7 Proof For Stochastic HSIC

We define the Cross Covariance operator as

$$
\begin{aligned}
C_{X,Y} &= E_{P(X,Y)}[(\phi(X) - \mathcal{U}_X)(\phi(Y) - \mathcal{U}_Y)^T] && (65)\\
&= E_{P(X,Y)}[(\phi(X)\phi(Y)^T] - \mathcal{U}_X\mathcal{U}_Y && (66)
\end{aligned}
$$

To shorten the notation, we will remove $P(X,Y)$ without losing its meaning. We next obtain the norm of this operator

$$
\begin{aligned}
||C_{X,Y}||_H &= \mathrm{Tr}\left(E[(\phi(X)\phi(Y)^T] - \mathcal{U}_X\mathcal{U}_Y^T\right)^T \left(E[(\phi(X)\phi(Y)^T] - \mathcal{U}_X\mathcal{U}_Y^T\right) && (67)\\
&= \underbrace{\mathrm{Tr}\left(E[(\phi(X)\phi(Y)^T]^T E[(\phi(X)\phi(Y)^T]\right)}_{\mathcal{A}} - \underbrace{\mathrm{Tr}\left(E[(\phi(X)\phi(Y)^T]^T \mathcal{U}_X\mathcal{U}_Y^T\right)}_{\mathcal{B}} && (68)\\
&\quad - \underbrace{\mathrm{Tr}\left(\mathcal{U}_Y\mathcal{U}_X^T E[(\phi(X)\phi(Y)^T]\right)}_{\mathcal{C}} + \underbrace{\mathrm{Tr}\left(\mathcal{U}_Y\mathcal{U}_X^T\mathcal{U}_X\mathcal{U}_Y^T\right)}_{\mathcal{D}} && (69)
\end{aligned}
$$

We can now look at the terms individually.

$$
\begin{aligned}
\mathcal{A} &= \mathrm{Tr}\left(E[(\phi(X)\phi(Y)^T]^T E[(\phi(X)\phi(Y)^T]\right) && (70)\\
&\approx \mathrm{Tr}\left[\left(\frac{1}{n}\sum_i \phi(x_i)\phi(y_i)^T\right)^T \left(\frac{1}{n}\sum_j \phi(x_j)\phi(y_j)^T\right)\right] && (71)\\
&\approx \frac{1}{n^2}\,\mathrm{Tr}\left[\left(\sum_i \phi(y_i)\phi(x_i)^T\right)\left(\sum_j \phi(x_j)\phi(y_j)^T\right)\right] && (72)\\
&\approx \frac{1}{n^2}\,\mathrm{Tr}\left[\left(\sum_i\sum_j \phi(y_i)\phi(x_i)^T\phi(x_j)\phi(y_j)^T\right)\right] && (73)\\
&\approx \frac{1}{n^2}\sum_i\sum_j \mathrm{Tr}\left(\phi(y_i)\phi(x_i)^T\phi(x_j)\phi(y_j)^T\right) && (74)\\
&\approx \frac{1}{n^2}\sum_i\sum_j \mathrm{Tr}\left(\phi(x_i)^T\phi(x_j)\phi(y_j)^T\phi(y_i)\right) && (75)\\
&\approx \frac{1}{n^2}\sum_i\sum_j k(x_i,x_j)k(y_i,y_j) && (76)
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{B} &= \mathrm{Tr}\left(E[(\phi(X)\phi(Y)^T]^T \mathcal{U}_X \mathcal{U}_Y^T\right) && (77)\\
&\approx \mathrm{Tr}\left(\frac{1}{n}\sum_i [\phi(x_i)\phi(y_i)^T]^T \mathcal{U}_X \mathcal{U}_Y^T\right) && (78)\\
&\approx \frac{1}{n}\,\mathrm{Tr}\left(\left[\sum_i \phi(y_i)\phi(x_i)^T\right] \mathcal{U}_X \mathcal{U}_Y^T\right) && (79)\\
&\approx \frac{1}{n}\,\mathrm{Tr}\left(\left[\sum_i \phi(y_i)\phi(x_i)^T\right]\left[\frac{1}{n}\sum_k \phi(x_k)\right]\left[\frac{1}{n}\sum_l \phi(y_l)\right]^T\right) && (80)\\
&\approx \frac{1}{n^3}\,\mathrm{Tr}\left(\left[\sum_i \phi(y_i)\phi(x_i)^T\right]\left[\sum_k \phi(x_k)\right]\left[\sum_l \phi(y_l)\right]^T\right) && (81)\\
&\approx \frac{1}{n^3}\,\mathrm{Tr}\left(\left[\sum_i\sum_k \phi(y_i)\phi(x_i)^T\phi(x_k)\right]\left[\sum_l \phi(y_l)\right]^T\right) && (82)\\
&\approx \frac{1}{n^3}\,\mathrm{Tr}\left(\left[\sum_l \phi(y_l)\right]^T\left[\sum_i\sum_k \phi(y_i)\phi(x_i)^T\phi(x_k)\right]\right) && (83)\\
&\approx \frac{1}{n^3}\,\mathrm{Tr}\left(\left[\sum_i\sum_k\sum_l \phi(y_l)^T\phi(y_i)\phi(x_i)^T\phi(x_k)\right]\right) && (84)\\
&\approx \frac{1}{n^3}\,\mathrm{Tr}\left(\left[\sum_i\sum_k\sum_l k(y_l,y_i)k(x_k,x_i)\right]\right) && (85)\\
& && (86)
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{C} &= \mathrm{Tr}\left(\mathcal{U}_Y \mathcal{U}_X^T E[(\phi(X)\phi(Y)^T]\right) && (87)\\
&\approx \frac{1}{n}\,\mathrm{Tr}\left(\left[\frac{1}{n}\sum_k \phi(y_k)\right]\left[\frac{1}{n}\sum_l \phi(x_l)^T\right]\left[\sum_j \phi(x_j)\phi(y_j)^T\right]\right) && (88)\\
&\approx \frac{1}{n^3}\,\mathrm{Tr}\left(\left[\sum_k \phi(y_k)\right]\left[\sum_l \phi(x_l)^T\right]\left[\sum_j \phi(x_j)\phi(y_j)^T\right]\right) && (89)\\
&\approx \frac{1}{n^3}\,\mathrm{Tr}\left(\left[\sum_l\sum_j \phi(x_l)^T\phi(x_j)\phi(y_j)^T\right]\left[\sum_k \phi(y_k)\right]\right) && (90)\\
&\approx \frac{1}{n^3}\,\mathrm{Tr}\left(\left[\sum_l\sum_j\sum_k \phi(x_l)^T\phi(x_j)\phi(y_j)^T\phi(y_k)\right]\right) && (91)\\
&\approx \frac{1}{n^3}\sum_l\sum_j\sum_k k(x_l,x_j)k(y_j,y_k) && (92)
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{D} &= \mathrm{Tr}\left(\mathcal{U}_Y \mathcal{U}_X^T \mathcal{U}_X \mathcal{U}_Y^T\right) && (93)\\
&\approx \mathrm{Tr}\left(\left[\frac{1}{n}\sum_k \phi(y_k)\right]\left[\frac{1}{n}\sum_l \phi(x_l)^T\right]\left[\frac{1}{n}\sum_s \phi(x_s)\right]\left[\frac{1}{n}\sum_t \phi(y_t)^T\right]\right) && (94)\\
&\approx \frac{1}{n^4}\left[\sum_l \phi(x_l)^T\right]\left[\sum_s \phi(x_s)\right]\left[\sum_t \phi(y_t)^T\right]\left[\sum_k \phi(y_k)\right] && (95)\\
&\approx \frac{1}{n^4}\left[\sum_s\sum_l k(x_s,x_l)\right]\left[\sum_t\sum_k k(y_t,y_k)\right] && (96)\\
&\approx \frac{1}{n^4}\left[\sum_s\sum_l\sum_t\sum_k k(x_s,x_l)k(y_t,y_k)\right] && (97)
\end{aligned}
$$

Putting all the terms together, HSIC can alternatively be computed with the following formulation (which we can then apply SGD). Note that a batch defines i, j indices, and the samples should loop through the entire dataset with k, l, s, t.

$$||C_{X,Y}||_H = \frac{1}{n^2}\sum_i\sum_j k(x_i, x_j)k(y_i, y_j) - \frac{1}{n^3}\sum_i\sum_k\sum_l k(y_l, y_i)k(x_k, x_i) \tag{98}$$

$$- \frac{1}{n^3}\sum_l\sum_j\sum_k k(x_l, x_j)k(y_j, y_k) + \frac{1}{n^4}\left[\sum_s\sum_l\sum_t\sum_k k(x_s, x_l)k(y_t, y_k)\right] \tag{99}$$