Moving least-square method in learning theory

Hong-Yan Wang\(^a\), Dao-Hong Xiang\(^b,\)*, Ding-Xuan Zhou\(^c\)

\(^a\) School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou, Zhejiang 310018, China
\(^b\) Department of Mathematics, Chinese University of Hong Kong, Shatin, N. T., Hong Kong, China
\(^c\) Department of Mathematics, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China

Received 15 March 2009; received in revised form 11 November 2009; accepted 14 December 2009
Available online 23 December 2009
Communicated by Martin Buhmann

Abstract

Moving least-square (MLS) is an approximation method for data interpolation, numerical analysis and statistics. In this paper we consider the MLS method in learning theory for the regression problem. Essential differences between MLS and other common learning algorithms are pointed out: lack of a natural uniform bound for estimators and the pointwise definition. The sample error is estimated in terms of the weight function and the finite dimensional hypothesis space. The approximation error is dealt with for two special cases for which convergence rates for the total \(L^2\) error measuring the global approximation on the whole domain are provided.

\(\text{c}\)\(\text{⃝}\) 2009 Elsevier Inc. All rights reserved.

Keywords: Learning theory; Moving least-square method; Sample error; Norming condition; Approximation error

1. Introduction

Moving least-square method (MLS) is an approximation method for data smoothing [8], numerical analysis [10], statistics [14] and many other fields. In statistics MLS has the advantage of adapting to various types of designs such as random and fixed designs with partial data used for data reducing calculations. In this paper we apply MLS to the regression problem in
learning theory. It has advantages over classical learning algorithms in the sense that its involved hypothesis space can be very simple such as the space of linear functions or a polynomial space.

Let $X$ be a compact subset of $\mathbb{R}^n$ (input space) and $Y = \mathbb{R}$. A Borel probability measure $\rho$ on $Z := X \times Y$ is used to model the regression problem [13,3]. Let $\rho_X$ be the marginal distribution of $\rho$ on $X$ and $\rho(y|x)$ be the conditional distribution at $x \in X$. Then the regression function $f_\rho : X \to Y$ is defined by

$$f_\rho(x) = \int_Y y \, d\rho(y|x).$$

In this paper, we study the learning of the regression function $f_\rho$ from samples by the moving least-square method. We consider the case when the hypothesis space $\mathcal{H}$ for learning is a finite dimensional subspace of $C(X)$, the space of all continuous functions on $X$. The most important example of hypothesis space $\mathcal{H}$ is the space $\Pi_l$ of polynomials of degree at most $l$.

The moving least-square method involves a MSL weight function.

**Definition 1.** A function $\Phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ is called a MSL weight function if there exists some constant $c_q > 0$ such that

$$\Phi(x, t) \leq 1 \quad \forall x, t \in \mathbb{R}^n \quad (1.1)$$

and

$$\Phi(x, t) \geq c_q \quad \forall |x - t| \leq 1. \quad (1.2)$$

With a sample $z = \{(x_i, y_i)\}_{i=1}^m \in Z^m$, a hypothesis space $\mathcal{H}$ and a MSL weight function $\Phi$, we define the estimator $f_z$ of $f_\rho$ by MLS pointwisely: for each $x \in X$ let $f_z(x) = f_{z,\sigma,x}(x)$ where $f_{z,\sigma,x}$ is a solution of the following minimization problem

$$f_{z,\sigma,x} = \arg \min_{f \in \mathcal{H}} \left\{ \frac{1}{m} \sum_{i=1}^m \Phi \left( \frac{x}{\sigma}, \frac{x_i}{\sigma} \right) (f(x_i) - y_i)^2 \right\}. \quad (1.3)$$

Here $\sigma > 0$ is a scaling parameter corresponding to standard deviation in statistics. Throughout the paper we assume that the sample $z$ is drawn independently according to the measure $\rho$.

There has been extensive study on the order of local approximation $|f_{z,\sigma,x}(x) - f_\rho(x)|$ in the literature of statistics [14,4] and approximation theory [7,16,15] for which the sample $\{x_i\}$ is deterministic and well distributed. In learning theory we are interested in the global approximation of $f_\rho$ by $f_z$ on the whole input space $X$ and the sample $z$ is random. In particular, the error $\|f_z - f_\rho\|_{L^2_{\rho_X}}$ and its convergence rates as $m \to \infty$ are used to measure the performance of the learning algorithm, which is often stated under some choice of the parameter $\sigma = \sigma(m)$ and conditions on the measure $\rho$ and hypothesis space $\mathcal{H}$.

Mathematical analysis for the learning algorithm with the scheme (1.3) is different from that for two types of well-understood learning schemes for regression in the literature: empirical risk minimization scheme [13] and Tikhonov regularization scheme [3,11,12]. One obvious difference is the hypothesis spaces. The space $\mathcal{H}$ for (1.3) is only finite dimensional but the moving weights in (1.3) support the learning ability.

There are two other technical essential differences. The first is the lack of a natural uniform bound for $\|f_{z,\sigma,x}\|$. The second is the pointwise definition of the function $f_z$ which makes the total error $\|f_z - f_\rho\|_{L^2_{\rho_X}}$ difficult to estimate. By imposing two mild conditions on $\rho_X$ and $\mathcal{H}$ we can overcome the essential difficulty for the mathematical analysis.
The first condition is about regularity of the measure $\rho_X$. When $\rho_X$ is very irregular, it may happen that the sampling points \( \{x_i\}_{i=1}^m \) lie totally on a zero set of some function \( f \in \mathcal{H} \) (shown in Example 1 below) or they are all far away from some point \( x \in X \) (meaning that \( |x_i - x|/\sigma \) is very large for each \( i \)). Such situations make \( f_{\mathcal{L}_\sigma,x} \) less informative for learning \( f_\rho \) at \( x \). We impose a regularity condition for the marginal distribution \( \rho_X \) which governs the location of \( \{x_i\}_{i=1}^m \). This condition was introduced in the literature of harmonic analysis for studying function spaces [6]. Denote \( B(x, r) = \{u \in X : |u - x| \leq r\} \) for \( r > 0 \).

**Definition 2.** We say that a probability measure $\rho_X$ on $X$ satisfies the condition $L_\tau$ with exponent $\tau > 0$ if there are constants $r_0 > 0$ and $c_\tau > 0$ such that

$$
\rho_X(B(x, r)) \geq c_\tau r^\tau, \quad \forall 0 < r \leq r_0, \ x \in X.
$$

(1.4)

**Remark 1.** Condition (1.4) holds with $\tau = n$ and $c_\tau$ depending on $X$ if $\rho$ is the uniform distribution on $X$ and $X$ satisfies an interior cone condition [1] saying that there exist an angle $\theta \in (0, \pi/2)$, a radius $r > 0$, and a unit vector $\xi(x)$ for every $x \in X$ such that the cone

\[ C(x, \xi(x), \theta, r) = \left\{ x + ty : y \in \mathbb{R}^n, |y| = 1, y^T \xi(x) \geq \cos \theta, t \in [0, r] \right\} \]

is contained in $X$.

The second condition is a norming condition about the hypothesis space $\mathcal{H}$.

**Definition 3.** We say that $\mathcal{H}$ satisfies the norming condition with exponents $\zeta > 0$ and $d \in \mathbb{N}$ if there exist some constants $\sigma_0 > 0$ and $c_\mathcal{H} > 0$ such that for every $x \in X$ and $0 < \sigma \leq \sigma_0$, we can find points \( \{u_i\}_{i=1}^d \subset B(x, \sigma) \) satisfying \( |u_i - u_j| \geq 2c_\mathcal{H}\sigma \) for $i \neq j$ and

$$
\left( \sum_{i=1}^{d} |f(u_i)|^2 \right)^{1/2} \geq c_\mathcal{H}\sigma^\zeta \|f\|_{C(X)} \ \forall f \in \mathcal{H}.
$$

(1.5)

Condition (1.5) required $d$ to be at least $\widetilde{d}$, the dimension of $\mathcal{H}$. The point set \( \{u_i\}_{i=1}^d \) is closely related to the concept of norming set [5] which plays an important role in the study of scattered data interpolation [16,9]. The norming condition is satisfied by some finite dimensional spaces generated by many radial basis functions and by the polynomial hypothesis space $\Pi_1$.

Throughout the paper we assume that $|y| \leq M$ almost surely and $X \subseteq B(0, B_X)$ for some $M > 0$ and $B_X > 0$. We also assume that all functions from the hypothesis space $\mathcal{H}$ are Lipschitz on $X$. Since $\mathcal{H}$ is finite dimensional, the Lipschitz norm $\|f\|_{Lip} = \|f\|_{C(X)} + \sup_{x \neq t} \frac{|f(x) - f(t)|}{|x - t|}$ of $\mathcal{H}$ is equivalent to $\|f\|_{C(X)}$. So there exists a constant $C_{\mathcal{H},0} \geq 1$ such that

$$
\sup_{x \neq t} \frac{|f(x) - f(t)|}{|x - t|} \leq C_{\mathcal{H},0} \|f\|_{C(X)} \ \forall f \in \mathcal{H}.
$$

(1.6)

Now we can state two results on learning rates for the MLS learning algorithm which will follow from Theorem 5 in Section 5 with constants $C_1^*$ and $C_2^*$ given by (5.2) explicitly (depending $\mathcal{H}$ and $\tau$, $r_0$, but not on $\delta$, $m$ or $\varepsilon$).

**Theorem 1.** Assume condition $L_\tau$ with exponent $\tau > 0$ for $\rho_X$ and norming condition with exponents $\zeta > 0$ and $d \in \mathbb{N}$ for $\mathcal{H}$. Let $0 < \varepsilon < 1/4$. If $f_\rho \in \mathcal{H}$, then there exist constants $C_1^*$
and $C_2^*$ such that for $0 < \delta < 1$ and $m$ satisfying
\[ m \geq C_1^* + 4(\log(4/\delta))^2 + \sigma_0^{-((2\xi + \max(\tau, \tau_\xi))/\varepsilon)} \quad \text{and} \quad m^{1/2 - 2\varepsilon} \geq C_1^* \log m \quad (1.7) \]
we have with confidence $1 - \delta$,
\[ \| f_z - f_\rho \|_{L_{\rho X}^2} \leq C_2^* (\log(4/\delta) \log m)^{1/4} m^{e - 1/4}. \]

The above theorem is for the special case when $f_\rho \in H$. The next result is about another special case of 1-dimensional hypothesis space.

**Theorem 2.** Under the assumption of Theorem 1 and with $0 < \varepsilon < 1/4$, if $H$ is 1-dimensional with a basis function $\varphi$ and $f_\rho$ is Lipschitz, then for $0 < \delta < 1$ and $m$ satisfying (1.7), we have
\[ \| f_z - f_\rho \|_{L_{\rho X}^2} \leq C_2^* (\log(4/\delta) \log m)^{1/4} m^{e - 1/4} + \| \varphi \|_{Lip} \| f_\rho \|_{Lip} \Lambda_\sigma, \]
where
\[ \Lambda_\sigma = \sup_{x \in X} \left\{ \int_X \Phi \left( \frac{x}{\sigma} - \frac{u}{\sigma} \right) |x - u| |\varphi(u)| d\rho_X(u) / \int_X \Phi \left( \frac{x}{\sigma} - \frac{u}{\sigma} \right) (\varphi(u))^2 d\rho_X(u) \right\}. \]

2. Special error decomposition

Mathematical analysis for most classical least-square learning algorithms for regression can be conducted by error decompositions and the relation $\| f - f_\rho \|_{L_{\rho X}^2}^2 = \mathcal{E}(f) - \mathcal{E}(f_\rho)$ involving the generalization error $\mathcal{E}(f) = \int_Z (f(x) - y)^2 d\rho$.

The pointwise nature of the MLS scheme causes various errors and minimizers defined pointwisely in a moving way.

For $x \in X$ and $f : X \rightarrow \mathbb{R}$, we denote the **moving empirical error** (depending on $\sigma$) as
\[ \mathcal{E}_{Z,x}(f) = \frac{1}{m} \sum_{i=1}^m \Phi \left( \frac{x}{\sigma} - \frac{x_i}{\sigma} \right) (f(x_i) - y_i)^2 \quad (2.1) \]
and the **moving generalization error** as
\[ \mathcal{E}_x(f) = \int_Z \Phi \left( \frac{x}{\sigma} - \frac{u}{\sigma} \right) (f(u) - y)^2 d\rho(u, y). \quad (2.2) \]

**Definition 4.** Given the hypothesis space $\mathcal{H}$ and the measure $\rho$, we define a function $f_{\mathcal{H}}$ on $X$ by
\[ f_{\mathcal{H}}(x) = f_{\mathcal{H},\sigma,x}(x), \quad x \in X, \quad (2.3) \]
where
\[ f_{\mathcal{H},\sigma,x} = \arg \min_{f \in \mathcal{H}} \mathcal{E}_x(f), \quad x \in X. \]

We call the quantity $\| f_z - f_{\mathcal{H}} \|_{L_{\rho X}^2}$ the **sample error** which will be estimated in Section 4. Our sample error bound is valid in general as long as the norming condition and condition $L_\tau$ hold.
The approximation error \( \| f_\mathcal{H} - f_\rho \|_{L^2_{P_X}} \) involves relations between \( f_\rho \) and \( \mathcal{H} \). To see this, we notice that
\[
\mathcal{E}_\mathcal{H}(f) - \mathcal{E}_\mathcal{H}(f_\rho) = \int_X \phi \left( \frac{X}{\sigma}, \frac{u}{\sigma} \right) (f(u) - f_\rho(u))^2 \, d\rho_X(u) \quad \forall f : X \to \mathbb{R} \tag{2.4}
\]
and
\[
\mathcal{E}_\mathcal{H}(f) - \mathcal{E}_\mathcal{H}(f_{\mathcal{H}, \sigma, x}) = \int_X \phi \left( \frac{X}{\sigma}, \frac{u}{\sigma} \right) (f(u) - f_{\mathcal{H}, \sigma, x}(u))^2 \, d\rho_X(u) \quad \forall f \in \mathcal{H}. \tag{2.5}
\]
These expressions can be used to bound the approximation error which is beyond this paper.

When \( f_\rho \in \mathcal{H} \), we see from (2.4) that \( f_\rho = f_{\mathcal{H}} \). Hence the total error \( \| f_X - f_\rho \|_{L^2_{P_X}} \) is reduced to the sample error \( \| f_X - f_{\mathcal{H}} \|_{L^2_{P_X}} \).

Estimating the approximation error for the MLS scheme with a general regression function is a difficult and interesting topic. We shall provide an example in the next section to point out some difficulty and another example dealing with the approximation error in Section 5.

3. Bounding the MLS estimator

For empirical risk minimization schemes, the hypothesis space is a bounded set of functions, often compact, satisfying some further conditions such as the finiteness of VC dimension or being a uniform Glivenko–Cantelli class. So the estimators have natural uniform bounds.

For regularization schemes, a penalty term like \( \lambda \| f \|^2 \) (with a regularization parameter \( \lambda > 0 \)) is added to an empirical error, which yields an immediate uniform bound for functions involved in the optimization process such as \( \| f \| \leq \frac{M}{\sqrt{\lambda}} \) when \( |y| \leq M \).

Thus the above two learning schemes can be analyzed by uniform laws of large numbers or theory of uniform convergence for bounded set of functions together with approximation theory.

A key feature of MLS scheme (1.3) is the lack of a natural uniform bound for optimizing functions \( f_{z, \sigma, x} \), which causes difficulty for mathematical analysis and the choice of parameter \( \sigma \). Let us show this by an example where \( \mathcal{H} \) is a polynomial space.

Example 1. Let \( l \in \mathbb{N} \) and \( \mathcal{H} = P_l = \{ \sum_{|\alpha| \leq l} a_\alpha x^\alpha : a_\alpha \in \mathbb{R} \} \) where for \( \alpha = (\alpha^1, \ldots, \alpha^n) \in \mathbb{Z}_+^n \), and \( x = (x^1, \ldots, x^n) \in \mathbb{R}^n \), we denote \( |\alpha| = \sum_{i=1}^n \alpha^i \) and \( x^\alpha = \Pi_{i=1}^n (x^i)^{\alpha^i} \). If the sample \( z \) corresponds to a nonzero polynomial \( p \in \mathcal{H} \) such that \( p(x_i) = 0 \) for \( i = 1, \ldots, m \), then \( f_{z, \sigma, x} + cp \) is also a solution to (1.3) for any \( c \in \mathbb{R} \). Such solutions are not uniformly bounded when \( c \to \infty \).

The above example tells us that we should not expect a uniform bound for \( f_{z, \sigma, x} \) even though the probability for the sample \( z \) in Example 1 is small.

Our approach here is to use the assumptions on \( \rho_X \) and \( \mathcal{H} \) to obtain bounds with large confidence. Denote a constant
\[
C_{\mathcal{H}, \zeta} = \min \left\{ \frac{c_{\mathcal{H}}}{2^{\zeta+1} \sqrt{d} C_{\mathcal{H}, 0}}, \frac{c_{\mathcal{H}}}{2 \cdot 2} \right\} \tag{3.1}
\]
and
\[
A_{\tau, \zeta} = \left[ 2^{2\tau+3} + 2^{2\tau+3} n \log \left( 1 + 4B_X / C_{\mathcal{H}, \zeta} \right) \right] c_{\tau}^{-2} C_{\mathcal{H}, \zeta}^{-2\tau}. \tag{3.2}
\]
Theorem 3. Assume that \( \mathcal{H} \) satisfies the norming condition with exponents \( \zeta > 0 \) and \( d \in \mathbb{N} \), and \( \rho_X \) satisfies the condition \( L_\tau \) with exponent \( \tau > 0 \). Let \( 0 < \sigma \leq \min\{\sigma_0, 1, (r_0/C_{\mathcal{H}, \zeta})^{\text{max}(\zeta, 1)} \} \) and \( 0 < \delta < 1 \). If
\[
m \geq -A_{\tau, \zeta} \log(\delta \sigma) \sigma^{-2 \tau \max(\zeta, 1)},
\]
then with confidence \( 1 - \delta \), we have
\[
\|f_{x, \sigma, x}\|_{C(X)} \leq \frac{2^{3+\tau+\zeta} M}{\sqrt{\zeta} c_{\mathcal{H}, \zeta} C_{\mathcal{H}, \zeta} C_{\mathcal{H}}} \sigma^{-\max\left\{ \frac{3}{\zeta}, \frac{\tau}{\zeta} \right\}} \quad \forall x \in X.
\]

To prove Theorem 3, we need two lemmas. The first lemma bounds the \( C(X) \)-norm of \( f \in \mathcal{H} \) by local \( L^2 \)-norms which can be used to analyze the \( L^2 \)-error by means of excess generalization errors (Theorem 4).

Lemma 1. Let \( x \in X \) and \( \sigma > 0 \). Assume that for some \( \zeta > 0 \) and \( c_{\mathcal{H}} > 0 \), a point set \( \{u_i\}_{i=1}^d \subset B(x, \sigma/2) \) satisfies \( |u_i - u_j| \geq c_{\mathcal{H}} \sigma \) for \( i \neq j \) and (1.5) with \( \sigma \) replaced by \( \sigma/2 \). If
\[
r \leq \frac{c_{\mathcal{H}}}{2^{\zeta+1}} \sigma \zeta \quad \text{and} \quad \{v_i\}_{i=1}^d \text{satisfies} \; v_i \in B(u_i, r) \text{ for each} \; i = 1, \ldots, d,
\]
then
\[
\|\{f(v_i)\}\|_{L^2(\mathbb{R}^d)} \geq \frac{c_{\mathcal{H}} \sigma \zeta}{2^{\zeta+1}} \|f\|_{C(X)} \quad \forall f \in \mathcal{H}.
\]

Proof. Let \( f \in \mathcal{H} \). We know from the Lipschitz condition (1.6) and \( v_i \in B(u_i, r) \) that
\[
|f(u_i) - f(v_i)| \leq C_{\mathcal{H}, 0} |u_i - v_i| \|f\|_{C(X)} \leq C_{\mathcal{H}, 0} r \|f\|_{C(X)}.
\]
Hence
\[
\|\{f(u_i)\} - \{f(v_i)\}\|_{L^2(\mathbb{R}^d)} \leq C_{\mathcal{H}, 0} \sqrt{d} \|f\|_{C(X)}.
\]
It follows from (1.5) with \( \sigma \) replaced by \( \sigma/2 \) and the restriction on \( r \) that
\[
\left( \sum_{i=1}^d |f(v_i)|^2 \right)^{1/2} \geq \|\{f(u_i)\}\|_{L^2(\mathbb{R}^d)} - \|\{f(u_i)\} - \{f(v_i)\}\|_{L^2(\mathbb{R}^d)} \geq c_{\mathcal{H}} \left( \frac{\sigma}{2} \right)^\zeta \|f\|_{C(X)} - C_{\mathcal{H}, 0} \sqrt{d} \|f\|_{C(X)} \geq \frac{c_{\mathcal{H}} \sigma \zeta}{2^{\zeta+1}} \|f\|_{C(X)}.
\]
This proves the desired bound. \( \Box \)

The second lemma we need for proving Theorem 3 is an estimate for the number of sampling points lying in the neighborhood \( B(x, r) \) independent of the center \( x \in X \).

Lemma 2. If \( \rho_X \) satisfies (1.4), then for any \( 0 < r \leq r_0 \) and \( 0 < \delta < 1 \), we have with confidence \( 1 - \delta \),
\[
#(x \cap B(x, r)) / m \geq c_{\tau} r^\tau / 2^\tau - \sqrt{2 \log(1/\delta) + 2n \log(4B_X/r + 1)/\sqrt{m}} \quad \forall x \in X.
\] (3.5)

In particular, if \( m \) satisfies
\[
m \geq \left[ 2^{2\tau+3} \log(1/\delta) + 2^2 r^{\tau+3} n \log(4B_X/r + 1) \right] / (c_{\tau} r^\tau)^2,
\] (3.6)
then with confidence $1 - \delta$,
\[
\frac{\#(x \cap B(x, r))}{m} \geq \frac{c_{\tau}}{2^{\tau+1}}r^{\tau}, \quad \forall x \in X.
\] (3.7)

**Proof.** By a covering number estimate for the ball of $\mathbb{R}^n$ with radius $B_X$ (e.g. Theorem 5.3 in [2]), we know that there exists a subset $\{v_i\}_{i=1}^{\mathcal{N}}$ of $X$ with $\mathcal{N} \leq (4B_X/r + 1)^n$ such that $X \subseteq \bigcup_{i=1}^{\mathcal{N}} B(v_i, \frac{r}{2})$.

Fix $i \in \{1, \ldots, \mathcal{N}\}$. Consider a random variable $\xi_i : X \to \mathbb{R}$ which is the characteristic function of the set $B(v_i, \frac{r}{2})$. Its mean is $\mu_i = \mathbb{E}(\xi_i) = \int_X \xi_i(x) d\rho_X = \rho_X(B(v_i, \frac{r}{2}))$. It also satisfies $|\xi_i - \mu| \leq 1$. Apply the one-side Hoeffding’s inequality for a random variable $\xi$ with mean $\mu$ satisfying $|\xi - \mu| \leq M$:
\[
\text{Prob}_{x \in X} \left\{ \frac{1}{m} \sum_{j=1}^{m} \xi_j(x) - \mu \left( \xi \right) \leq -\varepsilon \right\} \leq \exp \left\{ -\frac{m\varepsilon^2}{2M^2} \right\} \quad \forall \varepsilon > 0.
\] (3.8)

We see that for $\varepsilon > 0$,
\[
\text{Prob}_{x \in X} \left\{ \frac{1}{m} \sum_{j=1}^{m} \xi_j(x) - \mu_i \leq -\varepsilon \right\} \leq \exp \left\{ -\frac{m\varepsilon^2}{2} \right\}.
\]

Taking the union of the above $\mathcal{N}$ events, we know that
\[
\text{Prob}_{x \in X} \left\{ \min_{1 \leq i \leq \mathcal{N}} \left\{ \frac{1}{m} \sum_{j=1}^{m} \xi_j(x) - \mu_i \right\} \leq -\varepsilon \right\} \leq \mathcal{N} \exp \left\{ -\frac{m\varepsilon^2}{2} \right\} =: \delta.
\]

Thus, for $\delta \in (0, 1)$, if we choose
\[
\varepsilon_{m, \delta} := \sqrt{-2\log(\delta/\mathcal{N})/m},
\]
we know that with confidence at least $1 - \delta$, there holds
\[
\min_{1 \leq i \leq \mathcal{N}} \left\{ \frac{1}{m} \sum_{j=1}^{m} \xi_j(x) - \mu_i \right\} > -\varepsilon_{m, \delta}.
\]

That is
\[
\frac{1}{m} \sum_{j=1}^{m} \xi_j(x) - \mu_i > -\varepsilon_{m, \delta} \quad \forall i = 1, \ldots, \mathcal{N}.
\]

Condition (1.4) yields $\mu_i = \rho_X(B(v_i, \frac{r}{2})) \geq c_\tau \left( \frac{r}{2} \right)^\tau$. Also $\xi_i(x_j) = 1$ if $x_j \in B(v_i, \frac{r}{2})$ and 0 otherwise. It follows that $\frac{1}{m} \sum_{j=1}^{m} \xi_i(x_j) = \#(x \cap B(v_i, \frac{r}{2})) / m$ is the proportion of those sampling points lying in $B(v_i, \frac{r}{2})$. Therefore we have
\[
\frac{\#(x \cap B(v_i, \frac{r}{2}))}{m} > c_\tau r^{\tau} / 2^{\tau} - \varepsilon_{m, \delta} \quad \forall i = 1, \ldots, \mathcal{N}.
\]

Observe from $X \subseteq \bigcup_{i=1}^{\mathcal{N}} B(v_i, \frac{r}{2})$ that for each $x \in X$, we can find some $i \in \{1, \ldots, \mathcal{N}\}$ such that $x \in B(v_i, \frac{r}{2})$, i.e., $|x_i - x| \leq \frac{r}{2}$. Since $x_j \in B(v_i, \frac{r}{2})$ implies $|x_j - x| \leq |x_j - v_i| + |v_i - x| \leq r$, we know that
\[
\frac{\#(x \cap B(x, r))}{m} \geq \frac{\#(x \cap B(v_i, \frac{r}{2}))}{m} \geq c_\tau r^{\tau} / 2^{\tau} - \varepsilon_{m, \delta}.
\]
Bounding \( N \) by \((4B\|x\| + 1)^n\), we see that
\[
e_{m,\delta} = \sqrt{-2 \log(\delta/N)/m} \leq \sqrt{2\log(1/\delta) + 2n\log(4B\|x\| + 1)/\sqrt{m}}.
\]
Then (3.5) holds true. This proves the lemma. \( \square \)

Now we can prove Theorem 3.

**Proof of Theorem 3.** First by taking \( f = 0 \) in (1.3), we see from \( f_{x,\sigma,x}(x_i) = f_{x,\sigma,x}(x_i) - y_i + y_i \) that for each \( x \in X \),
\[
\frac{1}{m} \sum_{i=1}^{m} \phi \left( \frac{x}{\sigma}, \frac{x_i}{\sigma} \right) (f_{x,\sigma,x}(x_i))^2 \leq \frac{2}{m} \sum_{i=1}^{m} \phi \left( \frac{x}{\sigma}, \frac{x_i}{\sigma} \right) \left( (0 - y_i)^2 + y_i^2 \right) \leq 4M^2.
\]

Next we take \( r = C_{\mathcal{H},\xi} \sigma^{\max(\xi,1)} \). Then \( r \leq r_0 \) and (3.6) follows from (3.3). By Lemma 2, (3.7) is valid.

Let \( x \in X \). By the norming condition for \( \mathcal{H} \), we can find \( \{u_i\}_{i=1}^{d} \subset B(x, \frac{\sigma}{2}) \) such that \( |u_i - u_j| \geq c_{\mathcal{H}} \sigma \) for \( i \neq j \) and (1.5) holds true with \( \sigma \) replaced by \( \sigma/2 \).

Now we apply (3.7) to the point \( u_i \) with \( i \in \{1, \ldots, d\} \). It follows that for each \( i \),
\[
\#(x \cap B(u_i, r))/m \geq c_r r^\tau/2^{\tau+1}.
\]
Denote the points in the set \( x \cap B(u_i, r) \) as \( \{x_{i,l}\}_{l=1}^{m_i} \). We know that \( m_i/m \geq c_r r^\tau/2^{\tau+1} \) and \( |x_{i,l} - u_i| \leq r \). Hence \( |x_{i,l} - x_{j,k}| \geq |u_i - u_j| - 2r \geq c_{\mathcal{H}} \sigma - 2r > 0 \) and \( x_{i,l} \neq x_{j,k} \) for \( i \neq j \).

Denote \( \hat{m} = \min_{1 \leq l \leq d} \{m_i\} \). Then \( \hat{m} \geq c_r r^\tau m^{\tau/2^{\tau+1}} \). For \( i = 1, \ldots, d \) and \( l = 1, \ldots, \hat{m} \), we have
\[
|x_{i,l} - x| \leq |x_{i,l} - u_i| + |u_i - x| \leq r + \sigma/2 \leq \sigma
\]
which implies \( \phi \left( \frac{x}{\sigma}, \frac{x_{i,l}}{\sigma} \right) \geq c_q \) by (1.2). It follows that
\[
\frac{1}{m} \sum_{j=1}^{m} \phi \left( \frac{x}{\sigma}, \frac{x_j}{\sigma} \right) (f_{x,\sigma,x}(x_j))^2 \geq \frac{1}{m} \sum_{i=1}^{d} \sum_{l=1}^{\hat{m}} \phi \left( \frac{x}{\sigma}, \frac{x_{i,l}}{\sigma} \right) (f_{x,\sigma,x}(x_{i,l}))^2 \\
\geq \frac{1}{m} \sum_{i=1}^{d} \sum_{l=1}^{\hat{m}} c_q (f_{x,\sigma,x}(x_{i,l}))^2.
\]

Note that \( x_{i,l} \in B(u_i, r) \). By Lemma 1, we know that the function \( f_{x,\sigma,x} \in \mathcal{H} \) satisfies
\[
\sum_{i=1}^{d} |f_{x,\sigma,x}(x_{i,l})|^2 \geq (c_{\mathcal{H}} \sigma^{\xi}/2^{\xi+1})^2 \|f_{x,\sigma,x}\|_{C(X)}^2
\]
for each \( l = 1, \ldots, \hat{m} \). Hence
\[
\frac{1}{m} \sum_{i=1}^{d} \sum_{l=1}^{\hat{m}} c_q (f_{x,\sigma,x}(x_{i,l}))^2 \geq \frac{\hat{m}}{m} c_q \left( \frac{c_{\mathcal{H}} \sigma^{\xi}}{2^{\xi+1}} \right)^2 \|f_{x,\sigma,x}\|_{C(X)}^2
\]
\[
\geq \frac{c_r r^\tau c_q}{2^{\tau+1}} \left( \frac{c_{\mathcal{H}} \sigma^{\xi}}{2^{\xi+1}} \right)^2 \|f_{x,\sigma,x}\|_{C(X)}^2.
\]

Combining the above two parts, we have
\[
\frac{c_r r^\tau c_q}{2^{\tau+1}} \left( \frac{c_{\mathcal{H}} \sigma^{\xi}}{2^{\xi+1}} \right)^2 \|f_{x,\sigma,x}\|_{C(X)}^2 \leq \frac{1}{m} \sum_{j=1}^{m} \phi \left( \frac{x}{\sigma}, \frac{x_j}{\sigma} \right) (f_{x,\sigma,x}(x_j))^2 \leq 4M^2.
\]

This yields the desired bound and completes the proof of Theorem 3. \( \square \)
4. Bounding the MLS sample error

We first overcome the technical difficulty of MLS caused by its pointwise definition.

**Lemma 3.** Assume that \( \rho_X \) satisfies condition \( L_\tau \) with exponent \( \tau > 0 \) and \( \mathcal{H} \) satisfies the norming condition with exponents \( \xi > 0 \) and \( d \in \mathbb{N} \). Then we have

\[
\int_{B(x, \sigma)} (f(u))^2 \, d\rho_X \geq \tilde{C}_H \sigma^{2\xi + \tau \max(\xi, 1)} \|f\|_{L_\xi(x)}^2,
\]

\[\forall x \in X, \ 0 < \sigma \leq \min\{\sigma_0, 1\}, \ f \in \mathcal{H}, \quad (4.1)\]

where

\[
\tilde{C}_H = c_\tau \left( \frac{C_H}{2\xi + 1} \right)^2 C_{H, \xi}^\tau.
\]

**Proof.** Let \( x \in X \) and \( 0 < \sigma \leq \min\{\sigma_0, 1\} \). By the norming condition for \( \mathcal{H} \), we can find \( \{u_i\}_{i=1}^d \subset B(x, \frac{\sigma}{2}) \) such that \( |u_i - u_j| > c_H \sigma \) for \( i \neq j \) and (1.5) valid with \( \sigma \) replaced by \( \sigma/2 \). Take \( r = C_{H, \xi} \sigma^{\max(\xi, 1)} \). From Lemma 1, inequality (3.4) holds for every \( v_i \in B(u_i, r) \) and \( i = 1, \ldots, d \). Hence by (4.4),

\[
\sum_{i=1}^d \int_{B(u_i, r)} |f(v_i)|^2 \, d\rho_X(v_i)
\]

\[
= \sum_{i=1}^d \int_{B(u_1, r)} \cdots \int_{B(u_d, r)} |f(v_i)|^2 \, d\rho_X(v_1) \cdots \, d\rho_X(v_d) \rho_X(B(u_i, r))
\]

\[
\geq \frac{\int_{B(u_1, r)} \cdots \int_{B(u_d, r)} |f(v_i)|^2 \, d\rho_X(v_1) \cdots \, d\rho_X(v_d) c_\tau r^\tau}{\prod_{j=1}^d \rho_X(B(u_j, r))}
\]

\[
= \left( \frac{c_H \sigma \xi}{2\xi + 1} \right)^2 \|f\|_{L_\xi(X)}^2 \frac{c_\tau r^\tau}{\prod_{j=1}^d \rho_X(B(u_j, r))}
\]

Since \( |u_i - u_j| > c_H \sigma \geq 2r \) for \( i \neq j \), we know that \( B(u_i, r) \cap B(u_j, r) = \emptyset \) and \( B(u_i, r) \subset B(x, \sigma) \). So

\[
\int_{B(x, \sigma)} |f(v)|^2 \, d\rho_X(v) \geq \sum_{i=1}^d \int_{B(u_i, r)} |f(v_i)|^2 \, d\rho_X(v_i) \geq c_\tau \left( \frac{c_H}{2\xi + 1} \right)^2 \sigma^{2\xi + \tau \max(\xi, 1)} \|f\|_{L_\xi(X)}^2.
\]

Thus we obtain the desired result. \( \square \)

**Theorem 4.** If \( \rho_X \) satisfies condition \( L_\tau \) with exponent \( \tau > 0 \) and \( \mathcal{H} \) satisfies the norming condition with exponents \( \xi > 0 \) and \( d \in \mathbb{N} \), then

\[
\|f_x - f_H\|_{L_\rho_X^2}^2 \leq \frac{\sigma^{-2\xi - \tau \max(\xi, 1)}}{c_q \tilde{C}_H} \int_X \mathcal{E}_x(f_{x, \sigma}, x) - \mathcal{E}_x(f_{H, \sigma}, x) \, d\rho_X(x).
\]

(4.3)
Proof. Let \( x \in X \). By (2.5) and (1.2) for \( f = f_{z, \sigma, x} \) we know

\[
\mathcal{E}_x(f_{z, \sigma, x}) - \mathcal{E}_x(f_{H, \sigma, x}) = \int_X \phi \left( \frac{x}{\sigma}, \frac{u}{\sigma} \right) \left( f_{z, \sigma, x}(u) - f_{H, \sigma, x}(u) \right)^2 d\rho_X(u)
\geq \int_{B(x, \sigma)} c_q \left( f_{z, \sigma, x}(u) - f_{H, \sigma, x}(u) \right)^2 d\rho_X(u).
\]

Now we apply Lemma 3 for the function \( f_{z, \sigma, x} - f_{H, \sigma, x} \in H \) and find

\[
\mathcal{E}_x(f_{z, \sigma, x}) - \mathcal{E}_x(f_{H, \sigma, x}) \geq c_q \tilde{C}_H \sigma^{-2} |\| f_{z, \sigma, x} - f_{H, \sigma, x} |\| _{C(X)}^2.
\]

Hence

\[
| f_{z, \sigma, x}(x) - f_{H, \sigma, x}(x) |^2 \leq \sigma^{-2} \max \{ \| f_{z, \sigma, x} - f_{H, \sigma, x} \| _{C(X)} \} / (c_q \tilde{C}_H).
\]

Finally we recall the pointwise definition of the functions \( f_z \) and \( f_H \) and see that

\[
\| f_z - f_H \| _{L^2_X}^2 = \int_X (f_z(x) - f_{H, \sigma, x}(x))^2 d\rho_X(x).
\]

Therefore

\[
\| f_z - f_H \| _{L^2_X}^2 \leq \int_X \sigma^{-2} \max \{ \| f_{z, \sigma, x} - f_{H, \sigma, x} \| _{C(X)} \} d\rho_X(x) / (c_q \tilde{C}_H).
\]

This proves Theorem 4. \( \square \)

According to Theorem 4, the sample error \( \| f_z - f_H \| _{L^2_X}^2 \) can be bounded by estimating \( \int_X \mathcal{E}_x(f_{z, \sigma, x}) - \mathcal{E}_x(f_{H, \sigma, x}) \) \( d\rho_X(x) \). This can be regarded as an integral form of the excess generalization error \( \mathcal{E}(f_z) - \mathcal{E}(f_H) \) in the literature. In this section, we estimate this quantity. For each \( x \in X \), the quantity \( \mathcal{E}_x(f_{z, \sigma, x}) - \mathcal{E}_x(f_{H, \sigma, x}) \) equals

\[
\mathcal{E}_x(f_{z, \sigma, x}) - \mathcal{E}_{z, x}(f_{z, \sigma, x}) + \mathcal{E}_{z, x}(f_{z, \sigma, x}) - \mathcal{E}_{z, x}(f_{H, \sigma, x}) + \mathcal{E}_{z, x}(f_{H, \sigma, x}) - \mathcal{E}_x(f_{H, \sigma, x}).
\]

Since \( f_{z, \sigma, x} \) minimizes \( \mathcal{E}_{z, x} \) in \( H \), \( \mathcal{E}_{z, x}(f_{z, \sigma, x}) - \mathcal{E}_{z, x}(f_{H, \sigma, x}) \leq 0 \). Thus when the sample \( z \) satisfies \( \sup_{x \in X} \| f_{z, \sigma, x} \| _{C(X)} \leq R \) and \( \sup_{x \in X} \| f_{H, \sigma, x} \| _{C(X)} \leq R \) for some \( R > 0 \), we have

\[
\int_X \mathcal{E}_x(f_{z, \sigma, x}) - \mathcal{E}_x(f_{H, \sigma, x}) \ d\rho_X(x) \leq 2 \sup_{\| f \| _{C(X)} \leq R} \left| \int_X \mathcal{E}_x(f) - \mathcal{E}_{z, x}(f) \ d\rho_X(x) \right| . \quad (4.4)
\]

We can use a covering number argument to estimate this quantity. To this end, besides bounds for \( \| f_{z, \sigma, x} \| _{C(X)} \) in Theorem 3, we need to bound \( \| f_{H, \sigma, x} \| _{C(X)} \).

Lemma 4. If (4.1) holds true, then

\[
\| f_{H, \sigma, x} \| _{C(X)} \leq M \sigma^{-2} \max \{ \| f \| _{C(X)} \} \sqrt{C_{\tilde{H} q}} \quad \forall x \in X.
\]

Proof. We know from (2.5) with \( f = 0 \) that for each \( x \in X \),

\[
\int_X \phi \left( \frac{x}{\sigma}, \frac{u}{\sigma} \right) \left( f_{H, \sigma, x}(u) \right)^2 d\rho_X(u) = \mathcal{E}_x(0) - \mathcal{E}_x(f_{H, \sigma, x}) \leq \mathcal{E}_x(0)
= \int_Z \phi \left( \frac{x}{\sigma}, \frac{u}{\sigma} \right) (0 - y)^2 d\rho(u, y) \leq M^2,
\]
where the last inequality follows from (1.1) and the assumption $|y| \leq M$. By condition (1.2) for $\Phi$, we have
\[
\int_X \Phi \left( \frac{x}{\sigma}, \frac{u}{\sigma} \right) (f_{\mathcal{H}, \sigma, x}(u))^2 \, d\rho_X(u) \geq \int_{B(x, \sigma)} \Phi \left( \frac{x}{\sigma}, \frac{u}{\sigma} \right) (f_{\mathcal{H}, \sigma, x}(u))^2 \, d\rho_X(u) \geq c_q \int_{B(x, \sigma)} (f_{\mathcal{H}, \sigma, x}(u))^2 \, d\rho_X(u).
\]
Therefore, combining with (4.1), we have
\[
\tilde{C}_H \sigma^{2\xi + \max(\xi, 1)} \| f_{\mathcal{H}, \sigma, x} \|_{C(X)}^2 \leq \int_{B(x, \sigma)} (f_{\mathcal{H}, \sigma, x}(u))^2 \, d\rho_X(u) \leq M^2/c_q.
\]
Then the desired bound for $\| f_{\mathcal{H}, \sigma, x} \|_{C(X)}$ follows.

Denote $B_R = \{ f \in \mathcal{H} : \| f \|_{C(X)} \leq R \}$ the ball of $\mathcal{H}$ with radius $R > 0$. Since $\mathcal{H}$ is finite dimensional, its ball $B_R$ can be regarded as a compact subset of $C(X)$. Its covering number $\mathcal{N}(B_R, \eta)$ for $\eta > 0$ is defined to be the minimal $l \in \mathbb{N}$ such that there exist $l$ open balls in $B_R$ with radius $\eta$ covering $B_R$. Since $\mathcal{H}$ has finite dimension $\tilde{d}$, the covering number can be bounded as
\[
\mathcal{N}(B_R, \eta) \leq \left( \frac{2R}{\eta} + 1 \right)^{\tilde{d}} \quad \forall \eta > 0. \tag{4.5}
\]

The following elementary lemma can be essentially found in [2, p. 179]. For completeness we give a proof.

**Lemma 5.** Let $p, C_0, \Delta_1, \Delta_2$ be positive numbers and $m \in \mathbb{N}$, $0 < \delta < 1$. If $m \geq \max\{3, \Delta_1\}$, then the smallest positive number $\varepsilon^*$ satisfying
\[
C_0 \left( \log \frac{\Delta_1}{\varepsilon} \right)^p - \frac{m\varepsilon}{\Delta_2} \leq \log \delta
\]
is bounded as
\[
\varepsilon^* \leq \max \left\{ \Delta_2 \left( 2^p C_0 + \log(1/\delta) \right), 1 \right\} \frac{(\log m)^p}{m}.
\]

**Proof.** The function $h : \mathbb{R}_+ \rightarrow \mathbb{R}$ defined by
\[
h(\varepsilon) = C_0 \left( \log \frac{\Delta_1}{\varepsilon} \right)^p - \frac{m\varepsilon}{\Delta_2}
\]
is strictly decreasing and $\lim_{\varepsilon \to \infty} h(\varepsilon) = -\infty$, $\lim_{\varepsilon \to 0_+} h(\varepsilon) = +\infty$. Hence $\varepsilon^*$ exists.

Let $\varepsilon_m = \max\{\Delta_2 (2^p C_0 + \log(1/\delta)), 1\} \frac{(\log m)^p}{m}$. If $m \geq \max\{3, \Delta_1\}$, we see that $\varepsilon_m \geq \frac{(\log m)^p}{m} \geq \frac{1}{m}$ and $\log \Delta_1 + \log m \leq 2 \log m$. It follows that
\[
h(\varepsilon_m) \leq C_0 \left( \log \Delta_1 + \log(1/\varepsilon_m) \right)^p - (\log m)^p \left( 2^p C_0 + \log(1/\delta) \right) \leq C_0 \left( \log \Delta_1 + \log m \right)^p - (\log m)^p \left( 2^p C_0 \right) - (\log m)^p \log(1/\delta) \leq -\log(1/\delta) = \log \delta.
\]

Therefore, $\varepsilon^* \leq \varepsilon_m$ and the desired bound is proved.
Proposition 1. Let \( R > 0 \), \( 0 < \delta < 1 \). If \( m \geq 81 \) and \( 2 \log m/m \leq 1/(\tilde{d} + \log(2/\delta)) \), then with confidence at least \( 1 - \delta \), we have

\[
\sup_{f \in B_R} \left| \int_X \mathcal{E}_x(f) - \mathcal{E}_{x,x}(f) \, d\rho_X(x) \right| \leq 2(R + M)^2 \sqrt{2(\tilde{d} + \log(2/\delta))} \log m / \sqrt{m}.
\]

Proof. Let \( 0 < \varepsilon \leq (R + M)^2 \) and \( l = \mathcal{N}(B_R, \frac{\varepsilon}{4(R + M)}) \) be the covering number of \( B_R \), where the disks \( D_j (1 \leq j \leq l) \) centered at \( f_j \in B_R \) with radius \( \frac{\varepsilon}{4(R + M)} \) cover \( B_R \). For every \( f \in D_j \),

\[
\left| \int_X (\mathcal{E}_{z,x}(f) - \mathcal{E}_{x}(f)) - (\mathcal{E}_{z,x}(f_j) - \mathcal{E}_{x}(f_j)) \, d\rho_X(x) \right|
\]

\[
= \int_X \left\{ \frac{1}{m} \sum_{i=1}^m \Phi \left( \frac{X_i}{\sigma}, \frac{u_i}{\sigma} \right) (f(x_i) - f_j(x_i))(f(x_i) + f_j(x_i) - 2y_i) \right. \\
- \left. \int_Z \Phi \left( \frac{X}{\sigma}, \frac{u}{\sigma} \right) (f(u) - f_j(u))(f(u) + f_j(u) - 2y)d\rho(u, y) \right\} \, d\rho_X(x).
\]

By (1.1) and the assumption \(|y| \leq M\), this can be bounded by

\[
2(\|f\|_{C(X)} + \|f_j\|_{C(X)} + 2M)\|f - f_j\|_{C(X)} \leq 4(R + M) \frac{\varepsilon}{4(R + M)} = \varepsilon.
\]

For each \( j \in \{1, \ldots, l\} \), the random variable \( \xi = \int_X \Phi \left( \frac{X}{\sigma}, \frac{u}{\sigma} \right) (f(u) - f_j(u)) \, d\rho_X(x) \) with \((u, y) \in (Z, \rho)\) can be bounded as \(|\xi - E\xi| \leq (R + M)^2\). By the Hoeffding’s inequality we deduce that for \( j = 1, \ldots, l\),

\[
\text{Prob} \left\{ \sup_{x \in Z^m} \left| \int_X \mathcal{E}_{z,x}(f) - \mathcal{E}_{x}(f) \, d\rho_X(x) \right| \geq 2\varepsilon \right\} \leq \text{Prob} \left\{ \left| \int_X \mathcal{E}_{z,x}(f_j) - \mathcal{E}_{x}(f_j) \, d\rho_X(x) \right| \geq \varepsilon \right\} \leq 2 \exp \left\{ -\frac{m\varepsilon^2}{2(R + M)^4} \right\}.
\]

Since \( \bigcup_{j=1}^l D_j = B_R \), we see from (4.5) and the restriction \( \varepsilon \leq (R + M)^2 \) that

\[
\text{Prob} \left\{ \sup_{x \in Z^m} \left| \int_X \mathcal{E}_{z,x}(f) - \mathcal{E}_{x}(f) \, d\rho_X(x) \right| \leq 2\varepsilon \right\} \geq 1 - 2l \exp \left\{ -\frac{m\varepsilon^2}{2(R + M)^4} \right\} \geq 1 - 2 \left(8R(R + M)/\varepsilon + 1\right)^d \exp \left\{ -\frac{m\varepsilon^2}{2(R + M)^4} \right\} \geq 1 - 2 \exp \left\{ \frac{d}{2} \log \frac{81(R + M)^4}{\varepsilon^2} - \frac{m\varepsilon^2}{2(R + M)^4} \right\}.
\]

Now we need to find \( \varepsilon \). Choose \( \varepsilon^*_R \) to be the smallest positive number \( \varepsilon \) satisfying

\[
\frac{d}{2} \log \frac{81(R + M)^4}{\varepsilon^2} - \frac{m\varepsilon^2}{2(R + M)^4} \leq \log \frac{\delta}{2}.
\]
By Lemma 5 with $C_0 = \tilde{d}/2$, $\Delta_1 = 81$, $p = 1$, $\Delta_2 = 2$ and $\delta$ replaced by $\delta/2$, we know that when $m \geq 81$,
\[
\left(\frac{\epsilon_R^*}{(R + M)^2}\right)^2 \leq 2(\tilde{d} + \log(2/\delta)) \log m/m.
\]
When $2(\tilde{d} + \log(2/\delta)) \log m/m \leq 1$, we also have $\epsilon_R^* \leq (R + M)^2$. Therefore, with confidence $1 - \delta$, we have
\[
\sup_{f \in B_R} \left| \int_X \mathcal{E}_{Z,x}(f) - \mathcal{E}_x(f) \, d\rho_X(x) \right| \leq 2\epsilon_R^* \leq 2(R + M)^2 \sqrt{2(\tilde{d} + \log(2/\delta)) \log m} / \sqrt{m}.
\]
This proves the desired bound. □

Now we can bound the sample error $\|f_x - f_{\mathcal{H}}\|_{L^2_{\rho_X}}^2$ by restriction
\[
m \geq \max \left\{81, \sigma_0^{-1/\gamma}, (C_{\mathcal{H}, \xi} / r_0)^{1/\gamma \max[\xi, 1]} \right\}
\]
with $\gamma > 0$ to be specified.

**Proposition 2.** Assume that $\mathcal{H}$ satisfies the norming condition with exponents $\xi > 0$ and $d \in \mathbb{N}$, and $\rho_X$ satisfies the condition $L_\tau$ with exponent $\tau > 0$. Let $\gamma > 0$ and $0 < \delta < 1$. If $m$ satisfies (4.7) and
\[
m^{1 - 2\gamma \tau \max[\xi, 1]} \geq A_{\tau, \xi} (\log(2/\delta) + \gamma \log m) \quad \text{and} \quad 2(\tilde{d} + \log(4/\delta)) \log m \leq m
\]
then with confidence $1 - \delta$, we have
\[
\|f_x - f_{\mathcal{H}}\|_{L^2_{\rho_X}}^2 \leq \left[ 16 \sqrt{2(\tilde{d} + \log(4/\delta)) / (c_q \tilde{c}_{\mathcal{H}})} \right] m^{4\gamma(\xi + \max\left\{\frac{\xi}{2}, \frac{\xi}{\tau} \right\})} \frac{1}{\sqrt{\log m}}.
\]

**Proof.** Take $\sigma = m^{-\gamma}$ and
\[
R = \max \left\{2^{3 + \tau + \xi} M / \left[ \sqrt{c_{\tau, q} c_{\mathcal{H}, \xi}^{1/2} c_{\mathcal{H}}} \right], M \left[ \sqrt{\tilde{c}_{\mathcal{H}} c_{q, \xi}} / M \right] m^{\gamma(\xi + \max\left\{\frac{\xi}{2}, \frac{\xi}{\tau} \right\})} \right\}.
\]
First, (4.7) and the first restriction in (4.8) tell us that $\sigma \leq \min\{\sigma_0, 1, (r_0 / C_{\mathcal{H}, \xi})^{1/\max\{\xi, 1\}}\}$ and (3.3) with $\sigma$ replaced by $\sigma/2$ is valid. So we apply Theorem 3 and know that there exists a subset $Z_1$ of $Z^m$ with measure at least $1 - \delta/2$ such that
\[
\|f_{x, \sigma, x}\|_{C(X)} \leq 2^{3 + \tau + \xi} M m^{\gamma(\xi + \max\left\{\frac{\xi}{2}, \frac{\xi}{\tau} \right\})} \left[ \sqrt{c_{\tau, q} c_{\mathcal{H}, \xi}^{1/2} c_{\mathcal{H}}} \right] \leq R, \quad \forall x \in X, \ z \in Z_1.
\]
That is, $f_{x, \sigma, x} \in B_R$ for any $x \in X, z \in Z_1$.

Next we apply Lemmas 3 and 4 and find that $f_{\mathcal{H}, \sigma, x} \in B_R$ for any $x \in X$.

Then we apply Proposition 1. Since $\tilde{d} \leq d$, the second restriction of (4.8) ensures the condition $2(\tilde{d} + \log(4/\delta)) \log m \leq m$. So Proposition 1 with $\delta$ replaced by $\delta/2$ tells us that there exists another subset $Z_2$ of $Z^m$ with measure at least $1 - \delta/2$ such that
\[
\sup_{f \in B_R} \left| \int_X \mathcal{E}_x(f) - \mathcal{E}_{Z,x}(f) \, d\rho_X(x) \right| \leq 2(R + M)^2 \sqrt{2(\tilde{d} + \log(4/\delta)) \log m / \sqrt{m}}
\]
\[
\forall z \in Z_2.
\]
Finally we apply Theorem 4 and (4.4) and conclude that for $z \in Z_1 \cap Z_2$, there holds
\[
\|f_z - f_{\mathcal{H}}\|_{L^2_{\rho_X}}^2 \leq \left[16\sqrt{2} (d + \log(4/\delta)) \log m / (c_q \widetilde{C}_{\mathcal{H}})\right] \frac{m^{4\gamma \left(1 + \frac{\zeta}{\xi} \right)}}{\left(\zeta + \max \left\{\frac{\tau_2}{\xi}, \frac{\tau_\zeta}{\xi}\right\}\right)}^{1/2}.
\]
Since the measure of $Z_1 \cap Z_2$ is at least $1 - \delta$, our conclusion follows. \( \square \)

5. Bounding total error and discussion

Proposition 2 provides a general bound for the sample error. It leads to the following convergence rates. Recall Definitions 2 and 3 involving constants $\tau$, $r_0$, $c$, $\tau_\zeta$ and $c_{\mathcal{H}}$. Recall also the constants $C_{\mathcal{H},\xi}$, $A_{\tau,\zeta}$ and $\widetilde{C}_{\mathcal{H}}$ defined by (3.1), (3.2) and (4.2) respectively.

**Theorem 5.** Assume condition $L_\tau$ with exponent $\tau > 0$ for $\rho_X$ and norming condition with exponents $\zeta > 0$ and $d \in \mathbb{N}$ for $\mathcal{H}$. Let $0 < \varepsilon < 1/4$. Then for $0 < \delta < 1$ and $m$ satisfying (1.7), we have with confidence $1 - \delta$,
\[
\|f_z - f_{\mathcal{H}}\|_{L^2_{\rho_X}} \leq C_2^* (\log(4/\delta) \log m)^{1/2} m^{\varepsilon - 1/2}.
\]

Here the constants $C_1^*$ appearing in the restriction (1.7) and $C_2^*$ are given in terms of constants $C_{\mathcal{H},\xi}$, $A_{\tau,\zeta}$ and $\widetilde{C}_{\mathcal{H}}$ as
\[
C_1^* = \max \left\{81, (C_{\mathcal{H},\xi}/r_0)^2, A_{\tau,\zeta} (1 + 1/\zeta)\right\}, \quad C_2^* = 4 (2(d + 1))^{1/4} / \sqrt{c_q \widetilde{C}_{\mathcal{H}}}.
\]

**Proof.** Take $\gamma = \varepsilon / [2\zeta + \max \{\tau, \tau_\zeta\}] > 0$ in Proposition 2. Then $2\gamma\tau \max \{\zeta, 1\} < 2\varepsilon < 1/2$ and
\[
m^{1 - 2\gamma\tau \max \{\zeta, 1\}} \geq m^{1 - 2\varepsilon}.
\]
Also, $\gamma \max \{\zeta, 1\} < \frac{\varepsilon}{\tau} < \frac{1}{2\tau}$. So (4.7) is satisfied when
\[
m \geq \max \left\{81, \sigma_0^{-2\zeta + \max \{\tau, \tau_\zeta\}} / (C_{\mathcal{H},\xi}/r_0)^2, (C_{\mathcal{H},\xi}/r_0)^2\right\}.
\]

When $m \geq 4(d + \log(4/\delta))^2$ and
\[
m^{1/2 - \varepsilon} \geq A_{\tau,\zeta} (1 + 1/\zeta) \log m,
\]
we see that
\[
m^{1 - 2\varepsilon} \geq \log(2/\delta) A_{\tau,\zeta} (1 + \gamma) \log m \geq A_{\tau,\zeta} (\log(2/\delta) + \gamma \log m)
\]
Moreover,
\[
2 (d + \log(4/\delta)) \log m \leq \log m \sqrt{m} m \leq m.
\]
Thus both (4.7) and (4.8) are valid when (1.7) is satisfied. So we apply Proposition 2 and conclude that (5.1) holds with confidence $1 - \delta$. This proves Theorem 5. \( \square \)

Theorem 1 follows immediately from Theorem 5 because the assumption $f_\rho \in \mathcal{H}$ yields $f_\rho = f_{\mathcal{H}}$. 

Proof of Theorem 2. Since $\mathcal{H}$ is 1-dimensional with a basis function $\varphi$, we see from the definition of $f_{\mathcal{H}, \sigma, x}$ that
\[
 f_{\mathcal{H}, \sigma, x} = \int_{X} \Phi \left( \frac{x - u}{\sigma} \right) f_{\rho}(u) \varphi(u) \, d\rho_{X}(u) / \int_{X} \Phi \left( \frac{x - u}{\sigma} \right) (\varphi(u))^2 \, d\rho_{X}(u). 
\]
It follows that
\[
 f_{\mathcal{H}}(x) - f_{\rho}(x) = f_{\mathcal{H}, \sigma, x}(x) - f_{\rho}(x)
 = \int_{X} \Phi \left( \frac{x - u}{\sigma} \right) [f_{\rho}(u)\varphi(x) - f_{\rho}(x)\varphi(u)] \varphi(u) \, d\rho_{X}(u) / \int_{X} \Phi \left( \frac{x - u}{\sigma} \right) (\varphi(u))^2 \, d\rho_{X}(u). 
\]
Since $\varphi$ and $f_{\rho}$ are Lipschitz, we know that
\[
 |f_{\rho}(u)\varphi(x) - f_{\rho}(x)\varphi(u)| \leq \|\varphi\|_{Lip} \|f_{\rho}\|_{Lip} |x - u|. 
\]
Hence
\[
 |f_{\mathcal{H}}(x) - f_{\rho}(x)| \leq \int_{X} \Phi \left( \frac{x - u}{\sigma} \right) |x - u| \varphi(u) \, d\rho_{X}(u) / \int_{X} \Phi \left( \frac{x - u}{\sigma} \right) (\varphi(u))^2 \, d\rho_{X}(u) \|\varphi\|_{Lip} \|f_{\rho}\|_{Lip}
 \leq \Lambda \|\varphi\|_{Lip} \|f_{\rho}\|_{Lip}. 
\]
This bound holds true for every $x \in X$. Then the conclusion of Theorem 2 follows from Theorem 5. □

We have dealt with the approximation error only in two special cases: $f_{\rho} \in \mathcal{H}$ and 1-dimensional hypothesis space. The latter includes the example of $\mathcal{H}$ being the space of constant functions. Our Theorem 5 verifies the convergence of the MLS scheme with this very special but important hypothesis space. In particular, when $X$ satisfies a cone condition [1] and $\rho_{X}$ is the uniform distribution, explicit convergence rates are derived in Theorem 5. Estimating the approximation error for the MLS scheme is an interesting topic that deserves further study.

Acknowledgments

The work described in this paper is supported partially by the Research Grants Council of Hong Kong [Project No. CityU 104007]. The authors would like to thank the referees for their careful and constructive suggestions.

References


