Theoretical Proofs for Fast Higher-order Count Sketch: Algorithm and Application to Tensor Contraction Approximations

Xingyu Cao, Yipeng Liu *Senior Member, IEEE*, Jiani Liu, Zhen Long, Ce Zhu *Fellow, IEEE*

Abstract

This supplement contains the derivation of FHCS for generalized tensors and the proofs of Proposition 1 and Corollary 1.

I. Derivation of FHCS for General Tensors

For more convenience of representation, we assume the index of arrays starts at 0. Given $\mathcal{T} = \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \cdots \circ \mathbf{u}^{(N)} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, assign $l = \sum_{n=1}^{N} i_n \prod_{j=1}^{n-1} I_j$, we have $\text{vec}(\mathcal{T})_l = \mathbf{u}_{i_1}^{(1)} \cdots \mathbf{u}_{i_N}^{(N)}$. By definition of HCS, we have:

$$\begin{aligned}
&\sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} \text{vec}(\mathcal{T})_l \mathbf{s}_1(i_1) \cdots \mathbf{s}_N(i_N) w^{t(\mathbf{h}_1(i_1)+\cdots+\mathbf{h}_N(i_N))} \\
&= \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} \mathbf{u}_{i_1}^{(1)} \cdots \mathbf{u}_{i_N}^{(N)} \mathbf{s}_1(i_1) \cdots \mathbf{s}_N(i_N) w^{t(\mathbf{h}_1(i_1)+\cdots+\mathbf{h}_N(i_N))} \\
&= \sum_{i_1} \mathbf{u}_{i_1}^{(1)} \mathbf{s}_1(i_1) w^{t\mathbf{h}_1(i_1)} \cdots \sum_{i_N} \mathbf{u}_{i_N}^{(N)} \mathbf{s}_N(i_N) w^{t\mathbf{h}_N(i_N)} \\
&= \mathbb{F}(\text{CS}_1(\mathbf{u}^{(1)})) * \cdots * \mathbb{F}(\text{CS}_N(\mathbf{u}^{(N)})).
\end{aligned} \tag{1}$$

Let $\mathbf{s}_{N+1}(l) = \mathbf{s}_1(i_1) \cdots \mathbf{s}_N(i_N)$, $\mathbf{h}_{N+1}(l) = \mathbf{h}_1(i_1) + \cdots + \mathbf{h}_N(i_N)$, we have

$$\begin{aligned}
&\sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} \text{vec}(\mathcal{T})_l \mathbf{s}_1(i_1) \cdots \mathbf{s}_N(i_N) w^{t(\mathbf{h}_1(i_1)+\cdots+\mathbf{h}_2(i_N))} \\
&= \sum_{l} \text{vec}(\mathcal{T})_l \mathbf{s}_{N+1}(l) w^{t\mathbf{h}_{N+1}(l)} \\
&= \mathbb{F}(\text{FHCS}(\mathcal{T})).
\end{aligned} \tag{2}$$

Therefore, we have

$$\text{FHCS}(\mathcal{T}) = \mathbb{F}^{-1}(\mathbb{F}(\text{CS}_1(\mathbf{u}^{(1)})) * \cdots * \mathbb{F}(\text{CS}_N(\mathbf{u}^{(N)}))) \tag{3}$$

which completes the proof.

II. Proofs for Proposition 1 and Corollar 1

A. Proof for Proposition 1

First we prove consistency of FHCS. We proceed by proving a more general case, i.e., for any 3-order [1] tensors $\mathcal{M}$, $\mathcal{N}$ with the same size, $\langle \text{HCS}(\mathcal{M}), \text{HCS}(\mathcal{N}) \rangle$ is a consistent estimator of $\langle \mathcal{M}, \mathcal{N} \rangle$. To this end we prove it is unbiased with bounded variance.

- Unbiasedness

$$\begin{aligned}
\langle \text{HCS}(\mathcal{M}), \text{HCS}(\mathcal{N}) \rangle &= \sum_{t_1,t_2,t_3} \text{HCS}(\mathcal{M})_{t_1,t_2,t_3} \text{HCS}(\mathcal{N})_{t_1,t_2,t_3} \\
&= \sum_{t_1,t_2,t_3} \Big(\sum_{\substack{\mathbf{h}_1(i_1)=t_1 \\ \mathbf{h}_2(i_2)=t_2 \\ \mathbf{h}_3(i_3)=t_3}} \mathbf{s}_1(i_1)\mathbf{s}_2(i_2)\mathbf{s}_3(i_3)\mathcal{M}_{i_1,i_2,i_3} \Big) \Big(\sum_{\substack{\mathbf{h}_1(j_1)=t_1 \\ \mathbf{h}_2(j_2)=t_2 \\ \mathbf{h}_3(j_3)=t_3}} \mathbf{s}_1(j_1)\mathbf{s}_2(j_2)\mathbf{s}_3(j_3)\mathcal{N}_{j_1,j_2,j_3} \Big),
\end{aligned} \tag{4}$$

$$\begin{aligned}
\mathbb{E}_{\mathbf{h},\mathbf{s}} \langle \text{HCS}(\mathcal{M}), \text{HCS}(\mathcal{N}) \rangle = &\sum_{\substack{i_1,i_2,i_3 \\ j_1,j_2,j_3}} \mathbb{E}\left[\delta(\mathbf{h}_1(i_1),\mathbf{h}_1(j_1))\delta(\mathbf{h}_2(i_2),\mathbf{h}_2(j_2))\delta(\mathbf{h}_3(i_3),\mathbf{h}_3(j_3))\right] \\
&\mathbb{E}[\mathbf{s}_1(i_1)\mathbf{s}_2(i_2)\mathbf{s}_3(i_3)\mathbf{s}_1(j_1)\mathbf{s}_2(j_2)\mathbf{s}_3(j_3)]\mathcal{M}_{i_1,i_2,i_3}\mathcal{N}_{j_1,j_2,j_3}
\end{aligned} \tag{5}$$

[1]For simplicity we focus on 3-order tensors. The conclusion generalizes to tensors with any order.

Given $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$ are 2-wise independent, we have

$$\mathbb{E}[\mathbf{s}_1(i_1)\mathbf{s}_2(i_2)\mathbf{s}_3(i_3)\mathbf{s}_1(j_1)\mathbf{s}_2(j_2)\mathbf{s}_3(j_3)] = \begin{cases} 1,\ i_1 = j_1, i_2 = j_2, i_3 = j_3 \\ 0, \text{ otherwise} \end{cases} \tag{6}$$

Obviously $\mathbb{E}[\delta(\mathbf{h}_1(i_1), \mathbf{h}_1(j_1))]|_{i_1=j_1} = \mathbb{E}[\delta(\mathbf{h}_2(i_2), \mathbf{h}_2(j_2))]|_{i_2=j_2} = \mathbb{E}[\delta(\mathbf{h}_3(i_3), \mathbf{h}_3(j_3))]|_{i_3=j_3} = 1$. Hence

$$\mathbb{E}_{\mathbf{h},\mathbf{s}} \langle \mathrm{HCS}(\mathcal{M}), \mathrm{HCS}(\mathcal{N}) \rangle = \sum_{i_1,i_2,i_3} \mathcal{M}_{i_1,i_2,i_3}\mathcal{N}_{i_1,i_2,i_3} = \langle \mathcal{M}, \mathcal{N} \rangle \tag{7}$$

- Variance boundedness

$$\begin{aligned} \langle \mathrm{HCS}(\mathcal{M}), \mathrm{HCS}(\mathcal{N}) \rangle^2 = &\sum_{\substack{i_1,i_2,i_3 \\ j_1,j_2,j_3 \\ i'_1,i'_2,i'_3 \\ j'_1,j'_2,j'_3}} \delta(\mathbf{h}_1(i_1), \mathbf{h}_1(j_1))\delta(\mathbf{h}_2(i_2), \mathbf{h}_2(j_2))\delta(\mathbf{h}_3(i_3), \mathbf{h}_3(j_3)) \\ &\delta(\mathbf{h}_1(i'_1), \mathbf{h}_1(j'_1))\delta(\mathbf{h}_2(i'_2), \mathbf{h}_2(j'_2))\delta(\mathbf{h}_3(i'_3), \mathbf{h}_3(j'_3)) \\ &\mathbf{s}_1(i_1)\mathbf{s}_2(i_2)\mathbf{s}_3(i_3)\mathbf{s}_1(j_1)\mathbf{s}_2(j_2)\mathbf{s}_3(j_3)\mathbf{s}_1(i'_1)\mathbf{s}_2(i'_2)\mathbf{s}_3(i'_3)\mathbf{s}_1(j'_1)\mathbf{s}_2(j'_2)\mathbf{s}_3(j'_3) \\ &\mathcal{M}_{i_1,i_2,i_3}\mathcal{N}_{j_1,j_2,j_3}\mathcal{M}_{i'_1,i'_2,i'_3}\mathcal{N}_{j'_1,j'_2,j'_3} \end{aligned} \tag{8}$$

Denote $G_1 = \{i_1, j_1, i'_1, j'_1\}, G_2 = \{i_2, j_2, i'_2, j'_2\}, G_3 = \{i_3, j_3, i'_3, j'_3\}$. We say G_k has *best match* if $i_k = j_k, i'_k = j'_k$ for $k \in [3]$. then (8) can be grouped into following cases:

Case 1 G_1, G_2, G_3 all have best matches, then

$$\mathrm{S}_1 = \sum_{\substack{i_1,i_2,i_3 \\ i'_1,i'_2,i'_3}} \mathcal{M}_{i_1,i_2,i_3}\mathcal{N}_{i_1,i_2,i_3}\mathcal{M}_{i'_1,i'_2,i'_3}\mathcal{N}_{i'_1,i'_2,i'_3} = \langle \mathcal{M}, \mathcal{N} \rangle^2 \tag{9}$$

Case 2 One out of G_1, G_2, G_3 does not have best matches, take G_1 for example. Suppose $i_1 = i'_1 \neq j_1 = j'_1$, then

$$\begin{aligned} \mathrm{S}_2 &= \binom{3}{1}\binom{2}{1} \sum_{\substack{i_1,i_2,i_3 \\ j_1,i'_2,i'_3}} \mathbb{E}[\delta(\mathbf{h}_1(i_1), \mathbf{h}_1(j_1))^2]\mathcal{M}_{i_1,i_2,i_3}\mathcal{N}_{j_1,i_2,i_3}\mathcal{M}_{i_1,i'_2,i'_3}\mathcal{N}_{j_1,i'_2,i'_3} \\ &= \frac{6}{J} \sum_{\substack{i_1,i_2,i_3 \\ j_1,i'_2,i'_3}} \mathcal{M}_{i_1,i_2,i_3}\mathcal{N}_{j_1,i_2,i_3}\mathcal{M}_{i_1,i'_2,i'_3}\mathcal{N}_{j_1,i'_2,i'_3} \\ &= \frac{6}{J} \sum_{i_1,j_1} \langle \mathcal{M}(\mathbf{e}_{i_1}, \mathbf{I}, \mathbf{I}), \mathcal{N}(\mathbf{e}_{j_1}, \mathbf{I}, \mathbf{I}) \rangle^2 \\ &\leq \frac{6}{J} \sum_{i_1,j_1} \|\mathcal{M}(\mathbf{e}_{i_1}, \mathbf{I}, \mathbf{I})\|_F^2 \|\mathcal{N}(\mathbf{e}_{j_1}, \mathbf{I}, \mathbf{I})\|_F^2 \\ &= \frac{6}{J} \|\mathcal{M}\|_F^2 \|\mathcal{N}\|_F^2, \end{aligned} \tag{10}$$

where the inequality holds due to Cauchy-Schwartz inequality.

Case 3 One out of G_1, G_2, G_3 has best matches, take G_1 for example. Suppose $i_2 = i'_2 \neq j_2 = j'_2, i_3 = i'_3 \neq j_3 = j'_3$, then

$$\begin{aligned} \mathrm{S}_3 &= \binom{3}{1}\binom{2}{1}\binom{2}{1} \sum_{\substack{i_1,i_2,i_3 \\ j_2,j_3,i'_1}} \mathbb{E}[\delta(\mathbf{h}_2(i_2), \mathbf{h}_2(j_2))^2][\delta(\mathbf{h}_3(i_3), \mathbf{h}_3(j_3))^2]\mathcal{M}_{i_1,i_2,i_3}\mathcal{N}_{i_1,j_2,j_3}\mathcal{M}_{i'_1,i_2,i_3}\mathcal{N}_{i'_1,j_2,j_3} \\ &= \frac{12}{J^2} \sum_{\substack{i_1,i_2,i_3 \\ j_2,j_3,i'_1}} \mathcal{M}_{i_1,i_2,i_3}\mathcal{N}_{i_1,j_2,j_3}\mathcal{M}_{i'_1,i_2,i_3}\mathcal{N}_{i'_1,j_2,j_3} \\ &= \frac{12}{J^2} \sum_{\substack{i_2,i_3 \\ j_2,j_3}} \langle \mathcal{M}(\mathbf{I}, \mathbf{e}_{i_2}, \mathbf{e}_{i_3}), \mathcal{N}(\mathbf{I}, \mathbf{e}_{j_2}, \mathbf{e}_{j_3}) \rangle^2 \\ &\leq \frac{12}{J^2} \sum_{\substack{i_2,i_3 \\ j_2,j_3}} \|\mathcal{M}(\mathbf{I}, \mathbf{e}_{i_2}, \mathbf{e}_{i_3})\|_F^2 \|\mathcal{N}(\mathbf{I}, \mathbf{e}_{j_2}, \mathbf{e}_{j_3})\|_F^2 \\ &= \frac{12}{J^2} \|\mathcal{M}\|_F^2 \|\mathcal{N}\|_F^2 \end{aligned} \tag{11}$$

Case 4 None of G_1, G_2, G_3 has best matches. Suppose $i_1 = i_1' \neq j_1 = j_1', i_2 = i_2' \neq j_2 = j_2', i_3 = i_3' \neq j_3 = j_3'$, then

$$\begin{aligned}
\mathrm{S}_4 &= \binom{2}{1}\binom{2}{1}\binom{2}{1} \sum_{\substack{i_1,i_2,i_3 \\ j_1,j_2,j_3}} \mathbb{E}[\delta(\mathbf{h}_1(i_1),\mathbf{h}_1(j_1))^2][\delta(\mathbf{h}_2(i_2),\mathbf{h}_2(j_2))^2][\delta(\mathbf{h}_3(i_3),\mathbf{h}_3(j_3))^2]\mathcal{M}^2_{i_1,i_2,i_3}\mathcal{N}^2_{j_1,j_2,j_3} \\
&= \frac{8}{J^3} \sum_{\substack{i_1,i_2,i_3 \\ j_1,j_2,j_3}} \mathcal{M}^2_{i_1,i_2,i_3}\mathcal{N}^2_{j_1,j_2,j_3} \\
&= \frac{8}{J^3}\|\mathcal{M}\|_F^2\|\mathcal{N}\|_F^2
\end{aligned} \tag{12}$$

As a result, we have

$$\begin{aligned}
\mathbb{E}_{\mathbf{h},\mathbf{s}}[\langle \mathrm{HCS}(\mathcal{M}), \mathrm{HCS}(\mathcal{N})\rangle^2] &= \mathrm{S}_1 + \mathrm{S}_2 + \mathrm{S}_3 + \mathrm{S}_4 \\
&\leq \langle \mathcal{M}, \mathcal{N}\rangle^2 + \frac{6J^2 + 12J + 8}{J^3}\|\mathcal{M}\|_F^2\|\mathcal{N}\|_F^2
\end{aligned} \tag{13}$$

Therefore

$$\begin{aligned}
\mathbb{V}_{\mathbf{h},\mathbf{s}}[\langle \mathrm{HCS}(\mathcal{M}), \mathrm{HCS}(\mathcal{N})\rangle] &= \mathbb{E}_{\mathbf{h},\mathbf{s}}[\langle \mathrm{HCS}(\mathcal{M}), \mathrm{HCS}(\mathcal{N})\rangle^2] - \mathbb{E}_{\mathbf{h},\mathbf{s}}[\langle \mathrm{HCS}(\mathcal{M}), \mathrm{HCS}(\mathcal{N})\rangle]^2 \\
&< \frac{6J^2 + 12J + 8}{J^3}\|\mathcal{M}\|_F^2\|\mathcal{N}\|_F^2.
\end{aligned} \tag{14}$$

Combining (7, 14) the consistency is proved.

We further prove HCS computes a better estimator for tensor inner product than TS:

$$\begin{aligned}
\langle \mathrm{TS}(\mathcal{M}), \mathrm{TS}(\mathcal{N})\rangle^2 = \sum_{\substack{i_1,i_2,i_3 \\ j_1,j_2,j_3 \\ i_1',i_2',i_3' \\ j_1',j_2',j_3'}} & \delta(\mathbf{h}_1(i_1)+\mathbf{h}_2(i_2)+\mathbf{h}_3(i_3), \mathbf{h}_1(j_1)+\mathbf{h}_2(j_2)+\mathbf{h}_3(j_3) \;\%\; J) \\
& \delta(\mathbf{h}_1(i_1')+\mathbf{h}_2(i_2')+\mathbf{h}_3(i_3'), \mathbf{h}_1(j_1')+\mathbf{h}_2(j_2')+\mathbf{h}_3(j_3') \;\%\; J) \\
& \mathbf{s}_1(i_1)\mathbf{s}_2(i_2)\mathbf{s}_3(i_3)\mathbf{s}_1(j_1)\mathbf{s}_2(j_2)\mathbf{s}_3(j_3)\mathbf{s}_1(i_1')\mathbf{s}_2(i_2')\mathbf{s}_3(i_3')\mathbf{s}_1(j_1')\mathbf{s}_2(j_2')\mathbf{s}_3(j_3') \\
& \mathcal{M}_{i_1,i_2,i_3}\mathcal{N}_{j_1,j_2,j_3}\mathcal{M}_{i_1',i_2',i_3'}\mathcal{N}_{j_1',j_2',j_3'}.
\end{aligned} \tag{15}$$

Define G_1, G_2, G_3 similarly as abovementioned. Then (15) can be grouped into following cases:

Case 1 G_1, G_2, G_3 all have best matches, then

$$\mathrm{S}_1 = \sum_{\substack{i_1,i_2,i_3 \\ i_1',i_2',i_3'}} \mathcal{M}_{i_1,i_2,i_3}\mathcal{N}_{i_1,i_2,i_3}\mathcal{M}_{i_1',i_2',i_3'}\mathcal{N}_{i_1',i_2',i_3'} = \langle \mathcal{M}, \mathcal{N}\rangle^2 \tag{16}$$

Case 2 One out of G_1, G_2, G_3 does not have best matches, take G_1 for example. Suppose $i_1 = i_1' \neq j_1 = j_1'$, then

$$\begin{aligned}
\mathrm{S}_2 &= \binom{3}{1}\binom{2}{1} \sum_{\substack{i_1,i_2,i_3 \\ j_1,i_2',i_3'}} \mathbb{E}[\delta(\mathbf{h}_1(i_1),\mathbf{h}_1(j_1))^2]\mathcal{M}_{i_1,i_2,i_3}\mathcal{N}_{j_1,i_2,i_3}\mathcal{M}_{i_1,i_2',i_3'}\mathcal{N}_{j_1,i_2',i_3'} \\
&= \frac{6}{J} \sum_{\substack{i_1,i_2,i_3 \\ j_1,i_2',i_3'}} \mathcal{M}_{i_1,i_2,i_3}\mathcal{N}_{j_1,i_2,i_3}\mathcal{M}_{i_1,i_2',i_3'}\mathcal{N}_{j_1,i_2',i_3'} \\
&\leq \frac{6}{J}\|\mathcal{M}\|_F^2\|\mathcal{N}\|_F^2
\end{aligned} \tag{17}$$

Case 3 One out of G_1, G_2, G_3 has best matches, take G_1 for example. Suppose $i_2 = i_2' \neq j_2 = j_2', i_3 = i_3' \neq j_3 = j_3'$, then

$$\mathrm{S}_3 = \binom{3}{1}\binom{2}{1}\binom{2}{1} \sum_{\substack{i_1,i_2,i_3 \\ j_2,j_3,i_1'}} \mathbb{E}[\delta(\mathbf{h}_2(i_2)+\mathbf{h}_3(i_3), \mathbf{h}_2(j_2)+\mathbf{h}_3(j_3)\% m)]\mathcal{M}_{i_1,i_2,i_3}\mathcal{N}_{i_1,j_2,j_3}\mathcal{M}_{i_1',i_2,i_3}\mathcal{N}_{i_1',j_2,j_3}$$

From 2-wise independence of $\mathbf{h}_2$ and $\mathbf{h}_3$, we have

$$\mathbb{P}[\mathbf{h}_2(i_2)+\mathbf{h}_3(i_3) = t] = \begin{cases} \displaystyle\sum_{k=0}^{t} \mathbb{P}[\mathbf{h}_2(i_2) = k, \mathbf{h}_3(i_3) = t-k] = \frac{t+1}{J^2}, & t \leq J-1 \\ \displaystyle\sum_{k=t-J+1}^{J-1} \mathbb{P}[\mathbf{h}_2(i_2) = k, \mathbf{h}_3(i_3) = t-k] = \frac{2J-1-t}{J^2}, & t \geq J \end{cases}$$

$$\mathbb{P}[\mathbf{h}_2(j_2)+\mathbf{h}_3(j_3)\ \%\ J=t]=\begin{cases}\mathbb{P}[\mathbf{h}_2(j_2)+\mathbf{h}_3(j_3)=t]+\mathbb{P}[\mathbf{h}_2(j_2)+\mathbf{h}_3(j_3)=t+J]=\frac{1}{J}, & t<J-1\\ \mathbb{P}[\mathbf{h}_2(j_2)+\mathbf{h}_3(j_3)=t]=\frac{1}{J}, & t=J-1\end{cases}$$

i.e. $\mathbb{P}[\mathbf{h}_2(j_2)+\mathbf{h}_3(j_3)\ \%\ J=t]=\frac{1}{J}$ for $t=0,\cdots,J-1$. Hence

$$\mathbb{P}[\delta(\mathbf{h}_2(i_2)+\mathbf{h}_3(i_3),\mathbf{h}_2(j_2)+\mathbf{h}_3(j_3)\%J)=1]=\frac{1}{J^3}(1+2+\cdots+J)=\frac{1+J}{2J^2}$$

Therefore we have

$$\mathrm{S}_3=\frac{6(1+J)}{J^2}\sum_{\substack{i_1,i_2,i_3\\ j_2,j_3,i_1'}}\mathcal{M}_{i_1,i_2,i_3}\mathcal{N}_{i_1,j_2,j_3}\mathcal{M}_{i_1',i_2,i_3}\mathcal{N}_{i_1',j_2,j_3}\leq\frac{6(1+J)}{J^2}\|\mathcal{M}\|_F^2\|\mathcal{N}\|_F^2 \tag{18}$$

Case 4 None of G_1,G_2,G_3 has best matches. Suppose $i_1=i_1'\neq j_1=j_1', i_2=i_2'\neq j_2=j_2', i_3=i_3'\neq j_3=j_3'$, then

$$\mathrm{S}_4=\binom{2}{1}\binom{2}{1}\binom{2}{1}\sum_{\substack{i_1,i_2,i_3\\ j_1,j_2,j_3}}\mathbb{E}[\delta(\mathbf{h}_1(i_1)+\mathbf{h}_2(i_2)+\mathbf{h}_3(i_3),\mathbf{h}_1(j_1)+\mathbf{h}_2(j_2)+\mathbf{h}_3(j_3)\ \%\ J)]\mathcal{M}^2_{i_1,i_2,i_3}\mathcal{N}^2_{j_1,j_2,j_3} \tag{19}$$

Denote $\mathbf{h}_4(i_2,i_3)=\mathbf{h}_2(i_2)+\mathbf{h}_3(i_3)$. From **Case 3**:

$$\mathbb{P}[\mathbf{h}_4(i_2,i_3)=t]=\begin{cases}\frac{t+1}{J^2}, t\leq J-1\\ \frac{2J-1-t}{J^2}, t\geq J\end{cases}$$

Since $\mathbb{P}[\mathbf{h}_1(i_1)=t]=\frac{1}{J}$ for $t=0,\cdots,J-1$, we have

$$\begin{aligned}&\mathbb{P}[\mathbf{h}_1(i_1)+\mathbf{h}_2(i_2)+\mathbf{h}_3(i_3)=t]=\mathbb{P}[\mathbf{h}_1(i_1)+\mathbf{h}_4(i_2,i_3)=t]\\ &=\begin{cases}\sum_{k=0}^{t}\mathbb{P}[\mathbf{h}_4(i_2,i_3)=k,\mathbf{h}_1(i_1)=t-k]=\frac{(t+1)(t+2)}{2J^3}, & t\leq J-1\\ \sum_{k=t-J+1}^{t}\mathbb{P}[\mathbf{h}_4(i_2,i_3)=k,\mathbf{h}_1(i_1)=t-k]=\frac{-2t^2+6(J-1)t+9J-3J^2-4}{2J^3}, J\leq & t\leq 2J-2\\ \sum_{k=t-J+1}^{J-1}\mathbb{P}[\mathbf{h}_4(i_2,i_3)=k,\mathbf{h}_1(i_1)=t-k]=\frac{(3J-1-t)(3J-2-t)}{2J^3}, & t\geq 2J-1\end{cases}\end{aligned}$$

$$\begin{aligned}&\mathbb{P}[\mathbf{h}_1(i_1)+\mathbf{h}_2(i_2)+\mathbf{h}_3(i_3)\ \%\ J=t]=\mathbb{P}[\mathbf{h}_1(i_1)+\mathbf{h}_4(i_2,i_3)\ \%\ J=t]\\ &=\begin{cases}\sum_{k=0}^{2}\mathbb{P}[\mathbf{h}_4(i_2,i_3)+\mathbf{h}_1(i_1)=t+kJ]=\frac{1}{J}, & t\leq J-3\\ \sum_{k=0}^{1}\mathbb{P}[\mathbf{h}_4(i_2,i_3)+\mathbf{h}_1(i_1)=t+kJ]=\frac{1}{J}, & t\geq J-2\end{cases}\end{aligned}$$

i.e. $\mathbb{P}[\mathbf{h}_1(i_1)+\mathbf{h}_2(i_2)+\mathbf{h}_3(i_3)\ \%\ J=t]=\frac{1}{J}$ for $t=0,\cdots,J-1$. Hence we have

$$\begin{aligned}\mathbb{P}[\delta(\mathbf{h}_1(i_1)+\mathbf{h}_2(i_2)+\mathbf{h}_3(i_3),\mathbf{h}_1(j_1)+\mathbf{h}_2(j_2)+\mathbf{h}_3(j_3)\ \%\ J)=1]&=\frac{1}{J}(\frac{1}{J^3}+\frac{3}{J^3}+\frac{6}{J^3}+\cdots+\frac{1+J}{2J^2})\\ &=\frac{(J+1)(J+2)}{6J^3}\end{aligned}$$

Hence

$$\mathrm{S}_4=\binom{2}{1}\binom{2}{1}\binom{2}{1}\frac{(J+1)(J+2)}{6J^3}\|\mathcal{M}\|_F^2\|\mathcal{N}\|_F^2=\frac{4(J+1)(J+2)}{3J^3}\|\mathcal{M}\|_F^2\|\mathcal{N}\|_F^2. \tag{20}$$

As a result, we have

$$\begin{aligned}\mathbb{E}_{\mathbf{h},\mathbf{s}}[\langle\mathrm{TS}(\mathcal{M}),\mathrm{TS}(\mathcal{N})\rangle^2]&=\mathrm{S}_1+\mathrm{S}_2+\mathrm{S}_3+\mathrm{S}_4\\ &\leq\langle\mathcal{M},\mathcal{N}\rangle^2+\frac{40J^2+30J+8}{3J^3}\|\mathcal{M}\|_F^2\|\mathcal{N}\|_F^2\end{aligned} \tag{21}$$

Therefore

$$\begin{aligned}\mathbb{V}_{\mathbf{h},\mathbf{s}}[\langle\mathrm{TS}(\mathcal{M}),\mathrm{TS}(\mathcal{N})\rangle]&=\mathbb{E}_{\mathbf{h},\mathbf{s}}[\langle\mathrm{TS}(\mathcal{M}),\mathrm{TS}(\mathcal{N})\rangle^2]-\mathbb{E}_{\mathbf{h},\mathbf{s}}[\langle\mathrm{TS}(\mathcal{M}),\mathrm{TS}(\mathcal{N})\rangle]^2\\ &\leq\frac{40J^2+30J+8}{3J^3}\|\mathcal{M}\|_F^2\|\mathcal{N}\|_F^2.\end{aligned} \tag{22}$$

Clearly we have $\mathbb{V}_{\mathbf{h},\mathbf{s}}[\langle \mathrm{HCS}(\mathcal{M}), \mathrm{HCS}(\mathcal{N})\rangle] < \mathbb{V}_{\mathbf{h},\mathbf{s}}[\langle \mathrm{TS}(\mathcal{M}), \mathrm{TS}(\mathcal{N})\rangle]$, especially when the sketch dimension is small. The conclusion fits in FHCS

B. Proof for Corollary 1

(14) can be represented as $\mathbb{V}_{\mathbf{h},\mathbf{s}}[\langle \mathrm{HCS}(\mathcal{M}), \mathrm{HCS}(\mathcal{N})\rangle] = O(\frac{\|\mathcal{M}\|_F^2 \|\mathcal{N}\|_F^2}{J})$. By substituting $\mathcal{M} = \mathcal{T}$, $\mathcal{N} = \mathbf{u} \circ \mathbf{u} \circ \mathbf{u}$, and $\mathcal{M} = \mathcal{T}$, $\mathcal{N} = \mathbf{e}_i \circ \mathbf{u} \circ \mathbf{u}$, respectively, from Chebychev's inequality we prove the corollary immediately since $\mathbf{u}$ is unit vector and hence both $\mathbf{u} \circ \mathbf{u} \circ \mathbf{u}$ and $\mathbf{e}_i \circ \mathbf{u} \circ \mathbf{u}$ equals 1.