Supplement to "Kernel Measures of Conditional Dependence"

This supplementary material provides the technical proofs omitted in the submitted paper. The reference and equation numbers in this material follow the paper.

A Proof of Theorem 2

Proof of Theorem 2. Without loss of generality we can assume $\phi(0) = 1$. From the positive definiteness of k, we have $|\phi(z)|^2 \leq \phi(0)^2 = 1$. Recall that the RKHS associated with k has an explicit expression

$$\mathcal{H} = \Big\{ f \in L^2(\mathbb{R}^m) \,\Big|\, \int \frac{|\tilde{f}(u)|^2}{\tilde{\phi}(u)} du < \infty \Big\}, \tag{12}$$

where $\tilde{\phi}$ and $\tilde{f}$ are the Fourier transforms of ϕ and f, respectively.

Let P be an arbitrary Borel probability on $\mathbb{R}^m$, and $\xi \in \mathbb{R}^m$ be arbitrary. Since the Fourier transform of $e^{-\sqrt{-1}\xi^T z}\phi(z/\tau)$ is equal to $\tilde{\phi}(\tau(u+\xi))$, from Eq. (12) and the assumption of the theorem the function $e^{-\sqrt{-1}\xi^T z}\phi(z/\tau)$ belongs to $\mathcal{H}$ for any $\tau > \tau_0$. Thus, the bounded convergence theorem guarantees

$$E_P\big|e^{-\sqrt{-1}\xi^T z} - e^{-\sqrt{-1}\xi^T z}\phi(z/\tau)\big|^2 \longrightarrow 0 \qquad (\tau \to \infty).$$

This implies that we have only to prove that the linear hull of $\{e^{-\sqrt{-1}\xi^T z} \mid \xi \in \mathbb{R}^m\}$ is dense in $L^2(P)$.

Let f be an arbitrary function in $L^2(P)$. We can assume f is continuously differentiable with a compact support, because those functions are dense in $L^2(P)$. Let $\varepsilon > 0$ be arbitrary, $M = \sup_{x\in\mathbb{R}} |f(x)|$, and A be a positive number such that $[-A, A]^m$ contains the support of f and $P([-A, A]^m) > 1 - \varepsilon/4M^2$. It is well known [13, Theorem II.8] that the series of functions

$$f_N(z) = \sum_{n_1=-N}^{N} \cdots \sum_{n_m=-N}^{N} c_n e^{\frac{\pi\sqrt{-1}}{A} n^T z} \qquad (N \in \mathbb{N})$$

($n = (n_1, \ldots, n_m) \in (\mathbb{N} \cup \{0\})^m$) with the Fourier coefficients

$$c_n = \frac{1}{(2A)^m} \int_{[-A,A]^m} f(z)\, e^{-\frac{\pi\sqrt{-1}}{A} n^T z}\, dz$$

converges uniformly to $f(z)$ on $[-A, A]^m$, as $N \to \infty$. Thus, for sufficiently large N, we have $|f(z) - f_N(z)|^2 < \varepsilon/2$ on $[-A, A]^m$, and the periodicity of $f_N(z)$ ensures $\sup_{z\in\mathbb{R}^m} |f_N(z)|^2 < (M + \sqrt{\varepsilon/2})^2 < 2M^2$. We obtain $E_P|f - f_N|^2 < \varepsilon$, which completes the proof. □

B Proof of Theorem 4

We start with a lemma.

Lemma 6. *Assume that the kernels and the random variables satisfy (A-1), and $\mathcal{H}_\mathcal{Z} + \mathbb{R}$ is dense in $L^2(P_Z)$. Then,*

$$\langle g, \Sigma_{YY}^{1/2} V_{YZ} V_{ZX} \Sigma_{XX}^{1/2} f \rangle_{\mathcal{H}_\mathcal{Y}} = E\Big[\Big(E[f(X)|Z] - E[f(X)]\Big)\Big(E[g(X)|Z] - E[g(X)]\Big)\Big].$$

Proof. Since it is known [8] that Σ_{ZZ} is Hilbert-Schmidt under the assumption (A-1), there exists a CONS $\{\phi_i\}_{i=1}^\infty$ of $\mathcal{H}_\mathcal{Z}$ such that $\Sigma_{ZZ}\phi_i = \lambda_i \phi_i$ with an eigenvalue $\lambda_i \geq 0$. Let $I_+ = \{i \in \mathbb{N} \mid \lambda_i > 0\}$, and define

$$\tilde{\phi}_i = \frac{1}{\sqrt{\lambda_i}}(\phi_i - E[\phi(Z)])$$

for $i \in I_+$. Because $\mathcal{R}(V_{ZY})$ and $\mathcal{R}(V_{ZX})$ are orthogonal to $\mathcal{N}(\Sigma_{ZZ})$, we have

$$
\begin{aligned}
&\langle g, \Sigma_{YY}^{1/2} V_{YZ} V_{ZX} \Sigma_{XX}^{1/2} f \rangle_{\mathcal{H}_\mathcal{Y}} \\
&= \sum_{i=1}^{\infty} \big\langle \phi_i, V_{ZY}\Sigma_{YY}^{1/2} g \big\rangle_{\mathcal{H}_\mathcal{Y}} \big\langle \phi_i, V_{ZX}\Sigma_{XX}^{1/2} f \big\rangle_{\mathcal{H}_\mathcal{X}} \\
&= \sum_{i \in I_+} \langle \phi_i, V_{ZY}\Sigma_{YY}^{1/2} g \rangle_{\mathcal{H}_\mathcal{Y}} \langle \phi_i, V_{ZX}\Sigma_{XX}^{1/2} f \rangle_{\mathcal{H}_\mathcal{X}} \\
&= \sum_{i \in I_+} \Big\langle \frac{1}{\sqrt{\lambda_i}} \phi_i, \Sigma_{ZY} g \Big\rangle_{\mathcal{H}_\mathcal{Y}} \Big\langle \frac{1}{\sqrt{\lambda_i}} \phi_i, \Sigma_{ZX} f \Big\rangle_{\mathcal{H}_\mathcal{X}} \\
&= \sum_{i \in I_+} \big(\tilde{\phi}_i, E[g(Y)|Z] - E[g(Y)] \big)_{L_2(P_\mathcal{Z})} \big(\tilde{\phi}_i, E[f(X)|Z] - E[f(X)] \big)_{L_2(P_\mathcal{Z})}.
\end{aligned}
$$

Obviously, $\{\tilde{\phi}_i\}_{i \in I_+}$ is an orthonormal system in $L_2(P_\mathcal{Z})$. Furthermore, from the assumption that $\mathcal{H}_\mathcal{Z} + \mathbb{R}$ is dense in $L^2(P_Z)$, the system $\{\tilde{\phi}_i\}_{i \in I_+} \cup \{1\}$ is a CONS in $L^2(P_Z)$. This implies

$$
\begin{aligned}
&\langle g, \Sigma_{YY}^{1/2} V_{YZ} V_{ZX} \Sigma_{XX}^{1/2} f \rangle_{\mathcal{H}_\mathcal{Y}} \\
&= \big(E[g(Y)|Z] - E[g(Y)], E[f(X)|Z] - E[f(X)] \big)_{L_2(P_\mathcal{Z})} \\
&\qquad - \big(1, E[g(Y)|Z] - E[g(Y)] \big)_{L_2(P_\mathcal{Z})} \big(1, E[f(X)|Z] - E[f(X)] \big)_{L_2(P_\mathcal{Z})} \\
&= E_Z\Big[\big(E[g(Y)|Z] - E[g(Y)] \big) \big(E[f(X)|Z] - E[f(X)] \big) \Big],
\end{aligned}
$$

which proves the lemma. □

Using the above lemma, Theorem 4 is proved as follows.

Proof of Theorem 4. Let $\{\phi_i\}_{i=1}^{\infty}$ and $\{\psi_j\}_{j=1}^{\infty}$ be complete orthonormal systems of $\mathcal{H}_\mathcal{X}$ and $\mathcal{H}_\mathcal{Y}$, respectively, consisting of the eigenfunctions of Σ_{XX} and Σ_{YY} with $\Sigma_{XX}\phi_i = \lambda_i \phi_i$ ($\lambda_i \geq 0$) and $\Sigma_{YY}\psi_j = \nu_j \psi_j$ ($\nu_j \geq 0$). Then, we have the expansion

$$
\begin{aligned}
\|V_{YX|Z}\|_{HS}^2 = \sum_{i,j=1}^{\infty} \langle \psi_j, V_{YX}\phi_i \rangle_{\mathcal{H}_\mathcal{Y}}^2 - 2 \sum_{i,j=1}^{\infty} \langle \psi_j, V_{YX}\phi_i \rangle_{\mathcal{H}_\mathcal{Y}} \langle \psi_j, V_{YZ}V_{ZX}\phi_i \rangle_{\mathcal{H}_\mathcal{Y}} \\
+ \sum_{i,j=1}^{\infty} \langle \psi_j, V_{YZ}V_{ZX}\phi_i \rangle_{\mathcal{H}_\mathcal{Y}}^2. \qquad (13)
\end{aligned}
$$

Let $I_+^X = \{i \in \mathbb{N} \mid \lambda_i > 0\}$ and $I_+^Y = \{i \in \mathbb{N} \mid \nu_i > 0\}$. In a similar manner to the proof of Lemma 6, with the notations $\tilde{\phi}_i = (\phi_i - E[\phi_i(X)])/\sqrt{\lambda_i}$ and $\tilde{\psi}_j = (\psi_j - E[\psi_j(Y)])/\sqrt{\nu_j}$ for $i \in I_+^X$ and $j \in I_+^Y$, the first term of Eq. (13) is rewritten as

$$
\begin{aligned}
\sum_{i \in I_+^X, j \in I_+^Y} \langle \tilde{\psi}_j, \Sigma_{YX}\tilde{\phi}_i \rangle_{\mathcal{H}_\mathcal{Y}}^2 &= \sum_{i \in I_+^X, j \in I_+^Y} E_{YX}\big[\tilde{\psi}_j(Y)\tilde{\phi}_i(X)\big]^2 \\
&= \sum_{i \in I_+^X, j \in I_+^Y} \Big(\int\int_{\mathcal{X}\times\mathcal{Y}} p_{XY}(x,y)\tilde{\phi}_i(x)\tilde{\psi}_j(y) d\mu_\mathcal{X} d\mu_\mathcal{Y} \Big)^2.
\end{aligned}
$$

Let $\tilde{\phi}_0 = 1$ and $\tilde{\psi}_0 = 1$. From the assumption that $(\mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{Y}) + \mathbb{R}$ is dense in $L^2(P_X \otimes P_Y)$, the class $\{\tilde{\phi}_i\tilde{\psi}_j\}_{i \in I_+^X \cup \{0\}, j \in I_+^Y \cup \{0\}}$ is a CONS of $L_2(P_X \otimes P_Y)$. Thus, the last line of the above equations is further rewritten as

$$
\begin{aligned}
&\Big\| \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)} \Big\|_{L_2(P_X \otimes P_Y)}^2 - \sum_{i \in I_+^X} E\big[\tilde{\phi}_i(X)\big] - \sum_{j \in I_+^Y} E\big[\tilde{\psi}_j(Y)\big] - 1 \\
&= \int\int_{\mathcal{X}\times\mathcal{Y}} \frac{p_{XY}^2(x,y)}{p_X(x)p_Y(y)} d\mu_\mathcal{X} d\mu_\mathcal{Y} - 1.
\end{aligned}
$$

For the second term of Eq. (13), Lemma 6 implies

$$
\begin{aligned}
&\sum_{i,j=1}^{\infty} \langle \psi_j, V_{YX}\phi_i\rangle_{\mathcal{H}_\mathcal{Y}} \langle \psi_j, V_{YZ}V_{ZX}\phi_i\rangle_{\mathcal{H}_\mathcal{Y}} \\
&= \sum_{i\in I_+^X, j\in I_+^Y} \langle \tilde\psi_j, \Sigma_{YX}\tilde\phi_i\rangle_{\mathcal{H}_\mathcal{Y}} \langle \tilde\psi_j, \Sigma_{YY}^{1/2}V_{YZ}V_{ZX}\Sigma_{XX}^{1/2}\tilde\phi_i\rangle_{\mathcal{H}_\mathcal{Y}} \\
&= \sum_{i\in I_+^X, j\in I_+^Y} E\big[\tilde\psi_j(Y)\tilde\phi_i(X)\big] E\big[E[\tilde\psi_j(Y)|Z]\,E[\tilde\phi_i(X)|Z]\big] \\
&= \sum_{i\in I_+^X, j\in I_+^Y} \int\int_{\mathcal{X}\times\mathcal{Y}} \tilde\psi_j(y)\tilde\phi_i(x)p_{XY}(x,y)d\mu_\mathcal{X}d\mu_\mathcal{Y} \int\int_{\mathcal{X}\times\mathcal{Y}} \tilde\psi_j(y)\tilde\phi_i(x)p_{X\perp\!\!\!\perp Y|Z}(x,y)d\mu_\mathcal{X}d\mu_\mathcal{Y} \\
&= \sum_{i\in I_+^X, j\in I_+^Y} \left(\tilde\psi_j\tilde\phi_i, \frac{p_{XY}}{p_Xp_Y}\right)_{L_2(P_X\otimes P_Y)} \left(\tilde\phi_i\tilde\psi_j, \frac{p_{X\perp\!\!\!\perp Y|Z}}{p_Xp_Y}\right)_{L_2(P_X\otimes P_Y)}.
\end{aligned}
$$

By a similar argument to the case of the first term, the second term of Eq. (13) equals

$$
-2\left(\frac{p_{XY}}{p_Xp_Y}, \frac{p_{X\perp\!\!\!\perp Y|Z}}{p_Xp_Y}\right)_{L_2(P_X\otimes P_Y)} + 2,
$$

and the third term of Eq. (13) equals

$$
\left\|\frac{p_{X\perp\!\!\!\perp Y|Z}}{p_Xp_Y}\right\|^2_{L_2(P_X\otimes P_Y)} - 1.
$$

This completes the proof. □

C Proof of Theorem 5

Proof. From the expressions in Eq. (3) and Eq. (6), it is sufficient to prove $\|\widehat V_{YX}^{(n)} - V_{YX}\|_{HS} \to 0$ in probability. The proof is analogous to those of Lemma 6 and 7 in [4], though considering the Hilbert-Schmidt norm is more involved.

From the trivial decomposition

$$
\begin{aligned}
\|\widehat V_{YX}^{(n)} - V_{YX}\|_{HS} \le \|\widehat V_{YX}^{(n)} &- (\Sigma_{YY}+\varepsilon_n I)^{-1/2}\Sigma_{YX}(\Sigma_{XX}+\varepsilon_n I)^{-1/2}\big\|_{HS} \\
&+ \big\|(\Sigma_{YY}+\varepsilon_n I)^{-1/2}\Sigma_{YX}(\Sigma_{XX}+\varepsilon_n I)^{-1/2} - V_{YX}\big\|_{HS},
\end{aligned}
$$

it suffices to show

$$
\big\|\widehat V_{YX}^{(n)} - (\Sigma_{YY}+\varepsilon_n I)^{-1/2}\Sigma_{YX}(\Sigma_{XX}+\varepsilon_n I)^{-1/2}\big\|_{HS} = O_p(\varepsilon_n^{-3/2}n^{-1/2}), \tag{14}
$$

and

$$
\big\|(\Sigma_{YY}+\varepsilon_n I)^{-1/2}\Sigma_{YX}(\Sigma_{XX}+\varepsilon_n I)^{-1/2} - V_{YX}\big\|_{HS} = o(1). \tag{15}
$$

The operator on the left hand side of Eq. (14) is decomposed as

$$
\begin{aligned}
&\big\{(\widehat\Sigma_{YY}^{(n)}+\varepsilon_n I)^{-1/2} - (\Sigma_{YY}+\varepsilon_n I)^{-1/2}\big\}\widehat\Sigma_{YX}^{(n)}(\widehat\Sigma_{XX}^{(n)}+\varepsilon_n I)^{-1/2} \\
&\quad + (\Sigma_{YY}+\varepsilon_n I)^{-1/2}\big\{\widehat\Sigma_{YX}^{(n)} - \Sigma_{YX}\big\}(\widehat\Sigma_{XX}^{(n)}+\varepsilon_n I)^{-1/2} \\
&\quad + (\Sigma_{YY}+\varepsilon_n I)^{-1/2}\Sigma_{YX}\big\{(\widehat\Sigma_{XX}^{(n)}+\varepsilon_n I)^{-1/2} - (\Sigma_{XX}+\varepsilon_n I)^{-1/2}\big\}.
\end{aligned} \tag{16}
$$

From the equality

$$
A^{-1/2} - B^{-1/2} = A^{-1/2}\big(B^{3/2} - A^{3/2}\big)B^{-3/2} + (A-B)B^{-3/2},
$$

the first term of Eq. (16) is equal to

$$
\begin{aligned}
\big\{(\widehat\Sigma_{YY}^{(n)}+\varepsilon_n I)^{-1/2}\big\{\big(\Sigma_{YY}+\varepsilon_n I\big)^{3/2} - \big(\widehat\Sigma_{YY}^{(n)}+\varepsilon_n I\big)^{3/2}\big\} + \big(\widehat\Sigma_{YY}^{(n)} - \Sigma_{YY}\big)\big\} \\
\times \big(\widehat\Sigma_{YY}^{(n)}+\varepsilon_n I\big)^{-3/2}\widehat\Sigma_{YX}^{(n)}(\widehat\Sigma_{XX}^{(n)}+\varepsilon_n I)^{-1/2}.
\end{aligned}
$$

Using $\left\|(\widehat{\Sigma}_{YY}^{(n)}+\varepsilon_n I)^{-1/2}\right\| \le \frac{1}{\sqrt{\varepsilon_n}}$, $\left\|\left(\widehat{\Sigma}_{YY}^{(n)}+\varepsilon_n I\right)^{-1/2}\widehat{\Sigma}_{YX}^{(n)}\left(\widehat{\Sigma}_{XX}^{(n)}+\varepsilon_n I\right)^{-1/2}\right\| \le 1$, and Lemma 7 below, the HS norm of the above operator is bounded from above by

$$\frac{1}{\varepsilon_n}\Big\{\frac{3}{\sqrt{\varepsilon_n}}\max\big\{\|\Sigma_{YY}+\varepsilon_n I\|^{1/2}, \|\widehat{\Sigma}_{YY}^{(n)}+\varepsilon_n I\|^{1/2}\big\}+1\Big\}\|\widehat{\Sigma}_{YY}^{(n)}-\Sigma_{YY}\|_{HS}.$$

A similar bound also applies to the HS norm of the third term in Eq. (16). An upper bound on the HS norm of the second term is simply $\frac{1}{\varepsilon_n}\|\Sigma_{YX}-\widehat{\Sigma}_{YX}^{(n)}\|_{HS}$. Thus, Eq. (14) is obtained by noticing $\|\widehat{\Sigma}_{XX}^{(n)}\| = \|\Sigma_{XX}\|+o_p(1)$, $\|\widehat{\Sigma}_{YY}^{(n)}\| = \|\Sigma_{YY}\|+o_p(1)$, and the fact $\|\widehat{\Sigma}_{YX}^{(n)}-\Sigma_{YX}\|_{HS} = O_p(1/\sqrt{n})$ [4, Lemma 5].

To prove Eq. (15), take CONS's $\{\phi_i\}_{i=1}^\infty$ and $\{\psi_j\}_{j=1}^\infty$ for $\mathcal{H}_\mathcal{X}$ and $\mathcal{H}_\mathcal{Y}$, respectively, in the same manner as in the proof of Theorem 4. We have

$$\begin{aligned}
&\big\|(\Sigma_{YY}+\varepsilon_n I)^{-1/2}\Sigma_{YX}(\Sigma_{XX}+\varepsilon_n I)^{-1/2}-V_{YX}\big\|_{HS}^2\\
&=\sum_{i,j=1}^\infty\big\langle\psi_j,\big\{(\Sigma_{YY}+\varepsilon_n I)^{-1/2}\Sigma_{YX}(\Sigma_{XX}+\varepsilon_n I)^{-1/2}-V_{YX}\big\}\phi_i\big\rangle_{\mathcal{H}_\mathcal{Y}}^2\\
&=\sum_{i,j=1}^\infty\left\{\sqrt{\frac{\lambda_i}{\lambda_i+\varepsilon_n}}\sqrt{\frac{\nu_i}{\nu_i+\varepsilon_n}}-1\right\}^2\langle\psi_j,V_{YX}\phi_i\rangle_{\mathcal{H}_\mathcal{X}}^2.
\end{aligned}$$

Because each summand in the last line is upper bounded by $\langle\psi_j,V_{YX}\phi_i\rangle_{\mathcal{H}_\mathcal{Y}}^2$ and the series $\sum_{i,j=1}^\infty\langle\psi_j,V_{YX}\phi_i\rangle_{\mathcal{H}_\mathcal{Y}}^2$ is finite from the assumption that V_{YX} is Hilbert-Schmidt, the dominated convergence theorem guarantees the interchangeability of the series and the limit $\varepsilon_n\to 0$; thus, we have

$$\begin{aligned}
&\lim_{\varepsilon\to 0}\big\|(\Sigma_{YY}+\varepsilon_n I)^{-1/2}\Sigma_{YX}(\Sigma_{XX}+\varepsilon_n I)^{-1/2}-V_{YX}\big\|_{HS}^2\\
&\qquad=\sum_{i,j=1}^\infty\lim_{\varepsilon_n\to 0}\left\{\sqrt{\frac{\lambda_i}{\lambda_i+\varepsilon_n}}\sqrt{\frac{\nu_i}{\nu_i+\varepsilon_n}}-1\right\}^2\langle\psi_j,V_{YX}\phi_i\rangle_{\mathcal{H}_\mathcal{X}}^2 \quad = 0.
\end{aligned}$$

This shows Eq. (15) and completes the proof. □

Lemma 7. *Suppose A and B are positive, self-adjoint, Hilbert-Schmidt operators on a Hilbert space. Then,*

$$\|A^{3/2}-B^{3/2}\|_{HS}\le 3\big(\max\{\|A\|,\|B\|\}\big)^{1/2}\,\|A-B\|_{HS}.$$

Proof. Without loss of generality we can assume $\max\{\|A\|,\|B\|\}=1$. Then, we have $O\le A,B\le I$. Define functions f and g on $\{z\in\mathbb{C}\mid |z|\le 1\}$ by $f(z)=(1-z)^{3/2}$ and $g(z)=(1-z)^{1/2}$. Let

$$f(z)=\sum_{n=1}^\infty b_n z^n \qquad\text{and}\qquad g(z)=\sum_{n=0}^\infty c_n z^n$$

be the power series expansions. They converge absolutely for $|z|\le 1$. Using the facts $b_0=1$, $b_1=-\frac{3}{2}$, and $b_n>0$ ($n\ge 2$), the inequality

$$\begin{aligned}
\sum_{n=0}^N |b_n| &= 1+\frac{3}{2}+\sum_{n=2}^N b_n = 1+\frac{3}{2}+\lim_{x\uparrow 1}\sum_{n=2}^N b_n x^n\\
&\le 1+\frac{3}{2}+\lim_{x\uparrow 1}\Big\{f(x)-1+\frac{3}{2}\Big\} \;= 3
\end{aligned}$$

is obtained for all N. The bound $\sum_{n=0}^\infty|c_n|\le 2$ can be proved similarly. Obviously, $|nb_n|=|c_n|$ for $n\ge 2$ and $|b_1|=\frac{3}{2}|c_0|$.

Define a series of operators $\{D_N\}_{N=1}^\infty$ by

$$D_N=\sum_{n=0}^N b_n\big((I-A)^n-(I-B)^n\big).$$

Expansion of the right hand side shows that D_N is Hilbert-Schmidt. The Hilbert-Schmidt norm of D_N satisfies

$$\|D_N\|_{HS} \le \sum_{n=0}^{N} |b_n| \|(I-A)^n - (I-B)^n\|_{HS}.$$

We can prove by induction that $\|(I-A)^n - (I-B)^n\|_{HS} \le n\|A-B\|_{HS}$ holds for all n; in fact, this follows from the inequality

$$\begin{aligned}
&\|(I-A)^n - (I-B)^n\|_{HS} \\
&= \left\| \big((I-A)-(I-B)\big)(I-A)^{n-1} + (I-B)\big((I-A)^{n-1} - (I-B)^{n-1}\big) \right\|_{HS} \\
&\le \|(I-A)^{n-1}\| \, \|A-B\|_{HS} + \|I-B\| \left\|(I-A)^{n-1} - (I-B)^{n-1}\right\|_{HS}.
\end{aligned}$$

Thus, we obtain

$$\|D_N\|_{HS} \le \sum_{n=0}^{N} n|b_n| \|A-B\|_{HS} \le \frac{3}{2} \sum_{n=0}^{N} |c_n| \|A-B\|_{HS},$$

which implies D_N is a Cauchy sequence in the Hilbert space of the Hilbert-Schmidt operators. Thus, there is a Hilbert-Schmidt operator D_* such that $\|D_N - D_*\|_{HS} \to 0$. On the other hand, from the fact $O \le I-A, I-B \le I$, in the expression

$$D_N = \sum_{n=0}^{N} b_n (I-A)^n - \sum_{n=0}^{N} b_n (I-B)^n,$$

the two terms in the right hand side converge in operator norm to $A^{3/2}$ and $B^{3/2}$, respectively; hence $D_N \to A^{3/2} - B^{3/2}$ in operator norm. This necessarily means $D_* = A^{3/2} - B^{3/2}$, and we have

$$\|A^{3/2} - B^{3/2}\|_{HS} = \lim_{N\to\infty} \|D_N\|_{HS} \le \frac{3}{2} \lim_{N\to\infty} \sum_{n=0}^{N} |c_n| \|A-B\|_{HS} = 3\|A-B\|_{HS}.$$

$\square$