Homework 4: Solutions

November 30, 2015

1 A warm-up: from tails to expectations

(a)

$$\begin{aligned}
\mathbb{E}[X^p] &= \int_0^\infty x^p P(\mathrm{d}x) \\
\text{(using the hint)} &= \int_0^\infty p \int_0^x t^{p-1} \mathrm{d}t P(\mathrm{d}x) \\
\text{(Fubini's theorem)} &= \int_0^\infty \int_t^\infty p t^{p-1} P(\mathrm{d}x) \mathrm{d}t \\
&= p \int_0^\infty t^{p-1} \mathbb{P}[X \geq t] \mathrm{d}t
\end{aligned}$$

(b) For the upper bound, choose x to be the lower-left corner of the unit square, then

$$\begin{aligned}
\mathbb{E}\Big[\min_{1\leq i\leq n} \|X_i - x\|\Big] &= \int_0^{\sqrt{2}} \mathbb{P}\Big[\min_{1\leq i\leq n} \|X_i - x\| \geq t\Big] \mathrm{d}t = \int_0^{\sqrt{2}} \mathbb{P}[\|X_1 - x\| \geq t]^n \mathrm{d}t \\
&\leq \int_0^{\sqrt{2}} \Big(1 - \frac{2}{\pi}\frac{\pi t^2}{4}\Big)^n \mathrm{d}t \leq \int_0^{\sqrt{2}} e^{-nt^2/2} \mathrm{d}t \leq \frac{c_2}{\sqrt{n}}.
\end{aligned}$$

For the lower bound, choose x to be the center of the unit square, then

$$\begin{aligned}
\mathbb{E}\Big[\min_{1\leq i\leq n} \|X_i - x\|\Big] &= \int_0^{\sqrt{1/2}} \mathbb{P}\Big[\min_{1\leq i\leq n} \|X_i - x\| \geq t\Big] \mathrm{d}t = \int_0^{\sqrt{1/2}} \mathbb{P}[\|X_1 - x\| \geq t]^n \mathrm{d}t \\
&\geq \int_0^{\sqrt{1/\pi}} \mathbb{P}[\|X_1 - x\| \geq t]^n \mathrm{d}t = \int_0^{\sqrt{1/\pi}} (1 - \pi t^2)^n \mathrm{d}t \\
(\text{let } \pi t^2 = u) &\geq \frac{1}{2\sqrt{\pi}} \int_0^1 \Big(1 - u\Big)^n u^{-1/2} \mathrm{d}u = \frac{\Gamma(1+n)}{2\Gamma(3/2+n)} \\
\text{(property of Gamma function)} &\geq \frac{c_1}{\sqrt{n}}.
\end{aligned}$$

Therefore, for any x in the unit square,

$$\frac{c_1}{\sqrt{n}} \leq \mathbb{E}\Big[\min_{1\leq i\leq n} \|X_i - x\|\Big] \leq \frac{c_2}{\sqrt{n}}.$$

It follows that

$$\frac{c_1}{\sqrt{n}} \leq \frac{c_1}{\sqrt{n-1}} \leq \mathbb{E}\Big[\min_{2\leq i\leq n} \|X_i - X_1\|\Big] \leq \frac{c_2}{\sqrt{n-1}} \leq \frac{2c_2}{\sqrt{n}}.$$

2 The traveling salesman problem

(a) For the lower bound, let σ be the optimal order of $\{1, \ldots, n\}$, then

$$L_n = \|X_{\sigma(1)} - X_{\sigma(2)}\| + \ldots + \|X_{\sigma(n)} - X_{\sigma(1)}\| \geq \sum_{i=1}^{n} \min_{j \neq i} \|X_j - X_{\sigma(i)}\|.$$

From Problem 1(b), we have $\mathbb{E}[L_n] \geq c_1\sqrt{n}$.

For the upper bound, suppose that σ is the optimal order of $\{1, \ldots, n-1\}$. Now create a suboptimal order of $\{1, \ldots, n\}$ by picking a $j = \arg\min_{1 \leq i \leq n-1} \|X_{\sigma_{(j)}} - X_n\|$, and replacing the path $(X_{\sigma_{(j)}}, X_{\sigma_{(j+1)}})$ by the path $(X_{\sigma_{(j)}}, X_n, X_{\sigma_{(j+1)}})$. By triangle inequality,

$$\begin{aligned} L_n &\leq L_{n-1} + \|X_{\sigma(j)} - X_n\| + \|X_n - X_{\sigma_{(j+1)}}\| - \|X_{\sigma_{(j)}} - X_{\sigma_{(j+1)}}\| \\ &\leq L_{n-1} + 2\|X_{\sigma(j)} - X_n\| = L_{n-1} + 2 \min_{1 \leq i \leq n-1} \|X_{\sigma(i)} - X_n\|. \end{aligned}$$

From Problem 1(b), and by induction, we have $\mathbb{E}[L_n] \leq c_2\sqrt{n}$.

(b) The martingale difference $\Delta_i = \Lambda_i - \Lambda_{i-1}$ can be written as

$$\begin{aligned} \Delta_i &= \mathbb{E}[L_n(X_1, \ldots, X_i, \ldots, X_n)|X_1, \ldots, X_i] - \mathbb{E}[L_n(X_1, \ldots, X_i', \ldots, X_n)|X_1, \ldots, X_i] \\ &= \mathbb{E}[L_n(X_1, \ldots, X_i, \ldots, X_n) - L_n(X_1, \ldots, X_i', \ldots, X_n)|X_1, \ldots, X_i], \end{aligned}$$

where X_i' is an i.i.d. copy of X_i. Using a similar argument as in Part(a), we can create a suboptimal route over $\{x_1, \ldots, x_i', \ldots, x_n\}$ based on the optimal route over $\{x_1, \ldots, x_n\}$ by modifying the local path at x_i, and vice-versa. Since the distance between any two points is upper bounded by $\sqrt{2}$, we have

$$|L_n(x_1, \ldots, x_i, \ldots, x_n) - L_n(x_1, \ldots, x_i', \ldots, x_n)| \leq 2\sqrt{2}$$

for all $x_1, \ldots, x_n$ and x_i' in the unit square. Hence $|\Delta_i| \leq 2\sqrt{2}$, and by Azuma-Hoeffding inequality,

$$\mathbb{P}[|L_n - \mathbb{E}L_n| \geq t] \leq 2\exp\left(-\frac{t^2}{16n}\right).$$

This upper bound is not quite useful because Part(a) tells us that $c_1\sqrt{n} \leq \mathbb{E}L_n \leq c_2\sqrt{n}$, and to get an probability upper bound that does not increase in n, the deviation t must also be on the order of $\sqrt{n}$. We can simply use Markov inequality to get $\mathbb{P}[L_n - \mathbb{E}L_n \geq t] \leq c$, for some constant c when t is on the order of $\sqrt{n}$.

(c) Using the same argument as the one for proving the upper bound in Part(a), we can prove that

$$L(S) \leq L(S \cup \{x\}) \leq L(S) + 2\min_{y \in S} \|x - y\|, \qquad \text{for any } x \notin S.$$

(d) Let $S = \{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$. As in Part(b), write the martingale difference as

$$\begin{aligned}
\Delta_i &= \mathbb{E}[L_n(X_1, \ldots, X_i, \ldots, X_n)|X_1, \ldots, X_i] - \\
&\quad \mathbb{E}[L_n(X_1, \ldots, X_i', \ldots, X_n)|X_1, \ldots, X_i] \\
&= \mathbb{E}[L_n(S \cup \{X_i\}) - L_n(S \cup \{X_i'\})|X_1, \ldots, X_i] \\
\text{(from Part(c))} &\leq \mathbb{E}\Big[2 \min_{X_j \in S} \|X_j - X_i\| + 2 \min_{X_j \in S} \|X_j - X_i'\| \Big| X_1, \ldots, X_i\Big] \\
\text{(smaller range of minimization)} &\leq \mathbb{E}\Big[2 \min_{j>i} \|X_j - X_i\| + 2 \min_{j>i} \|X_j - X_i'\| \Big| X_1, \ldots, X_i\Big] \\
&\leq 2\mathbb{E}\Big[\min_{j>i} \|X_j - X_i\| \Big| X_i\Big] + 2\mathbb{E}\Big[\min_{j>i} \|X_j - X_i'\|\Big] \\
\text{(from Problem 1(b))} &\leq \frac{c}{\sqrt{n-1}}.
\end{aligned}$$

(e) Now apply Azuma-Hoeffding inequality with the improved martingale difference upper bound obtained in Part(d), we get

$$\mathbb{P}[|L_n - \mathbb{E}L_n| \geq t] \leq 2\exp\left(-\frac{t^2}{2c^2 \sum_{i=1}^n (n-i)^{-1}}\right) \mathbb{P}[|L_n - \mathbb{E}L_n| \geq t] \leq 2\exp\left(-\frac{Bt^2}{\log n}\right).$$

With this improved upper bound, we can let the deviation $t = \alpha \log n$, and get

$$\mathbb{P}[|L_n - \mathbb{E}L_n| \geq \alpha \log n] \leq \frac{2}{n^\beta}.$$

Compared with the bound in Part(b), this is a much tighter concentration result.

3 Gaussian integration-by-parts formula and the softmax function

(a)

$$\mathbb{E}[YF(Y)] = \int_{-\infty}^{\infty} \frac{y}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} F(y)\mathrm{d}y = -\int_{-\infty}^{\infty} \sqrt{\frac{\sigma^2}{2\pi}} F(y)\mathrm{d}e^{-\frac{y^2}{2\sigma^2}} \tag{1}$$

$$\text{(integration by parts)} = -\sqrt{\frac{\sigma^2}{2\pi}} \left(F(y)e^{-\frac{y^2}{2\sigma^2}}\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma^2}} F'(y)\mathrm{d}y\right) \tag{2}$$

$$= \sqrt{\frac{\sigma^2}{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma^2}} F'(y)\mathrm{d}y = \sigma^2 \mathbb{E}[F'(Y)]. \tag{3}$$

(b) Using the hint, for each fixed i, let $\rho_{j,i} = \frac{\mathbb{E}[Y_i Y_j]}{\mathbb{E}[Y_i^2]}$ and define the innovations $\tilde{\mathbf{Y}}$ by

$$\tilde{Y}_j = \begin{cases} Y_i, & \text{if } j = i \\ Y_j - Y_i \rho_{j,i}, & \text{if } j \neq i. \end{cases}$$

Define $\tilde{\mathbf{Y}}_{(i)} = (\tilde{Y}_1, \ldots, \tilde{Y}_{i-1}, \tilde{Y}_{i+1}, \ldots, \tilde{Y}_n)$. Since $Y_1, \ldots, Y_n$ are zero-mean and jointly Gaussian, Y_i is independent of $\tilde{\mathbf{Y}}_{(i)}$. For the fixed i and for each $j \neq i$, Y_j can be expressed as

$$Y_j = \tilde{Y}_j + Y_i \rho_{j,i}. \tag{4}$$

By iterated expectation,

$$\mathbb{E}[Y_i F(Y)] = \mathbb{E}\big[\mathbb{E}[Y_i F(Y)|\tilde{\mathbf{Y}}_{(i)}]\big].$$

From the independence between Y_i and $\tilde{\mathbf{Y}}_{(i)}$, and from (4),

$$\begin{aligned}
\mathbb{E}[Y_i F(Y)|\tilde{\mathbf{Y}}_{(i)} = \tilde{\mathbf{y}}_{(i)}] &= \mathbb{E}[Y_i F(\tilde{y}_1 + Y_i\rho_{1,i}, \ldots, Y_i, \ldots, \tilde{y}_n + Y_i\rho_{n,i})] \\
\text{(from Part(a))} &= \mathbb{E}[Y_i^2]\mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}Y_i} F(\tilde{y}_1 + Y_i\rho_{1,i}, \ldots, Y_i, \ldots, \tilde{y}_n + Y_i\rho_{n,i})\right] \\
\text{(law of total derivative)} &= \mathbb{E}[Y_i^2]\mathbb{E}\left[\sum_{j=1}^n \frac{\partial F(\mathbf{Y})}{\partial Y_j}\frac{\mathrm{d}Y_j}{\mathrm{d}Y_i}\right] \\
&= \mathbb{E}[Y_i^2]\sum_{j=1}^n \rho_{j,i}\mathbb{E}\left[\frac{\partial F(\mathbf{Y})}{\partial Y_j}\right] = \sum_{j=1}^n \mathbb{E}[Y_i Y_j]\mathbb{E}\left[\frac{\partial F(\mathbf{Y})}{\partial Y_j}\right]
\end{aligned}$$

(c)

$$\max_i x_i = \max_i \frac{1}{\beta}\log e^{\beta x_i} \le \frac{1}{\beta}\log\sum_{i=1}^n e^{\beta x_i} = F_\beta(\mathbf{x}) \le \frac{1}{\beta}\log\left(ne^{\beta \max_i x_i}\right) = \max_i x_i + \frac{\log n}{\beta}.$$

Hence

$$\lim_{\beta\to\infty} F_\beta(\mathbf{x}) = \max_i x_i.$$

(d)

$$p_i(\mathbf{x}) = \frac{\partial F_\beta}{\partial x_i}(\mathbf{x}) = \frac{e^{\beta x_i}}{\sum_i e^{\beta x_i}} \ge 0.$$

Thus $\sum_i p_i(\mathbf{x}) = 1$, and $(p_1(\mathbf{x}), \ldots, p_n(\mathbf{x}))$ is a probability distribution on $\{1, \ldots, n\}$. Furthermore,

$$\frac{\partial^2 F_\beta}{\partial x_i \partial x_j}(\mathbf{x}) = \frac{\partial}{\partial x_i}p_j(\mathbf{x}) = \begin{cases} \beta p_i(\mathbf{x}) - \beta p_i(\mathbf{x})^2 = \beta p_i(\mathbf{x})\sum_{j=1,j\neq i}^n p_j(\mathbf{x}), & i = j \\ -\beta p_i(\mathbf{x})p_j(\mathbf{x}), & i \neq j \end{cases}$$

4 A Gaussian comparison inequality

(a) Since both $\mathbb{E}[\sup_{t\in[0,1]} F_\beta(\mathbf{Z}_t)]$ and $\mathbb{E}[\sup_{t\in[0,1]} \frac{\mathrm{d}}{\mathrm{d}t}F_\beta(\mathbf{Z}_t)]$ exist, we can interchange the order of differentiation and expectation, and get

$$\begin{aligned}
\varphi'(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[F_\beta(\mathbf{Z}_t)] = \mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}t}F_\beta(\mathbf{Z}_t)\right] \\
\text{(law of total derivative)} &= \mathbb{E}\left[\sum_{i=1}^n \frac{\partial F_\beta(\mathbf{Z}_t)}{\partial Z_{t,i}}\frac{\mathrm{d}Z_{t,i}}{\mathrm{d}t}\right] \\
&= \frac{1}{2}\sum_{i=1}^n\left(\frac{1}{\sqrt{t}}\mathbb{E}\Big[W_i\frac{\partial F_\beta(\mathbf{Z}_t)}{\partial Z_{t,i}}\Big] - \frac{1}{\sqrt{t-1}}\mathbb{E}\Big[V_i\frac{\partial F_\beta(\mathbf{Z}_t)}{\partial Z_{t,i}}\Big]\right).
\end{aligned}$$

Using the result of Problem 3(b),

$$\begin{aligned}
\mathbb{E}\left[W_i \frac{\partial F_\beta(\mathbf{Z}_t)}{\partial Z_{t,i}}\Big|\mathbf{V}\right] &= \sum_{j=1}^n \mathbb{E}[W_i W_j|\mathbf{V}]\mathbb{E}\left[\frac{\partial}{\partial W_j}\frac{\partial F_\beta(\mathbf{Z}_t)}{\partial Z_{t,i}}\Big|\mathbf{V}\right] \\
(\text{chain rule of derivative}) &= \sum_{j=1}^n \mathbb{E}[W_i W_j|\mathbf{V}]\mathbb{E}\left[\frac{\partial^2 F_\beta(\mathbf{Z}_t)}{\partial Z_{t,j}\partial Z_{t,i}}\frac{\partial Z_{t,j}}{\partial W_j}\Big|\mathbf{V}\right] \\
(\mathbf{W} \text{ and } \mathbf{V} \text{ are independent}) &= \sum_{j=1}^n \kappa_{ij}^{\mathbf{W}}\sqrt{t}\mathbb{E}\left[\frac{\partial^2 F_\beta(\mathbf{Z}_t)}{\partial Z_{t,j}\partial Z_{t,i}}\Big|\mathbf{V}\right].
\end{aligned}$$

Therefore,

$$\mathbb{E}\left[W_i \frac{\partial F_\beta(\mathbf{Z}_t)}{\partial Z_{t,i}}\right] = \mathbb{E}\left[\mathbb{E}\Big[W_i \frac{\partial F_\beta(\mathbf{Z}_t)}{\partial Z_{t,i}}\Big|\mathbf{V}\Big]\right] = \sum_{j=1}^n \kappa_{ij}^{\mathbf{W}}\sqrt{t}\mathbb{E}\left[\frac{\partial^2 F_\beta(\mathbf{Z}_t)}{\partial Z_{t,j}\partial Z_{t,i}}\right].$$

By the same token,

$$\mathbb{E}\left[V_i \frac{\partial F_\beta(\mathbf{Z}_t)}{\partial Z_{t,i}}\right] = \mathbb{E}\left[\mathbb{E}\Big[V_i \frac{\partial F_\beta(\mathbf{Z}_t)}{\partial Z_{t,i}}\Big|\mathbf{W}\Big]\right] = \sum_{j=1}^n \kappa_{ij}^{\mathbf{V}}\sqrt{t-1}\mathbb{E}\left[\frac{\partial^2 F_\beta(\mathbf{Z}_t)}{\partial Z_{t,j}\partial Z_{t,i}}\right].$$

Therefore,

$$\varphi'(t) = \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n (\kappa_{ij}^{\mathbf{W}} - \kappa_{ij}^{\mathbf{V}})\mathbb{E}\left[\frac{\partial^2 F_\beta(\mathbf{Z}_t)}{\partial Z_{t,j}\partial Z_{t,i}}\right]$$

(b)

$$\begin{aligned}
\varphi'(t) &= \frac{1}{2}\sum_{i=1}^n\sum_{j\neq i} (\kappa_{ij}^{\mathbf{W}} - \kappa_{ij}^{\mathbf{V}})\mathbb{E}\left[\frac{\partial^2 F_\beta(\mathbf{Z}_t)}{\partial Z_{t,j}\partial Z_{t,i}}\right] + \sum_{i=1}^n (\kappa_{ii}^{\mathbf{W}} - \kappa_{ii}^{\mathbf{V}})\mathbb{E}\left[\frac{\partial^2 F_\beta(\mathbf{Z}_t)}{\partial Z_{t,i}^2}\right] \\
(\text{Problem 3(d)}) &= \frac{\beta}{2}\sum_{i=1}^n\sum_{j\neq i}(-\mathbb{E}[W_iW_j] + \mathbb{E}[V_iV_j])\mathbb{E}[p_{i,t}p_{j,t}] + \frac{\beta}{2}\sum_{i=1}^n(\mathbb{E}[W_i^2] - \mathbb{E}[V_i^2])\mathbb{E}\Big[p_{i,t}\sum_{j\neq i}p_{j,t}\Big] \\
&= \frac{\beta}{2}\sum_{i=1}^n\sum_{j\neq i}(-\mathbb{E}[W_iW_j] + \mathbb{E}[V_iV_j] + \mathbb{E}[W_i^2] - \mathbb{E}[V_i^2])\mathbb{E}[p_{i,t}p_{j,t}] \\
&= \frac{\beta}{4}\sum_{i=1}^n\sum_{j\neq i}(\mathbb{E}[W_i^2] + \mathbb{E}[W_j^2] - 2\mathbb{E}[W_iW_j] - \mathbb{E}[V_i^2] - \mathbb{E}[V_j^2] + 2\mathbb{E}[V_iV_j])\mathbb{E}[p_{i,t}p_{j,t}] \\
&= \frac{\beta}{4}\sum_{i=1}^n\sum_{j\neq i}(\sigma_{ij}^{\mathbf{W}} - \sigma_{ij}^{\mathbf{V}})\mathbb{E}[p_{i,t}p_{j,t}] = \frac{\beta}{4}\sum_{i=1}^n\sum_{j=1}^n(\sigma_{ij}^{\mathbf{W}} - \sigma_{ij}^{\mathbf{V}})\mathbb{E}[p_{i,t}p_{j,t}],
\end{aligned}$$

where in the fourth line we used the fact that

$$\sum_{i=1}^n\sum_{j\neq i}\mathbb{E}[W_i^2]\mathbb{E}[p_{i,t}p_{j,t}] = \sum_{j=1}^n\sum_{i\neq j}\mathbb{E}[W_i^2]\mathbb{E}[p_{i,t}p_{j,t}] = \sum_{i=1}^n\sum_{j\neq i}\mathbb{E}[W_j^2]\mathbb{E}[p_{j,t}p_{i,t}].$$

(c) From the result of Part(b), we know that if $\sigma_{ij}^{\mathbf{W}} \leq \sigma_{ij}^{\mathbf{V}}$ for $1 \leq i, j \leq n$, then $\varphi'(t) \geq 0$ for all $t \in [0, 1]$. Therefore,

$$\mathbb{E}[F_\beta(\mathbf{V})] = \varphi(0) \leq \varphi(1) = \mathbb{E}[F_\beta(\mathbf{W})], \qquad \forall \beta > 0.$$

(d) From the results of Problem 3, we know that $F_\beta(\mathbf{V}) \le \max_i V_i + \frac{\log n}{\beta}$ for all $\beta > 0$ and $\lim_{\beta\to\infty} F_\beta(\mathbf{V}) = \max_i V_i$. Thus by the dominated convergence theorem,

$$\mathbb{E}\Big[\max_i V_i\Big] = \mathbb{E}\Big[\lim_{\beta\to\infty} F_\beta(\mathbf{V})\Big] = \lim_{\beta\to\infty} \mathbb{E}[F_\beta(\mathbf{V})].$$

By the same token,

$$\mathbb{E}\Big[\max_i W_i\Big] = \mathbb{E}\Big[\lim_{\beta\to\infty} F_\beta(\mathbf{W})\Big] = \lim_{\beta\to\infty} \mathbb{E}[F_\beta(\mathbf{W})].$$

From the result of Part(c),

$$\mathbb{E}\Big[\max_i V_i\Big] \le \mathbb{E}\Big[\max_i W_i\Big].$$

5 Gaussian comparison in action

(a) Using the hint, let $\mathbf{V} = \frac{\sigma_*}{\sqrt{2}}\mathbf{Y}$ where $\mathbf{Y} = (Y_1, \dots, Y_n)$ are i.i.d. standard Gaussian. Then

$$\sigma_{ij}^{\mathbf{W}} \ge \sigma_* = \sigma_{ij}^{\mathbf{V}}, \qquad \forall\, 1 \le i, j \le n.$$

Thus from Problem 4,

$$\mathbb{E}\Big[\max_i W_i\Big] \ge \mathbb{E}\Big[\max_i V_i\Big] = \frac{\sigma_*}{\sqrt{2}}\mathbb{E}\Big[\max_i Y_i\Big] \ge \frac{\sigma_*}{2}\sqrt{\log n}.$$

(b) To apply Slepian's lemma, we need to show

$$\mathbb{E}[(Y_{v,w} - Y_{v',w'})^2] \le \mathbb{E}[(Z_{v,w} - Z_{v',w'})^2], \qquad \forall (v, w), (v', w') \in S^{n-1} \times S^{m-1}.$$

From the definition of matrix $\mathbf{A}$ and the fact that $\|v\| = \|w\| = \|v'\| = \|w'\| = 1$,

$$\mathbb{E}[(Y_{v,w} - Y_{v',w'})^2] = \mathbb{E}\left[\Big(\sum_{i,j}(v_i w_j - v_i' w_j')A_{ij}\Big)^2\right] = \sum_{i,j}(v_i w_j - v_i' w_j')^2 = 2 - 2\langle v, v'\rangle\langle w, w'\rangle.$$

Similarly, we have

$$\begin{aligned}\mathbb{E}[(Z_{v,w} - Z_{v',w'})^2] &= \mathbb{E}\left[\Big(\sum_i (v_i - v_i')Y_i + \sum_j (w_j - w_j')Y_j\Big)^2\right] \\ &= \sum_i (v_i - v_i')^2 + \sum_j (w_j - w_j')^2 = 4 - 2\langle v, v'\rangle - 2\langle w, w'\rangle.\end{aligned}$$

Thus

$$\mathbb{E}[(Z_{v,w} - Z_{v',w'})^2] - \mathbb{E}[(Y_{v,w} - Y_{v',w'})^2] = 2(1 - \langle v, v'\rangle)(1 - \langle w, w'\rangle \ge 0.$$

By Slepian's lemma,

$$\begin{aligned}\mathbb{E}\|\mathbf{A}\| &\le \mathbb{E}\sup_{v\in S^{n-1}, w\in S^{m-1}}\big(\langle v, \mathbf{Y}\rangle + \langle w, \mathbf{Y}\rangle\big) \le \mathbb{E}\sup_{v\in S^{n-1}}\langle v, \mathbf{Y}\rangle + \mathbb{E}\sup_{w\in S^{m-1}}\langle w, \mathbf{Y}'\rangle \\ &\le \mathbb{E}\|\mathbf{Y}\| + \mathbb{E}\|\mathbf{Y}'\| = \sqrt{n} + \sqrt{m}.\end{aligned}$$