n$ dimensional matrix. Let $\boldsymbol{x} \in \mathbb{R}^{2n}$ be an unknown k-sparse vector, and $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ the observed system output. This problem is concerned with ℓ_1 minimization (or basis pursuit) in recovering $\boldsymbol{x}$, i.e.

$$\text{minimize}_{\boldsymbol{z}\in\mathbb{R}^{2n}} \quad \|\boldsymbol{z}\|_1 \quad \text{s.t. } \boldsymbol{A}\boldsymbol{z} = \boldsymbol{y}. \tag{10}$$

(a) An optimization problem is called a linear program (LP) if it has the form

$$\begin{aligned} \text{minimize}_{\boldsymbol{z}} \quad & \boldsymbol{c}^\top \boldsymbol{z} + \boldsymbol{d} \\ \text{s.t.} \quad & \boldsymbol{G}\boldsymbol{z} \leq \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{z} = \boldsymbol{b} \end{aligned}$$

where $\boldsymbol{c}, \boldsymbol{d}, \boldsymbol{G}, \boldsymbol{h}, \boldsymbol{A}$, and $\boldsymbol{b}$ are known. Here, for any two vectors $\boldsymbol{r}$ and $\boldsymbol{s}$, we say $\boldsymbol{r} \leq \boldsymbol{s}$ if $r_i \leq s_i$ for all i. Show that (10) can be converted to a linear program.

Solution: *The given problem is*

$$\text{minimize}_{\boldsymbol{z}\in\mathbb{R}^{2n}} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{z} = \boldsymbol{y}. \tag{11}$$

This can be converted to

$$\text{minimize}_{\boldsymbol{z}\in\mathbb{R}^{2n}, \boldsymbol{s}\in\mathbb{R}^{2n}} \mathbf{1}^T \boldsymbol{s} \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{z} = \boldsymbol{y},\ |z_i| \leq s_i\ \forall\ i. \tag{12}$$

Denote optimal solutions achieving two problems as $\hat{\boldsymbol{z}}$ and $\bar{\boldsymbol{z}}, \bar{\boldsymbol{s}}$, respectively. Then,

$$\|\hat{\boldsymbol{z}}\|_1 \overset{(a)}{\leq} \|\bar{\boldsymbol{z}}\|_1 \overset{(b)}{\leq} \mathbf{1}^T \bar{\boldsymbol{s}},$$

where (a) holds because $\hat{\boldsymbol{z}}$ is minimizing the L1 norm, and (b) holds because of the feasibility condition of (12)*. Furthermore, we can prove that*

$$\|\hat{\boldsymbol{z}}\|_1 \geq \mathbf{1}^T \bar{\boldsymbol{s}}.$$

Take solution of (12) *as $z_i = \hat{z}_i$ and $s_i = |z_i|$. Then, these vectors are feasible, and the objective value becomes $\|\hat{\boldsymbol{z}}\|$. Thus, the objective value of $\bar{\boldsymbol{s}}$, which is an optimal solution, should be lower than or equal to $\|\hat{\boldsymbol{z}}\|$.*

It was shown that

$$\|\hat{\boldsymbol{z}}\|_1 = \mathbf{1}^\top \bar{\boldsymbol{s}},$$

which implies that the problem (11) *can be converted to* (12)*. The problem* (12) *can be rewritten as*

$$\text{minimize}_{\boldsymbol{z}\in\mathbb{R}^{2n}, \boldsymbol{s}\in\mathbb{R}^{2n}} \mathbf{1}^T \boldsymbol{s} \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{z} = \boldsymbol{y},\ \boldsymbol{z} - \boldsymbol{s} \leq \mathbf{0},\ \ -\boldsymbol{z} - \boldsymbol{s} \leq 0,$$

which is a linear program.

(b) Set $n = 256$, and let k range between 1 and 128. For each choice of k, run 10 independent numerical experiments: in each experiment, generate $\boldsymbol{A} = [a_{i,j}]_{1\leq i\leq n, 1\leq j\leq 2n}$ as a random matrix such that the $a_{i,j}$'s are i.i.d. standard Gaussian random variables, generate $\boldsymbol{x} \in \mathbb{R}^{2n}$ as a random k-sparse signal (e.g. you may generate the support of $\boldsymbol{x}$ uniformly at random, with each non-zero entry drawn from the standard Gaussian distribution), and solve (10) with $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. An experiment is claimed successful if the solution $\boldsymbol{z}$ returned by (10) obeys $\|\boldsymbol{x} - \boldsymbol{z}\|_2 \leq 0.001\|\boldsymbol{x}\|_2$. Report the empirical success rates (averaged over 10 experiments) for each choice of k.

Solution: *The support of $\boldsymbol{x}$ can be randomly chosen by using MATLAB built-in function randperm. Using cvx program, $\boldsymbol{z}$ can be computed, where feasibility conditions are arranged as*

$$\begin{aligned}
&\boldsymbol{A}\boldsymbol{z} = \boldsymbol{y}, \\
&\begin{bmatrix} \boldsymbol{I} & -\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{z} \\ \boldsymbol{s} \end{bmatrix} \leq \boldsymbol{0}, \\
&\begin{bmatrix} -\boldsymbol{I} & -\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{z} \\ \boldsymbol{s} \end{bmatrix} \leq \boldsymbol{0}.
\end{aligned}$$

Success rate for each k is given in Fig. 2 This graph shows that success rate is higher for sparse signals.

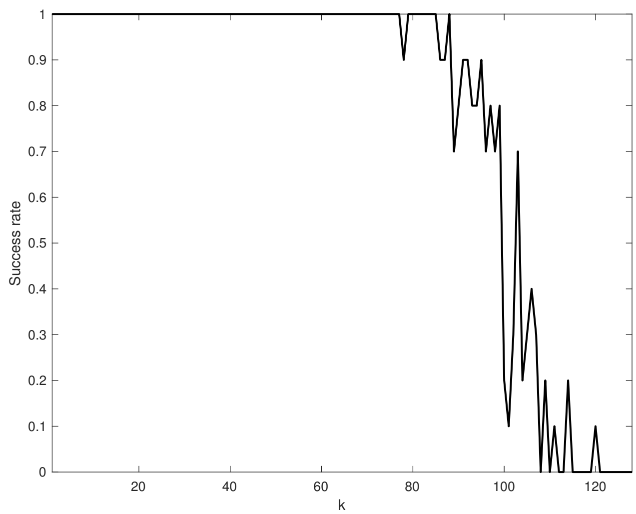

Figure 2: Success rate for each $k \in \{1, \cdots, 128\}$ with 10 expriments.

3. Restricted isometry properties (*30 points*)

Recall that the restricted isometry constant $\delta_s \geq 0$ of $\boldsymbol{A}$ is the smallest constant such that

$$(1-\delta_s)\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \leq (1+\delta_s)\|\boldsymbol{x}\|_2^2 \tag{13}$$

holds for all s-sparse vector $\boldsymbol{x} \in \mathbb{R}^p$.

(a) Show that

$$|\langle \boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{A}\boldsymbol{x}_2 \rangle| \leq \delta_{s_1+s_2}\|\boldsymbol{x}_1\|_2\|\boldsymbol{x}_2\|_2$$

for all pairs of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ that are supported on disjoint subsets $S_1, S_2 \subset \{1, \cdots, n\}$ with $|S_1| \leq s_1$ and $|S_2| \leq s_2$.

Solution: *WLOG, assume* $\|\boldsymbol{x}_1\| = \|\boldsymbol{x}_2\| = 1$. *Since* $\boldsymbol{x}_1$ *and* $\boldsymbol{x}_2$ *have disjoint support, we get*

$$\begin{aligned}
|\langle \boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{A}\boldsymbol{x}_2\rangle| &= \frac{1}{4}\left|\|\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{A}\boldsymbol{x}_2\|_2^2 - \|\boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{A}\boldsymbol{x}_2\|_2^2\right| \\
&= \frac{1}{4}\left|\left\|\boldsymbol{A}\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix}\right\|^2 - \left\|\boldsymbol{A}\begin{bmatrix}\boldsymbol{x}_1\\ -\boldsymbol{x}_2\end{bmatrix}\right\|^2\right| \\
&\le \frac{1}{4}|2(1+\delta_{s_1+s_2}) - 2(1-\delta_{s_1+s_2})| \\
&\le \delta_{s_1+s_2}.
\end{aligned}$$

(b) For any $\boldsymbol{u}$ and $\boldsymbol{v}$, show that

$$|\langle \boldsymbol{u},\ (\boldsymbol{I} - \boldsymbol{A}^\top \boldsymbol{A})\boldsymbol{v}\rangle| \le \delta_s \|\boldsymbol{u}\|_2 \cdot \|\boldsymbol{v}\|_2,$$

where s is the cardinality of support $(\boldsymbol{u}) \cup$ support $(\boldsymbol{v})$.

Solution: *Let* $\mathcal{S} = \textit{support}\,(\boldsymbol{u}) \cup \textit{support}\,(\boldsymbol{v})$.

$$\begin{aligned}
|\langle \boldsymbol{u}, (\boldsymbol{I} - \boldsymbol{A}^\top \boldsymbol{A})\boldsymbol{v}\rangle| &= |\langle \boldsymbol{u}, \boldsymbol{v}\rangle - \langle \boldsymbol{A}\boldsymbol{u}, \boldsymbol{A}\boldsymbol{v}\rangle| \\
&= |\langle \boldsymbol{u}_{\mathcal{S}}, \boldsymbol{v}_{\mathcal{S}}\rangle - \langle \boldsymbol{A}_{\mathcal{S}}\boldsymbol{u}_{\mathcal{S}}, \boldsymbol{A}_{\mathcal{S}}\boldsymbol{v}_{\mathcal{S}}\rangle| \\
&= |\langle \boldsymbol{u}_{\mathcal{S}}, (\boldsymbol{I} - \boldsymbol{A}_{\mathcal{S}}^\top \boldsymbol{A}_{\mathcal{S}})\boldsymbol{v}_{\mathcal{S}}\rangle| \\
&\le \|\boldsymbol{u}_{\mathcal{S}}\|_2 \|\boldsymbol{I} - \boldsymbol{A}_{\mathcal{S}}^\top \boldsymbol{A}_{\mathcal{S}}\|_{\mathsf{op}} \|\boldsymbol{v}_{\mathcal{S}}\|_2,
\end{aligned}$$

where $\|\cdot\|_{\mathsf{op}}$ *denotes the operator norm of a matrix as*

$$\|\boldsymbol{A}\|_{\mathsf{op}} = \max_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{A}\boldsymbol{x}\|_2.$$

By the definition of the restricted isometry constant,

$$|\langle (\boldsymbol{A}_{\mathcal{S}}^\top \boldsymbol{A}_{\mathcal{S}} - \boldsymbol{I})\boldsymbol{x}, \boldsymbol{x}\rangle| = |\langle \boldsymbol{A}_{\mathcal{S}}\boldsymbol{x}, \boldsymbol{A}_{\mathcal{S}}\boldsymbol{x}\rangle - \langle \boldsymbol{x}, \boldsymbol{x}\rangle| = |\|\boldsymbol{A}_{\mathcal{S}}\boldsymbol{x}\|^2 - \|\boldsymbol{x}\|_2^2| \le \delta_s \|\boldsymbol{x}\|_2^2.$$

Therefore,

$$\|\boldsymbol{I} - \boldsymbol{A}_{\mathcal{S}}^\top \boldsymbol{A}_{\mathcal{S}}\|_{\mathsf{op}} \le \delta_s$$

and

$$|\langle \boldsymbol{u}, (\boldsymbol{I} - \boldsymbol{A}^\top \boldsymbol{A})\boldsymbol{v}\rangle| \le \delta_s \|\boldsymbol{u}_{\mathcal{S}}\|_2 \|\boldsymbol{v}_{\mathcal{S}}\|_2 = \delta_s \|\boldsymbol{u}\|_2 \|\boldsymbol{v}\|_2.$$

(c) Suppose that each column of $\boldsymbol{A}$ has unit norm. Show that $\delta_2 = \mu(\boldsymbol{A})$, where $\mu(\boldsymbol{A})$ is the mutual coherence of $\boldsymbol{A}$.

Solution: *Given that*

$$|\langle (\boldsymbol{A}_{\mathcal{S}}^\top \boldsymbol{A}_{\mathcal{S}} - \boldsymbol{I})\boldsymbol{x}, \boldsymbol{x}\rangle| = |\langle \boldsymbol{A}_{\mathcal{S}}\boldsymbol{x}, \boldsymbol{A}_{\mathcal{S}}\boldsymbol{x}\rangle - \langle \boldsymbol{x}, \boldsymbol{x}\rangle| = |\|\boldsymbol{A}_{\mathcal{S}}\boldsymbol{x}\|^2 - \|\boldsymbol{x}\|^2| \le \delta_s \|\boldsymbol{x}\|^2,$$

δ_s *is the same as*

$$\delta_s = \max_{|\mathcal{S}| \le s} \|\boldsymbol{A}_{\mathcal{S}}^\top \boldsymbol{A}_{\mathcal{S}} - \boldsymbol{I}\|_{\mathsf{op}}.$$

When $s = 2$,

$$\delta_2 = \max_{i \neq j} \left\| \begin{bmatrix} \boldsymbol{a}_i & \boldsymbol{a}_j \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{a}_i & \boldsymbol{a}_j \end{bmatrix} - \boldsymbol{I} \right\|_{\mathsf{op}}.$$

The eigenvalues of the following matrix

$$\begin{bmatrix} \boldsymbol{a}_i & \boldsymbol{a}_j \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{a}_i & \boldsymbol{a}_j \end{bmatrix} - \boldsymbol{I} = \begin{bmatrix} 1 & \langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle \\ \langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle & 1 \end{bmatrix} - \boldsymbol{I} = \begin{bmatrix} 0 & \langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle \\ \langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle & 0 \end{bmatrix}$$

are $\pm \langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle$, and accordingly

$$\delta_2 = \max_{i \neq j} |\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle| = \mu(\boldsymbol{A}).$$