

# Math 888 HDPS Homework

Daniel Szabo

November 21, 2021

2.7.2 To prove the equivalence of the properties (up to constants) I will show (a)  $\implies$  (b), (b)  $\implies$  (c), (c)  $\implies$  (d), and (d)  $\implies$  (a)  
Assume the tail probabilities are bounded, namely

$$\mathbb{P}[|X| \geq t] \leq 2 \exp\left(\frac{-t}{K_1}\right).$$

Then

$$\mathbb{E}[|X|^p] = \int_0^\infty \mathbb{P}(|X| \geq t) p t^{p-1} dt \leq \int_0^\infty 2 \exp\left(\frac{-t}{K_1}\right) p t^{p-1} dt$$

To calculate this, use the known moments of the exponential distribution, which says if  $\mathbb{P}[Y \geq t] = e^{-\lambda t}$

$$\mathbb{E}[Y^p] = \int_0^\infty e^{-\lambda t} p t^{p-1} dt = \frac{p!}{\lambda^p}.$$

We have  $\lambda = \frac{1}{K_1}$ , so

$$(\mathbb{E}[|X|^p])^{1/p} \leq 2 (K_1^p p!)^{1/p} \approx 2 K_1 (2\pi p)^{\frac{1}{2p}} \left(\frac{p}{e}\right)^{p/p} \leq K_2 p$$

With this, we also show (d):

$$\mathbb{E}[e^{|X|/t}] = 1 + \sum_{p=1}^\infty \frac{1}{t^p p!} \mathbb{E}[|X|^p] \lesssim 1 + \sum_{p=1}^\infty \frac{1}{t^p p!} K_1^p p! = \sum_{p=0}^\infty \left(\frac{K_1}{t}\right)^p \leq 2$$

for  $t \geq K_4$  for some constant  $K_4$ .

This also shows (c), let  $\lambda = \frac{1}{t} \implies$

$$\mathbb{E}[e^{\lambda|X|}] \lesssim \sum_{p \geq 0} (K_1 \lambda)^p \leq \sum_{p \geq 0} \frac{(K_1 \lambda)^p}{p!} \leq e^{K_3 \lambda}$$

if the series converges, ie if  $\lambda \leq \frac{1}{K_3}$ .

That (c)  $\implies$  (d) is implicit in the above calculation.

To show (d)  $\implies$  (a), we just apply Markov's inequality to see

$$\mathbb{P}(|X| \geq t) = \mathbb{P}(e^{|X|/K_4} \geq e^{t/K_4}) \leq \frac{\mathbb{E}[e^{|X|/K_4}]}{e^{t/K_4}} \leq 2e^{-t/K_4}.$$

2.8.5 First, we show  $e^z \leq 1 + z + \frac{z^2/2}{1-|z|/3}$ :

$$\begin{aligned}
e^z &= 1 + z + \frac{z^2}{2} \left( 1 + \frac{z}{3} + \frac{z^2}{12} + \dots \right) \\
&= 1 + z + \frac{z^2}{2} \left( \sum_{k=0}^{\infty} \frac{z^k}{(k+2)!/2} \right) \\
&\leq 1 + z + \frac{z^2}{2} \left( \sum_{k=0}^{\infty} \left( \frac{z}{3} \right)^k \right) \\
&\leq 1 + z + \frac{z^2}{2} \left( \sum_{k=0}^{\infty} \left( \frac{|z|}{3} \right)^k \right) \\
&= 1 + z + \frac{z^2/2}{1-|z|/3}.
\end{aligned}$$

Now we just plug in  $z := \lambda X$  to get the bound:

$$\begin{aligned}
\mathbb{E}[e^{\lambda X}] &\leq \mathbb{E}\left[1 + \lambda X + \frac{\lambda^2/2}{1-|\lambda X|/3} X^2\right] \\
&\leq 1 + \frac{\lambda^2/2}{1-\lambda K/3} \mathbb{E}[X^2] \\
&\leq \exp(g(\lambda) \mathbb{E}[X^2]).
\end{aligned}$$

2.8.6 Following the proof of Bernstein's inequality, start by applying Markov's inequality to get

$$\mathbb{P} \left( \left| \sum_{i=0}^N X_i \right| \geq t \right) = \mathbb{P} \left( \exp \left( \lambda \left| \sum_{i=0}^N X_i \right| \right) \geq e^{\lambda t} \right) \leq e^{-\lambda t} \mathbb{E} \left[ e^{\lambda \left| \sum_{i=0}^N X_i \right|} \right] \leq e^{-\lambda t} \prod_{i=1}^N \mathbb{E} \left[ e^{\lambda X_i} \right].$$

We then use our bound from 2.8.5 to get

$$\mathbb{P} \left( \sum_{i=0}^N X_i \geq t \right) \leq e^{-\lambda t} \prod_{i=1}^N \mathbb{E} \left[ e^{\lambda X_i} \right] \leq e^{-\lambda t} \prod_{i=1}^N e^{g(\lambda) \mathbb{E}[X_i^2]} = \exp(\sigma^2 g(\lambda) - \lambda t)$$

All that is left is to find the  $\lambda$  that minimizes  $\sigma^2 g(\lambda) - \lambda t$ . To this end, take the derivative w.r.t  $\lambda$ :

$$\frac{d}{d\lambda} (\sigma^2 g(\lambda) - \lambda t) = \sigma^2 g'(\lambda) - t$$

Some calculus shows that  $g'(\frac{t}{\sigma^2 + tK/3}) = \frac{t}{\sigma^2}$ , so  $\sigma^2 g(\lambda) - \lambda t$  has a local minimum at  $\lambda = \frac{t}{\sigma^2 + tK/3}$ . The value of the function at this  $\lambda$  is

$$\sigma^2 \frac{t^2/2}{\frac{\sigma^2}{\sigma^2 + tK/3} (\sigma^2 + tK/3)^2} - \frac{t^2}{\sigma^2 + tK/3} = \frac{t^2/2}{\sigma^2 + tK/3} - \frac{t^2}{\sigma^2 + tK/3} = \frac{-t^2/2}{\sigma^2 + tK/3},$$

so our bound improves to

$$\mathbb{P} \left( \sum_{i=0}^N X_i \geq t \right) \leq \exp \left( \frac{-t^2/2}{\sigma^2 + tK/3} \right)$$

5.4.12 We start by proving the hint, that  $\mathbb{E}[\exp(\lambda \varepsilon_i A_i)] \preceq \exp(\lambda^2 A_i^2/2)$ . This will be based on the inequality that  $e^x \leq x + e^{x^2}$  for all  $x$  real. Thus by property (b) of Exercise 5.4.5, the matrix version  $e^{\lambda \varepsilon_i A_i} \preceq \lambda \varepsilon_i A_i + e^{(\lambda \varepsilon_i A_i)^2}$  holds as well as  $A_i$  is a deterministic matrix. Exponentiating both sides and taking the expectation, we have

$$\mathbb{E} e^{\lambda \varepsilon_i A_i} \preceq \mathbb{E} \left[ \lambda \varepsilon_i A_i + e^{(\lambda \varepsilon_i A_i)^2} \right] = e^{(\lambda A_i)^2/2}.$$

Now we follow the proof of matrix Bernstein's inequality. Let  $S = \sum_{i=1}^N \varepsilon_i A_i$ , and using the symmetry of the cases where  $\|S\| = \lambda_{\max}(S)$  and where  $\|S\| = \lambda_{\max}(-S)$  we have

$$\begin{aligned} \mathbb{P}(\|S\| \geq t) &\leq 2\mathbb{P}(\lambda_{\max}(S) \geq t) \leq 2e^{-\lambda t} \mathbb{E} e^{\lambda \lambda_{\max}(S)} \leq 2e^{-\lambda t} \mathbb{E} \lambda_{\max}(e^{\lambda S}) \leq 2e^{-\lambda t} \mathbb{E} \operatorname{tr}(e^{\lambda S}) \\ &\leq 2e^{-\lambda t} \operatorname{tr} \left( \exp \left( \sum_{i=1}^N \log \mathbb{E} e^{\lambda \varepsilon_i A_i} \right) \right) \end{aligned}$$

The details of each of those steps are just the same as in the proof of matrix Bernstein's, so refer to that for explanations. The hint we proved implies also that  $\sum_{i=1}^N \log \mathbb{E} e^{\lambda \varepsilon_i A_i} \preceq \sum_{i=1}^N \log e^{(\lambda A_i)^2/2}$ , so applying property (e) of Exercise 5.4.5 we have (because  $e^x$  is strictly increasing) that

$$\begin{aligned} \mathbb{P}(\|S\| \geq t) &\leq 2e^{-\lambda t} \operatorname{tr} \left( \exp \left( \sum_{i=1}^N \log e^{(\lambda A_i)^2/2} \right) \right) \\ &= 2e^{-\lambda t} \operatorname{tr} \left( \exp \left( \lambda^2/2 \sum_{i=1}^N A_i^2 \right) \right) \\ &\leq 2e^{-\lambda t} n \left\| \exp \left( \lambda^2/2 \sum_{i=1}^N A_i^2 \right) \right\| \\ &= 2n \exp(\sigma^2 \lambda^2/2 - \lambda t) \end{aligned}$$

which is minimized at  $\lambda = \frac{t}{\sigma^2}$ , where it achieves

$$\mathbb{P}(\|S\| \geq t) \leq 2n \exp \left( \frac{t^2}{2\sigma^2} - \frac{t^2}{\sigma^2} \right) = 2n \exp(-t^2/2\sigma^2)$$

6.2.6 (a) Take a subgaussian vector  $X$  such that  $\|X\|_{\psi_2} \leq K$ , which means  $\|BX\|_2^2 = X^\top B^\top BX$  is subgaussian with subgaussian norm  $K$ , so  $\mathbb{E}_X[\lambda \|BX\|_2^2] \leq \exp(C\lambda^2 K^2 \|BX\|_2^2)$ .

Let  $A = B^\top B$ . Note that  $\mathbb{E}_g[\exp(\mu g^\top AX)|X] = \exp(\mu^2 \|BX\|_2^2/2)$  because the MGF of  $\mathbb{E}[e^{\lambda g}]$  is  $e^{\lambda^2/2}$ . We can replace  $\mu$  with  $\sqrt{2}CK\lambda$  to match the left hand side of our bound on the MGF. Thus

$$\mathbb{E}_X[\exp(\lambda^2 \|BX\|_2^2)] \leq \mathbb{E}_g[\exp(\sqrt{2}CK\lambda g^\top AX)|X].$$

We can take the expectation over  $X$  of this and repeat the same trick to see

$$\mathbb{E} \exp(\lambda^2 \|BX\|_2^2) \leq \mathbb{E} \exp(CK^2 \lambda^2 \|Bg\|_2^2).$$

- (b) Similarly to the proof of Lemma 6.2.2, we can take the singular value decomposition of  $A = B^\top B$  as  $A = \sum_i s_i u_i v_i^\top$ . Then

$$\|Bg\|_2^2 = g^\top Ag = \sum_i s_i \langle u_i, g \rangle \langle v_i, g \rangle = \sum_i s_i g_i g'_i$$

by rotational invariance of the normal distribution under multiplication. Although  $g_i$  and  $g'_i$  are not necessarily independent, we can use the same conditioning method to bound  $g_i g'_i$  by  $g_i^2$  to say

$$\mathbb{E} \exp(\lambda s_i g_i g'_i) \leq \mathbb{E} \exp(\lambda^2 s_i^2 g_i^2) \leq \exp(C \lambda^2 s_i^2)$$

when  $\lambda^2 s_i^2 \leq c$  for constants  $c$  and  $C$ . By the independence for each  $i$  we have

$$\mathbb{E} \exp(\lambda^2 \|Bg\|_2^2) \leq \exp \left( C \lambda^2 \sum_i s_i^2 \right) \leq \exp(C \lambda^2 \|B^\top B\|_F^2) \leq \exp(C' \lambda^2 \|B\|_F^2)$$

$$\text{when } \lambda \leq \frac{c}{\max_i s_i^2} \leq \frac{c}{\|B\|_2}.$$

6.3.5 Just calculate using the previous problem:

$$\begin{aligned} \mathbb{P}(\|BX\|_2 \geq CK\|B\|_F + t) &= \mathbb{P}(\exp(\lambda^2 \|BX\|_2^2) \geq \exp((\lambda CK\|B\|_F + \lambda t)^2)) \\ &= \mathbb{P}(\exp(\lambda^2 \|BX\|_2^2) \geq \exp(\lambda^2 CK^2\|B\|_F^2 + \lambda CK\|B\|_F t + \lambda^2 t^2)) \\ &\leq \mathbb{P}(\exp(\lambda^2 \|BX\|_2^2) \geq \exp(\lambda^2 CK^2\|B\|_F^2 + \lambda^2 t^2)) \\ &\leq \frac{\exp(\lambda^2 CK^2\|B\|_F^2)}{\exp(\lambda^2 CK^2\|B\|_F^2 + \lambda^2 t^2)} \\ &= \exp(-\lambda^2 t^2) \\ &\leq \exp\left(-\frac{ct^2}{K^2\|B\|_2}\right) \end{aligned}$$

by the condition on  $\lambda$  that  $|\lambda| \leq \frac{c}{K\|B\|_2}$ .

5.6.4 Follow the proof of Theorem 5.6.1, except instead of the expectation version of matrix Bernstein's we apply the tail bound version, which gives us

$$\|\Sigma_m - \Sigma\| \leq \frac{1}{m} \left\| \sum_{i=1}^m (X_i X_i^\top - \Sigma) \right\| \leq \frac{1}{m} t$$

with probability  $1 - 2ne^{-c \min\{t^2/\sigma^2, \frac{t}{K}\}}$ . Take  $t = \sqrt{\sigma^2(\log n + u)} + K^2 \text{tr}(\Sigma)(\log n + u)$  to have this hold with probability  $\geq 1 - 2e^{-u}$ . The same bound for  $\sigma^2$  applies, so  $\sigma^2 \leq K^2 m \text{tr}(\Sigma) \|\Sigma\|$  means

$$\begin{aligned} \|\Sigma_m - \Sigma\| &\leq \frac{1}{m} \left( \sqrt{\sigma^2(\log n + u)} + K^2 \text{tr}(\Sigma)(\log n + u) \right) \\ &\leq \frac{1}{m} \left( \sqrt{K^2 m \text{tr}(\Sigma) \|\Sigma\| (\log n + u)} + K^2 \text{tr}(\Sigma)(\log n + u) \right) \\ &= \left( \sqrt{\frac{K^2 r(\log n + u)}{m}} + \frac{K^2 r(\log n + u)}{m} \right) \|\Sigma\| \end{aligned}$$

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

4.7.6 (a) The covariance matrix is

$$\Sigma = \mathbb{E}[XX^\top] = \mathbb{E}[(\theta\mu + g)(\theta\mu + g)^\top] = \mathbb{E}[\theta^2\mu\mu^\top] + \mathbb{E}[\theta(\mu g^\top + g\mu^\top)] + \mathbb{E}[gg^\top] = \mu\mu^\top + I_n.$$

Because  $\|\mu\mu^\top\| \geq 0$ , the eigenvector corresponding to the maximal eigenvalue will be  $\frac{\mu}{\|\mu\|}$ .

(b) We saw in class that

$$\|\Sigma_m - \Sigma\| \lesssim K^2 \left( \sqrt{\frac{n+u}{m}} + \frac{n+u}{m} \right) \|\Sigma\|$$

with probability at least  $1 - 2e^{-u}$ . Our subgaussian bound is just  $K = 1$ , so plugging in our bound for  $m$  and taking  $u = n$  we see

$$\|\Sigma_m - \Sigma\| \lesssim \left( \sqrt{n^{1-c}\|\mu\|_2^c} + n^{1-c}\|\mu\|_2^c \right) \|\Sigma\|.$$

Using also that  $\|\Sigma\| = \|\mu\mu^\top + I_n\| \leq \|\mu\|_2 + 1$ , we have

$$\|\Sigma_m - \Sigma\| \lesssim \left( \sqrt{n^{1-c}\|\mu\|_2^{1/2+c}} + n^{1-c}\|\mu\|_2^{1+c} \right)$$

(c) We can now use Davis-Kahan. The maximal eigenvalue gap of  $\Sigma$ , which is an approximate isometry having  $\|\Sigma - I\| = \|\mu\mu^\top\| \leq \|\mu\|$ , is  $2\|\mu\| \implies \exists \theta \in \{\pm 1\}$  such that

$$\|v_1(\Sigma) - \theta v_1(\Sigma_m)\| \leq \frac{\|\Sigma - \Sigma_m\|}{2\|\mu\|_2} \lesssim \left( \sqrt{n^{1-c}\|\mu\|_2^c} + n^{1-c}\|\mu\|_2^c \right)$$

(d) If we knew the mean  $\mu$ , the sign of  $\mu^\top X_i$  would simply describe whether  $X_i$  is closer to  $\mu$  or  $-\mu$ . Thus the probability of misclassifying would be bounded by the probability of being further away from the mean than  $\mu$ , which is just

$$\mathbb{P}(\text{Misclassifying } X_i) = \mathbb{P}(X_i^\top \mu < 0) \leq e^{-\|\mu\|^2} \leq e^{-C \log(1/\varepsilon)} \leq \varepsilon^C.$$

(e) Since the empirical eigenvector  $v$  is close to  $\mu$  with  $\|v - \mu\| \lesssim n^{1-c}\|\mu\|_2^c$  with probability at least  $1 - 2e^{-n}$ ,

$$\begin{aligned} \mathbb{P}(M \geq m\varepsilon) &\leq \mathbb{P}\left(\left\|\sum_{j=1}^n |v_j - \mu_j|^2\right\| \geq m\varepsilon^{1+C}\right) \\ &\leq \mathbb{P}\left(\|v - \mu\|^2 \geq \left(\frac{n}{\|\mu\|_2}\right)^c \varepsilon^{1+C}\right) \\ &\leq \mathbb{P}(\|v - \mu\| \geq n^{1-c}\|\mu\|_2^c) \\ &\leq 4e^{-n}. \end{aligned}$$

Thus with probability at least  $1 - 4e^{-n}$ , the Spectral Clustering Algorithm identifies the communities correctly up to  $\varepsilon m$  misclassified points.

5.5.1 (a) Let  $Z_{ij}$  be the matrix with  $Z_{ij} = A_{ij}$ ,  $Z_{ji} = A_{ji}$ , and the rest zero. Because each edge in a random matrix is sampled independently, each such  $Z_{ij}$  is independent, and  $A = \sum_{i \leq j} Z_{ij}$ .

(b) Each  $Z_{ij}$  has a norm  $\leq \|Z_{ij}\|$ . The expectation version of Matrix Bernstein's tells us

$$\mathbb{E}\|A\| = \mathbb{E}\left\|\sum_{i \leq j} Z_{ij}\right\| \lesssim \left\|\sum_{i \leq j} \mathbb{E}Z_{ij}^2\right\|^{1/2} \sqrt{1 + \log n} + K(1 + \log n),$$

where  $K$  is a bound on the norm of  $\|Z_{ij}\|$ . The matrix  $Z_{ij}$  has norm equal to just  $D_{ij} + R_{ij}$ , both of which are bounded by a constant, so  $K$  can be renormalized to 1.

We can also bound  $\mathbb{E}\left\|\sum_{i \leq j} Z_{ij}^2\right\|^{1/2}$ . To do this, first notice that  $\mathbb{E}Z_{ij}^2$  just has  $D_{ij}^2$  at entries  $ii$  and  $jj$  and is 0 everywhere else. Summing over all  $i \leq j$  we get a diagonal matrix  $\Sigma$  with entry  $\Sigma_{ii} = D_i^\top D_i = \|D_i\|_2^2$ , where  $D_i$  is the  $i$ th row of  $D$ . This is bounded by the maximal eigenvalue of  $D$ , which, as we saw in section 4.5, is just  $d$ . Thus

$$\mathbb{E}\|R\| \leq \mathbb{E}\|A\| \lesssim \sqrt{d}\sqrt{1 + \log n} + (1 + \log n) \lesssim \sqrt{d \log n} + \log n$$

5.5.2 Similar to the derivation in section 4.5, we have the difference between the unit eigenvectors of  $D$  and  $A$ ,  $v$  and  $u$ , respectively, satisfy

$$\|v - \theta u\|_2 \leq \frac{\|R\|}{\mu n}$$

where  $\theta \in \{1, -1\}$  by Davis-Kahan. We can now apply our bound in 5.5.1 to say

$$\mathbb{E}\|v - \theta u\|_2 \leq \frac{\mathbb{E}\|R\|}{\mu n} \lesssim \frac{\sqrt{d \log n} + \log n}{\mu n}.$$

If the matrix is sparse and  $d \gg \log n$ , this is approximately

$$\mathbb{E}\|v - \theta u\|_2 \lesssim \frac{\sqrt{d \log n}}{\mu n} \approx \frac{p+q}{2} \frac{1}{\mu \sqrt{n}}.$$

Following the analysis in 4.5, we can bound the expectation of the distance between the full eigenvectors by  $\frac{p+q}{2\mu}$ . Thus the expected number of misclassifications is bounded by the square of this (every misclassification has difference in eigenvectors at least 1), which is  $\frac{(p+q)^2}{4\mu^2}$ . A tail bound version of this can be obtained similarly, but I hope expectation is enough.