

RESEARCH

Open Access



Error analysis for l^q -coefficient regularized moving least-square regression

Qin Guo^{1*} and Peixin Ye¹

*Correspondence:
guoqin_1985@163.com

¹School of Mathematical Sciences
and LPMC, Nankai University,
Tianjin, China

Abstract

We consider the moving least-square (MLS) method by the coefficient-based regression framework with l^q -regularizer ($1 \leq q \leq 2$) and the sample dependent hypothesis spaces. The data dependent characteristic of the new algorithm provides flexibility and adaptivity for MLS. We carry out a rigorous error analysis by using the stepping stone technique in the error decomposition. The concentration technique with the l^2 -empirical covering number is also employed in our study to improve the sample error. We derive the satisfactory learning rate that can be arbitrarily close to the best rate $O(m^{-1})$ under more natural and much simpler conditions.

MSC: 68T05; 68Q32; 60E15

Keywords: Moving least-square method; Data dependent hypothesis space; Regularization function; Uniform concentration inequality; Learning rate

1 Introduction

The least-square (LS) method is an important global approximate method based on the regular or concentrated data sample points. However, there are still some irregular or scattered samples which are obtained in many practical applications such as engineering and machine learning [1–4]. They also need to be analyzed to achieve their special usefulness. For example, in geographical contour drawing, it is important to derive a set of contours but the height is available only for some scattered data sample points. Therefore, it is vital to seek a suitable local approximation method to deal with scattered data. The moving least-square (MLS) method was introduced by McLain in [4] to draw a set of contours based on a cluster of scattered data sample points. The central idea of the MLS method consists of two steps: first, one takes an arbitrary fixed point and forms a local approximation formula; second, since the fixed point is arbitrary, therefore, one can let it move over the whole domain. It turns out that MLS method is a useful local approximation tool in various mathematics fields such as approximation theory, data smoothing [5], statistics [6] and numerical analysis [7]. In computer graphics, the MLS method is useful for reconstructing a surface from a set of points. Often it is used to create a 3D surface from a point cloud. Recently, a research effort has been made to study the regression learning algorithm by the MLS method; see [8–12]. 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.

We recall the regression learning problem by the MLS method briefly. Functions for learning are defined on a compact subset X (input space) of \mathbb{R}^n and take values in $Y = \mathbb{R}$ (output space). The sampling process is controlled by a unknown Borel probability measure on $Z = X \times Y$. The regression function is given by

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

where $\rho(\cdot|x)$ is the conditional probability measure induced by ρ on Y given $x \in X$. The goal of regression learning is to find a good approximation of the regression function f_ρ based on a set of random samples $\mathbf{z} = \{z_i\}_{i=1}^m = \{(x_i, y_i)\}_{i=1}^m \in Z^m$ drawn according to the measure ρ independently and identically.

In [11], Tong and Wu considered the following regularized MLS regression algorithm. The hypothesis space is a reproducing kernel Hilbert space (RKHS) \mathcal{H}_K induced by a Mercer kernel K , which is a continuous, symmetric, and positive semi-definite function on $X \times X$. The RKHS \mathcal{H}_K is the completion of the linear span of the set of functions $\{K_x := K(x, \cdot) : x \in X\}$ with respect to the inner product $\langle \sum_{i=1}^n \alpha_i K_{x_i}, \sum_{j=1}^m \beta_j K_{y_j} \rangle_K := \sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j K(x_i, y_j)$. The reproducing property in \mathcal{H}_K is

$$f(x) = \langle f, K_x \rangle_K, \quad \text{for all } f \in \mathcal{H}_K, x \in X. \tag{1.1}$$

Denote $C(X)$ as the space of continuous functions on X with the norm $\|\cdot\|_\infty$. Since K is continuous in X , $\mathcal{H}_K \subseteq C(X)$. Let $\kappa := \sup_{t,x \in X} |K(x, t)| < \infty$. Then, by (1.1), we have

$$\|f\|_\infty \leq \kappa \|f\|_K, \quad \forall f \in \mathcal{H}_K, \tag{1.2}$$

We define the approximation $f_{\mathbf{z},\lambda}$ of f_ρ pointwise:

$$f_{\mathbf{z},\lambda}(x) = f_{\mathbf{z},\sigma,\lambda,x}(x) = f_{\mathbf{z},\sigma,\lambda,x}(u)|_{u=x}, \tag{1.3}$$

$$f_{\mathbf{z},\sigma,\lambda,x} := \arg \min_{f \in \mathcal{H}_K} \left\{ \frac{1}{m} \sum_{i=1}^m \Phi\left(\frac{x}{\sigma}, \frac{x_i}{\sigma}\right) (y_i - f(x_i))^2 + \lambda \|f\|_K^2 \right\}, \tag{1.4}$$

where $\lambda = \lambda(m) > 0$ is a regularization parameter, $\sigma = \eta(m) > 0$ is a window width, and $\Phi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^+$ is called a MLS weight function which satisfies the conditions as follows:

$$(1) \quad 0 \neq \Phi(x, t) \leq 1, \quad \forall x, t \in \mathbb{R}^n, \tag{1.5}$$

$$(2) \quad \Phi(x, t) \geq c_q, \quad \forall |x - t| \leq 1, \tag{1.6}$$

$$(3) \quad |\Phi(x, t_1) - \Phi(x, t_2)| \leq c_\Phi |t_1 - t_2|^s, \quad \forall x, t_1, t_2 \in \mathbb{R}^n, \tag{1.7}$$

where the constants $q > n + 1, c_q, c_\Phi > 0$.

The scheme (1.3)–(1.4) shows that regularization not only ensures the computational stability but also preserves localization property for the algorithm. In this paper, we study

the new regularized version of the MLS regression algorithm. We adopt the coefficient-based l^q -regularization and the data dependent hypothesis space.

$$f_{z,\eta}(x) = f_{z,\sigma,\eta,x}(x) = f_{z,\sigma,\eta,x}(u)|_{u=x}, \tag{1.8}$$

$$f_{z,\sigma,\eta,x} = \arg \min_{f \in \mathcal{H}_{K,z}} \left\{ \frac{1}{m} \sum_{i=1}^m \Phi \left(\frac{x}{\sigma}, \frac{x_i}{\sigma} \right) (f(x_i) - y_i)^2 + \eta \Omega_z(f) \right\}, \quad \eta = \eta(m) > 0, \tag{1.9}$$

where

$$\mathcal{H}_{K,z} = \left\{ f(x) = \sum_{i=1}^m \alpha_i K(x, x_i) : \alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{R}^m, m \in \mathbb{N} \right\},$$

$$\Omega_z(f) = \inf \left\{ \sum_{i=1}^m |\alpha_i|^q \right\}, \quad 1 \leq q \leq 2.$$

The data dependence nature of the kernel-based hypothesis space provides flexibility for the learning algorithm such as choosing the l^q -norm regularizer of a function expansion involving samples. Compared with the scheme (1.3)–(1.4) in a reproducing kernel Hilbert space, the first advantage of the algorithm (1.9) is the effectivity of computations without any optimization processes. Another advantage is that we can choose the suitable parameter q according to the research interest such as smoothness and sparsity. To study the approximation quality of $f_{z,\eta}$, we derive the upper bound of the error $\|f_{z,\eta} - f_\rho\|_{\rho_X}$ with $\|f(\cdot)\|_{\rho_X} := (\int_X |f(\cdot)|^2 d\rho_X)^{\frac{1}{2}}$ and its convergence rates as $m \rightarrow \infty$; see [8–11, 13, 14]. The remainder of this paper is organized as follows. In Sect. 2, we will provide the main result. The error decomposition analysis and the upper bounds of the hypothesis error, the approximation error and the sample error will be given in Sects. 3. In Sect. 4, we will prove the main result. Finally, Sect. 5 concludes the paper with future research lines.

2 Main result

We firstly formulate some basic notations and assumptions.

Let ρ_X be the marginal distribution of ρ on X and $L^2_{\rho_X}(X)$ be the Hilbert space of functions from X to Y square-integrable with respect to ρ_X with the norm denoted by $\|\cdot\|_{\rho_X}$. The integral operator $L_K : L^2_{\rho_X}(X) \rightarrow L^2_{\rho_X}(X)$ is defined by

$$(L_K f)(x) = \int_X K(x, t) f(t) d\rho_X(t), \quad x \in X.$$

Since X is compact and K is continuous, L_K is a compact operator. Its fractional power operator $L_K^r : L^2_{\rho_X}(X) \rightarrow L^2_{\rho_X}(X), r > 0$ is defined by

$$L_K^r(f) = \sum_{i=1}^{\infty} \mu_i^r \langle f, e_i \rangle_{L^2_{\rho_X}} e_i, \quad f \in L^2_{\rho_X}(X),$$

where $\{\mu_i\}$ are the eigenvalues of the operator L_K and $\{e_i\}$ are the corresponding eigenfunctions which form an orthonormal basis of $L^2_{\rho_X}(X)$; see [15]. For $r > 0$, the function f_ρ is said to satisfy the regularity condition of order r provided that $L_K^{-r} f_\rho \in L^2_{\rho_X}$.

We show the following nice feature for the capacity of $\mathcal{H}_{K,z}$ when the l^2 -empirical covering number is used; see [16],

$$\log \mathcal{N}_2^{\epsilon}(B_1, \epsilon) \leq c_p \epsilon^{-p}, \quad \forall \epsilon > 0, \tag{2.1}$$

where $B_1 = \{f \in \mathcal{H}_{K,z} : \|f\|_K \leq 1\}$, the exponent $0 < p < 2$ and the constant $c_p > 0$.

Definition 2.1 The probability measure ρ_X on X is said to satisfy the condition L_τ with exponent $\tau > 0$ if

$$\rho_X(B(x, r)) \geq c_\tau r^\tau, \quad \forall 0 < r \leq r_0, x \in X, \tag{2.2}$$

where the constants $r_0 > 0, c_\tau > 0$ and $B(x, r) = \{u \in X : |u - x| \leq r\}$ for $r > 0$.

We use the projection operator to obtain the faster learning rate under the condition $|y| \leq M$ and $M \geq 1$ almost surely; see [17–19].

Definition 2.2 Fix $M > 0$, the projection operator π_M on the space of measurable functions $f : X \rightarrow \mathbb{R}$ is defined as

$$\pi_M(f)(x) = \begin{cases} M, & \text{if } f(x) > M, \\ f(x), & \text{if } |f(x)| \leq M, \\ -M, & \text{if } f(x) < -M. \end{cases} \tag{2.3}$$

We assume all the constants are positive and independent of δ, m, λ, η or σ . Now we are in a position to give the learning rates of the algorithm (1.9).

Theorem 2.1 Suppose $L_K^{-r} f_\rho \in L_{\rho_X}^2$ with $r > 0$, (2.1) with $0 < p < 2$ and (2.2) hold. If all the functions $f \in \mathcal{H}_K \cup \{f_\rho\}$ satisfy the Lipschitz condition on X , that is, for the constant $c_0 > 0$,

$$|f(u) - f(v)| \leq c_0 |u - v|, \quad \forall u, v \in X, \tag{2.4}$$

then, for any $0 < \delta < 1$, with confidence $1 - \delta$, we have

$$\|\pi_M(f_{z,\eta}) - f_\rho\|_{\rho_X}^2 \leq \tilde{D} \left(\frac{1}{m}\right)^{\theta(r)} \log\left(\frac{2}{\delta}\right), \tag{2.5}$$

where

$$\theta(r) = \begin{cases} \min\left\{\frac{q}{[r(2p+2q+pq)+pq]}, 1\right\} \left(\frac{2r}{1+\tau}\right), & 0 < r < \frac{1}{2}; \\ \frac{2q}{(2p+2q+3pq)(1+\tau)}, & r \geq \frac{1}{2}. \end{cases}$$

Remark 2.1 When $p \rightarrow 0$ and $r \geq \frac{1}{2}$, our convergence rate $m^{-\frac{2q}{(2p+2q+3pq)(1+\tau)}}$ tends to $m^{-\frac{1}{1+\tau}}$. In [11], the authors have derived the rate $m^{-\frac{1}{1+\tau}}$. In particular, assuming the unnatural norm condition in [8] holds, we can obtain the faster rate $m^{\tau\epsilon - \frac{2q}{2p+2q+3pq}}$ for $r \geq \frac{1}{2}$, which can be arbitrarily close to $O(m^{-1})$ as $\epsilon \rightarrow 0$ and $p \rightarrow 0$.

3 Error analysis

We only present the results of the main propositions in this section. All the proofs will be given in the appendix. To estimate $\|\pi_M(f_{z,\eta}) - f_\rho\|_{\rho_X}^2$, we invoke the following proposition, whose proof is completely similar to that of Theorem 3.3 in [11].

Proposition 3.1 *If ρ_X satisfies (2.1), and all the functions $f \in \mathcal{H}_K \cup \{f_\rho\}$ satisfy (2.4), then*

$$\|\pi_M(f_{z,\eta}) - f_\rho\|_{\rho_X}^2 \leq \frac{\sigma^{-\tau}}{c_q c_\tau} \int_X \{\mathcal{E}_x(\pi_M(f_{z,\sigma,\eta,x})) - \mathcal{E}_x(f_\rho)\} d\rho_X(x) + 8c_0 M \sigma, \tag{3.1}$$

where

$$\mathcal{E}_x(f) = \int_Z \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) (f(u) - y)^2 d\rho(u, y), \quad \forall f : X \rightarrow \mathbb{R} \tag{3.2}$$

is called the local moving expected risk.

Then we only need to provide the upper bound of the integral in (3.1). So to do this, we give its decomposition by using $f_{z,\lambda}$, which plays a stepping stone role between $f_{z,\eta}$ and the regularization function f_λ , while different regularization parameters λ and η are adopted. Here f_λ is given by

$$f_\lambda := \arg \min_{f \in \mathcal{H}_K} \{\|f - f_\rho\|_{\rho_X}^2 + \lambda \|f\|_K^2\}. \tag{3.3}$$

Proposition 3.2 *Let $f_{z,\sigma,\eta,x}$ be defined as in (1.9) and*

$$\mathcal{E}_{z,x}(f) = \frac{1}{m} \sum_{i=1}^m \Phi\left(\frac{x}{\sigma}, \frac{x_i}{\sigma}\right) (f(x_i) - y_i)^2 \tag{3.4}$$

be the local moving empirical risk. Then

$$\int_X \{\mathcal{E}_x(\pi_M(f_{z,\sigma,\eta,x})) - \mathcal{E}_x(f_\rho)\} d\rho_X(x) \leq \mathcal{S}(z, \lambda, \eta) + \mathcal{H}(z, \lambda, \eta) + \mathcal{D}(\lambda), \tag{3.5}$$

where

$$\begin{aligned} \mathcal{S}(z, \lambda, \eta) &= \int_X \{\mathcal{E}_x(\pi_M(f_{z,\sigma,\eta,x})) - \mathcal{E}_{z,x}(\pi_M(f_{z,\sigma,\eta,x})) \\ &\quad + \mathcal{E}_{z,x}(f_\lambda) - \mathcal{E}_x(f_\lambda)\} d\rho_X(x), \\ \mathcal{H}(z, \lambda, \eta) &= \int_X \{(\mathcal{E}_{z,x}(\pi_M(f_{z,\sigma,\eta,x})) + \eta \Omega_z(f_{z,\sigma,\eta,x})) \\ &\quad - (\mathcal{E}_{z,x}(f_\lambda) + \lambda \|f_\lambda\|_K^2)\} d\rho_X(x), \\ \mathcal{D}(\lambda) &= \|f_\lambda - f_\rho\|_{\rho_X}^2 + \lambda \|f_\lambda\|_K^2. \end{aligned}$$

$\mathcal{S}(z, \lambda, \eta)$ is known as the sample error. $\mathcal{H}(z, \lambda, \eta)$ is called the hypothesis error. $\mathcal{D}(\lambda)$ is called the approximation error.

The estimation of the hypothesis error can be conducted analogously to that in [18].

Proposition 3.3 *Under the assumptions of Theorem 2.1, we have*

$$\mathcal{H}(\mathbf{z}, \lambda, \eta) \leq \frac{m\eta M^2}{(m\lambda)^q}.$$

For the approximation error, we directly invoke the following result in [20].

Proposition 3.4 *Under the assumption $L_K^{-r} f_\rho \in L^2_{\rho_X}$ with $r > 0$, we have*

$$\mathcal{D}(\lambda) \leq C_1 \lambda^{\min\{2r, 1\}}. \tag{3.6}$$

For the sample error, we decompose it into two parts:

$$\begin{aligned} \mathcal{S}(\mathbf{z}, \lambda, \eta) &= \int_X \{ \mathcal{E}_x(\pi_M(f_{\mathbf{z}, \sigma, \eta, x})) - \mathcal{E}_x(f_\rho) \\ &\quad - \mathcal{E}_{\mathbf{z}, x}(\pi_M(f_{\mathbf{z}, \sigma, \eta, x})) + \mathcal{E}_{\mathbf{z}, x}(f_\rho) \} d\rho_X(x) \\ &\quad + \int_X \{ \mathcal{E}_{\mathbf{z}, x}(f_\lambda) - \mathcal{E}_{\mathbf{z}, x}(f_\rho) \\ &\quad - \mathcal{E}_x(f_\lambda) + \mathcal{E}_x(f_\rho) \} d\rho_X(x) \\ &:= \mathcal{S}_1(\mathbf{z}, \eta) + \mathcal{S}_2(\mathbf{z}, \lambda). \end{aligned}$$

We firstly give the upper bound of $\mathcal{S}_2(\mathbf{z}, \lambda)$ by using the Bernstein probability inequality in [14, 21].

Proposition 3.5 *Under the assumptions of Theorem 2.1, for any $0 < \delta < 1$, with confidence $1 - \delta/2$,*

$$\mathcal{S}_2(\mathbf{z}, \lambda) \leq \frac{\mathcal{D}(\lambda)}{2} + \frac{7(3M + \kappa \sqrt{\frac{\mathcal{D}(\lambda)}{\lambda}})^2 \log(2/\delta)}{3m}. \tag{3.7}$$

Next the estimation for $\mathcal{S}_1(\mathbf{z}, \eta)$ is more difficult in the sense that it involves the complexity of the function space $\mathcal{H}_{K, \mathbf{z}}$. Hence we need the uniform concentration inequality from [22].

Proposition 3.6 *Under the assumptions of Theorem 2.1, for any $0 < \delta < 1$, with confidence $1 - \delta/2$,*

$$\begin{aligned} \mathcal{S}_1(\mathbf{z}, \eta) &\leq \frac{1}{2} \int_X \{ \mathcal{E}_x(\pi_M(f_{\mathbf{z}, \sigma, \eta, x})) - \mathcal{E}_x(f_\rho) \} d\rho_X(x) \\ &\quad + \frac{176M^2}{m} \log\left(\frac{2}{\delta}\right) + C_{p, M} R_\eta^{\frac{2p}{2+p}} m^{-\frac{2}{2+p}}, \end{aligned} \tag{3.8}$$

where $R_\eta = \kappa m^{1-\frac{1}{q}} (\frac{M^2}{\eta})^{\frac{1}{q}}$.

4 Proof of the main result

Now we derive the learning rates.

Proof of Theorem 2.1 Combining the four bounds of Proposition 3.3, 3.4, 3.5 and 3.6 with (3.5), with confidence $1 - \delta$, we have

$$\begin{aligned} & \int_X \{ \mathcal{E}_x(\pi_M(f_{z,\sigma,\eta,x})) - \mathcal{E}_x(f_\rho) \} d\rho_X(x) \\ & \leq D_1 \log\left(\frac{2}{\delta}\right) \left\{ \lambda^{\min\{2r,1\}} + m^{-1} \lambda^{\min\{2r-1,0\}} \right. \\ & \quad \left. + m^{1-q} \eta \lambda^{-q} + m^{\frac{-2q-2p+2pq}{(2+p)q}} \eta^{-\frac{2p}{q(2+p)}} \right\}. \end{aligned} \tag{4.1}$$

By substituting (4.1) into (3.1), we have

$$\begin{aligned} \|\pi_M(f_{z,\eta}) - f_\rho\|_{\rho_X}^2 & \leq D_2 \log\left(\frac{2}{\delta}\right) \left\{ \sigma^{-\tau} \left\{ \lambda^{\min\{2r,1\}} + m^{-1} \lambda^{\min\{2r-1,0\}} \right. \right. \\ & \quad \left. \left. + m^{1-q} \eta \lambda^{-q} + m^{\frac{-2q-2p+2pq}{(2+p)q}} \eta^{-\frac{2p}{q(2+p)}} \right\} + \sigma \right\}. \end{aligned}$$

When $0 < r < 1/2$,

$$\begin{aligned} \|\pi_M(f_{z,\eta}) - f_\rho\|_{\rho_X}^2 & \leq D_2 \log\left(\frac{2}{\delta}\right) \left\{ \sigma^{-\tau} \left\{ \lambda^{2r} + m^{-1} \lambda^{2r-1} + m^{1-q} \eta \lambda^{-q} \right. \right. \\ & \quad \left. \left. + m^{\frac{-2q-2p+2pq}{(2+p)q}} \eta^{-\frac{2p}{q(2+p)}} \right\} + \sigma \right\}. \end{aligned}$$

Let $\lambda = m^{-\theta_1}$, $\eta = m^{-\theta_2}$ and $\sigma = m^{-\theta_3}$.

$$\|\pi_M(f_{z,\eta}) - f_\rho\|_{\rho_X}^2 \leq D_3 \log\left(\frac{2}{\delta}\right) m^{-\theta}, \tag{4.2}$$

where

$$\begin{aligned} \theta = \min \left\{ & -\tau\theta_3 + 2r\theta_1, -\tau\theta_3 + 1 + (2r - 1)\theta_1, \right. \\ & -\tau\theta_3 + q - 1 + \theta_2 - q\theta_1, \\ & \left. -\tau\theta_3 + \frac{2q + 2p - 2pq}{(2 + p)q} - \frac{2p}{q(2 + p)}\theta_2, \theta_3 \right\}. \end{aligned}$$

To maximize the learning rate, we take

$$\begin{aligned} \theta_{\max} = \max_{\theta_1, \theta_3} \min \left\{ & \max_{\theta_2} \min \left\{ -\tau\theta_3 + q - 1 + \theta_2 - q\theta_1, \right. \right. \\ & \left. -\tau\theta_3 + \frac{2q + 2p - 2pq}{(2 + p)q} - \frac{2p}{q(2 + p)}\theta_2 \right\}, \\ & \left. -\tau\theta_3 + 2r\theta_1, -\tau\theta_3 + 1 + (2r - 1)\theta_1, \theta_3 \right\}. \end{aligned}$$

Let

$$-\tau\theta_3 + q - 1 + \theta_2 - q\theta_1 = -\tau\theta_3 + \frac{2q + 2p - 2pq}{(2 + p)q} - \frac{2p}{q(2 + p)}\theta_2.$$

Then

$$\begin{aligned} \theta_{\max} &= \max_{\theta_1, \theta_3} \min \left\{ -\tau\theta_3 + q - 1 - q\theta_1 + \frac{-pq + 4q + 2p - 2q^2 - pq^2}{2p + 2q + pq} \right. \\ &\quad \left. + \frac{(2+p)q^2}{2p + 2q + pq} \theta_1, -\tau\theta_3 + 2r\theta_1, \right. \\ &\quad \left. -\tau\theta_3 + 1 + (2r - 1)\theta_1, \theta_3 \right\} \\ &\geq \max_{\theta_3} \min \left\{ \max_{\theta_1} \min \left\{ -\tau\theta_3 + q - 1 - q\theta_1 \right. \right. \\ &\quad \left. \left. + \frac{-pq + 4q + 2p - 2q^2 - pq^2}{2p + 2q + pq} \right. \right. \\ &\quad \left. \left. + \frac{(2+p)q^2}{2p + 2q + pq} \theta_1, -\tau\theta_3 + 2r\theta_1 \right\}, \right. \\ &\quad \left. \max_{\theta_1} \min \left\{ -\tau\theta_3 + 1 + (2r - 1)\theta_1, -\tau\theta_3 + 2r\theta_1 \right\}, \theta_3 \right\}. \end{aligned}$$

Let

$$\begin{aligned} &-\tau\theta_3 + q - 1 - q\theta_1 + \frac{-pq + 4q + 2p - 2q^2 - pq^2}{2p + 2q + pq} \\ &\quad + \frac{(2+p)q^2}{2p + 2q + pq} \theta_1 = -\tau\theta_3 + 2r\theta_1, \\ &-\tau\theta_3 + 1 + (2r - 1)\theta_1 = -\tau\theta_3 + 2r\theta_1. \end{aligned}$$

Then

$$\begin{aligned} \theta_{\max} &\geq \max_{\theta_3} \min \left\{ -\tau\theta_3 + \frac{4qr}{2r(2p + 2q + pq) + 2pq}, -\tau\theta_3 + 2r, \theta_3 \right\} \\ &\geq \min \left\{ \max_{\theta_3} \min \left\{ -\tau\theta_3 + \frac{4qr}{2r(2p + 2q + pq) + 2pq}, \theta_3 \right\}, \right. \\ &\quad \left. \max_{\theta_3} \min \left\{ -\tau\theta_3 + 2r, \theta_3 \right\} \right\} \\ &= 2r \min \left\{ -\frac{q\tau}{(1 + \tau)[r(2p + 2q + pq) + pq]} \right. \\ &\quad \left. + \frac{q}{r(2p + 2q + pq) + pq}, \frac{-\tau}{1 + \tau} + 1 \right\}. \end{aligned}$$

When $r \geq 1/2$,

$$\begin{aligned} \|\pi_M(f_{z,\eta}) - f_\rho\|_{\rho_X}^2 &\leq D_2 \log\left(\frac{2}{\delta}\right) \left\{ \sigma^{-\tau} \left\{ \lambda + m^{-1} + m^{1-q} \eta \lambda^{-q} \right. \right. \\ &\quad \left. \left. + m^{\frac{-2q-2p+2pq}{(2+p)q}} \eta^{-\frac{2p}{q(2+p)}} \right\} + \sigma \right\}. \end{aligned}$$

Similarly, we obtain

$$\theta_{\max} \geq \frac{2q}{(1 + \tau)(2p + 2q + 3pq)}.$$

So we choose

$$\theta(r) = \begin{cases} \min\{\frac{q}{[r(2p+2q+pq)+pq]}, 1\}(\frac{2r}{1+\tau}), & 0 < r < \frac{1}{2}; \\ \frac{2q}{(2p+2q+3pq)(1+\tau)}, & r \geq 1/2. \end{cases}$$

We complete the proof of Theorem 2.1. □

5 Conclusion and further discussion

We obtain the upper error bound of the algorithm (1.9) for the independent and identical samples with $1 \leq q \leq 2$. We decomposed the error quantity into the approximation error, the hypothesis error and the sample error and obtained their upper bounds using error analysis techniques developed in learning theory. In some practical applications, we may often encounter the non-i.i.d. sampling processes such as weakly dependent or non-identical processes; see [13, 15, 20]. It may be interesting to continue our error analysis for the non-i.i.d. samples.

Appendix 1: Error decomposition

Proof of Proposition 3.2 We have

$$\begin{aligned} & \int_X \{ \mathcal{E}_x(\pi_M(f_{z,\sigma,\eta,x})) - \mathcal{E}_x(f_\rho) \} d\rho_X(x) \\ & \leq \int_X \{ \mathcal{E}_x(\pi_M(f_{z,\sigma,\eta,x})) - \mathcal{E}_x(f_\rho) + \eta \Omega_z(f_{z,\sigma,\eta,x}) \} d\rho_X(x) \\ & = \mathcal{S}(z, \lambda, \eta) + \mathcal{H}(z, \lambda, \eta) + \int_X \{ \mathcal{E}_x(f_\lambda) - \mathcal{E}_x(f_\rho) + \lambda \|f_\lambda\|_K^2 \} d\rho_X(x). \end{aligned} \tag{A.1}$$

Moreover, by (1.5),

$$\begin{aligned} \mathcal{E}_x(f_\lambda) - \mathcal{E}_x(f_\rho) & = \int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) (f_\lambda(u) - f_\rho(u))^2 d\rho_X(u) \\ & \leq \|f_\lambda - f_\rho\|_{\rho_X}^2. \end{aligned} \tag{A.2}$$

This completes the proof of Proposition 3.2. □

Appendix 2: Estimates for the hypothesis error

Proof of Proposition 3.3 It is known from Theorem 2.1 in [11] that the coefficient $\mathbf{a}^z = (a_1^z, \dots, a_m^z)^T$ of $f_{z,\sigma,\lambda,x}$ satisfies

$$(m\lambda I + Q_x K[\mathbf{x}]) \mathbf{a}^z = Q_x \mathbf{y},$$

where I is the identity matrix, $\mathbf{y} = (y_1, y_2, y_3, \dots, y_m)^T$, $K[\mathbf{x}] = (K(x_i, x_j))_{i,j=1}^m$ and $Q_x = \text{diag}(\Phi(\frac{x}{\sigma}, \frac{x_i}{\sigma}) : i = 1, 2, \dots, m)$.

This implies

$$\lambda m \mathbf{a}^z = Q_x(\mathbf{y} - K[\mathbf{x}]\mathbf{a}^z).$$

Therefore, for $i = 1, 2, \dots, m$, we get

$$\begin{aligned} a_i^z &= \frac{1}{\lambda m} \Phi\left(\frac{x}{\sigma}, \frac{x_i}{\sigma}\right) \left(y_i - \sum_{j=1}^m a_j^z K(x_i, x_j) \right) \\ &= \frac{1}{\lambda m} \Phi\left(\frac{x}{\sigma}, \frac{x_i}{\sigma}\right) (y_i - f_{z,\sigma,\lambda,x}(x_i)). \end{aligned}$$

By the Hölder inequality, we have

$$\begin{aligned} \sum_{i=1}^m |a_i^z|^q &= \frac{1}{(\lambda m)^q} \sum_{i=1}^m \left| \Phi\left(\frac{x}{\sigma}, \frac{x_i}{\sigma}\right) (y_i - f_{z,\sigma,\lambda,x}(x_i)) \right|^q \\ &\leq \frac{1}{(\lambda m)^q} \left(\sum_{i=1}^m \Phi\left(\frac{x}{\sigma}, \frac{x_i}{\sigma}\right)^{\frac{q}{2-q}} \right)^{1-\frac{q}{2}} \\ &\quad \times \left(\sum_{i=1}^m \left(\frac{x}{\sigma}, \frac{x_i}{\sigma}\right) (y_i - f_{z,\sigma,\lambda,x}(x_i))^2 \right)^{\frac{q}{2}}. \end{aligned}$$

By (1.5), we have

$$\sum_{i=1}^m |a_i^z|^q \leq \frac{m}{(\lambda m)^q} (\mathcal{E}_{z,x}(f_{z,\sigma,\lambda,x}))^{\frac{q}{2}}.$$

Thus,

$$\begin{aligned} &\mathcal{E}_{z,x}(\pi_M(f_{z,\sigma,\eta,x})) + \eta \Omega_z(f_{z,\sigma,\eta,x}) \\ &\leq \mathcal{E}_{z,x}(f_{z,\sigma,\eta,x}) + \eta \Omega_z(f_{z,\sigma,\eta,x}) \\ &\leq \mathcal{E}_{z,x}(f_{z,\sigma,\lambda,x}) + \eta \Omega_z(f_{z,\sigma,\lambda,x}) \\ &\leq \mathcal{E}_{z,x}(f_{z,\sigma,\lambda,x}) + \frac{m\eta}{(\lambda m)^q} (\mathcal{E}_{z,x}(f_{z,\sigma,\lambda,x}))^{\frac{q}{2}} \\ &\leq \mathcal{E}_{z,x}(f_{z,\sigma,\lambda,x}) + \lambda \|f_{z,\sigma,\lambda,x}\|_K^2 \\ &\quad + \frac{m\eta}{(\lambda m)^q} (\mathcal{E}_{z,x}(f_{z,\sigma,\lambda,x}) + \lambda \|f_{z,\sigma,\lambda,x}\|_K^2)^{\frac{q}{2}}. \end{aligned}$$

Since

$$\mathcal{E}_{z,x}(f_{z,\sigma,\lambda,x}) + \lambda \|f_{z,\sigma,\lambda,x}\|_K^2 \leq \mathcal{E}_{z,x}(0) + \lambda \|0\|_K^2,$$

we get

$$\begin{aligned} & \mathcal{E}_{\mathbf{z},x}(\pi_M(f_{\mathbf{z},\sigma,\eta,x})) + \eta\Omega_{\mathbf{z}}(f_{\mathbf{z},\sigma,\eta,x}) \\ & \leq \mathcal{E}_{\mathbf{z},x}(f_{\mathbf{z},\sigma,\lambda,x}) + \lambda\|f_{\mathbf{z},\sigma,\lambda,x}\|_K^2 + \frac{m\eta M^2}{(\lambda m)^q}. \end{aligned} \tag{B.1}$$

It follows from (1.3)–(1.4) that

$$\mathcal{E}_{\mathbf{z},x}(f_{\mathbf{z},\sigma,\lambda,x}) + \lambda\|f_{\mathbf{z},\sigma,\lambda,x}\|_K^2 \leq \mathcal{E}_{\mathbf{z},x}(f_\lambda) + \lambda\|f_\lambda\|_K^2. \tag{B.2}$$

Combining (B.1) and (B.2), we obtain our desired estimation. □

Appendix 3: Estimates for the sample error

We estimate $\mathcal{S}_2(\mathbf{z}, \lambda)$ by using the lemma in [14, 21] below.

Lemma C.1 *Let $\{\xi(z_i)\}_{i=1}^m$ be independent random variables on a probability space Z with mean $\mathbb{E}(\xi)$ and variance $\sigma^2(\xi)$. Assume $|\xi(z) - \mathbb{E}\xi| \leq M$ almost surely. Then, for any $0 < \delta < 1$, with confidence $1 - \delta$, we have*

$$\frac{1}{m} \sum_{i=1}^m \xi(z_i) - \mathbb{E}\xi \leq \frac{2M \log(1/\delta)}{3m} + \sqrt{\frac{2\sigma^2(\xi) \log(1/\delta)}{m}}.$$

Proof of Proposition 3.5 Let $g(u, y) = \int_X \Phi(\frac{x}{\sigma}, \frac{y}{\sigma}) [(y - f_\lambda(u))^2 - (y - f_\rho(u))^2] d\rho_X(x)$, for any $z = (u, y) \in Z$. Then

$$\begin{aligned} \int_Z g d\rho &= \int_X \{\mathcal{E}_x(f_\lambda) - \mathcal{E}_x(f_\rho)\} d\rho_X(x); \\ \frac{1}{m} \sum_{i=1}^m g(z_i) &= \int_X \{\mathcal{E}_{\mathbf{z},x}(f_\lambda) - \mathcal{E}_{\mathbf{z},x}(f_\rho)\} d\rho_X(x). \end{aligned}$$

By (1.2) and (3.3), we have

$$\|f_\lambda\|_\infty \leq \kappa \|f_\lambda\|_K \leq \kappa \sqrt{\frac{\mathcal{D}(\lambda)}{\lambda}}.$$

Combining with (1.5), we have

$$\begin{aligned} |g(u, y)| &\leq \int_X \Phi\left(\frac{x}{\sigma}, \frac{y}{\sigma}\right) |(f_\lambda(u) - f_\rho(u)) \\ &\quad \times (f_\lambda(u) + f_\rho(u) - 2y)| d\rho_X(x) \\ &\leq (\|f_\lambda\|_\infty + M)(3M + \|f_\lambda\|_\infty) \\ &\leq \left(3M + \kappa \sqrt{\frac{\mathcal{D}(\lambda)}{\lambda}}\right)^2 := B_\lambda. \end{aligned}$$

Therefore,

$$\left\|g(u, y) - \int_Z g d\rho\right\|_\infty \leq 2B_\lambda$$

and

$$\begin{aligned} \sigma^2(g) &\leq \int_Z g^2 d\rho \\ &= \int_Z \left(\int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) d\rho_X(x) \right)^2 (f_\lambda(u) - f_\rho(u))^2 \\ &\quad \times (f_\lambda(u) + f_\rho(u) - 2y)^2 d\rho(u, y) \\ &\leq (3M + \|f_\lambda\|_\infty)^2 \|f_\lambda - f_\rho\|_{\rho_X}^2 \\ &\leq B_\lambda \mathcal{D}(\lambda). \end{aligned}$$

By Lemma C.1, with confidence $1 - \delta/2$, we have

$$\begin{aligned} \frac{1}{m} \sum_{i=1}^m g(z_i) - \int_Z g d\rho &\leq \frac{4B_\lambda \log(2/\delta)}{3m} + \sqrt{\mathcal{D}(\lambda) \times \frac{2B_\lambda \log(2/\delta)}{m}} \\ &\leq \frac{\mathcal{D}(\lambda)}{2} + \frac{7B_\lambda \log(2/\delta)}{3m}. \end{aligned}$$

This completes the proof of Proposition 3.5. □

We estimate $\mathcal{S}_1(\mathbf{z}, \eta)$ by using the proposition from [22] below.

Proposition C.1 *Let \mathcal{F} be a class of bounded measurable functions. Assume that there are constants $Q, \tau > 0$ and $\alpha \in [0, 1]$ such that $\|f\|_\infty \leq Q$ and $\mathbb{E}f^2 \leq \tau \mathbb{E}f^\alpha$ for every $f \in \mathcal{F}$. If for some $a > 0$ and $0 < p < 2$,*

$$\log \mathcal{N}_2(\mathcal{F}, \varepsilon) \leq a\varepsilon^{-p}, \quad \forall \varepsilon > 0, \tag{C.1}$$

then there exists a constant c'_p depending only on p such that, for any $t > 0$, with probability at least $1 - e^{-t}$, we have

$$\mathbb{E}f - \frac{1}{m} \sum_{i=1}^m f(z_i) \leq \frac{1}{2} \eta^{1-\alpha} (\mathbb{E}f)^\alpha + c'_p \eta + 2 \left(\frac{\tau t}{m} \right)^{\frac{1}{2-\alpha}} + \frac{18Qt}{m},$$

where

$$\eta := \max \left\{ \tau^{\frac{2-p}{4-2\alpha+p\alpha}} \left(\frac{a}{m} \right)^{\frac{2}{4-2\alpha+p\alpha}}, Q^{\frac{2-p}{2+p}} \left(\frac{a}{m} \right)^{\frac{2}{2+p}} \right\}.$$

Proof of Proposition 3.6 Consider the set of functions

$$\begin{aligned} \mathcal{G}_R = \left\{ g(u, y) = \int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) \left((y - \pi_M(f)(u))^2 \right. \right. \\ \left. \left. - (y - f_\rho(u))^2 \right) d\rho_X(x) : f \in B_R \right\}. \end{aligned}$$

By (1.5),

$$\begin{aligned}
 |g(u, y)| &\leq \int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) |(\pi_M(f)(u) - f_\rho(u)) \\
 &\quad \times (\pi_M(f)(u) + f_\rho(u) - 2y)| d\rho_X(x) \\
 &\leq 8M^2.
 \end{aligned}$$

It follows from the Schwarz inequality that

$$\begin{aligned}
 |g(u, y)|^2 &= \left| \int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) (\pi_M(f)(u) - f_\rho(u)) \right. \\
 &\quad \left. \times (\pi_M(f)(u) + f_\rho(u) - 2y) d\rho_X(x) \right|^2 \\
 &\leq 16M^2 \int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) (\pi_M(f)(u) \\
 &\quad - f_\rho(u))^2 d\rho_X(x) \int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) d\rho_X(x).
 \end{aligned}$$

Hence,

$$\mathbb{E}(g^2) \leq 16M^2 \int_X \left(\int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) (\pi_M(f)(u) - f_\rho(u))^2 d\rho_X(u) \right) d\rho_X(x).$$

It has been proved in [8] that

$$\begin{aligned}
 &\int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) (f(u) - f_\rho(u))^2 d\rho_X(u) \\
 &= \int_Z \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) [(f(u) - y)^2 - (f_\rho(u) - y)^2] d\rho(u, y),
 \end{aligned}$$

which implies

$$\begin{aligned}
 \mathbb{E}(g^2) &\leq 16M^2 \int_X \left(\int_Z \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) [(\pi_M(f)(u) - y)^2 \right. \\
 &\quad \left. - (f_\rho(u) - y)^2] d\rho(u, y) \right) d\rho_X(x) \\
 &= 16M^2 \int_Z \left(\int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) [(\pi_M(f)(u) - y)^2 \right. \\
 &\quad \left. - (f_\rho(u) - y)^2] d\rho_X(x) \right) d\rho(u, y) \\
 &= 16M^2 \mathbb{E}(g).
 \end{aligned}$$

Then, for any $g_1, g_2 \in \mathcal{G}_R$, we get

$$\begin{aligned}
 &|g_1(u, y) - g_2(u, y)| \\
 &= \left| \int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) ((\pi_M(f_1)(u) - y)^2 - (\pi_M(f_2)(u) - y)^2) d\rho_X(x) \right|
 \end{aligned}$$

$$\begin{aligned} &\leq \left| \int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) (\pi_M(f_1)(u)) - \pi_M(f_2)(u) \right. \\ &\quad \left. \times (\pi_M(f_1)(u) + \pi_M(f_2)(u) - 2y) d\rho_X(x) \right| \\ &\leq 4M|f_1(u) - f_2(u)|, \end{aligned}$$

which implies

$$\mathcal{N}_2(\mathcal{G}_R, \epsilon) \leq \mathcal{N}_2\left(B_R, \frac{\epsilon}{4M}\right) = \mathcal{N}_2\left(B_1, \frac{\epsilon}{4MR}\right).$$

Thus from the capacity condition (2.1), we have

$$\log \mathcal{N}_2(\mathcal{G}_R, \epsilon) \leq c_p(4M)^p R^p \epsilon^{-p}.$$

Now we can apply Proposition C.1 to \mathcal{G} with $Q = 8M^2, \alpha = 1, \tau = 16M^2$ and $a = c_p(4M)^p R^p$.

Thus for any $0 < \delta < 1$, with confidence $1 - \delta/2$, we have

$$\mathbb{E}g - \frac{1}{m} \sum_{i=1}^m g(z_i) \leq \frac{\mathbb{E}g}{2} + \frac{176M^2}{m} \log\left(\frac{2}{\delta}\right) + C_{p,M} R^{\frac{2p}{2+p}} m^{-\frac{2}{2+p}},$$

where $C_{p,M} = c'_p(4M)^{\frac{4}{2+p}} c_p^{\frac{2}{2+p}}$.

Moreover, we take $f = f_{z,\sigma,\eta,x}$ and derive the following bound of $f_{z,\sigma,\eta,x}$ by using the same method as in Lemma 3 of [18] and (1.5):

$$\|f_{z,\eta}\|_K \leq \kappa m^{1-\frac{1}{q}} \left(\frac{M^2}{\eta}\right)^{\frac{1}{q}}.$$

Finally, we complete the proof of Proposition 3.6 by taking $R = R_\eta = \kappa m^{1-\frac{1}{q}} \left(\frac{M^2}{\eta}\right)^{\frac{1}{q}}$. □

Funding

This work was supported by the Natural Science Foundation of China (grant nos. 11271199, 11671213).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors conceived of the study, participated in its design and coordination, drafted the manuscript, participated in the sequence alignment, and read and approved the final manuscript.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 20 June 2018 Accepted: 18 September 2018 Published online: 25 September 2018

References

1. Cerny, M., Antoch, J., Hladik, M.: On the possibilistic approach to linear regression models involving uncertain, indeterminate or interval data. *Inf. Sci.* **244**(7), 26–47 (2013)
2. Fasshauer, G.E.: Toward approximate moving least squares approximation with irregularly spaced centers. *Comput. Methods Appl. Mech. Eng.* **193**(12–14), 1231–1243 (2004)
3. Komargodski, Z., Levin, D.: Hermite type moving-least-squares approximations. *Comput. Math. Appl.* **51**(8), 1223–1232 (2006)
4. McLain, D.H.: Drawing contours from arbitrary data points. *Comput. J.* **17**(17), 318–324 (1974)

5. Savitzky, A., Golay, M.J.E.: Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**(8), 1627–1639 (1964)
6. Wand, M.P., Jones, M.C.: *Kernel Smoothing*, Chapman & Hall, New York (1995)
7. Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data. In: *ACM National Conference*, pp. 517–524 (1968)
8. Wang, H.Y., Xiang, D.H., Zhou, D.X.: Moving least-square method in learning theory. *J. Approx. Theory* **162**(3), 599–614 (2010)
9. Wang, H.Y.: Concentration estimates for the moving least-square method in learning theory. *J. Approx. Theory* **163**(9), 1125–1133 (2011)
10. He, F.C., Chen, H., Li, L.Q.: Statistical analysis of the moving least-squares method with unbounded sampling. *Inf. Sci.* **268**(1), 370–380 (2014)
11. Tong, H.Z., Wu, Q.: Learning performance of regularized moving least square regression. *J. Comput. Appl. Math.* **325**, 42–55 (2017)
12. Guo, Q., Ye, P.: Error analysis of the moving least-squares method with non-identical sampling. *Int. J. Comput. Math.* 1–15 (2018, in press). <https://doi.org/10.1080/00207160.2018.1469748>
13. Guo, Q., Ye, P.X.: Coefficient-based regularized regression with dependent and unbounded sampling. *Int. J. Wavelets Multiresolut. Inf. Process.* **14**(5), 1–14 (2016)
14. Wu, Q., Ying, Y.M., Zhou, D.X.: Learning rates of least-square regularized regression. *Found. Comput. Math.* **6**(2), 171–192 (2006)
15. Pan, Z.W., Xiao, Q.W.: Least-square regularized regression with non-iid sampling. *J. Stat. Plan. Inference* **139**(10), 3579–3587 (2009)
16. Shi, L., Feng, Y.L., Zhou, D.X.: Concentration estimates for learning with l^1 -regularizer and data dependent hypothesis spaces. *Appl. Comput. Harmon. Anal.* **31**(2), 286–302 (2011)
17. Lv, S.G., Shi, D.M., Xiao, Q.W., Zhang, M.S.: Sharp learning rates of coefficient-based l^2 -regularized regression with indefinite kernels. *Sci. China Math.* **56**(8), 1557–1574 (2013)
18. Feng, Y.L., Lv, S.G.: Unified approach to coefficient-based regularized regression. *Comput. Math. Appl.* **62**(1), 506–515 (2011)
19. Nie, W.L., Wang, C.: Constructive analysis for coefficient regularization regression algorithms. *J. Math. Anal. Appl.* **431**(2), 1153–1171 (2015)
20. Guo, Q., Ye, P., Cai, B.: Convergence rate for l^2 -coefficient regularized regression with non-i.i.d. sampling. *IEEE Access* **6**, 18804–18813 (2018)
21. Cucker, F., Zhou, D.X.: *Learning Theory: An Approximation Theory Viewpoint*. Cambridge University Press, Cambridge (2007)
22. Wu, Q., Ying, Y., Zhou, D.X.: Multi-kernel regularized classifiers. *J. Complex.* **23**(1), 108–134 (2007)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
