# Solution 4

Euler Cat

July 28, 2023

## 1 Solutions to exercises in chapter 4

### 1.1 4.1.1

Note that

$$\begin{aligned} s_i(A^{-1}) &= \sqrt{\lambda_i(A^{-1}(A^{-1})^T)} && (1)\\ &= \sqrt{\lambda_i((A^TA)^{-1})} && (2)\\ &= \frac{1}{\sqrt{\lambda_i(A^TA)}} && (3)\\ &= \frac{1}{s_i(A)} && (4)\end{aligned}$$

Moreover, by definition

$$AA^T * u_i = \lambda * u_i \tag{5}$$

So

$$\frac{1}{\lambda} * u_i = (A^{-1})^T A^{-1} * u_i \tag{6}$$

So $u_i$ is the right singular vector of matrix $A^{-1}$. Similarly, $v_i$ is the left singular vector of matrix $A^{-1}$.

## 1.2 4.1.2

Note that we arrange the singular values in descending orders:

$$s_1 \geq s_2 \geq s_2 \geq ... \geq 0 \tag{7}$$

As a result of this,

$$\begin{aligned} ||A||_F^2 &= \sum_{i=1}^{r} s_i(A)^2 && (8) \\ &\geq s_1(A)^2 + ... + s_i(A)^2 && (9) \\ &\geq i * s_i(A)^2 && (10) \end{aligned}$$

And the conclusion follows.

## 1.3 4.1.3

### 1.3.1

Since

$$s_1(A) = ||A|| \tag{11}$$

And the largest singular value of $A - A_k$ is $s_{k+1}(A)$, so

$$||A - A_k||^2 = s_{k+1}(A)^2 \tag{12}$$

### 1.3.2

Since the singular values of $A_k$ are $s_i, i = 1, ..., k$. So the singular values of $A - A_k$ will only have $s_i, i = k+1, ..., n$. Then

$$||A - A_k||_F^2 = \sum_{i=k+1}^{n} s_i(A)^2 \tag{13}$$

## 1.4 4.1.4

The flow of our proof consists of:

- (a) $\longleftrightarrow$ (b)
- (a) $\longleftrightarrow$ (d)
- (c) $\longleftrightarrow$ (d)

### 1.4.1 (a) $\longleftrightarrow$ (b)

If $A^TA = I_n$, then

$$\begin{aligned} P^2 &= AA^TAA^T && (14)\\ &= AA^T && (15)\\ &= P && (16) \end{aligned}$$

Moreover, if $y = Px$ and $Pz = 0$ then we have

$$\begin{aligned} < y, z > &=< Px, z > && (17)\\ &=< P^2x, z > && (18)\\ &=< Px, P^Tz > && (19)\\ &=< Px, Pz > && (20)\\ &= 0 && (21) \end{aligned}$$

Where we use the fact that $P$ here is a symmetric matrix. Finally, note that $n = rank(A^TA) = rank(AA^T)$. This tells $P$ projects onto a subspace of dimension $n$.

On the other hand, if we know $P$ is an orthogonal projection onto a subspace of dimension $n$, then $A$ has full rank on its columns, which in turn implies $A^TA$ is invertible. So

$$AA^TA = A \tag{22}$$

Where we use the fact that $P^2 = P$ and $AA^T$ and $A$ share the same column space. Then

$$A^TAA^TA = A^TA \tag{23}$$

Multiply the $(A^TA)^{-1}$ will get us

$$A^TA = I_n \tag{24}$$

### 1.4.2 (a)$\longleftrightarrow$(d)

From (a) to (d) is trivial. On the other hand, since $s_1(A) = ... = s_n(A) = 1$ and $A^T A$ is a symmetric matrix, we can decompose it into

$$\begin{aligned} A^T A &= U^T \Sigma U & (25) \\ &= U^T I_n U & (26) \\ &= U^T u & (27) \\ &= I_n & (28) \end{aligned}$$

Where in the last step we use the fact that $U$ is an orthogonal matrix.

### 1.4.3 (c) $\longleftrightarrow$ (d)

Without loss of generality, we assume that $||x|| = 1$. Then we just need to note that

$$max_{x \in S^{n-1}} ||Ax|| = s_1(A) \tag{29}$$

And

$$min_{x \in S^{n-1}} ||Ax||| = s_n(A) \tag{30}$$

### 1.5 4.1.6

Note that

$$|||Ax||_2^2 - 1| \leq max(|s_1(A)^2 - 1|, |s_n(A)^2 - 1|) \tag{31}$$

So we only need to prove that

$$|z^2 - 1| \leq max(\sigma, \sigma^2), \forall z, 1 - \sigma \leq z \leq 1 + \sigma \tag{32}$$

We need to consider different settings of $\sigma$: $\sigma \leq 1$ and $\sigma > 1$.

If $\sigma \leq 1$, then

$$(1 - \sigma)^2 - 1 \leq z^2 - 1 \leq (1 + \sigma)^2 - 1 \tag{33}$$

$$\sigma^2 - 2 * \sigma \leq z^2 - 1 \leq \sigma^2 + 2 * \sigma \tag{34}$$

Clearly, $\sigma^2 + 2*\sigma \leq 3*\sigma$ due to the fact that $0 < \sigma \leq 1$. And $-3*\sigma \leq \sigma^2 - 2*\sigma$ since $\sigma > 0$. So

$$|z^2 - 1| \leq 3 * \sigma \tag{35}$$

On the other hand, when $\sigma > 1$ we have:

$$\sigma^2 + 2 * \sigma \leq 3 * \sigma^2 \tag{36}$$

$$-3 * \sigma^2 \leq \sigma^2 - 2 * \sigma \tag{37}$$

So

$$|z^2 - 1| \leq 3 * \sigma^2 \tag{38}$$

Combined all together, we can have

$$|z^2 - 1| \leq 3 * max(\sigma, \sigma^2) \tag{39}$$

## 1.6 4.1.8

Suppose we extract $m$ columns from $U$ denoted by $U' = (u_1, ..., u_m)$, then

$$\begin{aligned} < U'x, U'x > &= < \sum_i x_i * u_i, \sum_j x_j * u_j > && (40) \\ &= \sum_{ij} x_i x_j < u_i, u_j > && (41) \\ &= \sum_i x_i^2 && (42) \\ &= ||x||_2^2 && (43) \end{aligned}$$

So it is an isometry if we extract several columns from $U$.

On the other hand, we use $u_i, i = 1, ..., m$ to denote the row extracted. Then, we can have

$$U'(U')^T = \begin{bmatrix} u_1 u_1^T & ... & u_1 u_m^T \\ ... & ... & ... \\ u_m u_1 & ... & u_m u_m^T \end{bmatrix} \tag{44}$$

Which is equal to $I$. So $P = U'(U')^T$ is an orthogonal projection, which means $U'$ here is a projection.

## 1.7 4.2.5

### 1.7.1

Let $N_1 = P(K, d, \epsilon)$ and use $y_1, ..., y_{N_1}$ denote the points in this $\epsilon$-separated subset of $K$ while we use $N_2$ to denote the largest number of closed disjoint balls with centers in $K$ and radii $\frac{\epsilon}{2}$. Then,

Since $d(y_i, y_j) > \epsilon, \forall i \neq j$. Then $B(y_i, \frac{\epsilon}{2}) \bigcap B(y_j, \frac{\epsilon}{2}) = \emptyset, \forall i \neq j$, we have

$$N_2 \geq N_1 \tag{45}$$

On the other hand, suppose we have $N_2$ disjoint balls with centers $y_1, ..., y_{N_2}$ in $K$ and radii $\frac{\epsilon}{2}$, then clearly $d(y_i, y_j) > \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$. So

$$N_1 \geq N_2 \tag{46}$$

By which we have

$$N_2 = N_1 \tag{47}$$

### 1.7.2

The counterexample comes from link.

Suppose we have a finite metric space $T = \{a, b\}$ where

$$d(a, b) = \epsilon \tag{48}$$

Then we have two disjoint balls with radii $\frac{\epsilon}{2}$: $B(a, \frac{\epsilon}{2}), B(b, \frac{\epsilon}{2})$. And $P(K, d, \epsilon) = 1$ where here $K = T$.

## 1.8 4.2.9

The definition of exterior covering number is the smallest possible cardinality of an $\epsilon$-net of $K$ while $x_i, i = 1, ..., N^{ext}$ may not be in $K$. By definition, each covering number of $K$ is a special case of exterior covering number. So we must have

$$N^{ext}(K, d, \epsilon) \leq N(K, d, \epsilon) \tag{49}$$

On the other hand, we have a sequence of points $x_i, i = 1, ..., N^{ext}$ where $B(x_i, \frac{\epsilon}{2})$ covers $K$ with minimal cardinality. Then we do the following process:

- If $x_i \in K$, then we keep this point while enlarge $B(x_i, \frac{\epsilon}{2})$ to $B(x_i, \epsilon)$.
- Otherwise, we extract another point $x_i' \in K \bigcap B(x_i, \frac{\epsilon}{2})$. And build a new ball $B(x_i', \epsilon)$

We claim this new family of balls are actually a $\epsilon$-net of $K$. Indeed, since $d(x_i, x_i') \leq \frac{\epsilon}{2}$ then $B(x_i', \epsilon)$ will contain $B(x_i, \frac{\epsilon}{2})$ completely. By definition of covering number, we have

$$N(K, d, \epsilon) \leq N^{ext}(K, d, \frac{\epsilon}{2}) \tag{50}$$

## 1.9 4.2.10

### 1.9.1

The counter example comes from link.

Let $K = \{-1, 1\}$ and $T = [-1, 1]$ and $\epsilon = 1$, then
$N(K, d, 1) = 2$ and $N(T, d, 1) = 1$.

### 1.9.2

Since $L \subset K$, then the covering points of $K$ may not locate in $L$. So similar to the above exercise, we do the following process to the points $x_i, i = 1, ..., N(K, d, \frac{\epsilon}{2})$:

- If $x_i \in L$, then we enlarge the ball to $B(x_i, \epsilon)$.
- Otherwise, if $x_i \notin L$ and $B(x_i, \epsilon) \bigcap L \neq \emptyset$ choose $x_i' \in L \bigcap B(x_i, \frac{\epsilon}{2})$. And enlarge the original $B(x_i, \frac{\epsilon}{2})$ to $B(x_i', \epsilon)$.
- If $B(x_i, \frac{\epsilon}{2}) \bigcap L = \emptyset$ otherwise, we just remove this ball.

Now we use $N'$ to denote the number of balls remained after the above process. Clearly, this $N'$ balls together form a $\epsilon$ covering. So,

$$N(L, d, \epsilon) \leq N' \leq N(K, d, \frac{\epsilon}{2}) \tag{51}$$

### 1.10 4.2.15

Clearly, $d(x, x) = 0$. And if $d(x, y) = 0$ then by definition $x_i = y_i, \forall i$ so $x = y$. The remained is to prove

$$d(x, y) \leq d(x, z) + d(z, y) \tag{52}$$

We only need to observe the $i$th element of $x, y, z$.

- If $x_i = y_i$, then $z_i = x_i = y_i$ will give us an equality otherwise it will give use an inequality.
- If $x_i \neq y_i$, then either $z_i = x_i$ or $z_i = y_i$ will give us an inequality.

So LHS of (52) will always be less than or equal to the RHS of (52).

## 1.11 4.2.16

By definition of $N$ and $P$, we directly know that

$$N(K, d_H, m) \leq P(K, d_H, m) \tag{53}$$

So we only need to focus on the two sides' inequalities.

Based on proposition 4.2.12 in the book, we can have

$$\frac{|K|}{|mB_2^n|} \leq N \leq P \leq \frac{|K + \frac{m}{2}B_2^n|}{|\frac{m}{2}B_2^n|} \tag{54}$$

Also note that

$$|K| = |\{0,1\}^n| = 2^n \tag{55}$$

$$|mB_2^n| = |\{x \in \{0,1\}^n : d(0,x) \leq m\}| \tag{56}$$

$$= \sum_{k=0}^{m} C_n^k \tag{57}$$

Where in the last step we use the fact that $x$ is a binary sequence and we are counting how many terms are 1.

Clearly,

$$|K + \frac{m}{2}B_2^n| \leq |K| = 2^n \tag{58}$$

Since $\frac{m}{2} \geq \lfloor \frac{m}{2} \rfloor$, we have

$$|\frac{m}{2}B_2^n| \geq \sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} C_n^k \tag{59}$$

So the conclusion follows.

## 1.12 4.3.7(TODO)

### 1.12.1

### 1.12.2

### 1.13 4.4.2

By using Cauchy-Schwarz inequality, we can have

$$| < x, y > | \leq ||x||_2||y||_2 = ||x||_2 \tag{60}$$

Where in the last step we use the fact $y$ lives on a sphere. So the first inequality follows.

On the other hand, select a $y \in S^{n-1}$ such that

$$sup_{y \in N} < x, y >=< x, y > \tag{61}$$

Then find a $y_0 \in N$ such that $||y - y_0|| \leq \epsilon$. So we can have

$$| < x, y > - < x, y_0 > | = | < x, y - y_0 > | \leq ||x||_2||y - y_0||_2 \leq ||x||_2 * \epsilon \tag{62}$$

Then we have

$$| < x, y_0 > | \geq | < x, y > | - | < x, y - y_0 > | \geq (1 - \epsilon) * sup_{y \in N} < x, y > \tag{63}$$

## 1.14 4.4.3

### 1.14.1

Since $N \subset S^{n-1}, M \subset S^{m-1}$, then the first inequality follows directly.

On the other hand, pick $x \in S^{n-1}$ and $y \in S^{m-1}$ such that

$$< Ax, y >= ||A|| \tag{64}$$

Then by definition of $\epsilon$-net, we select two points $x_0 \in N$ and $y_0 \in M$ respectively such that

$$||x - x_0|| \leq \epsilon \tag{65}$$
$$||y - y_0|| \leq \epsilon \tag{66}$$

As a result of this,

$$
\begin{align}
& | < Ax, y > - < Ax_0, y_0 > | \tag{67}\\
& = | < Ax, y > - < Ax_0, y > + < Ax_0, y > - < Ax_0, y_0 > | \tag{68}\\
& \leq | < A(x - x_0), y > | + | < Ax_0, y - y_0 > | \tag{69}\\
& \leq 2 * \epsilon * ||A|| \tag{70}\\
& \tag{71}
\end{align}
$$

Then

$$||A|| =< Ax, y >\leq 2 * \epsilon * ||A|| + < Ax_0, y_0 > \tag{72}$$

And the conclusion follows.

### 1.14.2

This part of conclusion follows from the first part by letting $y = x$ and $y = -x$ respectively.

## 1.15 4.4.4

Without loss of generality, we assume that $\mu = 1$.

Then we can have

$$||Ax||_2^2 - 1 = x^T A^T Ax - 1 =< Rx, x > \tag{73}$$

Where $R = A^T A - I_n$. Then, from part (b) of the above exercise we can know that

$$sup_{x \in N}| < Rx, x > | \leq ||R|| \leq \frac{1}{1-2\epsilon} sup_{x \in N}| < Rx, x > | \tag{74}$$

Recall a useful numerical inequality:

$$|z - 1| \leq |z^2 - 1|, \forall z > 0 \tag{75}$$

Then

$$\begin{aligned}
sup_{x \in S^{n-1}}|||Ax||_2 - 1| &\leq sup_{x \in S^{n-1}}|||Ax||_2^2 - 1| && (76)\\
&= sup_{x \in S^{n-1}}| < Rx, x > | && (77)\\
&\leq ||R|| && (78)\\
&\leq \frac{1}{1-2*\epsilon} sup_{x \in N}| < Rx, x > | && (79)\\
&= \frac{1}{1-2*\epsilon} * sup_x|||Ax||_2^2 - 1| && (80)\\
&= \frac{1}{1-2*\epsilon} * sup_x|(||Ax||_2 - 1)(||Ax||_2 + 1)| && (81)\\
&\leq \frac{1}{1-2*\epsilon} * sup_x|||Ax||_2 - 1| * (1 + ||A||) && (82)\\
&= \frac{C}{1-2*\epsilon} * sup_x|||Ax||_2 - 1| && (83)
\end{aligned}$$

Note: The constant $C$ here depends on $A$. How to get rid of the dependence on $A$ needs further considerations(I have not yet come up with this).

## 1.16 4.4.6

Note that

$$
\begin{aligned}
E||A|| &= \int_0^\infty P(||A|| > t)dt && (84)\\
&= CK\int_{-\sqrt{m}-\sqrt{n}}^\infty P(||A|| > CK(\sqrt{m}+\sqrt{n}+t))dt && (85)\\
&\leq CK\int_{-\sqrt{m}-\sqrt{n}}^\infty 2e^{-t^2}dt && (86)\\
&= CK\int_{-\sqrt{m}-\sqrt{n}}^\infty e^{-t^2}dt && (87)\\
&\leq CK(\frac{\sqrt{\pi}}{2} + \int_0^{\sqrt{m}+\sqrt{n}} e^{-t^2}dt) && (88)\\
&= CK(\frac{\sqrt{\pi}}{2} + \int_0^{2\pi}\int_0^{\sqrt{m}+\sqrt{n}} re^{-r^2}drd\theta) && (89)\\
&= CK(\frac{\sqrt{\pi}}{2} + 2\pi\int_0^{\sqrt{m}+\sqrt{n}} re^{-r^2}dr) && (90)\\
&= CK(\frac{\sqrt{\pi}}{2} + \pi(1-e^{(\sqrt{m}+\sqrt{n})^2})) && (91)\\
&\leq CK(\pi(2-e^{(\sqrt{m}+\sqrt{n})^2})) && (92)\\
&\leq CK(\sqrt{m}+\sqrt{n}) && (93)
\end{aligned}
$$

Where in the last step, we use the fact that $2 - e^{-x^2} \leq x, x \geq 2$.

### 1.17 4.4.7

Following the hints, we can let $x = (1, 0, ..., 0)$ and $y = (\frac{a_{11}}{\sqrt{a_{11}^2+...+a_{m1}^2}}, ..., \frac{a_{m1}}{\sqrt{a_{11}^2+...+a_{m1}^2}})$. Then

$$
\begin{aligned}
||A|| &\geq < Ax, y > && (94)\\
&= \sqrt{a_{11}^2 + ... + a_{m1}^2} && (95)
\end{aligned}
$$

So

$$
\begin{aligned}
E||A|| &\geq E\sqrt{a_{11}^2 + ... + a_{m1}^2} && (96)\\
&\geq \sqrt{n} - CK^2 && (97)
\end{aligned}
$$

Where in the last step we use the conclusion from exercise 3.1.4.

Similarly, we have

$$E||A|| \geq \sqrt{m} - CK^2 \tag{98}$$

Now

$$
\begin{aligned}
E||A|| &\geq \frac{1}{2}(\sqrt{m} + \sqrt{n} - CK^2) && (99)\\
&\geq \frac{1}{4}(\sqrt{m} + \sqrt{n}) && (100)
\end{aligned}
$$

Where in the last step, we use the inequality

$$\sqrt{m} + \sqrt{n} \geq CK^2 \tag{101}$$

As long as $m, n$ are large enough.

But pay attention to our formulation: here we assume that $K$ does not depend on $m$ or $n$.

### 1.18 4.5.2

Since the rows are actually two vectors: $(p, p, q, q)$ and $q, q, p, p$, so the rank is two(note that the book assumes $p \neq q$).

Following the definitions of eigen-vector and eigen-value, we can have

$$Dx = \lambda x \tag{102}$$

We can have

$$\begin{aligned} px_1 + px_2 + qx_3 + qx_4 &= \lambda x_1 && (103) \\ px_1 + px_2 + qx_3 + qx_4 &= \lambda x_2 && (104) \\ px_1 + px_2 + qx_3 + qx_4 &= \lambda x_3 && (105) \\ px_1 + px_2 + qx_3 + qx_4 &= \lambda x_4 && (106) \end{aligned}$$

Since we only need to focus on $\lambda \neq 0$, then we have

$$\begin{aligned} & x_1 = x_2 && (107) \\ & x_3 = x_4 && (108) \\ & p(x_1 + x_2) + q(x_3 + x_4) = \lambda x_1 && (109) \\ & q(x_1 + x_2) + p(x_3 + x_4) = \lambda x_3 && (110) \end{aligned}$$

So we can have

$$\begin{aligned} px_1 + qx_2 &= \frac{\lambda}{2} x_1 && (111) \\ qx_1 + px_2 &= \frac{\lambda}{2} x_2 && (112) \end{aligned}$$

Sum up we can have,

$$\lambda = 2(p - q) \tag{113}$$

If $x_1 = -x_3$ And

$$\lambda = 2(p + q) \tag{114}$$

If $x_1 = x_3$. Similarly, we can generalize it to $n$ dimension case where

$$\lambda = n(\frac{p-q}{2}), n(\frac{p+q}{2}) \tag{115}$$

## 1.19 4.5.4

The proof comes from link.

Note that

$$S = S + T - S \tag{116}$$

So we have(for $||v|| = 1$)

$$\begin{aligned} v^T S v &= v^T (T + S - T) v & (117) \\ &= v^T T v + v^T (S - T) v & (118) \\ &\leq v^T T v + ||S - T|| & (119) \end{aligned}$$

And

$$min_{||v||=1} u^T S u \leq v^T S v \tag{120}$$

So

$$max_{dim(U)=i} min_{u \in U, ||u||=1} u^T S u \leq max_{dim(V)=i} min_{v \in V, ||v||=1} v^T T v + ||S - T|| \tag{121}$$

Which is just

$$\lambda_i(S) \leq \lambda_i(T) + ||S - T|| \tag{122}$$

Similarly, we can prove

$$\lambda_i(T) \leq \lambda_i(S) + ||S - T|| \tag{123}$$

So the conclusion follows.

## 1.20 4.6.2

By using the hints, we can have

$$E||\frac{1}{m}A^TA - I_n|| = \int_0^\infty P(||\frac{1}{m}A^TA - I_n|| > t)dt \tag{124}$$

$$\tag{125}$$

Let $t = K^2 max(\sigma, \sigma^2)$, then we can have

$$\int_0^\infty P(||\frac{1}{m}A^TA - I_n|| > t)dt = \int P(||\frac{1}{m}A^TA - I_n|| > K^2\sigma)dt \tag{126}$$

$$+ \int P(||\frac{1}{m}A^TA - I_n|| > K^2\sigma^2)dt \tag{127}$$

$$= K^2 \int_0^1 P(||\frac{1}{m}A^TA - I_n|| > K^2\sigma)d\sigma \tag{128}$$

$$+ 2 * K^2 \int_1^\infty P(||\frac{1}{m}A^TA - I_n|| > K^2\sigma^2)\sigma d\sigma \tag{129}$$

$$\leq K^2 \int 2 * e^{-t^2} d\sigma + 2 * K^2 \int 2 * e^{-t^2} \sigma d\sigma \tag{130}$$

$$= \frac{CK^2}{\sqrt{m}} \int e^{-t^2} dt + \frac{CK^2}{\sqrt{m}} \int e^{-t^2} (\sqrt{\frac{n}{m}} + \frac{t}{\sqrt{m}})dt \tag{131}$$

$$= \frac{CK^2}{\sqrt{m}} \int_{-\sqrt{n}}^{\frac{\sqrt{m}}{C} - \sqrt{n}} e^{-t^2} dt \tag{132}$$

$$+ \frac{CK^2}{\sqrt{m}} \int_{\frac{\sqrt{m}}{C} - \sqrt{n}}^\infty e^{-t^2} (\sqrt{\frac{n}{m}} + \frac{t}{\sqrt{m}})dt \tag{133}$$

$$\leq \frac{CK^2}{\sqrt{m}} \int_{-\infty}^\infty e^{-t^2} dt + \frac{CnK^2}{m} \int_{-\infty}^\infty e^{-t^2} dt \tag{134}$$

$$+ \frac{CK^2}{m} \int_{\frac{\sqrt{m}}{C} - \sqrt{n}}^\infty e^{-t^2} t dt \tag{135}$$

$$= \frac{C\sqrt{\pi}K^2}{\sqrt{m}} + \frac{CnK^2}{m} + \frac{CK^2}{m}(\frac{m}{C^2} - 2\frac{\sqrt{mn}}{C} + n) \tag{136}$$

$$\leq \frac{K^2}{C} + \frac{(C\sqrt{\pi} - 2)K^2\sqrt{n}}{\sqrt{m}} + \frac{CK^2n}{m} \tag{137}$$

We only need to pick a large enough $C$, so that the last term will be upper bounded by

$$CK^2(\sqrt{\frac{n}{m}} + \frac{n}{m}) \tag{138}$$

## 1.21 4.6.3

In the same way as before, we have

$$\begin{aligned}
Es_1(A) &= \int_0^\infty P(s_1(A) > t)dt && (139)\\
&\leq CK^2 \int_{-\frac{\sqrt{m}}{CK^2}-\sqrt{n}}^\infty P(s_1(A) > \sqrt{m} + CK^2(t+\sqrt{n}))dt && (140)\\
&\leq CK^2(\frac{\sqrt{\pi}}{2} + \int_0^{\sqrt{n}+\frac{\sqrt{m}}{CK^2}} e^{-t^2}dt) && (141)\\
&\leq CK^2(\frac{\sqrt{\pi}}{2} + \sqrt{n} + \frac{\sqrt{m}}{CK^2}) && (142)\\
&\leq \sqrt{m} + CK^2(\sqrt{n} + \frac{\sqrt{\pi}}{2}) && (143)\\
&\leq \sqrt{m} + CK^2(\sqrt{n} + \sqrt{n}) && (144)\\
&= CK^2\sqrt{n} + \sqrt{m} && (145)
\end{aligned}$$

Similarly, we note that within small probability,

$$s_n(A) \leq \sqrt{m} - CK^2(\sqrt{n} + t) \tag{146}$$

So,

$$\begin{aligned}
E[\sqrt{m} - CK^2\sqrt{n} - s_n(A)] &= \int_0^\infty P(\sqrt{m} - CK^2\sqrt{n} - s_n(A) > t)dt && (147)\\
&= CK^2 \int_0^\infty P(\sqrt{m} - CK^2\sqrt{n} - s_n(A) > CK^2 t)dt && (148)\\
&\leq CK^2 \int_0^\infty e^{-t^2}dt && (149)\\
&= CK^2\frac{\sqrt{\pi}}{2} && (150)
\end{aligned}$$

So

$$Es_n(A) \geq \sqrt{m} - CK^2(\sqrt{n} + \frac{\sqrt{\pi}}{2}) \geq \sqrt{m} - 2CK^2\sqrt{n} \tag{151}$$

### 1.22 4.6.4(TODO)

## 1.23 4.7.3

We focus on equation (4.22) in the book. There, we set $t = \sqrt{u}$ so that

$$||\frac{1}{m}A^TA - I_n|| \leq K^2 max(\sigma, \sigma^2) \tag{152}$$

holds with probability at least $1 - 2e^{-u}$. Where

$$\sigma = C(\sqrt{\frac{n}{m}} + \sqrt{\frac{u}{m}}) \tag{153}$$

Note that

$$\sigma \leq 2 * C * \sqrt{\frac{n+u}{m}} \tag{154}$$

And

$$\sigma^2 = \frac{n + u + 2\sqrt{nu}}{m} \tag{155}$$

$$\leq \frac{2(n+u)}{m} \tag{156}$$

So

$$K^2 max(\sigma, \sigma^2) \leq CK^2(\frac{n+u}{m} + \sqrt{\frac{n+u}{m}}) \tag{157}$$

## 1.24 4.7.6

### 1.24.1

Note that

$$X = \theta\mu + g \tag{158}$$

Where $g \in N(0, I_n)$, $\theta$ is a symmetric Bernoulli random variable and $\theta, g$ are independent.

So

$$\begin{aligned}
\Sigma &= EXX^T && (159)\\
&= E(\theta\mu + g)(\theta\mu^T + g^T) && (160)\\
&= E[\mu\mu^T + \theta\mu g^T + \theta g\mu^T + gg^T] && (161)\\
&= E\mu\mu^T + Egg^T + E\theta(\mu g^T + g\mu^T) && (162)\\
&= \mu\mu^T + I_n && (163)
\end{aligned}$$

Clearly, if we restrict $x$ sit in $||x|| \leq 1$ then the eigenvector which corresponds to the largest eigenvalue should be like

$$x = (\frac{\mu_1}{||\mu||}, ..., \frac{\mu_n}{||\mu||}) \tag{164}$$

Which is parallel to $\mu$.

### 1.24.2

Now we go back to equation (4.22) and let $t = \sqrt{n}$ there. So with probability at least $1 - 2e^{-n}$, we can have

$$||\Sigma_m - \Sigma|| \leq C\sqrt{\frac{n}{m}}||\Sigma|| \tag{165}$$

Let $m \geq (\frac{n}{||\mu||_2})$, we can have

$$||\Sigma_m - \Sigma|| \leq C||\mu||_2^{\frac{1}{2}}||\Sigma|| = C\frac{||\mu||_2^{\frac{5}{2}}}{\sqrt{n}} \tag{166}$$

So as long as $m$ is large enough, the sample covariance matrix will be close to $\Sigma$. Now we need to calculate the spectral gap of $\mu\mu^T + I_n$. Note that the eigenvalues of $\mu\mu^T$ consist of one $||\mu||_2^2$ and $n-1$ zeros. So the eigenvalues of $\mu\mu^T + I_n$ consist of one $||\mu||_2^2 + 1$ and $n-1$ ones. So the spectral gap is $||\mu||_2^2$.

### 1.24.3

Now using Davis-Kahan inequality, we have

$$||v_1(\Sigma_m) - \theta v_1(\Sigma)||_2 \leq \frac{2^{\frac{3}{2}}||\Sigma_m - \Sigma||}{||\mu||_2^2} \leq C * 2^{\frac{3}{2}}||\mu||_2^{\frac{1}{2}} \tag{167}$$

$$sin\angle(v_1(\Sigma_m), v_1(\Sigma)) \leq 2C||\mu||_2^{\frac{1}{2}} \tag{168}$$

Where $\theta \in \{-1, 1\}$. Since $v_1(\Sigma) = \frac{\mu}{||\mu||_2}$, we can further have

$$||v_1(\Sigma_m) - \theta\frac{\mu}{||\mu||_2}||_2 \leq C * 2^{\frac{3}{2}}||\mu||_2^{\frac{1}{2}} \tag{169}$$

So

$$||||\mu||_2 v_1(\Sigma_m) - \theta\mu||_2 \leq C * 2^{\frac{3}{2}}||\mu||_2^{\frac{3}{2}} \tag{170}$$

From these two inequalities, we can see that $v_1(\Sigma_m)$ is actually very close to the direction of $\mu$.

### 1.24.4

Note that

$$< \mu, X_i >= \theta_i||\mu||_2^2 + < \mu, g > \tag{171}$$

Without loss of generality, we assume $\theta = 1$. And note that, $< \mu, g > \in N(0, ||\mu||_2^2)$. So in order to make a good prediction, we should let

$$| < \mu, g > | \leq ||\mu||_2^2 \tag{172}$$

But on the other hand, equation (2.14) and example 2.5.8 in the book(for sub-gaussian random variables) tells us

$$P(|X| \geq t) \leq 2e^{-\frac{ct^2}{||X||_{\phi_2}^2}} \leq 2e^{-\frac{ct^2}{||\mu||_2^2}} \tag{173}$$

So with high probability, we can get

$$| < \mu, g > | \leq ||\mu||_2^2 \tag{174}$$

So that is why the signs of $< \mu, X_i >$ work well. More formally, let $t = ||\mu||_2^2$:

$$P(|X| > ||\mu||_2^2) \leq 2e^{-c||\mu||_2^2} \leq 2\epsilon^c \tag{175}$$

Where we use the condition $||\mu||_2 \geq \sqrt{log(\frac{1}{\epsilon})}$. By applying a union bound, we can see that we will make mistakes at most

$$m\epsilon^c \tag{176}$$

Where $c$ here is an absolute number. Actually, by using $P(|X| > c_0||\mu||_2^2)$ for some absolute $c_0$ we can have at most $\epsilon m$ mis-classified points.

#### 1.24.5

Firstly, note that the sign of $< v_1(\Sigma_m), X_i >$ is the same as that of $< ||\mu||_2 * v_1(\Sigma_m), X_i >$. And recall the relationship reflected in inequality (170):

$$
\begin{aligned}
| < ||\mu||_2 * v_1(\Sigma_m), X_i > | &= | < ||\mu||_2 * v_1(\Sigma_m) - \mu, X_i > + < \mu, X_i > | && (177) \\
&\leq | < ||\mu||_2 * v_1(\Sigma_m) - \mu, X_i > | + | < \mu, X_i > | && (178)
\end{aligned}
$$

As a result of this, we can have

$$
\begin{aligned}
& P(| < ||\mu||_2 * v_1(\Sigma_m), X_i > | \geq ||\mu||_2^2) && (179) \\
& \leq P(| < ||\mu||_2 * v_1(\Sigma_m) - \mu, X_i > | \geq \frac{||\mu||_2^2}{2}) + P(| < \mu, X_i > | \geq \frac{||\mu||_2^2}{2}) && (180) \\
& \leq e^{-C||\mu||_2} + \epsilon^c && (181) \\
& = \epsilon^{\frac{C}{\sqrt{log(\frac{1}{\epsilon})}}} + \epsilon^c && (182)
\end{aligned}
$$

Where in the second to last step, we use the inequality (2.14) from the book and inequality (170). As a result of this, the union bound will imply with probability at least $1 - e^{-n}$, the error we make will be at most

$$
m\epsilon^c + m\epsilon^{\frac{C}{\sqrt{log(\frac{1}{\epsilon})}}} \tag{183}
$$

Where $C$ and $c$ are absolute constants.