Solution 3

Euler Cat

July 22, 2023

1 Solutions to exercises in chapter 3

1.1 3.1.4

1.1.1

The proof comes from link.

Note that $||X||_1 \leq ||X||_{\phi_2}$ since $Ee^{X^2/||X||_1^2} \geq 1 + \frac{EX^2}{||X||_1^2} \geq 2$ where we use the inequality $e^x \geq 1 + x$ and Cauchy-Schwarz inequality.

As a result of this, we can have

$$\begin{aligned} |EX - \sqrt{n}| &\leq E|X - \sqrt{n}| && (1) \\ &\leq ||X - \sqrt{n}||_{\phi_2} && (2) \\ &\leq C * K^2 && (3) \end{aligned}$$

1.1.2

This part of proof is adapted from link.

By calculus, we can show the following numerical inequality:

$$\frac{1 + u - (u-1)^2}{2} \leq \sqrt{u} \leq \frac{u+1}{2} \tag{4}$$

For $\forall u \geq 0$. So we can just let $u = \frac{1}{n}\sum_{i=1}^n X_i^2$ and take the expectation:

$$\frac{1}{\sqrt{n}} E||X||_2 \leq 1 \tag{5}$$

And

$$\frac{1}{\sqrt{n}} E||X||_2 \geq \frac{2 - max_i \frac{Var(X_i^2)}{n}}{2} \tag{6}$$

So I think in general, unless $Var(X_i^2)$ is bounded for $\forall i$ or X_i is iid we can not replace $C * K^2$ with $o(1)$ since $Var(X_i^2)$ may evolve with n.

1.2 3.1.5

Following the hints,

$$
\begin{aligned}
E(||X||_2^2 - n)^2 &= E(\sum_{i=1}^{n} X_i^2 - n)^2 && (7)\\
&= E(\sum_{i=1}^{n} X_i^2)^2 - 2*n*\sum_{i=1}^{n} EX_i^2 + n^2 && (8)\\
&= E(\sum_{i=1}^{n} X_i^2)^2 - n^2 && (9)\\
&= \sum_{i=1}^{n} EX_i^4 + 2*\frac{n*(n-1)}{2}*1 - n^2 \quad \leq n*max_i||X_i||_4^4 - n && (10)\\
&\leq n*C*2*K^4 && (11)
\end{aligned}
$$

Where in the last step we use the inequality $||X||_{L^p} \leq C*||X||_{\phi_2}*\sqrt{p}$ from part 2.5.2 of book.

Next, note that

$$E(||X||_2^2 - n)^2 = E(||X|| - \sqrt{n})^2(||x|| + \sqrt{n})^2 \leq C*K^4*n \tag{12}$$

So

$$E(||X|| - \sqrt{n})^2 \leq E(||X|| - \sqrt{n})^2(\frac{||X||}{\sqrt{n}} + 1)^2 \leq C*K^4 \tag{13}$$

Finally,

$$
\begin{aligned}
Var(||X||_2) &= E(||X||_2 - E||X||_2)^2 && (14)\\
&= E(||X||_2 - \sqrt{n} + \sqrt{n} - E||X||_2)^2 && (15)\\
&= E(||X||_2 - \sqrt{n})^2 - 2*E(||X||_2 - \sqrt{n})(\sqrt{n} - E||X||_2) + (\sqrt{n} - E||X||_2)^2 && (16)\\
&= E(||X||_2 - \sqrt{n})^2 - (\sqrt{n} - E||X||_2)^2 && (17)\\
&\leq E(||X||_2 - \sqrt{n})^2 && (18)\\
&\leq C*K^4 && (19)
\end{aligned}
$$

1.3 3.1.6

In equation (10), we just use the inequality $EX_i^4 \leq C * K^4$ and we can get the same result as before.

1.4 3.1.7

Note that

$$P(||X||_2 \leq \epsilon * \sqrt{n}) = P(||X||_2^2 \leq \epsilon^2 * n) \tag{20}$$

$$= P(\sum_{i=1}^{n} X_i^2 \leq \epsilon^2 * n) \tag{21}$$

$$= P(-\sum_{i=1}^{n} t * X_i^2 \geq -t * \epsilon^2 * n) \tag{22}$$

$$\leq e^{t*\epsilon^2*n} * \prod_{i=1}^{n} Ee^{-t*X_i^2} \tag{23}$$

Also note that

$$Ee^{-t*X^2} = \int_{-\infty}^{\infty} p(x) * e^{-t*x^2} dx \tag{24}$$

$$= \frac{1}{\sqrt{t}} * \int_{-\infty}^{\infty} p(x) * e^{-x^2} dx \tag{25}$$

$$\leq \frac{\sqrt{\pi}}{\sqrt{t}} \tag{26}$$

As a result of this, we know that

$$(23) \leq e^{t*\epsilon^2*n} * (\frac{\pi}{\sqrt{t}})^{\frac{n}{2}} \tag{27}$$

$$= e^{t*\epsilon^2*n+\frac{n}{2}*ln(\frac{\pi}{\pi})} \tag{28}$$

$$= e^{n*(t*\epsilon^2+\frac{1}{2}*ln(\frac{\pi}{t}))} \tag{29}$$

Let $t = \frac{1}{2*\epsilon^2}$, then we have

$$P(||X||_2 \leq \epsilon * \sqrt{n}) \leq e^{n*(ln\epsilon+ln\sqrt{2*\pi*e}} = (C * \epsilon)^n \tag{30}$$

1.5 3.2.2

1.5.1

$$EX = \mu + \Sigma^{\frac{1}{2}} EZ = \mu \tag{31}$$

$$cov(X) = E[(X-\mu)(X-\mu)^T] = \sigma E[ZZ^T] = \Sigma \tag{32}$$

1.5.2

$$EZ = \Sigma^{-\frac{1}{2}}(EX - \mu) = 0 \tag{33}$$

$$cov(Z) = cov(\Sigma^{-\frac{1}{2}}(X-\mu)) = \Sigma^{-1} cov(X-\mu) = \Sigma^{-1} * \Sigma = I \tag{34}$$

1.6 3.2.6

Note that

$$
\begin{aligned}
E||X-Y||_2^2 &= E[< X-Y, X^T - Y^T >] && (35)\\
&= E[< X, X^T >] - 2E[< X, Y^T >] + E[< Y, Y^T >] && (36)\\
&= E[XX^T] + E[YY^T] && (37)\\
&= 2 * n && (38)
\end{aligned}
$$

Where in (36) we use independence between X and Y while in the last step we use Lemma 3.2.4.

1.7 3.3.1

Note that

$$n * EX_i^2 = \sum_i EX_i^2 = n \tag{39}$$

$$\tag{40}$$

So the diagonal line of EXX^T consists of 1. For $i \neq j$, we have $E[X_i X_j] = 0$ since the sphere is a symmetric object. Since we need to make sure $\sum_i X_i^2 = n$, the random variables X_i are not independent.

1.8 3.3.3

1.8.1

$$< g, u >= \sum_{i=1}^{n} g_i * u_i \tag{41}$$

Since g_i corresponds to $N(0,1)$ we have $N(0, u_i^2)$. Gathering together, we have $N(0, ||u||_2^2)$.

1.8.2

This directly follows from exercise 3.3.3 by noting that $u = (\sigma_1, \sigma_2), g = (\frac{X}{\sigma_1}, \frac{Y}{\sigma_2})$

1.8.3

Let $g = Gu$, then

$$E(G_{i1}u_1 + ... + G_{in} * u_n)^2 \tag{42}$$

$$= \sum_{j=1}^{n} E[G_{ij}^2 u_j^2] \tag{43}$$

$$= E[G^2] \sum_{j=1}^{n} u_j^2 \tag{44}$$

$$= E[G^2] \qquad = 1 \tag{45}$$

Where in (43) we use the fact that $E[G_{im} * G_{in}] = E[G_{im}] * E[G_{in}] = 0, \forall m \neq n$ and in last step we use the fact $Var(X) = E(X - EX)^2 = EX^2 - (EX)^2$.

1.9 3.3.4

If X has a multivariate normal distribution, then from exercise 3.3.3 we have $< X, \theta >$ has a normal distribution.

On the other hand, we we first let $\theta_n = (..., 1, ...)$ where the element on nth slot is 1. Then we can have(from assumption) $X_i =< X, \theta_n >$ has a normal distribution. So X_i has mean and variance denoted by μ_i and σ_i respectively.

Now we calculate the covariance $cov_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$.

Finally, we construct a vector $\mu = (\mu_1, ..., \mu_n)$ and a matrix $\Sigma \in R^{n\times n}$ where $\Sigma_{ii} = \sigma_i$ and $\Sigma_{ij} = cov_{ij}, \forall i \neq j$.

Based on μ and Σ, we build a multivariate normal distribution, denoted by Y. Clearly, $< Y, \theta >=< X, \theta >, \forall \theta \in R^n$. So by Cramer-Wold theorem(link) we can have our conclusion.

1.10 3.3.5

1.10.1

$$< X, u >= \sum X_i * u_i \tag{46}$$

$$< X, v >= \sum x_j * v_j \tag{47}$$

So

$$< X, u >< X, v >= \sum_{i,j} X_i * X_j * u_i * v_j \tag{48}$$

Note that $E[X_i * X_j] = 0, \forall i \neq j$. So we only have

$$E[< X, u >< X, v >] = \sum E[X_i^2] * u_i * v_i =< u, v > \tag{49}$$

1.10.2

Note that $X_u - X_v =< X, u - v >$. So $Y = X_u - X_v$ has a normal distribution $N(0, ||u - v||_2^2)$.

Now, note that $||X_u - X_v||^2 = ||Y||^2 = \frac{1}{\sqrt{2 * \pi}} * \int y^2 * e^{-\frac{y^2}{2*||u-v||_2^2}} dy = ||u - v||_2$.

1.11 3.3.6

From exercise 3.3.3, we know that Gu and Gv are both normal distributions $N(0, I_m)$. Now we only need to prove the independence of these two distributions.

Since they are both normal distributions, uncorrelated implies independence. So We only need to check that $E[(Gu)_i(Gv)_j] = 0, \forall 1 \leq i, j \leq m$.

$$(Gu)_i = \sum_p G_{ip} u_p \tag{50}$$

$$(Gv)_j = \sum_q G_{jq} v_q \tag{51}$$

As a result of this, we have

$$(Gu)_i * (Gv)_j = \sum_{p,q} G_{ip} * G_{jq} * u_p * v_q \tag{52}$$

If $i \neq j$, then $E[G_{ip} * G_{jq}] = 0$. If $i = j$, then if $p \neq q$ we have $E[G_{ip} * G_{jq}] = 0$. So we will only have the terms like $G_{ip}^2 u_p * v_p$.

Note that

$$E[\sum_p G_{ip}^2 * u_p * v_p] = \sum_p u_p * v_P = 0 \tag{53}$$

Which means they are uncorrelated. And for gaussian distribution, uncorrelated implies independence.

1.12 3.3.7

Proof comes from link.

1.12.1

Note that

$$f(x) = \frac{1}{(2*\pi)^{\frac{n}{2}}} * \int_{R^n} exp^{-\frac{||x||_2^2}{2}} \tag{54}$$

$$= \int_0^{\infty} \int_{S^{n-1}} f(r,\theta) * r^{n-1} dr d\sigma(\theta) \tag{55}$$

$$= \frac{\omega_{n-1}}{(2*\pi)^{\frac{n}{2}}} \int_0^{\infty} exp^{-\frac{r^2}{2}} dr * \frac{1}{\omega_{n-1}} * \int_{S^{n-1}} d\sigma(\theta) \tag{56}$$

$$= f(r) * g(\theta) \tag{57}$$

So r and θ areactually independent.

1.12.2

From the above formulation, we can see that θ obeys a uniform distribution on the sphere.

1.13 3.3.9

Similar to the proof of lemma 3.2.3,

$$Axx^T = \sum_{i=1}^{N} | < u_i, x > |^2 = x \sum_{i=1}^{N} u_i u_i^T x^T \tag{58}$$

And this holds for any x, so the conclusion follows.

1.14 3.4.3

1.14.1

$$||X||_{\phi_2} = sup_{x \in S^{n-1}} || < X, x > ||_{\phi_2} \tag{59}$$
$$= sup_{x \in S^{n-1}} || \sum X_i * x_i ||_{\phi_2} \tag{60}$$
$$\leq sup_{x \in S^{n-1}} \sum |x_i| * ||X_i||_{\phi_2} \tag{61}$$
$$\leq (\sum ||X_i||^2_{\phi_2})^{\frac{1}{2}} \tag{62}$$
$$< \infty \tag{63}$$

Where in the last step, we use Cauchy-Schwarz inequality.

1.14.2 TODO

Let (67) = 2, then we can have $t = \sqrt{\frac{8}{3}}$.

Now assume that we have N random variables $X_1, ..., X_n$ such that all of them obey $N(0, 1)$. Let $x = (\frac{1}{\sqrt{N}}, ..., \frac{1}{\sqrt{N}})$. Then, we have

$$||X||_{\phi_2} \geq || < X, x > ||_{\phi_2} = ||\sqrt{N} * X_i||_{\phi_2} = \sqrt{\frac{8N}{3}} \tag{64}$$

Which approaches infinity when N goes into infinity.

1.15 3.4.4

Note that for $X = (X_1, ..., X_n)$, X_i is a random variable that takes value 0 and $\sqrt{n}$ with probability $\frac{n-1}{n}$ and $\frac{1}{n}$ respectively. Then,

$$Ee^{\frac{X^2}{t^2}} = \frac{n-1}{n} + \frac{1}{n} * e^{\frac{n}{t^2}} \tag{65}$$

Let (69) = 2, then we have

$$\frac{n}{t^2} = log(n+1) \tag{66}$$

$$t = \sqrt{\frac{n}{log(n+1)}} \tag{67}$$

Now, let $x = (1, 0, 0, 0, ..., 0)$, then we have

$$||X||_{\phi_2} \geq || < X, x > || = \sqrt{\frac{n}{log(n+1)}} \asymp \sqrt{\frac{n}{logn}} \tag{68}$$

For the upper bound, proof comes from link.

$$2 = Ee^{\frac{(<X,x>)^2}{||<X,x>||^2}} \tag{69}$$

$$= \frac{1}{n} \sum_{i=1}^{n} e^{\frac{nx_i^2}{||<X,x>||^2}} \tag{70}$$

$$\leq \frac{1}{n} * (n - 1 + e^{\frac{n}{||<X,x>||^2}}) \tag{71}$$

From which we can get

$$|| < X, x > || \leq \sqrt{\frac{n}{log(n+1)}} \asymp \sqrt{\frac{n}{logn}} \tag{72}$$

Note that in equation (75) we use the fact that $\sum e^{x_i}$ is a convex function and $\sum |x_i| \leq 1$ is a convex simplex. So the convex function must achieve its maximum at some corner point, which is one of the basis vectors.

1.16 3.4.5

The proof comes from a combination between link and link.

Firstly, I will give a high level idea of the whole proof. Since X is an isotropic random vector, then we have $n = E||X||_2^2$. Next, by using the upper bound for maximum of sub-gaussian random variables we can get something like $n \leq C * \sqrt{n} * \sqrt{logn}$ where C will be analyzed in detail.

Firstly, note that $X = (X_i)_{1 \leq i \leq n}$ is an isotropic random vector such that $||X||_{\phi_2} = O(1)$ then $||X_i||_{\phi_2} \leq C, \forall i$.

By exercise 2.5.10, we have

$$Emax_{i \leq n}|X_i| \leq C * C' \sqrt{logn} \tag{73}$$

Secondly, we let A be a number large enough such that the support T is contained in the ball $B(0, A * \sqrt{n})$. As a result of this,

$$\begin{aligned} n = E||X||_2^2 = E| < X, X > | & \quad (74) \\ \leq Esup_{x \in S}| < X, x > | & \quad (75) \\ = A * \sqrt{n} * Esup_{x \in S}| < X, \frac{x}{A * \sqrt{n}} > | & \quad (76) \\ \leq A * \sqrt{n} * C * C' * \sqrt{log|S|} & \quad (77) \end{aligned}$$

So

$$|S| \geq e^{\frac{n}{C * C' * A}} \tag{78}$$

In the third part, we will prove that the constant A is actually a permanent constant controlled by C, so our conclusion follows.

Since $||X||_{\phi_2} = O(1)$ then we have $|| < X, \theta > ||_{\phi_2} \leq C$ for every unit vector θ. By using 2.15 in section 2.5.2, we can know that

$$E < X, \theta >^4 \leq 2 * C * || < X, \theta > ||_{\phi_2}^4 \leq 2 * C^2 \leq 4 * C^2 \tag{79}$$

Now we create a new random variable Y defined by

$$Y = X * 1_{\{||X|| \leq 4 * C * \sqrt{n}\}} \tag{80}$$

Then we can have

$$\begin{aligned} \theta^T E Y Y^T \theta = E < Y, \theta >^2 & \quad (81) \\ = E[< X, \theta >^2 * 1_{\{||X|| \leq 4 * C * \sqrt{n}\}}] & \quad (82) \\ \preceq E < X, \theta >^2 & \quad (83) \\ = \theta^T E X X^T \theta & \quad (84) \\ = \theta^T * \theta & \quad (85) \end{aligned}$$

Where in the last step, we use the fact that X is an isotropic random vector. And $A \preceq B$ if and only if $A - B$ is a semi-positive definite matrix.

On the other hand, we have

$$\begin{aligned}
& E[< X, \theta >^2 * 1_{\{||X|| > 4*C*\sqrt{n}\}}] && (86)\\
& \preceq (E[< X, \theta >^4] * P(||X||^2 > 16 * C^2 * n))^{\frac{1}{2}} && (87)\\
& \leq \frac{4 * C^2}{16 * C^2} && (88)\\
& = \frac{1}{2} && (89)
\end{aligned}$$

As a result of this, we have

$$\begin{aligned}
\theta^T E[YY^T]\theta &= E < Y, \theta >^2 && (90)\\
&= E < X, \theta >^2 - E[< X, \theta >^2 * 1_{\{||X|| > 4*C*\sqrt{n}\}}] && (91)\\
&\succeq \frac{1}{2} * E < X, \theta >^2 = \frac{1}{2} * I_n && (92)
\end{aligned}$$

So

$$E[YY^T] \succeq \frac{1}{2} * I_n \tag{93}$$

Combined together, we have

$$\frac{1}{2} * I_n \preceq E[YY^T] \preceq I_n \tag{94}$$

In other words, $E[YY^T]$ can be inverse.

Finally, we make a replacement as follows:

$$\tilde{Y} = \frac{Y}{\sqrt{E[YY^T]}} \tag{95}$$

Note that the support of $\tilde{Y}$ is the same as Y. Moreover, we can prove that $\tilde{Y}$ is sub-gaussian as follows(let $T = \frac{1}{\sqrt{E[YY^T]}}$):

$$\begin{aligned}
||\tilde{Y}||_{\phi_2} &= sup_\theta || < \tilde{Y}, \theta > ||_{\phi_2} && (96)\\
&= sup_\theta || < Y, T * \theta > ||_{\phi_2} && (97)\\
&= ||T * \theta||_{\phi_2} * sup_\theta || < Y, \frac{T * \theta}{||T * \theta||_{\phi_2}} > ||_{\phi_2} && (98)\\
&\leq C * ||T * \theta||_{\phi_2} && (99)\\
&\leq C * \sqrt{2} && (100)
\end{aligned}$$

Where in (99) we use the fact that Y is a sub-gaussian random vector and in (100) we use (94).

Now we come to the final part. Note two things:

- Note that $\tilde{Y}$ is supported in the ball $B(0, 8 * C * \sqrt{n})$.
- The support of $\tilde{Y}$ is contained in T.
- Apply (78) to $\tilde{Y}$ can know that T contains at least $e^{\frac{n}{8*C^2*C'}}$ points.

Which is to say

$$|T| \geq e^{c*n} \tag{101}$$

This beautiful proof is due to Guillaume Aubrun and Iosif Pinelis.

1.17 3.4.7

The proof comes from link.

Since we are now in a ball, so in order to make connection between ball and sphere, we can first pick a point on the sphere then we select a point along the radius. More formally, we define the random variable

$$Y = U^{\frac{1}{n}} X \tag{102}$$

Where X is a random variable corresponds to uniform distribution on the sphere and U is a random variable corresponds to uniform distribution on interval $[0, 1]$. The $\frac{1}{n}$ appears since we need to make sure the probability(which is an integral for calculating the volume of $B(0, \sqrt{n})$) over whole ball is normalized into 1.

Now, note that(by rotational invariance)

$$\begin{aligned} P(|Y_1| \geq t) &= P(|U^{\frac{1}{n}} X_1| \geq t) && (103) \\ &\leq P(|X_1| \geq t) && (104) \end{aligned}$$

Where in the last step we use the fact that $U^{\frac{1}{n}} \leq 1$. Finally, applying the conclusion of theorem 3.4.6 will suffice.

1.18 3.4.9

Proof comes from link. The element of this proof lies in that we calculate the exact form of EX_1^2 and discuss when it will meet the requirements of isotropic random vector. To achieve this, we need to firstly calculate the density function of X_1.

$$P(X_1 \leq x) = \frac{Vol(S)}{Vol(B^n(0,r))} \tag{105}$$

Where Vol denotes the volume of geometry object under L_1 norm and S here denotes the area such that $X_1 \leq x$.

Note that if $X_1 = t$, then $||X||_1 \leq r$ is equivalent to $X_2 + ... + X_n \leq r - t$ which is a $n-1$ dimension ball, so multiply it by dt will give us some tiny volume elements, integrating them can have

$$Vol(S) = \int_{-r}^{x} Vol(B^{n-1}(r-|t|))dt \tag{106}$$

$$= (r-|t|)^{n-1} \int_{-r}^{x} Vol(B^{n-1}(r-|t|))dt \tag{107}$$

So that

$$P(X_1 \leq x) = \int_{-r}^{x} (r-|t|)^{n-1} \frac{Vol(B^{n-1}(0,1))}{Vol(B^n(0,r))} dt \tag{108}$$

Let $x = r$, then we have

$$1 = P(X_1 \leq r) \tag{109}$$

$$= \int_{-r}^{r} (r-|t|)^{n-1} dt * \frac{Vol(B^{n-1}(0,1))}{Vol(B^n(0,r))} \tag{110}$$

And from

$$\int_{-r}^{r} (r-|t|)^{n-1} dt = 2 * \int_{0}^{r} (r-t)^{n-1} dt \tag{111}$$

$$= \frac{2}{n} * r^n \tag{112}$$

we can know

$$\frac{Vol(B^{n-1}(0,1))}{Vol(B^n(0,r))} = \frac{n}{2 * r^n} \tag{113}$$

So the density function is $\frac{n}{2*r^n} * (r - |t|)^{n-1}$. Thus we can calculate the expectation of X^2 as follows:

$$EX^2 = \frac{n}{2*r^n} \int_{-r}^{r} x^2 (r - |x|)^{n-1} dx \tag{114}$$

$$= \frac{n}{r^n} \int_{0}^{r} x^2 * (r - x)^{n-1} dx \tag{115}$$

$$= n * r^2 \int_{0}^{1} x^2 (1 - x)^{n-1} dx \tag{116}$$

$$= n * r^2 * \frac{2}{n * (n+1) * (n+2)} \tag{117}$$

$$= \frac{r^2}{2 * (n+1) * (n+2)} \tag{118}$$

So in order to let X be sub-gaussian, we must have $r \asymp n$.

For the second question, note that the intensity function is of the form $(r - |t|)^{n-1} = r^{n-1}(1 - \frac{|t|}{r})^{n-1} \asymp r^{n-1} * e^{-C*|t|}$. And note that when n grows, r^{n-1} will grow as well, so now constant can be used for bounding this sub-gaussian norm.

1.19 3.4.10

The counterexample comes from link.

1.20 3.5.2

1.20.1

If

$$|\sum_{i,j} a_{ij}x_iy_j| \leq max_i|x_i|max_j|y_j| \tag{119}$$

for any real numbers x_i and y_j. Then we just need to restrict the values of x_i and y_j within the set $-1, 1$.

On the other hand, if

$$|\sum_{i,j} a_{ij}x_iy_j| \leq 1 \tag{120}$$

For any x_i and y_j in $-1, 1$.

Then without loss of generality, we assume that $max|a_i| \leq 1, max|b_i| \leq 1$. Now we enlarge a_i or b_j in the following way: if $a_{ij} \leq 0$ then set $a'_i * b'_j = -1$, otherwise set $a'_i * b'_j = 1$.

clearly, this would give us an upper bound

$$|\sum_{ij} a_{ij}a_ib_j| \leq |\sum_{ij} a_{ij}a'_ib'_j| \tag{121}$$

$$\leq 1 \tag{122}$$

Where in the last step, we use the inequality (120).

1.20.2

If we have

$$|\sum_{ij} a_{ij} < u_i, v_j > | \leq K * max_i||u_i|| * max_j||v_j|| \tag{123}$$

Then just set $||u_i|| = 1, ||v_j|| = 1$ can help us get the conclusion.

On the other hand, firstly note that since H is Hilbert space, then we have the Cauchy-Schwarz inequality

$$< u_i, v_j >\leq ||u_i||_2 * ||v_j||_2 \tag{124}$$

Without loss of generality, we assume that $max||u_i|| \leq 1, max||v_j|| \leq 1$. Then from (124) we can know that $< u_i, v_j >\leq 1$. As a result of this, we can create another two vectors u'_i, v'_j such that $< u'_i, v'_j >= 1$ if $a_{ij} \geq 0$ and $< u'_i, v'_j >= -1$ if $a_{ij} < 0$ while keeping $||u'_i|| = ||v'_j|| = 1$.

Clearly,

$$|\sum a_{ij} < u_i, v_j > | \leq |\sum a_{ij} < u'_i, v'_j > | \leq K \tag{125}$$

1.21 3.5.3

Note: In a new edition of this book link, the author puts another two constraints on the matrix A: it should be either positive-semidefinite or has zero diagonal.

For any vectors x and y where each of the elements takes value in $-1, 1$, we use another two variables $u = \frac{x-y}{2}$ and $v = \frac{x+y}{2}$.

So we can have $< Ax, y >=< Au, u > - < Av, v >$.

Note that we can not apply Grothendieck's inequality directly since u or v may contain some zero elements.

So we need to prove two lemmas borrowed from link:

Lemma 1.1. *Suppose that for any numbers $x_i \in \{-1, 1\}$ we have*

$$|\sum_{i,j} a_{ij}x_ix_j| \leq 1 \tag{126}$$

Then for every set $I \subset \{1, 2, 3, ..., n\}$, we will get

$$-1 \leq \sum_i a_{ii} + \sum_{i,j \in I, i \neq j} a_{ij}x_ix_j \leq 1 \tag{127}$$

Proof. Outside the set I, we use -1 or 1 to fill up the remained $n - |I|$ entries. So have $M = 2^{n-|I|}$ possible outcomes. Note that

$$\sum_{i,j} a_{ij}x_ix_j = \sum a_{ii} + \sum_{1 \neq j} a_{ij}x_ix_j \tag{128}$$

$$= \sum a_{ii} + \sum_{i,j \in I} a_{ij}x_ix_j + \sum_{i,j \notin I} a_{ij}x_ix_j \tag{129}$$

Also note that M here is an even number, so for each $\sum_{i,j \notin I} a_{ij}x_ix_j$ we will have another $-\sum_{i,j \notin I} a_{ij}x_ix_j$. So sum up all M inequalities, we will have

$$-M \leq M * \sum_i a_{ii} + M * \sum_{i,j \in I, i \neq j} a_{ij}x_ix_j \leq M \tag{130}$$

□

Lemma 1.2. *Suppose A is either PSD or has zero diagonal and*

$$| < Ax, x > | \leq 1, \forall x \in \{-1, 1\}^n \tag{131}$$

We can have

$$| < Ax, x > | \leq 1, \forall x \in \{-1, 0, 1\}^n \tag{132}$$

Proof. If A has zero diagonal, then from equation (127), we have for any set I,

$$-1 \leq \sum_{i,j \in I, i \neq j} a_{ij} x_i x_j \leq 1 \tag{133}$$

Now, we only need to take I as the support set of x to get the conclusion.
On the other hand, if A is a PSD matrix, then $0 \leq < Ax, x >$ trivially holds for any vector x. For another direction, note that we still let I as the support of vector x. Then

$$\begin{align}
< Ax, x > &= \sum a_{ij} x_i x_j \tag{134}\\
&= \sum_i a_{ii} x_i^2 + \sum_{i \neq j} a_{ij} x_i x_j \tag{135}\\
&\leq \sum_i a_{ii} + \sum_{j \in I, i \neq j} a_{ij} x_i x_j \tag{136}\\
&\leq 1 \tag{137}
\end{align}$$

Where in the last step, we use (127) again. □

By using lemma 2, we know(note that $u, v \in 1, 0, -1^n$)

$$\begin{align}
| < \frac{Ax}{2}, y > | &= | < \frac{Au}{2}, u > - < \frac{Av}{2}, v > | \tag{138}\\
&\leq 1 \tag{139}
\end{align}$$

So apply Grothendieck's inequality will get the result.

1.22 3.5.5

For one direction, we consider the Gram matrix G of vectors X_i and X_j:

$$G_{ij} =< X_i, X_j > \tag{140}$$

Clearly, G is a symmetric matrix and $G_{ii} = 1$ due to X_i is a normalized vector. Moreover, G is PSD:

$$\begin{aligned}
a^T G a &= \sum_{ij} G_{ij} a_i a_j && (141)\\
&= \sum_{ij} < X_i, X_j > a_i a_j && (142)\\
&=< \sum_i a_i * X_i, \sum_j a_j * X_j > && (143)\\
&\geq 0 && (144)
\end{aligned}$$

So

$$0 \preceq X \tag{145}$$

On the other hand, note that X is positive semi-definite, then we can know that X can be decomposed into

$$X = VV^T \tag{146}$$

We use v_i to denote the row vector of V, so

$$X_{ij} = v_i * v_j^T \tag{147}$$

Note that $< v_i, v_i >= X_{ii} = 1$. So the $v_i, \forall i$ is what we want.

Based on the above analysis, we know these two problems are actually equivalent.

1.23 3.5.7

Let

$$\tilde{A} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \tag{148}$$

And

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix} \tag{149}$$

Where X and Y are matrices with rows X_i^T and Y_i^T respectively.

Note that

$$\tilde{A}ZZ^T = \begin{bmatrix} AY^TX & AYY^T \\ A^TXX^T & A^TXY^T \end{bmatrix} \tag{150}$$

So $tr(\tilde{A}ZZ^T) = tr(AY^TX) + tr(A^TXY^T) = 2 * \sum_{ij} A_{ij} < X_i, Y_j >$.

As a result of this, we can re-formalize the objective function

$$\sum_{ij} A_{ij} < X_i, X_j > = \frac{1}{2} * tr(\tilde{A}ZZ^T) \tag{151}$$

$$= \frac{1}{2} < A, Y > \tag{152}$$

Where $Y = ZZ^T$. Note that

$$x^T ZZ^T x = x^T Z(x^T Z)^T = ||x^T Z||_2^2 \geq 0 \tag{153}$$

So $Y = ZZ^T$ is a PSD matrice. Also note that

$$ZZ^T = \begin{bmatrix} XX^T & XY^T \\ YX^T & YY^T \end{bmatrix} \tag{154}$$

So the diagonal line of ZZ^T consists of $< X_i, X_i >, < Y_j, Y_j >, i = 1, ..., m, j = 1, ..., n$. Which are all 1 elements.

Based on the above analysis, we can transform the orginal problem into

$$Maximize \quad \frac{1}{2} * < A, X > \tag{155}$$

$$0 \preceq X \tag{156}$$

$$< B_i, X >= 1, \quad for \quad i = 1, ..., m + n \tag{157}$$

Where B_i is a all-zero matrix except that at (i, i) it has a 1 entry.

1.24 3.6.4

The proof comes from link. The construction of this proof lies in repeatedly using the Proposition 3.6.3.

We sample k random vectors $x_1, ..., x_k \sim Unif(\{-1, 1\}^n)$ one by one until the kth vector x_k achieves $(0.5 - \epsilon) * |E|$ maximum-cut. In the following part, we will analyze the expected running time. Let y_i denote the number of cuts in ith experiment.

Then:

$$\begin{aligned} P(\forall i, y_i < (0.5 - \epsilon) * |E|) &= [P(y_i < (0.5 - \epsilon) * |E|)]^k && (158) \\ &= [P((\epsilon + 1) * |E| < |E| - y_i)]^k && (159) \\ &\leq (\frac{E[|E| - y_i]}{(\epsilon + 1) * |E|})^k && (160) \\ &= (\frac{\frac{|E|}{2}}{(\epsilon + 1) * |E|})^k && (161) \\ &= (\frac{1}{2 * (\epsilon + 1)})^k && (162) \end{aligned}$$

Where we use the Markov inequality in (160). So if we use T to denote the running time of this algorithm, we can calculate the expectation of T as follows:

$$\begin{aligned} E[T] &= \sum_n P(T \geq n) && (163) \\ &\leq \sum_n (\frac{1}{2 * (\epsilon + 1)})^n && (164) \\ &= \frac{1}{1 - \frac{1}{2*(\epsilon+1)}} && (165) \\ &= \frac{2 * \epsilon + 2}{2 * \epsilon + 1} && (166) \end{aligned}$$

So T is finite almost everywhere.

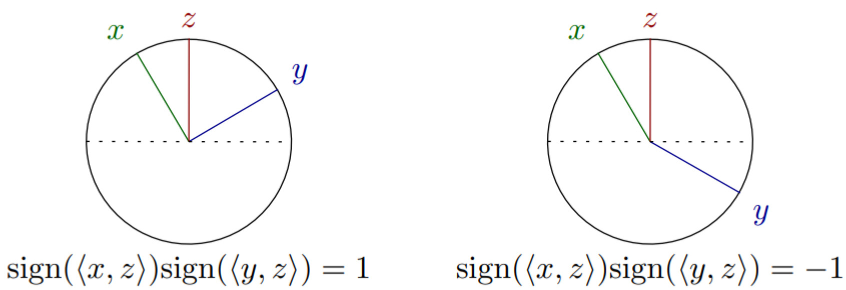

Figure 1: In Same/Different Arch

1.25 3.6.7

The proof comes from link. By rotational invariance, we can focus our attention to R^2 space.

In Figure 1, we can see that we need to find out when and where the sign will be different. Suppose that the angle between x and y is θ and note that the projection of normal distribution onto the sphere is a uniform distribution, we can have

$$P(x,y \quad in \quad different \quad archs) = \frac{2\theta}{2\pi} = \frac{\theta}{\pi} \tag{167}$$

Where the factor 2 here appears due to the fact that z and $-z$ has the same decomposition.

Finally, note that

$$\begin{aligned} E(sign < g,u > *sign < g,v >) &= (1-\frac{\theta}{\pi})*1+\frac{\theta}{\pi}*(-1) &(168)\\ &= 1-\frac{2*\theta}{\pi} &(169)\end{aligned}$$

And note that

$$\begin{aligned} \frac{2}{\pi}arcsin(< x,y >) &= \frac{2}{\pi}arcsin(cos\theta) &(170)\\ &= \frac{2}{\pi}arcsin(sin(\frac{\pi}{2}\theta)) &(171)\\ &= \frac{2}{\pi}*(\frac{\pi}{2}-\theta) &(172)\\ &= 1-\frac{2}{\pi}*\theta &(173)\end{aligned}$$

1.26 3.7.4

For the left part of the equation, we have

$$< u^{\otimes k}, v^{\otimes k} > = \sum_{i_1,\dots,i_k} a_{i_1\dots i_k} b_{i_1\dots i_k} \tag{174}$$

$$= \sum_{i_1,\dots,i_k} u_{i_1} * \dots * u_{i_k} * v_{i_1} \dots * v_{i_k} \tag{175}$$

$$= \sum_{i_1,\dots,i_k} (u_{i_1} v_{i_1}) * \dots * (u_{i_k} v_{i_k}) \tag{176}$$

Where each of the $i_1, ..., i_k$ sums up from 1 to n.

While on the right side, we have

$$< u, v >^k = (\sum_{i=1}^{n} u_i * v_i)^k \tag{177}$$

$$= \sum_{k_1+\dots+k_n=k, k_i \geq 0, \forall i} \frac{k!}{k_1! \dots k_n!} \prod_{t=1}^{k} (u_t * v_t)^{k_t} \tag{178}$$

Clearly, every term in (167) will appear in (169). Moreover, there will be repeated terms and the number of repeated times is exactly $\frac{k!}{k_1!\dots k_n!}$ where $k_1 + ... + k_n = k$.

1.27 3.7.5

1.27.1

Consider the Hilbert given by

$$H = R^{n\times n} \oplus R^{n\times n\times n} \tag{179}$$

Where the corresponding inner product is given by

$$< u, v >_H = 2* < u, v >_2 + 5* < u, v >_3 \tag{180}$$

Where

$$< u, v >_2 = \sum_{i_1 i_2} u_{i_1} u_{i_2} v_{i_1} v_{i_2} \tag{181}$$

And

$$< u, v >_2 = \sum_{i_1 i_2 i_3} u_{i_1} u_{i_2} u_{i_2} v_{i_1} v_{i_2} v_{i_3} \tag{182}$$

Clearly, this is an inner product that meets all the requirements: $< u, u >_H \geq 0$, $< u, u >_H = 0 \Leftrightarrow u = 0$ and $< u, \alpha * v_1 + \beta * v_2 > = \alpha * < u, v_1 > + \beta * < u, v_2 >$
Besides,

$$\Phi(u) = \sqrt{2} u_1 \oplus \sqrt{5} u_2 \tag{183}$$

Where

$$u = u_1 \oplus u_2, u_1 \in R^{n\times n}, u_2 \in R^{n\times n\times n} \tag{184}$$

1.27.2

Without loss of generality, we assume the highest degree of f is m. So

$$f(x) = \sum_{i=0}^{m} a_i x^i \tag{185}$$

Then

$$f(< u, v >) = \sum_{i=0}^{m} a_i (< u, v >)^i \tag{186}$$

Similar to the above sub-question, we consider the Hilbert space

$$H = R \oplus R^n ... \oplus R^{n\times ... \times n} \tag{187}$$

Where the inner product is defined as

$$< u, v >_H = \sum_{i=0}^{m} a_i * < u, v >_i \tag{188}$$

Where

$$< u, v >_i = \sum_{k_1 \dots k_i} u_{k_1} \dots u k_i v_{k_1} \dots v_{k_i} \tag{189}$$

Since $a_i, \forall i$ is a non-negative sequence, we can know that it is actually an inner product.

Besides,

$$\Phi(u) = \oplus_{i=1}^{m} \sqrt{a_i} u_i \tag{190}$$

1.27.3

In the almost same way, we define the inner product as

$$< u, v >_H = \sum_{a=0}^{\infty} a_k < u, v >_k \tag{191}$$

Where

$$< u, v >_k = \sum_{b_1 \dots b_k} u_{b_1} \dots u_{b_k} v_{b_1} \dots v_{b_k} \tag{192}$$

Note that the original series converge for each $x \in R$, so this inner product makes sense. Moreover, a_i is non-negative, this makes it a valid inner product.

Besides,

$$\Phi(u) = \oplus_{i=1}^{\infty} \sqrt{a_i} u_i \tag{193}$$

1.28 3.7.6

Similar to the above, we define

$$\Phi(u) = \oplus_{i=0}^{\infty} \sqrt{|a_i|} u_i \tag{194}$$

$$\Psi(v) = \oplus_{i=0}^{\infty} - \sqrt{|a_i|} v_i \tag{195}$$

Clearly,

$$\begin{aligned}
||\Phi||^2 &= ||\Psi||^2 & (196)\\
&= \sum_{i=0}^{\infty} |a_i| < u_i, u_i >^k & (197)\\
&= \sum_{i=0}^{\infty} |a_i| (u_1^2 + ... u_n^2)^k & (198)\\
&= \sum_{k=0} |a_k| ||u||_2^{2k} & (199)
\end{aligned}$$