Metric Learning via Cross-Validation

Linlin Dai, Kani Chen, Gang Li and Yuanyuan Lin

Southwestern University of Finance and Economics,
Hong Kong University of Science and Technology,
University of California, Los Angeles and
The Chinese University of Hong Kong

Supplementary Material

S1. Technical Proofs

S1.1 Proof of Theorem 1

It suffices to show $\|L_0^\top \hat{L}_1\| \to 0$ in probability, where $\hat{L}_1$ is the basis of the subspace orthogonal to that spanned by the columns of $\hat{L}_1$, which is obtained via the minimization in problem (2.2). First, we show that $\hat{f}^{(i)}(X_i)$ is not a consistent estimator of $f(X_i)$ when $\|L_0^\top \hat{L}_1\| \to 0$. On the one hand,
\[
\text{CM}_n(M) = \frac{1}{n} \sum_{i=1}^{n} \{ \hat{f}(-i)(X_i) - f(X_i) - \epsilon_i \}^2 w(X_i) \\
= \frac{1}{n} \sum_{i=1}^{n} w(X_i) \epsilon_i^2 + \frac{1}{n} \sum_{i=1}^{n} \{ \hat{f}(-i)(X_i) - f(X_i) \}^2 w(X_i) \\
- \frac{2}{n} \sum_{i=1}^{n} \{ \hat{f}(-i)(X_i) - f(X_i) \} w(X_i) \epsilon_i \\
\equiv \eta_0 + \Pi_1 + \Pi_2,
\]

where \(\eta_0\) is irrelevant to the minimization over \(M\). Since \(\Pi_2\) is mean 0 and

\[
E(\Pi_1|X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} w(X_i) \left\{ B(X_i)^2 + \sum_{j \neq i} K_{j,i}^* \sigma_j^2 \right\},
\]

\[
E(\Pi_2^2|X_1, \ldots, X_n) = \frac{4}{n^2} \sum_{i=1}^{n} \sigma_i^2 w(X_i) \left\{ B(X_i)^2 + \sum_{j \neq i} K_{j,i}^* \sigma_j^2 \right\},
\]

where \(B(X_i) = \tilde{f}(X_i) - f(X_i)\) and \(\tilde{f}(X_i) = \sum_{j=1}^{n} f(X_j) K_{j,i}^*\). We have

\(\Pi_2 = O_p(n^{-1} \sqrt{\sum_{i=1}^{n} \{ B(X_i)^2 + \sum_{j \neq i} K_{j,i}^* \sigma_j^2 \}})\). Hence, \(\Pi_1\) is the dominant term compared with \(\Pi_2\) in \(\text{CM}_n(M) - \eta_0\). On the other hand,

\[
\hat{f}(-i)(X_i) - f(X_i) = \sum_{j=1}^{n} f(X_j) K_{j,i}^* - f(X_i) + \sum_{j=1}^{n} \epsilon_j K_{j,i}^*.
\]

Suppose that there exists a subsequence of \(n = 1, 2, \ldots\), such that \(\hat{L}_1 \rightarrow L_1^\dagger\) but \(\|L_0^\dagger L_1^{\dagger\perp}\| \neq 0\). For notational simplicity, we still denote the subsequence as the original \(n\). With \(\|\hat{h}\| \rightarrow 0\),

\[
\sum_{j=1}^{n} f(X_j) K_{j,i}^* \rightarrow E\{f(X)|X \in x + L_1^{\dagger\perp}\}|_{x=X_i} \equiv f^\dagger(X_i)
\]
in probability, as \( n \to \infty \). We intend to show

\[ P\{f^\dagger(X) = f(X)\} < 1. \]

Indeed, suppose that \( P\{f^\dagger(X) = f(X)\} = 1 \) and then \( f^\dagger(x) = f(x) \) for all \( x \in \Omega \). Since \( f^\dagger(x) = f^\dagger(t) \) if \( x - t \in L_1^\perp \), we have \( f(x) = f(t) \) if \( x - t \in L_1^\perp \). It follows that \( f\{t + c(x - t)\} = f(t) \) for all \( c \in \mathbb{R} \). By the identifiability condition (C4), we have \( x - t \in F \) and thus \( L_1^\perp \subseteq F = L_0^\perp \) or equivalently \( S(L_0) \subseteq S(L_1^\perp) \). Since \( L_0 \) and \( L_1^\perp \) are both column orthogonal matrices of size \( p \times r_0 \), we have \( S(L_0) = S(L_1^\perp) \). This is in contradiction with the assumption that \( \|L_0^\top L_1^\perp\| \neq 0 \). Hence, \( P\{f^\dagger(X) = f(X)\} < 1 \).

Since \( f^\dagger(\cdot) \) and \( f(\cdot) \) are smooth functions, we write

\[
\frac{1}{n} \sum_{i=1}^{n} \{\hat{f}^{(-i)}(X_i) - f(X_i)\}^2 w(X_i) \\
= \frac{1}{n} \sum_{i=1}^{n} \{f^{(-i)}(X_i) - f^\dagger(X_i) + f^\dagger(X_i) - f(X_i)\}^2 w(X_i) \\
\geq \frac{1}{n} \sum_{i=1}^{n} \{f^\dagger(X_i) - f(X_i)\}^2 w(X_i) + \frac{1}{n} \sum_{i=1}^{n} \{\hat{f}^{(-i)}(X_i) - f^\dagger(X_i)\}^2 w(X_i) \\
- \frac{2}{n} \sum_{i=1}^{n} |\{f(\cdot) - f^\dagger(\cdot)\} \{f^\dagger(\cdot) - f(\cdot)\}| w(X_i) \\
\geq c_0 + o_p(1). \quad \text{(S1.1)}
\]

The last inequality is followed by the Cauchy-Schwarz inequality, the strong law of large number of \( W_i \equiv \{f^\dagger(X_i) - f(X_i)\}^2 w(X_i) \) and the consistency of \( \hat{f}^{(-i)}(X_i) \) with respect to \( f^\dagger(X_i) \). As a result, \( \text{CM}_n(M) - \eta_0 \) is at the
order of $O_p(1)$ with some positive lower bound $c_0 > 0$. Nevertheless, we
now show that with $\|L_0^T \hat{L}_i\| \to 0$ and $\|h\| \to 0$,
\[
\sup_{1 \leq i \leq n} |\hat{f}(X_i) - f(X_i)| \to 0 \quad \text{in probability,} \quad (\text{S1.2})
\]
as $n \to \infty$. As a consequence, $CM_n(M) - \eta_0$ is at the order of $o_p(1)$.

In fact, recall that $\bar{\Omega}^0$ is the support of $w(\cdot)$ and $f_{r_0}(u) = f(u_1, \ldots, u_{r_0})$
is the density function of $U = L_0^T X$. Now, define $\bar{\Omega}^{\delta} = \{y \in \mathbb{R}^p :$
\[
\inf_{x \in \Omega^0} \|y - x\| < \bar{\delta} \}$, where $\bar{\delta} > 0$ is a small constant such that $\min_{x \in \bar{\Omega}^{\delta}} f(x) > 0$. Hence, there exists $\tau > 0$ such that $\min_{x \in \bar{\Omega}^{\delta}} f_{r_0}(L_0^T x) \geq \tau$. To show
(S1.2), it is sufficient to prove that for any $\epsilon > 0$,
\[
P\left\{ \sup_{x \in \bar{\Omega}^{\delta}} |\hat{f}_n(x) - f(x)| > \epsilon \right\} \to 0, \quad \text{as } n \to \infty. \quad (\text{S1.3})
\]
For simplicity, denote $\nu(x) = f_{r_0}(L_0^T x)$. Let $\phi(x) = f(x)\nu(x) = g(L_0^T x)\nu(x)$,
\[
\phi_n(x) = \frac{1}{n} \sum_{j=1}^{n} Y_j K_M(X_j - x), \quad \nu_n(x) = \frac{1}{n} \sum_{j=1}^{n} K_M(X_j - x).
\]
And thus $\hat{f}_n(x) = \phi_n(x)/\nu_n(x)$. It is not hard to verify that
\[
P\left\{ \sup_{x \in \bar{\Omega}^{\delta}} |\hat{f}_n(x) - f(x)| > \epsilon \right\}
\leq P\left\{ \sup_{x \in \bar{\Omega}^{\delta}} |\phi_n(x) - f(x)\nu_n(x)| \geq \epsilon (\tau - \epsilon) \right\} + P\left\{ \sup_{x \in \bar{\Omega}^{\delta}} |\nu_n(x) - \nu(x)| > \epsilon \right\}
\leq P\left\{ \sup_{x \in \bar{\Omega}^{\delta}} |\nu_n(x) - \nu(x)| > \frac{\epsilon (\tau - \epsilon)}{2b} \right\} + P\left\{ \sup_{x \in \bar{\Omega}^{\delta}} |\nu_n(x) - \nu(x)| > \epsilon \right\}
+ P\left\{ \sup_{x \in \bar{\Omega}^{\delta}} |\phi_n(x) - f(x)\nu(x)| \geq \frac{\epsilon (\tau - \epsilon)}{2} \right\}, \quad (\text{S1.4})
where \( b = \sup_{x \in \Omega^5} |f(x)| < \infty \). Recall that
\[
L = \begin{pmatrix} L_1 & L_2 \end{pmatrix}
\]
is a \( p \times p \) orthonormal matrix, where \( L_1 \in \mathbb{R}^{p \times r_0} \) and \( L_2 \) is the augmented orthonormal basis in \( \mathbb{R}^p \). Define \( f_{L_2}(x) = \int_{s_2 \in \mathbb{R}^{p-r_0}} f_X(x + L_2 s_2) ds_2 \).

To proceed, we first show that as \( n \to \infty, \|h\| \to 0 \) and \( \|L_0^T L_2\| \to 0 \),
\[
\sup_{x \in \Omega^5} |E\{\phi_n(x)\} - \phi(x)| \to 0. \tag{S1.5}
\]
Define \( \tilde{\phi}(x) = f(x) f_X(x), \phi_{L_2}(x) = \int_{s_2 \in \mathbb{R}^{p-r_0}} \tilde{\phi}(x + L_2 s_2) ds_2 \) and \( I_{r_0} \) be the \( r_0 \times r_0 \) identity matrix. We have
\[
E\{f(X) K_M(X - x)\} \tag{S1.6}
\]
\[
= \frac{1}{h_1 \cdots h_{r_0}} \int_{t \in \mathbb{R}^p} K\{(t - x)^T M(t - x)\} \tilde{\phi}(t) dt
= \frac{1}{h_1 \cdots h_{r_0}} \int_{t \in \mathbb{R}^p} K(t^T L_1 H^{-2} L_1^T t) \tilde{\phi}(t + x) dt
= \frac{1}{h_1 \cdots h_{r_0}} \int_{s \in \mathbb{R}^p} K(s^T L_1 H^{-2} L_1^T L_2) \tilde{\phi}(L_1 s + x) ds
= \frac{1}{h_1 \cdots h_{r_0}} \int_{s_1 \in \mathbb{R}^{r_0}, s_2 \in \mathbb{R}^{(p - r_0)}} K\left\{\begin{pmatrix} s_1^T & s_2^T \end{pmatrix}\begin{pmatrix} I_{r_0} & 0 \\ 0 & I_{r_0} \end{pmatrix} H^{-2} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}\right\} \times \tilde{\phi}(L_1 s_1 + L_2 s_2 + x) ds_1 ds_2
= \frac{1}{h_1 \cdots h_{r_0}} \int_{s_1 \in \mathbb{R}^{r_0}, s_2 \in \mathbb{R}^{(p - r_0)}} K(s_1^T H^{-2} s_1) \tilde{\phi}(L_1 s_1 + L_2 s_2 + x) ds_1 ds_2
= \int_{s_1 \in \mathbb{R}^{r_0}} K(\|s_1\|^2) \phi_{L_2}(x + L_1 H s_1) ds_1,
= \phi_{L_2}(x) + \frac{R_1(K)}{2} \text{tr}\{H L_1^T \tilde{\phi}(L_2(x) L_1 H)\} + o(\|h\|^2),
where the last equality is due to the Taylor expansion of \( \phi_{L_2}(x + L_1 Hs_1) \) and the condition \( \int_{s_1 \in \mathbb{R}^p} s_1 K(\|s_1\|^2)ds_1 = 0 \). Therefore, we have

\[
\sup_{x \in \Omega^d} |E\{\phi_n(x)\} - \phi_{L_2}(x)| \leq \sup_{x \in \Omega^d} |R_1(K)tr\{HL_1^\top \phi_{L_2}(x)L_1H\}|\{1 + o(1)\} = O(\|h\|^2). \tag{S1.7}
\]

Recall that \( f(x) = g(L_0^\top x) \). According to the Taylor’s expansion,

\[
\sup_{x \in \Omega^d} |\phi_{L_2}(x) - \phi(x)| \\
\quad = \sup_{x \in \Omega^d} \left| \int_{s_2 \in \mathbb{R}^{p-r_0}} g(L_0^\top x + L_0^\top L_2s_2)f_X(x + L_2s_2)ds_2 - g(L_0^\top x)\nu(x) \right| \\
\quad = \sup_{x \in \Omega^d} \left| g(L_0^\top x)f_{L_2}(x) + \left( L_0^\top x \right)^\top L_0^\top L_2 \int_{s_2 \in \mathbb{R}^{p-r_0}} s_2f(x + L_2s_2)ds_2 - g(L_0^\top x)\nu(x) + o(\|L_0^\top L_2\|) \right| \\
\quad = \sup_{x \in \Omega^d} \left| g(L_0^\top x)f_{L_2}(x) - g(L_0^\top x)\nu(x) + O(\|L_0^\top L_2\|) \right| \\
\quad \leq \sup_{x \in \Omega^d} |g(L_0^\top x)|o(1) + O(\|L_0^\top L_2\|), \tag{S1.8}
\]

where the last inequality holds by the fact that \( f_{L_2}(x) = f_{r_0}(L_0^\top x)\{1 + o(1)\} \), as \( n \to \infty \) and \( \|L_0^\top L_2\| \to 0 \). This result can be derived using the condition (C1) and the Taylor expansion of \( f_{L_2}(x) \). In fact, recall that \( f_X(x) \) is the density of \( X \) and thus the density function \( f_Q(u) \) of \( U = QX \) satisfies \( f_Q(u) = f_X(x) \) for any \( p \times p \) rotation matrix \( Q \). By taking \( Q^\top = (L_0 \ L_0^+) \),
we have that as \( \|L_0^\top L_2\| \to 0 \),

\[
f_{L_2}(x) = \int_{s_2 \in \mathbb{R}^{n-r_0}} f_X(x + L_2s_2) ds_2
\]

\[
= \int_{s_2 \in \mathbb{R}^{n-r_0}} f_Q \left( \begin{pmatrix} L_0^\top x + L_0^\top L_2s_2 \\ L_0^\top x + L_0^\top L_2s_2 \end{pmatrix} \right) ds_2
\]

\[
= \int_{\tilde{s}_2 \in \mathbb{R}^{n-r_0}} f_Q \left( \begin{pmatrix} L_0^\top x \\ \tilde{s}_2 \end{pmatrix} \right) d\tilde{s}_2 \{1 + o(1)\}
\]

\[
= f_{r_0}(L_0^\top x) \{1 + o(1)\}.
\]

Hence, as \( n \to 0 \), if \( \|h\| \to 0 \) and \( \|L_0^\top L_2\| \to 0 \), (S1.7) in conjunction with (S1.8) yields (S1.5). On the other hand, following a similar proof of Lemma B.1 in Newey (1994) and applying condition (C5) and \( h_1 \cdots h_{r_0} > n^{-\delta} \) for some \( 0 < \delta < 1 \), we have

\[
\sup_{x \in \Omega^\delta} |\phi_n(x) - E\{\phi_n(x)\}| \to 0, \quad \text{in probability.} \quad \text{(S1.9)}
\]

Therefore, (S1.7), (S1.8) and (S1.9) yield

\[
P\{ \sup_{x \in \Omega^\delta} |\phi_n(x) - f(x)\nu(x)| \geq \frac{\epsilon(\tau - \epsilon)}{2} \} \to 0.
\]

Likewise, by replacing \( Y_i \) with 1, it can be shown that \( \sup_{x \in \Omega^\delta} |\nu_n(x) - \nu(x)| \to 0 \) in probability. As a result, combining inequalities (S1.4), we have proved (S1.3).

In conclusion, in case of \( \|L_0^\top L_2\| \to 0 \) and \( \|h\| \to 0 \), \( CM_n(M) - \eta_0 \) is at
the order of $o_p(1)$. This violates the definition that $CM_n(\hat{M}) \leq CM_n(M)$ for $M \in S_p^+$. The proof is complete.

S1.2 Proof of Theorem 2

For brevity, we write $w(X_i)$ by $w_i$. The following lemmas are needed to prove Theorem 2. The proofs of Lemmas 1–4 is given latter.

**Lemma 1.** Suppose conditions (C1)–(C5) hold. Then, $E\{K_M(X - x)\} = f_L(x) + O(\|h\|^2)$. Moreover, for any $i = 1, \ldots, n$, $\sum_{j \neq i} K_M(X_j - X_i) = nf_L(X_i)\{1 + o_p(1)\}$.

**Lemma 2.** Define $\sigma^2_{L_2}(x) = \int_{s \in \mathbb{R}} (p - r_0)f_X(x + L_2s)\sigma^2(x + L_2s)ds$. Under conditions (C1)–(C5), we have $E\{(K^*_j)^2\sigma^2_j w_i\} = \frac{R_2(K)V_0}{n^2h_1 \cdots h_r_0}\{1 + o(1)\}$, where $R_2(K) = \int_{s \in \mathbb{R}^{r_0}} K^2(\|s\|^2)ds$ and $V_0 = \int_{x \in \mathbb{R}^p} \sigma^2(L_0^+x)\frac{f_X(x)w(x)}{f_{r_0}(L_0^+x)}dx$.

**Lemma 3.** Under regularity conditions (C1)–(C5), suppose $\|L_0^+L_2\| \to 0$ and $\|h\| \to 0$. Then, for any $t \in \Omega$,

$$E[(f(X) - f(t))K_M(X - t)] = \psi(t, h, L_1) + o(\|h\|^2 + \|L_0^+L_2\|),$$

where the definition of $\psi(\cdot)$ is given in Theorem 2.
Lemma 4. Under regularity conditions (C1)–(C5), for any $t \in \Omega$, $\text{Var}[\{f(X) - f(t)\}K_M(X - t)] = O\{\|h\|^2/(h_1 \cdots h_{r_0})\}$. Consequently,

$$\text{Var}\left[\frac{1}{n} \sum_{j=1}^{n} \{f(X_j) - f(t)\}K_M(X_j - t)\right] = O\left(\frac{1}{nh_1 \cdots h_{r_0}}\right).$$

Proof of Theorem 2. Write

$$CM_n(M) = \frac{1}{n} \sum_{i=1}^{n} w_i \epsilon_i^2 + \frac{1}{n} \sum_{i=1}^{n} \{B(X_i)\}^2 w_i + \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \epsilon_j K_{j,i}^*\right)^2 w_i - \frac{2}{n} \sum_{i=1}^{n} B(X_i) w_i \epsilon_i - \frac{2}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \epsilon_i \epsilon_j K_{j,i}^* w_i + \frac{2}{n} \sum_{i=1}^{n} B(X_i) \sum_{j=1}^{n} \epsilon_j K_{j,i}^* w_i$$

$$\equiv \eta_0 + \eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5,$$

where $\tilde{f}(X_i) = \sum_{j=1}^{n} f(X_j)K_{j,i}^*$ and $B(X_i) = \tilde{f}(X_i) - f(X_i)$. Here $B(\cdot)$ stands for the bias. Observe the facts that

(a) $\eta_0 \equiv n^{-1} \sum_{i=1}^{n} w_i \epsilon_i^2$ is free of $M$ and thus it is irrelevant to the minimization over $M$.

(b) $\eta_1 \equiv n^{-1} \sum_{i=1}^{n} \{B(X_i)\}^2 w_i$ stands for the bias term and $\eta_1 \geq 0$.

(c) $\eta_2$ is viewed as the variance term and $\eta_2 \geq 0$. $E(\eta_2|X_1, \ldots, X_n) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} (K_{j,i}^*)^2 \sigma_j^2 w_i$.

(d) $E(\eta_3|X_1, \ldots, X_n) = 0$ and $E(\eta_3^2|X_1, \ldots, X_n) = 4n^{-2} \sum_{i=1}^{n} \{B(X_i)\}^2 \sigma_i^2 w_i^2$.

Hence, $\eta_3 = O_p(n^{-1} \sqrt{\sum_{i=1}^{n} \{B(X_i)\}^2})$. 

(e) \( E(\eta_4 | X_1, \ldots, X_n) = 0 \) and \( E(\eta_1^2 | X_1, \ldots, X_n) = 4n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left\{ (K_{j,i}^*)^2 w_i^2 + K_{j,i}^* K_{i,j}^* w_i w_j \right\} \sigma_i^2 \sigma_j^2 \). Hence, \( \eta_4 = O_p (n^{-1} \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} (K_{j,i}^*)^2}) \).

(f) \( E(\eta_5 | X_1, \ldots, X_n) = 0 \) and \( \eta_5 = O_p (n^{-1/2} \sup_{X \in \Omega^\delta} |B(X)|) \).

The above statements (a)–(e) are trivial and we only give one justification for statement (f).

Since \( \|h\| \to 0 \) as \( n \to \infty \), \( K_{j,i}^* w_i = 0 \) for all \((X_i, X_j)\) if \( X_i \) or \( X_j \) is outside \( \Omega^\delta \) for all large \( n \). Set \( a_n(x) = n^{-1} \sum_{t \neq i} K_M(X_i - x) \). By the Lemma 1, with probability one, for all large \( n \), there exists some constant \( C > 0 \) such that

\[
1/C \leq \inf_{x \in \Omega^\delta} a_n(x) \leq \sup_{x \in \Omega^\delta} a_n(x) \leq C.
\]

It follows that with probability one, for all large \( n \),

\[
\sup_{1 \leq j \leq n} \left| \sum_{i=1}^{n} B(X_i) K_{j,i}^* w_i \right| = \sup_{X_j \in \Omega^\delta} \left| \sum_{i=1}^{n} B(X_i) K_{j,i}^* w_i \right| \\
\leq \left\{ \sup_{X_j \in \Omega^\delta} |B(X_i) w_i| \right\} \left\{ \sup_{X_j \in \Omega^\delta} \sum_{j=1}^{n} K_{j,i}^* \right\} \\
= \left\{ \sup_{X_j \in \Omega^\delta} |B(X_i) w_i| \right\} \left\{ \sup_{X_j \in \Omega^\delta} \frac{1}{n} \sum_{j=1}^{n} K_M(X_j - X_i) \right\} \\
\leq C^2 \sup_{X_i \in \Omega^\delta} |B(X_i) w_i|.
\]

Hence,

\[
E(\eta_5^2 | X_1, \ldots, X_n) = 4n^{-2} \sum_{j=1}^{n} \sigma_j^2 \left( \sum_{i=1}^{n} B(X_i) K_{j,i}^* w_i \right)^2 = O_p (n^{-1} \sup_{X \in \Omega^\delta} (B(X))^2 w_i^2).
\]
In the following, we intend to show that $\eta_1$ and $\eta_2$ are the dominating terms, compared with $\eta_3$, $\eta_4$ and $\eta_5$.

Write

$$\{B(X_i)\}^2 = \{\tilde{f}(X_i) - f(X_i)\}^2$$

$$= \left( \frac{n^{-1} \left[ \sum_{j \neq i} \{f(X_j) - f(X_i)\} K_M(X_j - X_i) \} \right]^2}{\left\{ n^{-1} \sum_{j \neq i} K_M(X_j - X_i) \right\}^2} \right).$$

By Lemma 1, the above denominator is $f_2^2(X_i)\{1 + o_p(1)\}$. And it follows from Lemmas 3–4 that the above numerator is $\psi_2^2(X_i, h, L_1) + o_p\left( \|h\|^4 + \|L_0^\top L_2\|^2 + \frac{1}{nh_1 \cdots h_{r_0}} \right)$.

Hence, by the law of large number and the continuous mapping theorem, we have

$$\eta_1 = \frac{1}{n} \sum_{i=1}^n \{B(X_i)\}^2 w_i = \int_{x \in \mathbb{R}^p} \frac{\psi_2(x, h, L_1)}{f_2^2(L_0 x)} f_X(x) w(x) dx$$

$$+ o_p\left( \|L_0^\top L_2\|^2 + \|h\|^4 + \frac{1}{nh_1 \cdots h_{r_0}} \right).$$

Write

$$\eta_2 = \frac{1}{n} \sum_{j=1}^n \epsilon_j^2 \hat{a}_j + \frac{1}{n} \sum_{j_1=1}^{n} \sum_{j_2 \neq j_1} \epsilon_{j_1} \epsilon_{j_2} \hat{a}_{j_1,j_2},$$

where $\hat{a}_j = \sum_{i=1}^n (K_{j,i}^*)^2 w_i$ for all $i$ and $\hat{a}_{j_1,j_2} = \sum_{i=1}^n K_{j_1,i}^* K_{j_2,i}^* w_i$ for all $j_2 \neq j_1$. We now show that $\sum_{j=1}^n \epsilon_j^2 \hat{a}_j$ is the dominant term in $\eta_2$.

First, from Lemma 2, we have

$$E\left( \frac{1}{n} \sum_{j=1}^n \epsilon_j^2 \hat{a}_j \right) = \frac{R_2(K)V_0}{nh_1 \cdots h_{r_0}} \{1 + o(1)\}.$$
Second, it can be easily verified that $E\left(\sum_{j_1=1}^{n} \sum_{j_2\neq j_1}^{n} \epsilon_{j_1} \epsilon_{j_2} \tilde{a}_{j_1,j_2}\right) = 0$. By Lemmas 1–2, we have $\tilde{a}_{j_1,j_2} = O_p(n^{-1})$. And recall condition (C5), it follows that

$$E\left(\sum_{j_1=1}^{n} \sum_{j_2\neq j_1}^{n} \epsilon_{j_1} \epsilon_{j_2} \tilde{a}_{j_1,j_2}\right)^2 = E\left(\sum_{j_1=1}^{n} \sum_{j_2\neq j_1}^{n} \sigma_{j_1}^2 \sigma_{j_2}^2 \tilde{a}_{j_1,j_2}^2\right) = O(1).$$

As a result,

$$\eta_2 = \frac{1}{n} \sum_{j=1}^{n} \epsilon_j^2 \tilde{a}_j \{1 + o_p(1)\} = \frac{R_2(K) V_0}{nh_1 \cdots h_{r_0}} \{1 + o_p(1)\}.$$

Since $\eta_3, \eta_4$ and $\eta_5$ are of smaller order than $\|L_0^T L_2\|^2 + \|h\|^4 + 1/(nh_1 \cdots h_{r_0})$ and $\eta_0$ is free of $M$, then

$$CM_n(M) - \eta_0 = \eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5$$

$$= \int_{x \in \mathbb{R}^p} \frac{\psi^2(x, h, L_1)}{f_{L_0}(L_0 x)} f_{x}(x) w(x) dx + \frac{R_2(K) V_0}{nh_1 \cdots h_{r_0}}$$

$$+ o_p \left(\|L_0^T L_2\|^2 + \|h\|^4 + \frac{1}{nh_1 \cdots h_{r_0}}\right).$$

The proof is complete.

\[ \square \]

**S1.3 Proof of Corollary 1**

It is seen that $L_0^T L_2$ is only contained in the bias term of the asymptotic expansion shown in Theorem 2. And we can easily verify that

$$\psi(x, h, L_1) = -\text{vec}(T^T)^\top \text{vec}\{b(L_0^T t)\dot{g}(L_0^T t)^\top\} + R_1(K) \text{tr}(HL_1^T L_0^T A(L_0^T t)L_0^T L_1 H),$$
where \( T = L_0^\top L_2 \). Let \( \tilde{b}(t) = \text{vec}\{b(L_0^\top t)\dot{g}(L_0^\top t)^\top\} \) and

\[
\tilde{c}_1(t, h) = R_1(K)\text{tr}(HL_1^\top L_0 A(L_0^\top t)L_0^\top L_1 H).
\]

Since the objective function is quadratic, the optimization procedure over \( \text{vec}(T^\top) \) yields the solution

\[
\begin{aligned}
&\left\{ -\int_{t \in \mathbb{R}_p} \tilde{b}(t)\tilde{b}(t)^\top f_X(t) f_r^2(L_0^\top t) dt \right\}^+ \int_{t \in \mathbb{R}_p} \tilde{c}_1(t, h)\tilde{b}(t) f_X(t) f_r^2(L_0^\top t) dt,
\end{aligned}
\]

where \( A^+ \) denotes the generalized inverse of a matrix \( A \). By some simple calculations, it can be shown that the order of (S1.10) is \( O(\|h\|^2) \).

Since \( L_0^\top L_1 \) is asymptotically orthonormal, we obtain that the \( \|L_0^\top L_2\| = O(\|h\|^2) \).

Further, to find the optimal rate of the bandwidth \( h \), we also consider optimizing the asymptotic expansion. Let \( L_0^\top L_1 = (\tilde{\ell}_1, \ldots, \tilde{\ell}_{r_0}) \) and then

\[
\tilde{c}_1(t, h) = R_1(K) \sum_{j=1}^{r_0} h_j^2 \tilde{\ell}_j^\top A(L_0^\top t) \tilde{\ell}_j.
\]

Taking derivative over \( h_k \), \( k = 1, \ldots, r_0 \), we have

\[
\frac{\partial\{CM_n(M) - \eta_0\}}{\partial h_k} = 4h_k \tilde{C}_k(L_1) + 4h_k \sum_{j=1}^{r_0} C_j h_j^2 - \frac{R_2(K)V_0}{nh_k^2(\prod_{j \neq k} h_j)} = 0,
\]

where

\[
C_j = \{R_1(K)\}^2 \int_{t \in \mathbb{R}_p} \{\tilde{\ell}_k^\top A(L_0^\top t)\tilde{\ell}_k\}\{\tilde{\ell}_j^\top A(L_0^\top t)\tilde{\ell}_j\} f_X(t) f_r^2(L_0^\top t) dt = O(1)
\]
and

\[ \tilde{C}_k(L_1) = R_1(K) \int_{t \in \mathbb{R}^p} \hat{g}(L_0^T t) \tau^T T \hat{b}(L_0^T t) \hat{\ell}_k f_X(t) \frac{w(t)}{f_{r_0}^2(L_0^T t)} dt \]

\[ = R_1(K) \text{vec}(T^T) \tau^T \times \int_{t \in \mathbb{R}^p} \text{vec}\{(b(L_0^T t) \hat{g}(L_0^T t) \tau^T) \hat{\ell}_k A(L_0^T t) \hat{\ell}_k f_X(t) \frac{w(t)}{f_{r_0}^2(L_0^T t)} dt \}
\]

\[ = O(\|h\|^2). \]

As a result, we obtain that \( \hat{h} = O\{n^{-1/(r_0+4)}\} \). This completes the proof.

### S1.4 Proof of Proposition 1

Recall that from Theorem 1, we have \( \text{CM}_n(\hat{M}_{r_0}) - \tilde{\eta}_0 = o_p(1) \), where \( \tilde{\eta}_0 = E(w(X)\sigma^2(L_0^T X)) \) is irrelevant to \( r_0 \). When \( 1 \leq r < r_0 \), we can show that \( \text{CM}_n(\hat{M}_r) - \tilde{\eta}_0 \geq c_1 + o_p(1) \) for some constant \( c_1 > 0 \). Let \( \hat{L}_1(r) \in \mathbb{R}^{p \times r} \) be the CVML estimator when the dimension of CMS is set to be \( r \) and \( \hat{L}_1^+(r) \) be the augmented orthonormal basis in \( \mathbb{R}^p \). Since the column vectors of \( L_0 \) and \( L_0^\perp \) form a set of basis in \( \mathbb{R}^p \), there exists a unique decomposition of \( \hat{L}_1^+(r) \) such that

\[ \hat{L}_1^+(r) = L_0 A(r) + L_0^\perp B(r), \tag{S1.11} \]

where \( A(r) \) is a \( r_0 \times (p - r) \) matrix and \( B(r) \) is a \( (p - r_0) \times (p - r) \) matrix.

We now show that \( \|L_0^\top \hat{L}_1^+(r)\| \) does not converge to zero. Suppose that \( \|L_0^\top \hat{L}_1^+(r)\| \) converges to zero. Then by the decomposition (S1.11), we have
that \( \|L_0^\top \hat{L}_1^\top (r)\| = \|A(r)\| \rightarrow 0 \). Since the column vectors of \( \hat{L}_1^\top (r) \) are orthogonal, we have that \( \hat{L}_1^\top (r)^\top \hat{L}_1^\top (r) = A(r)^\top A(r) + B(r)^\top B(r) = I_{(p-r)} \), where \( I_{(p-r)} \) is a \((p-r) \times (p-r)\) identity matrix. It follows from \( \|A(r)\| \rightarrow 0 \) that \( B(r)^\top B(r) \rightarrow I_{(p-r)} \). However, due to \( r < r_0 \), the rank of \( B(r)^\top B(r) \) shall not exceed \((p - r_0)\) and is not able to attain \((p - r)\), which is in contradiction with \( B(r)^\top B(r) \rightarrow I_{(p-r)} \). Therefore, we have \( \|L_0^\top \hat{L}_1^\top (r)\| \nrightarrow 0 \).

Then it follows from similar proofs in Theorem 1 that \( \hat{f}^{(-i)}(X_i) \) with \( M \) set to be \( \hat{M}_r = \hat{L}_1(r) \hat{H}^{-2} \hat{L}_1(r)^\top \) is not a consistent estimator of \( f(X_i) \) when \( \|L_0^\top \hat{L}_1^\top (r)\| \nrightarrow 0 \). Moreover, by similar derivation of (S1.1), we obtain that for any \( 1 \leq r < r_0 \), there exists a positive constant \( c_1 \) such that \( \text{CM}_n(\hat{M}_r) - \bar{\eta}_0 \geq c_1 + o_p(1) \). As a result, \( \text{CM}_n(\hat{M}_r) > \text{CM}_n(\hat{M}_{r_0}) \) for all \( 1 \leq r < r_0 \) because of the lack of fit.

One the other hand, when \( r > r_0 \), let \( L_1(r) \) represent the column orthogonal matrix \( L_1 \) of order \( p \times r \) and \( L_2(r) \) be the augmented orthonormal basis in \( \mathbb{R}^p \). Then, we have

\[
\hat{f}^{(-i)}(X_i) = \frac{\sum_{j \neq i} Y_j K_M(X_j - X_i)}{\sum_{j \neq i} K_M(X_j - X_i)},
\]

where \( M = L_1(r) \hat{H}^{-2} L_1(r)^\top \). Following a similar derivation as Theorem 1, we have that as \( n \rightarrow \infty \), \( \|h\| \rightarrow 0 \) and \( \|L_0^\top L_2(r)\| \rightarrow 0 \), \( \hat{f}^{(-i)}(X_i) \) is also a consistent estimate for \( f(X_i) \). As a result, \( \text{CM}_n(\hat{M}_r) = \bar{\eta}_0 + o_p(1) \) for all \( r_0 \leq r \leq p \). Therefore, we have \( \text{CM}_n(\hat{M}_r)/\text{CM}_n(\hat{M}_{r_0}) \rightarrow_p 1 \), for all
Proof of Lemma 1. Recall that $L = (L_1 \quad L_2)$ is a $p \times p$ orthogonal matrix. Analogue to (S1.6), we have

$$E\{K_M(X - x)\} = \int_{s_1 \in \mathbb{R}^{r_0}} K(||s_1||^2)f_{L_2}(x + L_1Hs_1)ds_1. \quad (S1.12)$$

According to condition (C3) and the Taylor expansion of $f_{L_2}(x + L_1Hs_1)$, (S1.12) equals

\[ f_{L_2}(x) + \left\{ \hat{f}_{L_2}(x) \right\}^\top L_1H \int_{s_1 \in \mathbb{R}^{r_0}} s_1K(||s_1||^2)ds_1 \]
\[ \quad + \frac{1}{2} \int_{s_1 \in \mathbb{R}^{r_0}} s_1^\top H^\top L_1^\top \hat{f}_{L_2}(x)L_1HS_1K(||s_1||^2)ds_1 + o(||h||^2) \]
\[ = f_{L_2}(x) + \frac{R_1(K)}{2} \text{tr}\{HL_1^\top \hat{f}_{L_2}(x)L_1H\} + o(||h||^2) \]
\[ = f_{L_2}(x) + O(||h||^2), \]

where $\hat{f}_{L_2}(x) = \partial f_{L_2}(x)/\partial x$ and $\tilde{f}_{L_2}(x) = (\partial^2/\partial x^2)f_{L_2}(x)$.

On the other hand, a similar calculation to (S1.12) and (S1.13) yields

$$E\{K_M^2(X - x)\} = \frac{R_2(K)}{h_1 \cdots h_{r_0}} f_{L_2}(x)\{1 + o(1)\}. \tag{S1.14}$$

Consequently, $\text{Var}\{n^{-1}\sum_{i=1}^n K_M(X_i - x)\} = O\{(nh_1 \cdots h_{r_0})^{-1}\} = o(1)$. \hfill \Box

Proof of Lemma 2. By Lemma 1 and the continuous mapping theorem, for
where the last two equalities hold by invoking $R_2(K) = \int_{s \in \mathbb{R}^p} \{K(\|s\|^2)\}^2 ds$.

Noting that

$$
\sigma^2(x + L_2s_2) = \sigma^2(L_0^\top x + L_0^\top L_2s_2)
$$

$$
= \sigma^2(L_0^\top x) + \sigma^2(L_0^\top x)^\top L_0^\top L_2s_2 + o(||L_0^\top L_2||).
$$

It follows that

$$
E\{(K_{j,i}^*)^2 w_i \sigma_j^2\}
$$

$$
= \frac{R_2(K)}{n^2 h_1 \cdots h_{r_0}} \int_{x \in \mathbb{R}^p} \sigma^2(L_0^\top x) \frac{f(x) w(x)}{f^2_{L^2}(x)} dx \{1 + o(1)\}
$$

$$
= \frac{R_2(K)}{n^2 h_1 \cdots h_{r_0}} \int_{x \in \mathbb{R}^p} \sigma^2(L_0^\top x) \frac{f(x) w(x)}{f^2_{L^2}(x)} dx \{1 + o(1)\}
$$

$$
= \frac{R_2(K)}{n^2 h_1 \cdots h_{r_0}} \int_{x \in \mathbb{R}^p} \sigma^2(L_0^\top x) \frac{f(x) w(x)}{f_{L^2}(x)} dx \{1 + o(1)\}
$$

$$
= \frac{R_2(K) V_0}{n^2 h_1 \cdots h_{r_0}}
$$
Proof of Lemma 3. It follows from $f(x) = g(L_0^T x)$ that

\begin{align*}
E[\{f(X) - f(t)\}K_M(X - t)]
&= \frac{1}{h_1 \cdots h_{r_0}} \int_{s \in \mathbb{R}^p} \{f(Ls + t) - f(t)\}K(s_1^T H^{-2}s_1)f_X(Ls + t)ds \\
&= \int_{s_1 \in \mathbb{R}^{r_0}, s_2 \in \mathbb{R}^{(p-r_0)}} \{f(L_1 Hs_1 + L_2 s_2 + t) - f(t)\} \\
&\quad \times K(||s_1||^2)f_X(L_1 Hs_1 + L_2 s_2 + t)ds_1 ds_2. \\
&= \int_{s_1 \in \mathbb{R}^{r_0}, s_2 \in \mathbb{R}^{(p-r_0)}} \{g(L_0^T L_1 Hs_1 + L_0^T L_2 s_2 + L_0^T t) - g(L_0^T t)\} \\
&\quad \times K(||s_1||^2)f_X(L_1 Hs_1 + L_2 s_2 + t)ds_1 ds_2ds_2 ds_1. \quad (S1.15)
\end{align*}

Now expanding both $g(L_0^T L_1 Hs_1 + L_0^T L_2 s_2 + L_0^T t)$ and $f_X(t + L_1 Hs_1 + L_2 s_2)$ in Taylor expansions yield

\begin{align*}
g(L_0^T L_1 Hs_1 + L_0^T L_2 s_2 + L_0^T t) - g(L_0^T t)
&= \hat{g}(L_0^T t_1)^T (L_0^T L_1 Hs_1 + L_0^T L_2 s_2) \\
&\quad + \frac{1}{2} (L_0^T L_1 Hs_1 + L_0^T L_2 s_2)^T \hat{g}(L_0^T t)(L_0^T L_1 Hs_1 + L_0^T L_2 s_2) + o(||h||^2)
\end{align*}

and

\begin{align*}
f_X(t + L_1 Hs_1 + L_2 s_2) = f_X(t + L_2 s_2) + \hat{f}_X(t + L_2 s_2)^T L_1 Hs_1 + o(||h||).
\end{align*}
Therefore, (S1.15) equals

\[
\begin{align*}
\int_{s_1 \in \mathbb{R}^{p_0}, s_2 \in \mathbb{R}^{(p-r_0)}} \left\{ \dot{g}(L_0^T t)^\top (L_0^T L_1 H s_1 + L_0^T L_2 s_2) + \\
\frac{1}{2} (L_0^T L_1 H s_1 + L_0^T L_2 s_2)^\top \dot{g}(L_0^T t_1) (L_0^T L_1 H s_1 + L_0^T L_2 s_2) + o(\|h\|^2) \right\} K(\|s_1\|^2) \{f_X(t + L_2 s_2) + \dot{f}_X(t + L_2 s_2)^\top L_1 H s_1 + o(\|h\|)\} ds_1 ds_2.
\end{align*}
\]

\[
= \dot{g}(L_0^T t)^\top L_0^T \int_{s_1 \in \mathbb{R}^{p_0}, s_2 \in \mathbb{R}^{(p-r_0)}} L_2 s_2 f_X(t + L_2 s_2) ds_2 \\
+ R_1(K) \dot{g}(L_0^T t)^\top L_0^T L_1 H^2 L_1^\top \int_{s_2 \in \mathbb{R}^{(p-r_0)}} \dot{f}_X(t + L_2 s_2) ds_2 \\
+ \frac{1}{2} R_1(K) \text{tr} \{H L_1^T L_0 \dot{g}(L_0^T t) L_0^T L_1 H\} f_{L_2}(t) + o(\|h\|^2 + \|L_0^T L_2\|)
\equiv \Delta_1 + \Delta_2 + \Delta_3 + o(\|h\|^2 + \|L_0^T L_2\|)
\]

As \(\|L_0^T L_2\| \to 0\) and \(\|h\| \to 0\), taking \(Q^\top = (L_0 L_0^\perp)\), we have

\[
\Delta_1 = \dot{g}(L_0^T t)^\top L_0^T L_2 \int_{s_2 \in \mathbb{R}^{(p-r_0)}} s_2 f_X(t + L_2 s_2) ds_2
\]

\[
= \dot{g}(L_0^T t)^\top L_0^T L_2 \int_{s_2 \in \mathbb{R}^{(p-r_0)}} s_2 f_Q \left( \begin{array}{c} L_0^T t + L_0^T L_2 s_2 \\ L_0^\perp t + L_0^\perp L_2 s_2 \end{array} \right) ds_2
\]

\[
= \dot{g}(L_0^T t)^\top L_0^T L_2 \int_{s_2 \in \mathbb{R}^{(p-r_0)}} \left( s_2 - L_0^\perp t \right) f_Q(L_0^T t, s_2) ds_2 \{1 + o(1)\}
\]

\[
= \dot{g}(L_0^T t)^\top L_0^T L_2 \int_{s_2 \in \mathbb{R}^{(p-r_0)}} (s_2 - L_0^\perp t) f_{u_2 | u_1}(s_2 | L_0^T t) ds_2 f_{r_0}(L_0^T t) \{1 + o(1)\}
\]

\[
= \dot{g}(L_0^T t)^\top L_0^T L_2 E \left. u_2 | u_1 \right| (U_2 - L_0^\perp t) | U_1 = L_0^T t) f_{r_0}(L_0^T t) \{1 + o(1)\}
\]

\[
= \dot{g}(L_0^T t)^\top L_0^T L_2 b(L_0^T t) \{1 + o(1)\},
\]

(S1.16)
where $u_1 \in \mathbb{R}^{r_0}$, $u_2 \in \mathbb{R}^{(p-r_0)}$, $U_1 \in \mathbb{R}^{r_0}$, $U_2 \in \mathbb{R}^{(p-r_0)}$ and

$$b(L_0^\top t) = E_{u_2|u_1}(U_2 - L_0^\perp t|U_1 = L_0^\top t)f_{r_0}(L_0^\top t).$$

For $\Delta_2$, it is straightforward to show that

$$\Delta_2 = R_1(K) \text{tr}\{HL_1^\top L_0 \dot{g}(L_0^\top t)\hat{f}_{r_0}(L_0^\top t)^\top L_0^\top L_1 H\} \{1 + o(1)\}, \quad \text{and}$$

$$\Delta_3 = \frac{1}{2} R_1(K) \text{tr}\{HL_1^\top L_0 \ddot{g}(L_0^\top t)L_0^\top L_1 H\} f_{r_0}(L_0^\top t) \{1 + o(1)\}.$$

As a result,

$$E[f(X) - f(t)] K_M(X - t)$$

$$= \dot{g}(L_0^\top t)^\top L_0^\top L_2 b(L_0^\top t) + R_1(K) \text{tr}\{HL_1^\top L_0 \dot{g}(L_0^\top t)\hat{f}_{r_0}(L_0^\top t)^\top L_0^\top L_1 H\}$$

$$+ \frac{1}{2} R_1(K) \text{tr}\{HL_1^\top L_0 \ddot{g}(L_0^\top t)L_0^\top L_1 H\} f_{r_0}(L_0^\top t) + o(\|h\|^2 + \|L_0^\top L_2\|)$$

$$= g(L_0^\top t)^\top L_0^\top L_2 b(L_0^\top t) + R_1(K) \text{tr}\{HL_1^\top L_0 A(L_0^\top t)L_0^\top L_1 H\}$$

$$+ o(\|h\|^2 + \|L_0^\top L_2\|)$$

$$= \psi(t, h, L_1) + o(\|h\|^2 + \|L_0^\top L_2\|),$$

where

$$A(L_0^\top t) = \frac{1}{2} \ddot{g}(L_0^\top t) f_{r_0}(L_0^\top t) + \dot{g}(L_0^\top t) \hat{f}_{r_0}(L_0^\top t)^\top.$$
Proof of Lemma 4. With an analogue calculation to Lemma 3, we have

\[
E[\{f(X) - f(t)\} K_M(X - t)]^2 \\
= \frac{1}{h_1 \cdots h_{r_0}} \int_{s_1 \in \mathbb{R}^{r_0}, s_2 \in \mathbb{R}^{r_0}} \{g(L_0^T L_1 H s_1 + L_0^T L_2 s_2 + L_0^T t) - g(L_0^T t)\}^2 \\
\times K^2(\|s_1\|^2) f_X(L_1 H s_1 + L_2 s_2 + t) ds_1 ds_2 \\
= \frac{1}{h_1 \cdots h_{r_0}} \int_{s_1 \in \mathbb{R}^{r_0}, s_2 \in \mathbb{R}^{r_0}} \{\dot{g}(L_0^T t)^T (L_0^T L_1 H s_1 + L_0^T L_2 s_2) + O(\|h\| + \|L_0^T L_2\|)\}^2 \\
\times K^2(\|s_1\|^2) \{f_X(t + L_2 s_2) + O(\|h\|)\} ds_1 ds_2.
\]

Recall that \( \int_{s_1 \in \mathbb{R}^{r_0}} s_1 K^2(\|s_1\|^2) ds_1 = 0 \) and \( \int_{s_1 \in \mathbb{R}^{r_0}} s_1 s_1^T K^2(\|s_1\|^2) ds_1 \) exists.

Let

\[
\int_{s_1 \in \mathbb{R}^{r_0}} s_1 s_1^T K^2(\|s_1\|^2) ds_1 = c_2 I_{r_0 \times r_0}
\]

for some \( c_2 \geq 0 \). It follows that

\[
\text{Var}[\{f(X) - f(t)\} K_M(X - t)] \\
\leq E[\{f(X) - f(t)\} K_M(X - t)]^2 \\
= \frac{c_2}{h_1 \cdots h_{r_0}} \text{tr} \{H L_1^T L_0 \dot{g}(L_0^T t) \dot{g}(L_0^T t)^T L_0^T L_1 H\} f_{L_2}(t)\{1 + o(1)\} \\
= \frac{c_2}{h_1 \cdots h_{r_0}} \dot{g}(L_0^T t)^T L_0^T L_1 H^2 L_1^T L_0 \dot{g}(L_0^T t) f_{L_2}(t)\{1 + o(1)\} \\
= O \left( \frac{\|h\|^2}{h_1 \cdots h_{r_0}} \right).
\]

\( \square \)
Reference
