507A Final Solutions[1]

1. Question 1

This problem proves a dominated convergence theorem for conditional expectation. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. Let $X, Y, X_1, X_2, X_3, \ldots$ be $\mathcal{F}$-measurable random variables on $(\Omega, \mathcal{F}, \mathbf{P})$. Assume that for all $n \geq 1$, $|X_n| \leq Y$ almost surely, and $\mathbf{E}\,|Y| < \infty$. Let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. Assume that $X_1, X_2, \ldots$ converges almost surely to X. Conclude that

$$\mathbf{E}(X_1|\mathcal{G}), \mathbf{E}(X_2|\mathcal{G}), \ldots$$

converges almost surely to $\mathbf{E}(X|\mathcal{G})$.

You can freely use the conditional monotone convergence theorem: if $0 \leq X_1 \leq X_2 \leq \cdots$ are $\mathcal{F}$-measurable random variables that converge almost surely to X, then

$$\lim_{n\to\infty} \mathbf{E}(X_n|\mathcal{G}) = \mathbf{E}(X|\mathcal{G}).$$

(Hint: formulate and prove a conditional version of Fatou's Lemma, i.e. under some assumptions, show

$$\mathbf{E}(\liminf_{n\to\infty} X_n|\mathcal{G}) \leq \liminf_{n\to\infty} \mathbf{E}(X_n|\mathcal{G}).\;\;)$$

Solution. Let $0 \leq Y_1 \leq Y_2 \leq \ldots$ be $\mathcal{F}$-measurable random variables. We will show that

$$\mathbf{E}(\liminf_{n\to\infty} Y_n|\mathcal{G}) \leq \liminf_{n\to\infty} \mathbf{E}(Y_n|\mathcal{G}). \qquad (*)$$

For any $k \geq 1$, let $Z_k := \inf_{n\geq k} Y_k$. Note that $0 \leq Z_1 \leq Z_2 \leq \cdots$, and these random variables are increasing pointwise, so their pointwise limit exists almost surely (since a monotone sequence of real numbers converges, possible to ∞), so if $Z := \lim_{n\to\infty} Z_n = \liminf_{n\to\infty} Y_n$, then by the conditional Monotone Convergence Theorem,

$$\lim_{n\to\infty} \mathbf{E}(Z_n|\mathcal{G}) = \mathbf{E}(Z|\mathcal{G}) = \mathbf{E}(\liminf_{n\to\infty} Y_n|\mathcal{G}). \qquad (**)$$

Finally, by definition of Z_k, we have $Z_k \leq Y_n$ for all $n \geq k$, so monotonicity of condition expectation implies that $\mathbf{E}(Z_k|\mathcal{G}) \leq \mathbf{E}(Y_n|\mathcal{G})$ for all $n \geq k$, so that $\mathbf{E}(Z_k|\mathcal{G}) \leq \inf_{n\geq k} \mathbf{E}(Y_k|\mathcal{G})$. Letting $k \to \infty$, we get

$$\lim_{k\to\infty} \mathbf{E}(Z_k|\mathcal{G}) \leq \lim_{k\to\infty} \inf_{n\geq k} \mathbf{E}(Y_k|\mathcal{G}) = \liminf_{n\to\infty} \mathbf{E}(Y_n|\mathcal{G}).$$

Combining this equality with $(**)$ proves $(*)$.

Now, we apply $(*)$ to the sequences $(Y + X_n)_{n\geq 1}$ and $(Y - X_n)_{n\geq 1}$ separately (noting that both sequences are nonnegative by assumption) to get

$$\mathbf{E}(Y + X|\mathcal{G}) \leq \liminf_{n\to\infty} \mathbf{E}(Y + X_n|\mathcal{G}).$$

$$\mathbf{E}(Y - X|\mathcal{G}) \leq \liminf_{n\to\infty} \mathbf{E}(Y - X_n|\mathcal{G}).$$

Subtracting $\mathbf{E}(Y|\mathcal{G})$ from these inequalities and combining them,

$$\limsup_{n\to\infty} \mathbf{E}(X_n|\mathcal{G}) = -\liminf_{n\to\infty} \mathbf{E}(-X_n|\mathcal{G}) \leq \mathbf{E}(X|\mathcal{G}) \leq \liminf_{n\to\infty} \mathbf{E}(X_n|\mathcal{G}).$$

Since $\limsup \geq \liminf$, we see that all above quantities are equal and the limit exists, i.e.

$$\lim_{n\to\infty} \mathbf{E}(X_n|\mathcal{G})\mathbf{E}(X|\mathcal{G}).$$

[1]November 23, 2020,

2. Question 2

Let H be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$. As usual we denote $\|h\| := \langle h, h\rangle^{1/2}$ for all $h \in H$. We say a Hilbert space H is *separable* if there exists a countable set $h_1, h_2, \ldots \in H$ such that $\|h_i\| = 1$ for all $i \geq 1$, $\langle h_i, h_j \rangle = 0$ for all $i, j \geq 1$ with $i \neq j$, and such that, $\forall$ $h \in H$,

$$\lim_{n\to\infty} \left\| h - \sum_{i=1}^{n} \langle h, h_i \rangle h_i \right\| = 0.$$

Let K be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle'$. Assume that H and K are each separable. Show that H and K are linearly isometric. That is, $\exists$ a linear function $T\colon H \to K$ such that T is injective, T is surjective, and $\forall$ $g, h \in H$, we have

$$\langle T(g), T(h) \rangle' = \langle g, h \rangle.$$

Solution. Let $h_1, h_2, \ldots \subseteq H$ be the countable set defined in terms of separability for H. Let $k_1, k_2, \ldots \subseteq K$ be the countable set defined in terms of separability for K. Define a function $T\colon H \to K$ so that $T(h_i) = k_i$ for all $i \geq 1$. Below, we write $\lim_{n\to\infty}$ to denote limits with respect to the metric topology on the Hilbert space K. For any $h \in H$, we can then formally define

$$T(h) := \lim_{n\to\infty} \sum_{i=1}^{n} \langle h, h_i \rangle k_i \sum_{i=1}^{\infty} \langle h, h_i \rangle T(h_i) = \sum_{i=1}^{\infty} \langle h, h_i \rangle k_i.$$

From the Pythagorean Theorem, we have

$$\left\| \sum_{i=1}^{\infty} \langle h, h_i \rangle k_i \right\|^2 = \sum_{i=1}^{\infty} \langle h, h_i \rangle^2 = \|h\|^2, \qquad \forall\, h \in H. \qquad (*)$$

So, T is well-defined for all $h \in H$. Evidently T is also linear since it is the limit of a sum of linear functions.

T is injective since, if $T(h) = 0$, then the left side of $(*)$ is zero, so by $(*)$ we have $h = 0$.

T is surjective since, for any $k \in K$, the definition of separability says that

$$k = \sum_{i=1}^{\infty} \langle k, k_i \rangle k_i.$$

So, given this k, define

$$h := \sum_{i=1}^{\infty} \langle k, k_i \rangle h_i.$$

Then $h \in H$ exists since, again by the Pythagorean theorem, $\sum_{i=1}^{\infty} \langle k, k_i \rangle^2 = \|k\|^2 < \infty$, so $h \in H$ exists by completeness of the Hilbert space H. By definition of H, we have $T(h) = k$, so that T is surjective. Finally, to prove the isometry property, we have (using continuity of the inner product and the definition of each orthonormal basis)

$$\begin{aligned}\langle T(g), T(h)\rangle' &= \Big\langle \sum_{i=1}^{\infty}\langle g, h_i\rangle k_i, \sum_{j=1}^{\infty}\langle h, h_j\rangle k_j\Big\rangle' = \sum_{i,j=1}^{\infty}\langle g, h_i\rangle\langle h, h_j\rangle\langle k_i, k_j\rangle' \\ &= \sum_{i=1}^{\infty}\langle g, h_i\rangle\langle h, h_i\rangle = \sum_{i,j=1}^{\infty}\langle g, h_i\rangle\langle h, h_j\rangle\langle h_i, h_j\rangle' \\ &= \Big\langle \sum_{i=1}^{\infty}\langle g, h_i\rangle h_i, \sum_{j=1}^{\infty}\langle h, h_j\rangle h_j\Big\rangle = \langle g, h\rangle.\end{aligned}$$

3. Question 3

For any natural number n and a parameter $0 < p < 1$, define an Erdös-Renyi graph on n vertices with parameter p to be a random graph (V, E) on a (deterministic) vertex set $V = \{1, \ldots, n\}$ of n vertices (thus (V, E) is a random variable taking values in the discrete space of all $2^{\binom{n}{2}}$ possible undirected graphs one can place on V) such that the events $\{i, j\} \in E$ for unordered pairs with $i, j \in V$ are independent and each edge occurs with probability p.

A *stable set* in a graph is a subset S of the vertices of the graph such that no two of the vertices in S are connected by an edge.

For any $n \geq 1$, let X_n be the cardinality of the largest cardinality stable set in an Erdös-Renyi random graph on n vertices with parameter $p = 1/2$. Show that

$$\mathbf{P}(|X_n - \mathbf{E}X_n| > t) \leq 2e^{-\frac{t^2}{2n}}, \qquad \forall\, n \geq 1, \quad \forall\, t > 0.$$

(Hint: construct an increasing sequence of σ-algebras $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n = 2^{\Omega}$ such that $\mathcal{F}_j$ corresponds to all subsets of edges going between vertices $\{1, \ldots, j\}$, and define $Y_j := \mathbf{E}(X_n|\mathcal{F}_j)$, $\forall\, 0 \leq j \leq n$. Show that $Y_0, Y_1, \ldots, Y_n$ is a martingale.)

Solution. From the definition of Y_j and the tower property, we have, for all $0 \leq j \leq n-1$,

$$\mathbf{E}(Y_{j+1}|\mathcal{F}_|) = \mathbf{E}(\mathbf{E}(X_n|\mathcal{F}_j)|\mathcal{F}_j) = \mathbf{E}(X_n|\mathcal{F}_j) = Y_j$$

Also, $0 \leq Y_j \leq n$ for all $0 \leq j \leq n$, so that $\mathbf{E}\,|Y_j| < \infty$ for all $0 \leq j \leq n$, so $Y_0, \ldots, Y_n$ is a martingale.

By the definition of Y_j, we have $|Y_{j+1} - Y_j| \leq 1$ for all $0 \leq j \leq n-1$. To see this, let G be an undirected graph on n vertices (with no self-edges) and let G_j be the set of all undirected graphs G' on n vertices such that G and G' are the same induced subgraph on the vertices $\{1, \ldots, j\}$. Then $Y_j(G)$ is the average size of the largest stable set in G_j. And $Y_{j+1}(G)$ is the average size of the largest stable set in G_{j+1}. Note that $G_{j+1} \subseteq G_j$, and for any $G' \in G_{j+1}$, if we average over all graphs $G'' \in G_j$ such that G' and G'' have the same induced subgraph on $\{1, \ldots, j\}$, this corresponds to adding or subtracting a single vertex from the largest stable set of G'. So, by the total expectation theorem, we conclude that $|Y_{j+1}(G) - Y_j(G)| \leq 1$ for all $0 \le j \le n-1$. That is,

$$\begin{aligned}
Y_j(G) &= \frac{\sum_{G' \in G_j} X_n(G')\mathbf{P}(G')}{\sum_{G' \in G_j} \mathbf{P}(G')} \\
&= \frac{\sum_{S \subseteq \{1,\ldots,j\}} \sum_{G'' \colon \{i,j+1\} \in G'', \forall i \in S} X_n(G'')\mathbf{P}(G'')}{\sum_{G' \in G_j} \mathbf{P}(G')} \\
&= \frac{\sum_{S \subseteq \{1,\ldots,j\}} \frac{\sum_{G'' \colon \{i,j+1\} \in G'', \forall i \in S} X_n(G'')\mathbf{P}(G'')}{\sum_{G'' \in G_{j+1}} \mathbf{P}(G'')} \cdot \sum_{G'' \in G_{j+1}} \mathbf{P}(G'')}{\sum_{G' \in G_j} \mathbf{P}(G')} \\
&= 2^{-j} \sum_{S \subseteq \{1,\ldots,j\}} \frac{\sum_{G'' \colon \{i,j+1\} \in G'', \forall i \in S} X_n(G'')\mathbf{P}(G'')}{\sum_{G'' \in G_{j+1}} \mathbf{P}(G'')}
\end{aligned}$$

So,

$$\begin{aligned}
Y_j(G) - Y_{j+1}(G) &= 2^{-j} \sum_{S \subseteq \{1,\ldots,j\}} \Big(\frac{\sum_{G'' \colon \{i,j+1\} \in G'', \forall i \in S} X_n(G'')\mathbf{P}(G'')}{\sum_{G'' \in G_{j+1}} \mathbf{P}(G'')} - Y_{j+1}(G) \Big) \\
&= 2^{-j} \sum_{S \subseteq \{1,\ldots,j\}} \frac{\sum_{G'' \colon \{i,j+1\} \in G'', \forall i \in S} X_n(G'')\mathbf{P}(G'') - \sum_{G'' \in G_{j+1}} X_n(G'')\mathbf{P}(G'')}{\sum_{G'' \in G_{j+1}} \mathbf{P}(G'')}
\end{aligned}$$

The numerator can be arranged so that the difference of the X_n terms is at most one, so $|Y_j(G) - Y_{j+1}(G)| \le 1$. We conclude by Azuma's inequality.

4. Question 4

Let $X_0, X_1, \ldots$ be the symmetric simple random walk on the integers $\mathbf{Z}$. For any $k \in \mathbf{Z}$, let $\mathbf{P}_k$ denote the probability law on this random walk such that $X_0 := k$. Let $T_0 := \min\{n \ge 1 \colon X_n = 0\}$ be the first positive time that the random walk takes the value 0. For any positive integers j, k, r, show:

$$\mathbf{P}_k(T_0 < r,\ X_r = j) = \mathbf{P}_k(X_r = -j).$$

Solution. If the random walk takes the value zero, then after this visit to zero, the walk is independent of its previous movements, and we can then treat the walk as if it started at 0. That is, for any integers $0 < s < r$ and j,

$$\mathbf{P}_k(X_{T_0+(r-s)} = j \,|\, T_0 = s,\ X_s = 0) = \mathbf{P}_0(X_{r-s} = j).$$

Rearranging and simplifying,

$$\mathbf{P}_k(T_0 = s,\ X_r = j) = \mathbf{P}_k(T_0 = s)\mathbf{P}_0(X_{r-s} = j). \qquad (*)$$

When the walk starts at zero, it has equal probability of reaching j or $-j$ (that is, the random walk is symmetric with respect to zero). So, the right side is equal to

$$\mathbf{P}_k(T_0 = s)\mathbf{P}_0(X_{r-s} = -j) \overset{(*)}{=} \mathbf{P}_k(T_0 = s,\ X_r = -j).$$

Summing over all $1 \leq s < r$, and combining this equality with $(*)$ (with $j > 0$),

$$\mathbf{P}_k(T_0 < r,\, X_r = j) = \mathbf{P}_k(T_0 < r,\, X_r = -j) = \mathbf{P}_k(X_r = -j).$$

The last equality follows since a random walk started from $k > 0$ must pass through 0 before reaching a negative integer $-j$. That is, given $X_0 = k$, the event $X_r = -j$ is contained in the event $T_0 < r$.

5. Question 5

Let a, b be positive integers. Suppose there are c votes cast by c people in an election. Candidate 1 gets a votes and candidate 2 gets b votes. (So $c = a + b$.) Assume $a > b$. The votes are counted one by one. The votes are counted in a uniformly random ordering, and we would like to keep a running tally of who is currently winning. Suppose the first candidate eventually wins the election. We ask: with what probability will candidate 1 always be ahead in the running tally of who is currently winning the election? You are asked to show that the answer is $\frac{a-b}{a+b}$.

To prove this, for any positive integer k, let S_k be the number of votes for candidate 1, minus the number of votes for candidate 2, after k votes have been counted. Then, define $X_k := S_{c-k}/(c-k)$. Show that $X_0, X_1, \ldots$ is a martingale with respect to the σ-algebras generated by the (reversed) ordering of $S_c, S_{c-1}, S_{c-2}, \ldots$. Then, let T such that $T = \min\{0 \leq k \leq c \colon X_k = 0\}$, or $T = c-1$ if no such k exists. Apply the Optional Stopping theorem to X_T.

Solution.

$$\begin{aligned}
&\mathbf{E}(X_{k+1} - X_k \,|\, S_{c-k} = s_{c-k}, \ldots, S_c = s_c, X_0 = x_0) \\
&\quad = \mathbf{E}\left(\frac{S_{c-k-1}}{c-k-1} - \frac{s_{c-k}}{c-k} \,|\, S_{c-k} = s_{c-k}, \ldots, S_c = s_c, X_0 = x_0\right) \\
&\quad = \mathbf{E}\left(\frac{S_{c-k-1} - S_{c-k} + s_{c-k}}{c-k-1} - \frac{s_{c-k}}{c-k} \,|\, S_{c-k} = s_{c-k}, \ldots, S_c = s_c, X_0 = x_0\right)
\end{aligned}$$

Given that $S_{c-k} = s_{c-k}$, $c-k$ votes have been counted, and there are s_{c-k} more votes for candidate 1 than candidate 2, among the first $c-k$ counted votes. So, if there are x votes for candidate 1 after $c-k$ votes have been counted, then there are $c-k-x$ votes for candidate 2. And we know $x = s_{c-k} + (c-k-x)$ by definition of s_{c-k}, so $2x = s_{c-k} + c - k$, and $x = (1/2)(s_{c-k} + c - k)$.

Given that $S_{c-k} = s_{c-k}$, the expected value of $S_{c-k-1} - S_{c-k}$ is the change in the vote tally, with all $c-k$ votes equally likely to be chosen. (That is, we can think of counting the ballots "in reverse." Given the value of S_{c-k}, we can think of $c-k$ votes as sitting in a pile of "counted" votes. Then $S_{c-k-1} - S_{c-k}$ can be found by choosing any of these $c-k$ votes uniformly at random, and placing this vote into the pile of "uncounted" votes.) That is, this (conditional) expected value of $S_{c-k-1} - S_{c-k}$ is

$$(-1) \cdot \frac{x}{c-k} + (1)\frac{c-k-x}{c-k} = \frac{-2x + c - k}{c-k} = -\frac{s_{c-k}}{c-k}.$$

Therefore,

$$\begin{aligned}
&\mathbf{E}(X_{k+1}-X_k \,|\, S_{c-k}=s_{c-k},\ldots,S_c=s_c, X_0=x_0)\\
&\quad=\mathbf{E}\left(\frac{-\frac{s_{c-k}}{c-k}+s_{c-k}}{c-k-1}-\frac{s_{c-k}}{c-k}\,|\, S_{c-k}=s_{c-k},\ldots,S_c=s_c,X_0=x_0\right)\\
&\quad=s_{c-k}\left(\frac{-\frac{1}{c-k}+1}{c-k-1}-\frac{1}{c-k}\right)=s_{c-k}\left(\frac{\left(\frac{c-k-1}{c-k}\right)}{c-k-1}-\frac{1}{c-k}\right)=0.
\end{aligned}$$

We conclude that $X_1, X_2, \ldots$ is a martingale. Then

$$\mathbf{P}(\text{candidate 1 always leads the vote tally})=\mathbf{E}X_T=\mathbf{E}X_0=\mathbf{E}S_c/c=\frac{a-b}{a+b}.$$

(Since $a>b$, if the first vote is counted for candidate 2, then X_t will be zero for some t. So, $X_T=1$ if and only if $S_k>0$ for all $1\leq k\leq c$. And $X_T=0$ otherwise. So, $\mathbf{E}X_T=\mathbf{P}(S_k>0)$ for all $1\leq k\leq c$. That is, $\mathbf{E}X_T$ is the probability that candidate 1 always leads the vote tally.)

6. Question 6

For any $a=(a_1,a_2,\ldots)\in\mathbf{R}^{\mathbf{N}}$, $b=(b_1,b_2,\ldots)\in\mathbf{R}^{\mathbf{N}}$, define $\langle a,b\rangle:=\sum_{i=1}^{\infty}a_ib_i$ (if it exists). Let ℓ_2 denote the Hilbert space $\{a\in\mathbf{R}^{\mathbf{N}}\colon \langle a,a\rangle<\infty\}$ with respect to the inner product $\langle\cdot,\cdot\rangle$. (You can freely use that ℓ_2 is a Hilbert space.)

This problem proves Grothendieck's inequality: $\exists$ a constant $k>0$ such that, $\forall$ $n\geq 1$, $\forall$ real $n\times n$ matrices $(c_{ij})_{1\leq i,j\leq n}$,

$$\max_{\substack{x^{(1)},\ldots,x^{(n)},y^{(1)},\ldots,y^{(n)}\in\ell_2\\ \|x^{(i)}\|=\|y^{(i)}\|=1}}\sum_{i,j=1}^{n}c_{ij}\langle x^{(i)},y^{(j)}\rangle\leq k\cdot\max_{\varepsilon_1,\ldots,\varepsilon_n,\delta_1,\ldots,\delta_n\in\{-1,1\}}\sum_{i,j=1}^{n}c_{ij}\varepsilon_i\delta_j.$$

To prove this inequality, do the following. Let Γ denote the left side of the inequality, and let Δ denote the right side (without the k constant). We need to show that $\Gamma=O(\Delta)$.

- First, let $g_1,g_2,\ldots$ be a fixed sequence of i.i.d. standard Gaussians. Using a previous problem, we may replace ℓ_2 with the Hilbert space $H:=\{\sum_{i=1}^{\infty}a_ig_i\colon (a_i)_{i=1}^{\infty}\in\ell_2\}$ with respect to the inner product $\langle X,Y\rangle:=\mathbf{E}XY$ for all $X,Y\in H$. (You can freely use that H is itself a Hilbert space.)
- Fix $m>0$ and let $X\in H$. Denote $X_{\leq m}:=X1_{|X|\leq m}$. Fix $\varepsilon>0$. Pick $X^{(1)},\ldots,X^{(n)}$ and $Y^{(1)},\ldots,Y^{(n)}$ that come within ε of achieving the maximum in the definition of Γ. Compare $\sum_{i,j=1}^{n}c_{ij}\mathbf{E}X^{(i)}Y^{(j)}$ to $\sum_{i,j=1}^{n}c_{ij}\mathbf{E}X_{\leq m}^{(i)}Y_{\leq m}^{(j)}$ by adding and subtracting some terms inside the sum. You should be able to show that
$$\Gamma-\varepsilon\leq m^2\Delta+100e^{-m^2/4}\Gamma.$$
(Hint: Try dividing and multiplying some terms by m, and try dividing and multiplying some terms by $\max_{k=1,\ldots,n}\|X^{(k)}-X_{\leq m}^{(k)}\|$ or $\max_{k=1,\ldots,n}\|Y^{(k)}-Y_{\leq m}^{(k)}\|$.)
- To get the $e^{-m^2/4}$ bound, note that $X^{(k)}-X_{\leq m}^{(k)}=X^{(k)}1_{|X^{(k)}|>m}$, so you should be able to bound its L_2 norm.

Solution. Define

$$\Gamma := \max_{\substack{x^{(1)},\ldots,x^{(n)},y^{(1)},\ldots,y^{(n)}\in\ell_2\\ \|x^{(i)}\|=\|y^{(i)}\|=1}} \sum_{i,j=1}^{n} c_{ij}\langle x^{(i)}, y^{(j)}\rangle, \quad \Delta := \max_{\varepsilon_1,\ldots,\varepsilon_n,\delta_1,\ldots,\delta_n\in\{-1,1\}} \sum_{i,j=1}^{n} c_{ij}\varepsilon_i\delta_j.$$

We wish to show $\Gamma = O(\Delta)$. Let $\{g_i\}_{i=1}^{\infty}$ be standard iid gaussians on some probability space (Ω,μ). Define a space H by

$$H := \left\{\sum_{i=1}^{\infty} a_i g_i \colon \{a_i\}_{i=1}^{\infty}\in\ell_2\right\}$$

Note that H is a Hilbert space with respect to the inner product $\langle X,Y\rangle := \mathbf{E}(XY)$. Now, let $X^{(1)},\ldots,X^{(n)},Y^{(1)},\ldots,Y^{(n)}\in H$ satisfy $\sum_{i,j=1}^{n} a_{ij}\mathbf{E}(X^{(i)}Y^{(j)}) \geq \Gamma-\varepsilon$, $\mathbf{E}(X^{(i)})^2, \mathbf{E}(X^{(j)})^2 \leq 1$, $\varepsilon < \Gamma/4$. (Since all separable Hilbert spaces are isomorphic, it suffices to prove the theorem where H replaces ℓ_2).

We now use truncation. For $X\in L_2(\mu)$, and $m>0$ define $X_{\leq m} := X1_{|X|\leq m}$. Now,

$$\begin{aligned}
&\Gamma-\varepsilon \leq \sum_{i,j=1}^{n} a_{ij}\mathbf{E}(X^{(i)}Y^{(j)}) \quad \text{, by choice of } X^{(i)}, Y^{(j)}\\
&= \sum_{i,j=1}^{n} a_{ij}\mathbf{E}(X^{(i)}_{\leq m}Y^{(j)}_{\leq m}) + \sum_{i,j=1}^{n} a_{ij}\mathbf{E}((X^{(i)}-X^{(i)}_{\leq m})Y^{(j)}) + \sum_{i,j=1}^{n} a_{ij}\mathbf{E}(X^{(i)}_{\leq m}(Y^{(j)}-Y^{(j)}_{\leq m}))\\
&= M^2\sum_{i,j=1}^{n} a_{ij}\mathbf{E}\left(\frac{X^{(i)}_{\leq m}}{M}\frac{Y^{(j)}_{\leq m}}{M}\right) + \max_{k=1,\ldots,n}\|X^{(k)}-X^{(k)}_{\leq m}\|\sum_{i,j=1}^{n} a_{ij}\mathbf{E}\left(\frac{X^{(i)}-X^{(i)}_{\leq m}}{\max_{k=1,\ldots,n}\|X^{(k)}-X^{(k)}_{\leq m}\|}Y^{(j)}\right)\\
&\quad + \max_{k=1,\ldots,n}\|Y^{(k)}-Y^{(k)}_{\leq m}\|\sum_{i,j=1}^{n} a_{ij}\mathbf{E}\left(X^{(i)}_{\leq m}\frac{Y^{(j)}-Y^{(j)}_{\leq m}}{\max_{k=1,\ldots,n}\|Y^{(k)}-Y^{(k)}_{\leq m}\|}\right)\\
&\leq M^2\Delta + 2Ce^{-M^2/4}\Gamma \quad \text{, by definition of } \Delta,\ \Gamma
\end{aligned}$$

For the final inequality, we need two observations. First, the quantity Δ achieves its maximum over the larger set $\varepsilon_i,\delta_j\in[-1,1]$. To see this, observe that $\sum a_{ij}\varepsilon_i\delta_j$ is a linear function of the variables ε_i,δ_j. Therefore, this linear function achieves its maximum on the extreme points of the set $\{|\varepsilon_i|\leq 1, |\delta_j|\leq 1\}$. These extreme points are given by $\varepsilon_i,\delta_j\in\{\pm 1\}$. So, since $X^{(i)}_{\leq m}/M, Y^{(i)}_{\leq m}/M \in [-1,1]$, we take the pointwise maximum of $\sum_{i,j} a_{i,j}(X^{(i)}_{\leq m}Y^{(j)}_{\leq m})/(M\cdot M)$. (Recall that the $X^{(i)}_{\leq m}, Y^{(i)}_{\leq m}$ are real valued functions on a probability space.) The pointwise maximum is bounded by Δ, so we then take the expected values, giving our desired bound of $M^2\Delta$.

For our second observation, note that each $X_k = \sum b_i g_i$ satisfies $\mathbf{E}X_k^2 = \sum_{i\geq 1} b_i^2 \leq 1$ (by assumption). In particular, the distribution of X_k is $\|b\|_{\ell_2}\, g$ where g is a standard gaussian.

So, using Theorem 1.86 from the notes ("integration by parts") we have

$$\begin{aligned}
\|X^{(i)} - X^{(i)}_{\le m}\|^2 &= \mathbf{E}(X^{(i)} - X^{(i)}_{\le m})^2 = 2\int_M^\infty (t-M)\mathbb{P}(X^{(i)} > t)dt \\
&= 2\int_0^\infty t\mathbb{P}(X^{(i)} > t+M)dt = 2\int_0^\infty te^{-t^2/2}e^{-M^2/2}e^{-Mt}dt \\
&\le 2e^{-M^2/2}\int_0^\infty te^{-t^2/2}dt \le C^2e^{-M^2/2}
\end{aligned}$$

Since $\varepsilon < \Gamma/4$, if we let M large so that $2Ce^{-M^2/4} < 1/2$, we get $\Gamma = O(\Delta)$.