Network histograms and universality of blockmodel approximation

Sofia C. Olhede and Patrick J. Wolfe

Supporting Information

A Auxiliary results for the proof of Theorem 1

Throughout we assume that f is a symmetric function on $(0,1)^2$ that is also α-Hölder continuous for some $0 < \alpha \leq 1$, with $f \in \text{Hölder}^\alpha(M)$ meaning that

$$\sup_{(x,y)\neq(x',y')\in(0,1)^2} \frac{|f(x,y) - f(x',y')|}{|(x,y)-(x',y')|^\alpha} \leq M < \infty,$$

where $|\cdot|$ is the Euclidean metric on $\mathbb{R}^2$.

We also define a set of summation indices R_{ab}, which is the range of values of $i < j$ over which one must aggregate A_{ij} to retrieve $\bar{A}^*_{ab}$. We write

$$h_a := h\,\mathbb{I}(a < k) + (h+r)\,\mathbb{I}(a = k);$$

$$h^2_{ab} := |R_{ab}| = \begin{cases} h^2 & \text{if} \quad 1 \leq a < b < k, \\ \binom{h}{2} & \text{if} \quad 1 \leq a = b < k, \\ h \cdot (h+r) & \text{if} \quad 1 \leq a < b = k, \\ \binom{h+r}{2} & \text{if} \quad a = b = k. \end{cases}$$

Proposition 1 (Moments of $\bar{A}^*_{ab}$)**.** *Let $f \in$ Hölder$^\alpha(M)$ be symmetric on $(0,1)^2$, and let the labeling $\tilde{z}_i$ be determined from the latent vector ξ by*

$$\tilde{z}_i = \min\Big\{\lceil (i)^{-1}/h \rceil, k\Big\},$$

where $(i)^{-1}$ is the rank of ξ_i from smallest to largest. Thus (i) is defined as the index chosen so that $\xi_{(1)} \leq \xi_{(2)} \leq \cdots \leq \xi_{(n)}$, and $(i)^{-1}$ is its inverse function.

Assign $i_n = i/(n+1)$ for $i = 1, \ldots, n$, and define the oracle estimator of $f(x,y)$ based on knowledge of ξ in terms of the quantities

$$\bar{A}^*_{ab} = \frac{\sum_{i<j} A_{ij}\,\mathbb{I}(\tilde{z}_i = a)\,\mathbb{I}(\tilde{z}_j = b)}{\sum_{i<j} \mathbb{I}(\tilde{z}_i = a)\,\mathbb{I}(\tilde{z}_j = b)}, \quad 1 \leq a, b \leq k.$$

With these definitions, the means and variances of each oracle estimator component $\bar{A}^*_{ab}$ satisfy the following:

$$\left|\mathbb{E}\,\bar{A}^*_{ab} - \rho_n \bar{f}_{ab}\right| \le \rho_n M (2n)^{-\alpha/2}\{1+o(1)\},$$

$$\left|\operatorname{Var}\bar{A}^*_{ab} - \frac{\rho_n \bar{f}_{ab} - \rho_n^2 \overline{f^2}_{ab}}{h^2_{ab}}\right| \le \rho_n \frac{M}{h^2_{ab}(2n)^{\alpha/2}}\{1+o(1)\} + \rho_n^2 M^2 (2n)^{-\alpha};$$

where $\bar{f}_{ab}$ and $\overline{f^2}_{ab}$ are defined by

$$\bar{f}_{ab} = \frac{1}{|\omega_{ab}|}\iint_{\omega_{ab}} f(x,y)\,dx\,dy, \quad \overline{f^2}_{ab} = \frac{1}{|\omega_{ab}|}\iint_{\omega_{ab}} f^2(x,y)\,dx\,dy;$$

and the region ω_{ab} is given by

$$\omega_{ab} = \begin{cases} [(a-1)h/n, ah/n] \times [(b-1)h/n, bh/n] & \textit{if } a<k \textit{ and } b<k, \\ [(k-1)h/n, 1] \times [(b-1)h/n, bh/n] & \textit{if } a=k \textit{ and } b<k, \\ [(b-1)h/n, bh/n] \times [(k-1)h/n, 1] & \textit{if } a<k \textit{ and } b=k, \\ [(k-1)h/n, 1] \times [(k-1)h/n, 1] & \textit{if } a=k \textit{ and } b=k. \end{cases} \tag{1}$$

Proof. Note that the oracle sample proportion estimator takes the form

$$\begin{aligned} \bar{A}^*_{ab} &= \frac{\sum_{i<j} A_{ij}\,\mathbb{I}(\tilde{z}_i = a)\,\mathbb{I}(\tilde{z}_j = b)}{\sum_{i<j}\mathbb{I}(\tilde{z}_i = a)\,\mathbb{I}(\tilde{z}_j = b)} \\ &= \begin{cases} \dfrac{\sum_{j=h(b-1)+1}^{hb}\sum_{i=h(a-1)+1}^{ha\,\mathbb{I}(a\neq b)+(j-1)\,\mathbb{I}(a=b)} A_{(i)(j)}}{h^2_{ab}} & \text{if } a<k \text{ and } b<k, \\ \dfrac{\sum_{j=h(k-1)+1}^{n}\sum_{i=h(a-1)+1}^{ha\,\mathbb{I}(a\neq b)+(j-1)\,\mathbb{I}(a=b)} A_{(i)(j)}}{h^2_{ak}} & \text{if } a\le k \text{ and } b=k, \\ \bar{A}^*_{bk} & \text{if } a=k \text{ and } b\le k; \end{cases} \\ &= \frac{\sum_{(i,j)\in R_{ab}} A_{(i)(j)}}{h^2_{ab}}, \end{aligned}$$

where R_{ab} is defined implicitly to make the summation valid, and is non-random. Thus we may conclude that

$$\mathbb{E}\,\bar{A}^*_{ab} = \frac{1}{h^2_{ab}}\sum_{(i,j)\in R_{ab}} \mathbb{E}\,A_{(i)(j)}. \tag{2}$$

We define $\tilde{f}_{ab}$, for $i_n = i/(n+1)$ and $j_n = j/(n+1)$, as

$$\tilde{f}_{ab} = \frac{1}{h^2_{ab}}\sum_{(i,j)\in R_{ab}} f(i_n, j_n). \tag{3}$$

We then use (6) from Lemma 2 to obtain that

$$\left|\mathbb{E}\,\bar{A}^*_{ab} - \rho_n \tilde{f}_{ab}\right| \le \rho_n M\{2(n+2)\}^{-\alpha/2}. \tag{4}$$

We note from Lemma 4 that as $f \in \text{Hölder}^\alpha(M)$ on $(0,1)^2$,

$$|\tilde{f}_{ab} - \bar{f}_{ab}| < M\, 2^{\alpha/2} n^{-\alpha}\{1 + 2^\alpha\, \mathbb{I}(a = b)\}. \tag{5}$$

We then apply the triangle inequality to (2)–(5) to derive

$$\begin{aligned}\left|\mathbb{E}\,\bar{A}^*_{ab} - \rho_n \bar{f}_{ab}\right| &\leq \rho_n M \Big[\{2(n+2)\}^{-\alpha/2} + 2^{\alpha/2} n^{-\alpha}\{1 + 2^\alpha\, \mathbb{I}(a = b)\}\Big] \\ &\leq \rho_n M (2n)^{-\alpha/2}\{1 + o(1)\}.\end{aligned}$$

This establishes the form of $\mathbb{E}\,\bar{A}^*_{ab}$. We next calculate

$$\operatorname{Var} \bar{A}^*_{ab} = \frac{\sum_{(i,j)\in R_{ab}} \sum_{(m,l)\in R_{ab}} \operatorname{Cov}\big\{A_{(i)(j)}, A_{(m)(l)}\big\}}{h^4_{ab}}.$$

Referring to (7) of Lemma 2,

$$\begin{aligned}\operatorname{Var} \bar{A}^*_{ab} &= \frac{1}{h^4_{ab}} \sum_{(i,j)\in R_{ab}} \sum_{(m,l)\in R_{ab}} \operatorname{Cov}\{A_{(i)(j)}, A_{(m)(l)}\} \\ &\leq \frac{1}{h^4_{ab}} \sum_{(i,j)\in R_{ab}} \rho_n f(i_n, j_n)\{1 - \rho_n f(i_n, j_n)\} + \rho_n^2 M^2 [2(n+2)]^{-\alpha} \\ &\qquad + \frac{\rho_n}{h^2_{ab}} M\{2(n+2)\}^{-\alpha/2}\Big[1 + \rho_n M\{2(n+2)\}^{-\alpha/2}\Big].\end{aligned}$$

We may likewise determine the lower bound of

$$\begin{aligned}\operatorname{Var} \bar{A}^*_{ab} &\geq \frac{1}{h^4_{ab}} \sum_{(i,j)\in R_{ab}} \rho_n f(i_n, j_n)\{1 - \rho_n f(i_n, j_n)\} - \rho_n^2 M^2 [2(n+2)]^{-\alpha} \\ &\qquad - \frac{\rho_n}{h^2_{ab}} M\{2(n+2)\}^{-\alpha/2}\Big[1 + \rho_n M\{2(n+2)\}^{-\alpha/2}\Big].\end{aligned}$$

From Lemmas 4 and 5 below, writing $\overline{f^2}_{ab}$ for the normalized integral of $f^2(x,y)$ over the block ω_{ab}, we have respectively that

$$\left|\tilde{f}_{ab} - \bar{f}_{ab}\right| \leq \frac{M 2^{\alpha/2}}{n^\alpha}\{1 + 2^\alpha\, \mathbb{I}(a = b)\}$$

and

$$\left|\frac{1}{h^2_{ab}} \sum_{(i,j)\in R_{ab}} f^2(i_n, j_n) - \overline{f^2}_{ab}\right| \leq \frac{2\|f\|_\infty M 2^{\alpha/2}}{n^\alpha}\{1 + 2^\alpha\, \mathbb{I}(a = b)\}.$$

Together these results yield the claimed expression for the variance of $\bar{A}^*_{ab}$. □

Lemma 1. *Let $f \in \text{Hölder}^\alpha(M)$, with $\bar{f}_{ab} = |\omega_{ab}|^{-1} \iint_{\omega_{ab}} f(x,y)\,dx\,dy$ defined as its local average over ω_{ab}. Then*

$$\frac{1}{|\omega_{ab}|} \iint_{\omega_{ab}} \left|f(x,y) - \bar{f}_{ab}\right|^2 dx\,dy \leq M^2 2^\alpha (h/n)^{2\alpha}\big\{1 + 2^{2\alpha}\, \mathbb{I}(a = k \text{ or } b = k)\big\}.$$

Proof. Recall that ω_{ab} is given by (1), as before. Note from the definition of the set Hölder$^\alpha(M)$ that if $(x,y) \in \omega_{ab}$ and $a,b < k$, then

$$\begin{aligned}
\left|\bar{f}_{ab} - f(x,y)\right| &= \left|\frac{1}{|\omega_{ab}|}\iint_{\omega_{ab}} f(x',y')\,dx'\,dy' - f(x,y)\right| \\
\Rightarrow \left|\bar{f}_{ab} - f(x,y)\right| &\leq \frac{1}{|\omega_{ab}|}\iint_{\omega_{ab}} |f(x',y') - f(x,y)|\,dx'\,dy' \\
&\leq \frac{1}{|\omega_{ab}|}\iint_{\omega_{ab}} M|(x',y') - (x,y)|^\alpha\,dx'\,dy' \\
&\leq \frac{1}{|\omega_{ab}|}\iint_{\omega_{ab}} M\left[2(h/n)^2\right]^{\alpha/2}\,dx'\,dy' = M2^{\alpha/2}(h/n)^\alpha.
\end{aligned}$$

Thus

$$\begin{aligned}
\left|\bar{f}_{ab} - f(x,y)\right|^2 &\leq M^2 2^\alpha (h/n)^{2\alpha} \\
\Rightarrow \frac{1}{|\omega_{ab}|}\iint_{\omega_{ab}} \left|f(x,y) - \bar{f}_{ab}\right|^2 dx\,dy &\leq M^2 2^\alpha (h/n)^{2\alpha}.
\end{aligned}$$

If $a = k$ or $b = k$ then we replace h by $2h$ to obtain a bound. □

Lemma 1 has been adapted from Wolfe and Olhede [16].

Lemma 2 (Moments of $A_{(i)(j)}$)**.** *Let $i_n = i/(n+1)$ for $i = 1, \ldots, n$, and let (i) be defined as the index chosen so that $\xi_{(1)} \leq \xi_{(2)} \leq \cdots \leq \xi_{(n)}$. Then the means and variances of each $A_{(i)(j)}$ for $i < j$ satisfy the following:*

$$\left|\mathbb{E}\,A_{(i)(j)} - \rho_n f(i_n, j_n)\right| \leq \rho_n M\{2(n+2)\}^{-\alpha/2}, \tag{6}$$

$$\begin{aligned}
\left|\operatorname{Var} A_{(i)(j)} - \rho_n f(i_n,j_n)(1 - \rho_n f(i_n,j_n))\right| &\leq \rho_n \cdot M\{2(n+2)\}^{-\alpha/2} \qquad (7)\\
&\quad \cdot \left[1 + \rho_n M\{2(n+2)\}^{-\alpha/2}\right].
\end{aligned}$$

$$\textit{For } i \neq m \textit{ or } j \neq l, \qquad \operatorname{Cov}\left\{A_{(i)(j)}, A_{(m)(l)}\right\} \leq \rho_n^2 M^2 \{2(n+2)\}^{-\alpha}.$$

Proof. Equation (6) follows directly from the law of iterated expectation, with the first calculation following from conditioning on ξ:

$$\mathbb{E}\,A_{(i)(j)} = \mathbb{E}_\xi\left[\mathbb{E}_{A|\xi}\left\{A_{(i)(j)}\,|\,\xi\right\}\right] = \mathbb{E}_\xi\left\{\rho_n f\left(\xi_{(i)}, \xi_{(j)}\right)\right\}, \tag{8}$$

and the second calculation following by approximation of the latter expectation, as we now show. As $|\cdot|$ is convex, Jensen's inequality permits us to deduce that

$$\left|\mathbb{E}_\xi\,\rho_n f\left(\xi_{(i)}, \xi_{(j)}\right) - \rho_n f(i_n, j_n)\right| \leq \rho_n\,\mathbb{E}_\xi\left\{\left|f\left(\xi_{(i)}, \xi_{(j)}\right) - f(i_n, j_n)\right|\right\}. \tag{9}$$

We note that from Lemma 3, we have

$$\mathbb{E}_\xi\left|f\left(\xi_{(i)}, \xi_{(j)}\right) - f(i_n, j_n)\right| \leq M\{2(n+2)\}^{-\alpha/2}, \tag{10}$$

and so we can deduce (6) by combining (8)–(10).

Equation (7) is derived from the law of total variance by

$$\begin{aligned}\operatorname{Var} A_{(i)(j)} &= \mathbb{E}_\xi\left[\operatorname{Var}_{A|\xi}\left\{A_{(i)(j)}\right\}\right] + \operatorname{Var}_\xi\left[\mathbb{E}_{A|\xi}\left\{A_{(i)(j)}\right\}\right] \\ &= \mathbb{E}_\xi\left\{\rho_n f\left(\xi_{(i)}, \xi_{(j)}\right)\left(1-\rho_n f\left(\xi_{(i)}, \xi_{(j)}\right)\right)\right\} \qquad (11) \\ &\quad + \mathbb{E}_\xi\left\{\rho_n^2 f^2\left(\xi_{(i)}, \xi_{(j)}\right)\right\} - \mathbb{E}_\xi^2\left\{\rho_n f\left(\xi_{(i)}, \xi_{(j)}\right)\right\},\ 1 \le i < j \le n.\end{aligned}$$

The second and third terms in (11) cancel, and thus we obtain that

$$\operatorname{Var} A_{(i)(j)} = \rho_n\left\{\mathbb{E}_\xi\, f\left(\xi_{(i)}, \xi_{(j)}\right)\right\}\left\{1-\rho_n\,\mathbb{E}_\xi\, f\left(\xi_{(i)}, \xi_{(j)}\right)\right\}.$$

We now need to calculate expectations with respect to the latent vector ξ. Owing to (10), we can upper bound $\mathbb{E}_\xi\, f\left(\xi_{(i)}, \xi_{(j)}\right)$ by the quantity $\rho_n f(i_n, j_n) + \rho_n M\{2(n+2)\}^{-\alpha/2}$, and likewise the negative term $-\mathbb{E}_\xi\, f\left(\xi_{(i)}, \xi_{(j)}\right)$ by the quantity $-\rho_n f(i_n, j_n) + \rho_n M\{2(n+2)\}^{-\alpha/2}$. Similarly, $1-\rho_n\,\mathbb{E}_\xi\, f\left(\xi_{(i)}, \xi_{(j)}\right)$ and its negative can be lower bounded. Thus we may deduce the two inequalities

$$\begin{aligned}\operatorname{Var} A_{(i)(j)} &\le \rho_n\left[f(i_n, j_n) + M\{2(n+2)\}^{-\alpha/2}\right] \\ &\quad \cdot \left[1-\rho_n f(i_n, j_n) + \rho_n M\{2(n+2)\}^{-\alpha/2}\right], \\ \operatorname{Var} A_{(i)(j)} &\ge \rho_n\left[f(i_n, j_n) - M\{2(n+2)\}^{-\alpha/2}\right] \\ &\quad \cdot \left[1-\rho_n f(i_n, j_n) - \rho_n M\{2(n+2)\}^{-\alpha/2}\right].\end{aligned}$$

Combining these two relationships, we obtain (7).

From the law of total covariance, we have that since $i < j$ and $m < l$, when at least either $i \neq m$ or $j \neq l$, the conditional independence of the Bernoulli trials comprising A yields

$$\begin{aligned}\operatorname{Cov}\left\{A_{(i)(j)}, A_{(m)(l)}\right\} &= \mathbb{E}_\xi\left[\operatorname{Cov}_{A|\xi}\left\{A_{(i)(j)}, A_{(m)(l)}\right\}\right] \qquad (12) \\ &\quad + \operatorname{Cov}_\xi\left[\mathbb{E}_{A|\xi}\left\{A_{(i)(j)}\right\}, \mathbb{E}_{A|\xi}\left\{A_{(m)(l)}\right\}\right] \\ &= \rho_n^2 \operatorname{Cov}_\xi\left\{f\left(\xi_{(i)}, \xi_{(j)}\right), f\left(\xi_{(m)}, \xi_{(l)}\right)\right\}.\end{aligned}$$

We now simplify this expression further, working directly with the form in (12). We define $j_n = j/(n+1)$, as well as $m_n = m/(n+1)$ and $l_n = l/(n+1)$. We then use the shift-invariance of the covariance operator to write

$$\begin{aligned}&\left|\operatorname{Cov}_\xi\left\{f\left(\xi_{(i)}, \xi_{(j)}\right), f\left(\xi_{(m)}, \xi_{(l)}\right)\right\}\right| \\ &\qquad \le \left|\mathbb{E}_\xi\left[\left\{f\left(\xi_{(i)}, \xi_{(j)}\right) - f(i_n, j_n)\right\}\left\{f\left(\xi_{(m)}, \xi_{(l)}\right) - f(m_n, l_n)\right\}\right]\right|,\end{aligned}$$

where we have a bound, rather than equality, because we do *not* claim that $\mathbb{E}\left\{f\left(\xi_{(i)}, \xi_{(j)}\right)\right\} = f(i_n, j_n)$. We may use Jensen's inequality to deduce that

$$\begin{aligned}&\left|\mathbb{E}_\xi\left\{f\left(\xi_{(i)}, \xi_{(j)}\right) - f(i_n, j_n)\right\}\left\{f\left(\xi_{(m)}, \xi_{(l)}\right) - f(m_n, l_n)\right\}\right| \\ &\qquad \le \mathbb{E}_\xi\left|\left\{f\left(\xi_{(i)}, \xi_{(j)}\right) - f(i_n, j_n)\right\}\left\{f\left(\xi_{(m)}, \xi_{(l)}\right) - f(m_n, l_n)\right\}\right|.\end{aligned}$$

Now, because $f \in \text{Hölder}^{\alpha}(M)$ by hypothesis, there exists $M < \infty$ such that

$$|f(x, y) - f(x', y')| \leq M|(x, y) - (x', y')|^{\alpha},$$

and so we obtain that

$$\mathbb{E}_{\xi}\big|f\big(\xi_{(i)}, \xi_{(j)}\big) - f(i_n, j_n)\big|\big|f\big(\xi_{(m)}, \xi_{(l)}\big) - f(m_n, l_n)\big| \\ \leq M^2\, \mathbb{E}_{\xi}\big|\big(\xi_{(i)}, \xi_{(j)}\big) - (i_n, j_n)\big|^{\alpha}\big|\big(\xi_{(m)}, \xi_{(l)}\big) - (m_n, l_n)\big|^{\alpha}.$$

From the Cauchy–Schwarz inequality, it therefore follows that

$$\mathbb{E}_{\xi}\big|\big(\xi_{(i)}, \xi_{(j)}\big) - (i_n, j_n)\big|^{\alpha}\big|\big(\xi_{(m)}, \xi_{(l)}\big) - (m_n, l_n)\big|^{\alpha} \\ \leq \sqrt{\mathbb{E}_{\xi}\big|\big(\xi_{(i)}, \xi_{(j)}\big) - (i_n, j_n)\big|^{2\alpha}}\sqrt{\mathbb{E}_{\xi}\big|\big(\xi_{(m)}, \xi_{(l)}\big) - (m_n, l_n)\big|^{2\alpha}}.$$

We then calculate

$$\mathbb{E}_{\xi}\big|\big(\xi_{(m)}, \xi_{(l)}\big) - (m_n, l_n)\big|^{2\alpha} = \mathbb{E}_{\xi}\Big\{\big(\xi_{(m)} - m_n\big)^2 + \big(\xi_{(l)} - l_n\big)^2\Big\}^{\alpha}.$$

Applying Jensen's inequality, we find that for $\alpha \leq 1$,

$$\mathbb{E}_{\xi}\Big\{\big(\xi_{(m)} - m_n\big)^2 + \big(\xi_{(l)} - l_n\big)^2\Big\}^{\alpha} \leq \big[\text{Var}\{\xi_{(l)}\} + \text{Var}\{\xi_{(j)}\}\big]^{\alpha} \leq \{2(n+2)\}^{-2\alpha}.$$

Thus we may deduce that

$$\big|\text{Cov}_{\xi}\big\{f\big(\xi_{(i)}, \xi_{(j)}\big), f\big(\xi_{(m)}, \xi_{(l)}\big)\big\}\big| \leq M^2\Big[\{2(n+2)\}^{-\alpha}\{2(n+2)\}^{-\alpha}\Big]^{1/2}.$$

Combining this expression with (12) then yields the stated result. □

Lemma 3. *Let $f \in$ Hölder$^{\alpha}(M)$, and let $\{\xi_{(i)}\}_{i=1}^{n}$ be an ordered sample of independent* Uniform(0, 1) *random variables. Then for $1 \leq i, j \leq n$ we have*

$$\mathbb{E}_{\xi}\,\big|f\big(\xi_{(i)}, \xi_{(j)}\big) - f\big(i_n, j_n\big)\big| \leq M\{2(n+2)\}^{-\alpha/2}.$$

Proof. We note that as $f \in \text{Hölder}^{\alpha}(M)$,

$$\big|f\big(\xi_{(i)}, \xi_{(j)}\big) - f(i_n, j_n)\big| \leq M\big|(\xi_{(i)}, \xi_{(j)}) - (i_n, j_n)\big|^{\alpha}, \quad 1 \leq i, j \leq n.$$

Since $\text{Var}\,\xi_{(i)} = i_n(1 - i_n)/(n+2) \leq (1/4)/(n+2)$, by Jensen's inequality we have for any $0 < \alpha \leq 1$ that

$$\mathbb{E}_{\xi}\big\{(\xi_{(i)} - i_n)^2 + (\xi_{(j)} - j_n)^2\big\}^{\alpha/2} \leq \big(\text{Var}\,\xi_{(i)} + \text{Var}\,\xi_{(j)}\big)^{\alpha/2} \leq \{2(n+2)\}^{-\alpha/2}.$$

This completes the proof. □

Lemma 3 has been adapted from Wolfe and Olhede [16].

Lemma 4 (Linear quadrature bounds)**.** *Let $f \in$ Hölder$^\alpha(M)$ be a symmetric function on $(0,1)^2$, and define $i_n = i/(n+1)$, $j_n = j/(n+1)$. Then with*

$$\tilde{f}_{ab} = \frac{1}{h_{ab}^2} \sum_{(i,j)\in R_{ab}} f(i_n, j_n), \quad 1 \le a \le b \le k,$$

we have that

$$\left|\tilde{f}_{ab} - \bar{f}_{ab}\right| \le M\, 2^{\alpha/2} n^{-\alpha} \{1 + 2^\alpha\, \mathbb{I}(a=b)\}.$$

Proof. We start from the definition of

$$\tilde{f}_{ab} = \frac{1}{h_{ab}^2} \sum_{(i,j)\in R_{ab}} f(i_n, j_n).$$

Thus we may by simple expansion determine

$$\begin{aligned}\tilde{f}_{ab} &= \frac{n^2}{h_{ab}^2} \sum_{(i,j)\in R_{ab}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} [f(x,y) + f(i_n,j_n) - f(x,y)]\, dx\, dy \\ &= \frac{n^2}{h_{ab}^2} \sum_{(i,j)\in R_{ab}} \left[\int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} f(x,y)\, dx\, dy + \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \{f(i_n,j_n) - f(x,y)\}\, dx\, dy.\right]\end{aligned}$$

We now use the fact that $f(x,y) \in$ Hölder$^\alpha(M)$. Thus we may write

$$\begin{aligned}&\left|\frac{n^2}{h_{ab}^2} \sum_{(i,j)\in R_{ab}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \{f(i_n,j_n) - f(x,y)\}\, dx\, dy\right| \\ &\qquad \le \frac{n^2}{h_{ab}^2} \sum_{(i,j)\in R_{ab}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} |f(i_n,j_n) - f(x,y)|\, dx\, dy \le \frac{M 2^{\alpha/2}}{n^\alpha},\end{aligned} \tag{13}$$

with the last inequality following from the fact that f is a α-Hölder function on the domain of integration. Furthermore, we note directly if $a < b$ then, with ω_{ab} as defined in (1),

$$\frac{n^2}{h_{ab}^2} \sum_{(i,j)\in R_{ab}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} f(x,y)\, dx\, dy = \frac{n^2}{h_{ab}^2} \iint_{\omega_{ab}} f(x,y)\, dx\, dy. \tag{14}$$

If on the other hand $a = b$ then

$$\begin{aligned}&\frac{n^2}{\binom{h_b}{2}} \sum_{(i,j)\in R_{bb}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} f(x,y)\, dx\, dy \\ &\qquad = \frac{n^2}{\binom{h_b}{2}} \sum_{j=(b-1)h+1}^{hb\,\mathbb{I}(b<k)+n\,\mathbb{I}(b=k)} \sum_{i=(b-1)h+1}^{j-1} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} f(x,y)\, dx\, dy.\end{aligned} \tag{15}$$

This equation acknowledges that group a has size h_a, which is equal to h for $a = 1, \ldots, k-1$, and $h_k = h + r$ for $a = k$. We shall start by simplifying this expression. We note that the latter becomes:

$$
\begin{aligned}
&\frac{n^2}{\binom{h_b}{2}} \sum_{j=(b-1)h+1}^{hb\,\mathbb{I}(b<k)+n\,\mathbb{I}(b=k)} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{(b-1)h}{n}}^{\frac{j-1}{n}} f(x,y)\,dx\,dy \\
&\quad = \bar{f}_{bb} + \sum_{j=(b-1)h+1}^{hb\,\mathbb{I}(b<k)+n\,\mathbb{I}(b=k)} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \left\{ \left[\frac{n^2}{\binom{h_b}{2}} - \frac{2n^2}{h_b^2} \right] \int_{\frac{(b-1)h}{n}}^{y} - \frac{n^2}{\binom{h_b}{2}} \int_{\frac{j-1}{n}}^{y} \right\} f(x,y)\,dx\,dy \\
&\quad = \bar{f}_{bb} + \frac{2n^2}{h_b} \sum_{j=(b-1)h+1}^{hb\,\mathbb{I}(b<k)+n\,\mathbb{I}(b=k)} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \left\{ \left[\frac{1}{h_b-1} - \frac{1}{h_b} \right] \int_{\frac{(b-1)h}{n}}^{y} - \frac{1}{h_b-1} \int_{\frac{j-1}{n}}^{y} \right\} \\
&\qquad\qquad \cdot f(x,y)\,dx\,dy \\
&\quad = \bar{f}_{bb} + \frac{2n^2}{h_b} \sum_{j=(b-1)h+1}^{hb\,\mathbb{I}(b<k)+n\,\mathbb{I}(b=k)} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \left\{ \frac{1}{h_b(h_b-1)} \int_{\frac{(b-1)h}{n}}^{y} - \frac{1}{h_b-1} \int_{\frac{j-1}{n}}^{y} \right\} \\
&\qquad\qquad \cdot f(x,y)\,dx\,dy \\
&\quad = \bar{f}_{bb} + \frac{1}{(h_b-1)} \sum_{j=(b-1)h+1}^{hb\,\mathbb{I}(b<k)+n\,\mathbb{I}(b=k)} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \left\{ \frac{2n^2}{h_b^2} \int_{\frac{(b-1)h}{n}}^{y} - \frac{2n^2}{h_b} \int_{\frac{j-1}{n}}^{y} \right\} f(x,y)\,dx\,dy \\
&\quad = \bar{f}_{bb} + \frac{1}{(h_b-1)} \left\{ \bar{f}_{bb} - \frac{2n^2}{h_b} \sum_{j=(b-1)h+1}^{hb\,\mathbb{I}(b<k)+n\,\mathbb{I}(b=k)} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{j-1}{n}}^{y} f(x,y)\,dx\,dy \right\} \\
&\quad = \bar{f}_{bb} + \frac{1}{(h_b-1)} \left\{ \frac{2n^2}{h_b} \sum_{j=(b-1)h+1}^{hb\,\mathbb{I}(b<k)+n\,\mathbb{I}(b=k)} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{j-1}{n}}^{y} \left(\bar{f}_{bb} - f(x,y) \right) dx\,dy \right\}.
\end{aligned}
\tag{16}
$$

We note that

$$
\begin{aligned}
&\left| \frac{2n^2}{h_b} \sum_{j=(b-1)h+1}^{hb\,\mathbb{I}(b<k)+n\,\mathbb{I}(b=k)} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{j-1}{n}}^{y} \left(\bar{f}_{bb} - f(x,y) \right) dx\,dy \right| \\
&\qquad \le \frac{2n^2}{h_b} \sum_{j=(b-1)h+1}^{hb\,\mathbb{I}(b<k)+n\,\mathbb{I}(b=k)} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{j-1}{n}}^{y} \left| \bar{f}_{bb} - f(x,y) \right| dx\,dy \le M \left(\sqrt{2} h_b / n \right)^{\alpha}.
\end{aligned}
\tag{17}
$$

Thus, combining (17) with (13), (15), and (16), we have

$$
\left| \bar{f}_{bb} - \tilde{f}_{bb} \right| \le \frac{M 2^{\alpha/2}}{n^{\alpha}} + M \left(\sqrt{2} h_b / n \right)^{\alpha} \frac{1}{h_b - 1}.
$$

From the off-diagonal entries $a < b$ we may conclude from (13) and (14) that

$$\left|\bar{f}_{ab} - \tilde{f}_{ab}\right| \leq \frac{M2^{\alpha/2}}{n^{\alpha}}.$$

Thus it follows that

$$\left|\bar{f}_{ab} - \tilde{f}_{ab}\right| \leq \frac{M2^{\alpha/2}}{n^{\alpha}} + \begin{cases} 0 & a \neq b, \\ M\left(\sqrt{2}h_b/n\right)^{\alpha} \frac{1}{h_b - 1} & a = b. \end{cases}$$

Since $\frac{h_b^{\alpha}}{(h_b-1)} \leq 2^{\alpha}$ if $h_b \geq 2$, the expression follows. This concludes the proof. $\square$

Lemma 4 has been adapted from Wolfe and Olhede [16].

Lemma 5 (Square quadrature bounds)**.** *Let $f \in$ Hölder$^{\alpha}(M)$ be a symmetric function on $(0,1)^2$, and define $i_n = i/(n+1)$, $j_n = j/(n+1)$. Then with*

$$\widetilde{f^2}_{ab} = \frac{1}{h_{ab}^2} \sum_{(i,j) \in R_{ab}} f^2(i_n, j_n), \quad 1 \leq a \leq b \leq k,$$

we have that

$$\left|\widetilde{f^2}_{ab} - \overline{f^2}_{ab}\right| \leq \frac{2\|f\|_{\infty} M2^{\alpha/2}}{n^{\alpha}} \{1 + 2^{\alpha}\, \mathbb{I}(a = b)\}.$$

Proof. We start from

$$\begin{aligned}
\widetilde{f^2}_{ab} &= \frac{n^2}{h_{ab}^2} \sum_{(i,j) \in R_{ab}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \left[f^2(x,y) + f^2(i_n, j_n) - f^2(x,y)\right] dx\, dy \\
&= \frac{n^2}{h_{ab}^2} \sum_{(i,j) \in R_{ab}} \left\{ \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} f^2(x,y) + \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} f^2(i_n, j_n) - f^2(x,y) \right\} dx\, dy.
\end{aligned}$$

We now use that $f(x,y) \in$ Hölder$^{\alpha}(M)$. We write

$$\begin{aligned}
&\left| \frac{n^2}{h_{ab}^2} \sum_{(i,j) \in R_{ab}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \left\{f^2(i_n, j_n) - f^2(x,y)\right\} dx\, dy \right| \\
&\qquad \leq \frac{n^2}{h_{ab}^2} \sum_{(i,j) \in R_{ab}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \left|f^2(i_n, j_n) - f^2(x,y)\right| dx\, dy \\
&\qquad = \frac{n^2}{h_{ab}^2} \sum_{(i,j) \in R_{ab}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} |f(i_n, j_n) + f(x,y)||f(i_n, j_n) - f(x,y)|\, dx\, dy \\
&\qquad \leq \frac{2\|f\|_{\infty} n^2}{h_{ab}^2} \sum_{(i,j) \in R_{ab}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} |f(i_n, j_n) - f(x,y)|\, dx\, dy \\
&\qquad \leq \frac{2\|f\|_{\infty} M2^{\alpha/2}}{n^{\alpha}}, \qquad (18)
\end{aligned}$$

with the final inequality following from (13) of the previous lemma. We note directly if $a < b$ then

$$\frac{n^2}{h_{ab}^2} \sum_{(i,j)\in R_{ab}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} f^2(x,y)\,dx\,dy = \frac{n^2}{h_{ab}^2} \iint_{\omega_{ab}} f^2(x,y)\,dx\,dy. \tag{19}$$

From the off-diagonal entries, for which $a < b$, we may conclude directly from (18) and (19) that

$$\left|\overline{f^2}_{ab} - \widetilde{f^2}_{ab}\right| \leq \frac{2\|f\|_\infty M 2^{\alpha/2}}{n^\alpha}.$$

If on the other hand $a = b$ then

$$\begin{aligned}
&\frac{n^2}{\binom{h_b}{2}} \sum_{(i,j)\in R_{bb}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} f^2(x,y)\,dx\,dy \\
&\qquad = \frac{n^2}{\binom{h_b}{2}} \sum_{j=(b-1)h+1}^{hb\,\mathbb{I}(b<k)+n\,\mathbb{I}(b=k)} \sum_{i=(b-1)h+1}^{j-1} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{i-1}{n}}^{\frac{i}{n}} f^2(x,y)\,dx\,dy.
\end{aligned}$$

We shall start by simplifying this expression. We note that the latter becomes:

$$\begin{aligned}
&\frac{n^2}{\binom{h_b}{2}} \sum_{j=(b-1)h+1}^{hb\,\mathbb{I}(b<k)+n\,\mathbb{I}(b=k)} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{(b-1)h}{n}}^{\frac{j-1}{n}} f^2(x,y)\,dx\,dy \\
&\quad = \overline{f^2}_{bb} + \frac{1}{(h_b-1)} \frac{2n^2}{h_b} \sum_{j=(b-1)h+1}^{hb\,\mathbb{I}(b<k)+n\,\mathbb{I}(b=k)} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{j-1}{n}}^{y} \left(\overline{f^2}_{bb} - f^2(x,y)\right) dx\,dy.
\end{aligned}$$

We may note directly that

$$\frac{(b-1)h}{n} \leq x < y \leq \frac{bh\,\mathbb{I}(b<k) + n\,\mathbb{I}(b=k)}{n} = \frac{bh}{n}\,\mathbb{I}(b<k) + \mathbb{I}(b=k),$$

and so it follows that

$$\begin{aligned}
\left|\overline{f^2}_{bb} - f^2(x,y)\right| &\leq \left| \frac{\int_{\frac{h(b-1)}{n}}^{\frac{bh}{n}\mathbb{I}(b<k)+\mathbb{I}(b=k)} \int_{\frac{(b-1)h}{n}}^{\frac{bh}{n}\mathbb{I}(b<k)+\mathbb{I}(b=k)} f^2(x',y')\,dx'\,dy'}{\left\{\frac{h_b}{n}\right\}^2} - f^2(x,y) \right| \\
&\leq \frac{n^2}{h_b^2} \int_{\frac{(b-1)h}{n}}^{\frac{bh}{n}\mathbb{I}(b<k)+\mathbb{I}(b=k)} \int_{\frac{(b-1)h}{n}}^{\frac{bh}{n}\mathbb{I}(b<k)+\mathbb{I}(b=k)} \left|f^2(x',y') - f^2(x,y)\right| dx'\,dy' \\
&\leq \frac{2\|f\|_\infty n^2}{h_b^2} \int_{\frac{(b-1)h}{n}}^{\frac{bh}{n}\mathbb{I}(b<k)+\mathbb{I}(b=k)} \int_{\frac{(b-1)h}{n}}^{\frac{bh}{n}\mathbb{I}(b<k)+\mathbb{I}(b=k)} |f(x',y') - f(x,y)|\,dx'\,dy' \\
&\leq 2\|f\|_\infty M \left(\frac{\sqrt{2}h_b}{n}\right)^\alpha.
\end{aligned}$$

Thus

$$\left| \frac{n^2}{\binom{h_b}{2}} \sum_{j=(b-1)h+1}^{hb\,\mathbb{I}(b<k)+n\,\mathbb{I}(b=k)} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{(b-1)h}{n}}^{\frac{j-1}{n}} f^2(x,y)\,dx\,dy - \overline{f^2}_{bb} \right|$$
$$\leq \frac{1}{(h_b-1)} \frac{2n^2}{h_b} \sum_{j=(b-1)h+1}^{hb\,\mathbb{I}(b<k)+n\,\mathbb{I}(b=k)} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \int_{\frac{j-1}{n}}^{y} 2\|f\|_\infty M \left(\frac{\sqrt{2}h_b}{n}\right)^\alpha dx\,dy.$$
$$= \frac{1}{(h_b-1)} 2\|f\|_\infty M \left(\frac{\sqrt{2}h_b}{n}\right)^\alpha.$$

For the on-diagonal entries having $a = b$, it therefore follows that

$$\left|\overline{f^2}_{bb} - \widetilde{f^2}_{bb}\right| \leq \frac{2\|f\|_\infty M 2^{\alpha/2}}{n^\alpha} + \frac{1}{(h_b-1)} 2\|f\|_\infty M \left(\frac{\sqrt{2}h_b}{n}\right)^\alpha.$$

Note that $\frac{h_b^\alpha}{(h_b-1)} \leq 2^\alpha$ if $h_b \geq 2$, and so the expression follows. □