Hindawi Publishing Corporation
Abstract and Applied Analysis
Volume 2012, Article ID 436510, 16 pages
doi:10.1155/2012/436510

Research Article

Consistency Analysis of Spectral Regularization Algorithms

Yukui Zhu and Hongwei Sun

Shandong Provincial Key Laboratory of Network based Intelligent Computing, School of Mathematic Science, University of Jinan, Jinan 250022, China

Correspondence should be addressed to Hongwei Sun, ss_sunhw@ujn.edu.cn

Received 10 January 2012; Revised 30 March 2012; Accepted 16 April 2012

Academic Editor: Ondřej Došlý

We investigate the consistency of spectral regularization algorithms. We generalize the usual definition of regularization function to enrich the content of spectral regularization algorithms. Under a more general prior condition, using refined error decompositions and techniques of operator norm estimation, satisfactory error bounds and learning rates are proved.

1. Introduction

In this paper, we study the consistency analysis of spectral regularization algorithms in regression learning.

Let (X, d) be a compact metric space and ρ a probability distribution on $Z = X \times Y$ with $Y = \mathbb{R}$. The regression learning aims at estimating or approximating the regression function

$$f_\rho(x) = \int_Y y d\rho(y \mid x) \tag{1.1}$$

through a set of samples $\mathbf{z} = \{(x_i, y_i)\}_{i=1}^m \in Z^m$ drawn independently and identically according to ρ from Z.

In learning theory, a reproducing kernel Hilbert space (RKHS) associated with a Mercer kernel $K(x, y)$ is usually taken as the hypothesis space. Recall that a function $K : X \times X \to R$ is called a Mercer kernel if it is continuous, symmetric, and positive semidefinite. The reproducing kernel Hilbert space $\mathcal{H}_K$ is defined to be the closure of the linear span of $K_x := K(\cdot, x),\ x \in X$. The reproducing property takes the form

$$f(x) = \langle f, K_x \rangle_K, \quad \forall f \in \mathcal{H}_K,\ \forall x \in X. \tag{1.2}$$

For the Mercer kernel $K(x,y)$, we denote that

$$\kappa = \max_{x\in X}\sqrt{K(x,x)}. \tag{1.3}$$

Our first contribution is to generalize the definition of regularization in [1] such that many more learning algorithms can be included in the scope of spectral algorithms.

Definition 1.1. We say that a family of continuous functions $g_\lambda : [0,\kappa^2] \to R$, $\lambda \in (0,\ 1]$ is regularization, if the following conditions hold.

(i) There exists a constant D such that

$$\sup_{0<\sigma\le\kappa^2} \left|\sigma g_\lambda(\sigma)\right| \le D. \tag{1.4}$$

(ii) There exists a constant $B > 0,\ 0 < \alpha \le 1$ such that

$$\sup_{0<\sigma\le\kappa^2} \left|g_\lambda(\sigma)\right| \le \frac{B}{\lambda^\alpha}. \tag{1.5}$$

(iii) There exists a constant γ such that

$$\sup_{0<\sigma\le\kappa^2} \left|1 - g_\lambda(\sigma)\sigma\right| \le \gamma. \tag{1.6}$$

(iv) The qualification ν_0 of the regularization g_λ is the maximal ν such that

$$\sup_{0<\sigma\le\kappa^2} \left|1 - g_\lambda(\sigma)\sigma\right|\sigma^\nu \le \gamma_\nu \lambda^{\alpha\nu}, \tag{1.7}$$

where γ_ν does not depend on λ.

Our definition for regularization is different from that in [1]. In fact, the definition given by [1] is the special case when taking $\alpha = 1$ in (1.5) and (1.7). So from this viewpoint, our assumption is more mild and it is fit for more general situations, for example, coefficient regularization algorithms correspond to spectral algorithms with $\alpha = 1/2$, the relation between coefficient regularization algorithms and spectral algorithms had been explored in [2].

Let $\mathbf{x} = \{x_i\}_{i=1}^m$ and $\mathbf{y} = \{y_i\}_{i=1}^m$. The sample operator $S_\mathbf{x} : \mathscr{H}_K \to \mathbb{R}^m$ is defined as $S_\mathbf{x} f = \{f(x_i)\}_{i=1}^m$. The adjoint of $S_\mathbf{x}$ under $1/m$ times the Euclidean norm is $S_\mathbf{x}^T c = (1/m)\sum_{i=1}^m c_i K_{x_i}$. For simplicity, we use $T_\mathbf{x}$ to stand for $S_\mathbf{x}^T S_\mathbf{x}$.

The spectral regularization algorithm considered here is given by

$$f_z^\lambda = g_\lambda(T_\mathbf{x}) S_\mathbf{x}^T \mathbf{y}. \tag{1.8}$$

The regularization g_λ, $\lambda \in (0,1]$ in (1.8) was proposed originally to solve ill-posed inverse problems. The relation between learning theory and regularization of linear ill-posed problems has been well discussed in a series of articles, see [1, 3] and the references therein.

The analysis made in previous literatures provides us with a deep understanding of the connection between learning theory and regularization.

A large class of learning algorithms can be considered as spectral regularization algorithms in accordance with different regularizations.

Example 1.2. The regularized least square algorithm is given as

$$f_{\mathbf{z}}^{\lambda} = \arg\min_{f\in\mathcal{H}_K} \frac{1}{m}\sum_{i=1}^{m}(y_i - f(x_i))^2 + \lambda\|f\|_K^2. \tag{1.9}$$

It has been well understood due to a lot of literatures [4–11], and so forth. It is proved in [7] that

$$f_{\mathbf{z}}^{\lambda} = (T_{\mathbf{x}} + \lambda I)^{-1} S_{\mathbf{x}}^{T} y, \tag{1.10}$$

which corresponds to algorithm (1.8) with the regularization

$$g_{\lambda}(\sigma) = (\sigma + \lambda)^{-1}. \tag{1.11}$$

In this case, we have $B = D = \gamma = \gamma_{\nu_0} = \alpha = 1$, the qualification $\nu_0 = 1$.

Example 1.3. In regression learning, the coefficient regularization with l^2 norm becomes

$$f_{\mathbf{z}}^{\lambda} = f_{\alpha_{\mathbf{z}}}, \qquad \alpha_{\mathbf{z}} = \arg\min_{\alpha\in\mathbb{R}^m} \frac{1}{m}\sum_{i=1}^{m}(y_i - f_{\alpha}(x_i))^2 + \lambda m\sum_{i=1}^{m}\alpha_i^2, \tag{1.12}$$

where

$$f_{\alpha} = \sum_{i=1}^{m}\alpha_i K_{x_i}, \qquad \forall\alpha \in \mathbb{R}^m. \tag{1.13}$$

The coefficient regularization was first introduced by Vapnik [12] to design linear programming support vector machines. The consistency of this algorithm has been studied in literatures [2, 13, 14]. In [2], it is proved that the sample error has $O(1/\sqrt{m})$ decay, even for nonpositive semidefinite kernels, and

$$f_{\mathbf{z}}^{\lambda} = \left(\lambda I + T_{\mathbf{x}}^2\right)^{-1} T_{\mathbf{x}} S_{\mathbf{x}}^{T}\mathbf{y}. \tag{1.14}$$

Thus, it corresponds to algorithm (1.8) with the regularization

$$g_{\lambda}(\sigma) = \left(\sigma^2 + \lambda\right)^{-1}\sigma. \tag{1.15}$$

In this case, we have $B = D = \gamma = \gamma_{\nu_0} = 1$, the qualification $\nu_0 = 2$ and $\alpha = 1/2$.

Example 1.4. Landweber iteration is defined by $g_\lambda(\sigma) = \sum_{i=0}^{\lfloor 1/\lambda \rfloor - 1}(1-\sigma)^i$ where $\lfloor a \rfloor = \max\{m : m \in \mathbb{Z}, m \le a\}$. This corresponds to the gradient descent algorithm in Yao et al. [15] with constant step-size. In this case, we have that any $\nu \in [0, +\infty)$ can be considered as qualification of this method and $\gamma_\nu = 1$ if $0 < \nu \le 1$ and $\gamma_\nu = \nu^\nu$ otherwise.

Let $f_{\mathcal{H}}^+$ be the projection of f_ρ onto $\overline{\mathcal{H}_K}$, here $\overline{\mathcal{H}}_K$ denotes the closure of $\mathcal{H}_K$ in $L^2_{\rho_X}(X)$. The generalization error of f_z^λ is

$$\mathcal{E}\left(f_z^\lambda\right) = \int_Z \left(f_z^\lambda(x) - y\right)^2 d\rho = \int_X \left(f_z^\lambda(x) - f_{\mathcal{H}}^+(x)\right)^2 d\rho_X + \int_X \left(f_{\mathcal{H}}^+(x) - f_\rho(x)\right)^2 d\rho_X + \sigma^2, \tag{1.16}$$

where ρ_X is the marginal distribution of ρ on X, σ^2 is the variance of random variable $y - f_\rho(x)$. So the goodness of the approximation f_z^λ is measured by $\|f_z^\lambda - f_{\mathcal{H}}^+\|_{\rho_X}$, where we take the L^2 norm defined as

$$\|f\|_{\rho_X} = \left(\int_X |f(x)|^2 d\rho_X\right)^{1/2}, \quad \forall f \in L^2_{\rho_X}(X). \tag{1.17}$$

The integral operator L_K associated with kernel K from $L^2_{\rho_X}(X)$ to $L^2_{\rho_X}(X)$ is defined by

$$L_K f(x) = \int_X K(x,t) f(t) d\rho_X(t), \quad \forall f \in L^2_{\rho_X}(X). \tag{1.18}$$

L_K is a nonnegative self-adjoint compact operator [4]. If the domain of L_K is restricted to $\mathcal{H}_K$, it also is a nonnegative self-adjoint compact operator from $\mathcal{H}_K$ to $\mathcal{H}_K$, with norm $\|L_K\|_{\mathcal{H}_K \to \mathcal{H}_K} \le \kappa^2$ [16]. In the sequel, we simply write $\|L_K\|$ instead of $\|L_K\|_{\mathcal{H}_K \to \mathcal{H}_K}$ and assume that $|y| \le M$ almost surely.

As usual, we use the following error decomposition:

$$\left\|f_z^\lambda - f_{\mathcal{H}}^+\right\|_{\rho_X} \le \left\|f_z^\lambda - f_\lambda\right\|_{\rho_X} + \|f_\lambda - f_{\mathcal{H}}^+\|_{\rho_X}, \tag{1.19}$$

where

$$f_\lambda = g_\lambda(L_K) L_K f_{\mathcal{H}}^+. \tag{1.20}$$

The first term on the right-hand side of (1.19) is called sample error, and the second one is approximation error. Sample error depends on the sampling, and the law of large numbers would lead to its estimation; approximation error is independent of the sampling, and its estimation is mainly through the method of operator approximation.

In order to deduce the error bounds and learning rates, we have to set restriction on the class of possible probability measures that is usually called prior condition. In previous literatures, prior conditions are usually described through the smoothness of regression function f_ρ. We suppose the following prior condition:

$$f_{\mathscr{H}}^{+} = \varphi(L_K)h_0, \qquad h_0 \in L^2_{\rho_X}(X),\ \|h_0\|_{\rho_X} \leq R. \tag{1.21}$$

Here, φ called the index function is some continuous nondecreasing function defined on $[0, \kappa^2]$ with $\varphi(0) = 0$.

In the sequel, we request the qualification $\nu_0 > 1/2$, and there exists $\mu_0 > 0$ covering φ, which means that there is $c > 0$ such that

$$c\frac{\lambda^{\mu_0}}{\varphi(\lambda)} \leq \inf_{\lambda \leq \sigma \leq \kappa^2} \frac{\sigma^{\mu_0}}{\varphi(\sigma)}, \quad 0 < \lambda \leq \kappa^2. \tag{1.22}$$

It is easy to see that, for any $\mu \geq \mu_0$, μ covers φ.

Furthermore, we request that $\varphi(t)$ is operator monotone on $[0, \kappa^2]$, that is, there is a constant $c_\varphi < \infty$, such that for any pair U,V of nonnegative self-adjoint operators on some Hilbert space with norm less than κ^2, it holds

$$\left\|\varphi(U) - \varphi(V)\right\| \leq c_\varphi \varphi(\|U - V\|), \tag{1.23}$$

and, there is $d_\varphi > 0$ such that

$$d_\varphi \frac{\lambda}{\varphi(\lambda)} \leq \frac{\sigma}{\varphi(\sigma)}, \quad 0 < \lambda < \sigma \leq \kappa^2. \tag{1.24}$$

It is proved that $\varphi(t) = t^\alpha$ for $0 \leq \alpha \leq 1$ is operator monotone [8].

In [1], Bauer et al. consider the following prior condition:

$$f_{\mathscr{H}}^{+} \in \Omega_{\varphi,R}, \qquad \Omega_{\varphi,R} = \left\{f \in \mathscr{H}_K : f = \varphi(L_K)v,\ \|v\|_K \leq R\right\}. \tag{1.25}$$

This condition is somewhat restrictive, since it asks that $f_{\mathscr{H}}^{+}$ must belong to $\mathscr{H}_K$.

Our result shows that satisfactory error bound is available with a more general prior condition, this is our second main contribution. So from this view point, our work is meaningful. The main result of this paper is the following theorem.

Theorem 1.5. *Suppose the index function φ with covering number $\mu_0 > 0$ is operator monotone on $[0, \kappa^2]$. The qualification ν_0 satisfies $\nu_0 > \max\{1/2, \mu_0\}$ and that $m \geq 2\log(4/\delta)$ for $0 < \delta < 1$. Then, with confidence $1 - \delta$, there holds*

$$\begin{aligned}\left\|f_z^\lambda - f_{\mathscr{H}}^{+}\right\|_{\rho_X} \leq C_1\Big\{&\left(1 + \lambda^{-\alpha/2}\zeta^{1/2}\right)\left(\varphi(\lambda)\lambda^{(\alpha-1)\mu_0} + \varphi(\zeta) + \lambda^{-\alpha/2}\eta\right) \\ &+\left[\lambda^{(\alpha-1)}\left(\varphi(\lambda)\right)^{1/\mu_0}\right]^{\min\{\mu_0,\ \nu_0-1/2\}}\Big\},\end{aligned} \tag{1.26}$$

where

$$\zeta = 2\kappa^2\sqrt{\frac{2\log(4/\delta)}{m}},$$

$$\eta = \varphi(\lambda)\lambda^{-\mu_0+\min\{\alpha(\mu_0-1/2),\,0\}}m^{-1}\log\frac{4}{\delta} + \left(1+\varphi(\lambda)\lambda^{(\alpha-1)\mu_0}\right)m^{-1/2}\sqrt{\log\frac{4}{\delta}}$$
$$+\,\lambda^{\min\{\mu_0,\,\nu_0-1/2\}(\alpha-1)+\alpha/2}\left(\varphi(\lambda)\right)^{\min\{(2\nu_0-1)/2\mu_0,\,1\}}, \tag{1.27}$$

and C_1 is a constant independent of λ, m, δ.

This theorem shows the consistency of the spectral algorithms, gives the error bound, and also can lead to satisfactory learning rates by the explicit expression of φ.

This paper is prepared as follows. In Section 2, we will prove a basic lemma about estimation of operator norms related to the regularization and two concentration inequalities with vector value random variables. In Section 3, we give the proof of Theorem 1.5. In Section 4, we derive learning rate under the setting of several specific regularization.

2. Some Lemmas

We simply write γ_0 instead of γ_{ν_0} in (1.7) for qualification ν_0. To estimate the error $\|f_z^\lambda - f_{\mathcal{H}}^+\|_{\rho_X}$, we need the following lemma to bound the norms of some operators.

Lemma 2.1. *Let φ be an index function and $\nu_0 > \max\{1/2, \mu_0\}$. Then, the following inequalities hold true:*

$$\sup_{0<\sigma\le\kappa^2}\left|1-g_\lambda(\sigma)\sigma\right|\sigma^s \le \gamma^{1-s/\nu_0}\gamma_0^{s/\nu_0}\lambda^{\alpha s}, \quad \forall 0<s\le\nu_0, \tag{2.1}$$

$$\sup_{0<\sigma\le\kappa^2}\left|1-g_\lambda(\sigma)\sigma\right|\varphi^s(\sigma) \le \alpha_s\varphi^s(\lambda)\lambda^{(\alpha-1)\mu_0 s}, \quad \forall 0<s\le\frac{\nu_0}{\mu_0}, \tag{2.2}$$

$$\sup_{0<\sigma\le\kappa^2}\left|1-g_\lambda(\sigma)\sigma\right|\varphi(\sigma)\sqrt{\sigma} \le \beta_1\lambda^{\min\{\mu_0,\,\nu_0-1/2\}(\alpha-1)+\alpha/2}\left(\varphi(\lambda)\right)^{\min\{(2\nu_0-1)/2\mu_0,\,1\}}, \tag{2.3}$$

$$\sup_{0<\sigma\le\kappa^2}\left|g_\lambda(\sigma)\sigma^{1/2}\varphi(\sigma)\right| \le \beta_2\varphi(\lambda)\lambda^{\min\{\alpha(\mu_0-1/2),\,0\}-\mu_0}. \tag{2.4}$$

Here, $\alpha_s, \beta_1, \beta_2$ are constants only dependent on $\nu_0, \mu_0, \gamma, \gamma_0, c, \varphi(\kappa^2)$.

Proof. By (1.6) and (1.7), for any $0 < s \le \nu_0$, we have

$$\begin{aligned}\sup_{0<\sigma\le\kappa^2}\left|1-g_\lambda(\sigma)\sigma\right|\sigma^s &\le \sup_{0<\sigma\le\kappa^2}\left[\left|1-g_\lambda(\sigma)\sigma\right|\sigma^{\nu_0}\right]^{s/\nu_0}\times\left|1-g_\lambda(\sigma)\sigma\right|^{1-s/\nu_0}\\ &\le \gamma^{1-s/\nu_0}\gamma_0^{s/\nu_0}\lambda^{\alpha s}.\end{aligned} \tag{2.5}$$

Since $\mu_0 s \le \nu_0$ and φ is covered by μ_0, by (2.1) and (1.6), we get

$$\begin{aligned}
\sup_{0<\sigma\le\kappa^2} |1-g_\lambda(\sigma)\sigma|\varphi^s(\sigma) &= \max\left\{\sup_{0<\sigma<\lambda}|1-g_\lambda(\sigma)\sigma|\varphi^s(\sigma), \sup_{\lambda\le\sigma\le\kappa^2}|1-g_\lambda(\sigma)\sigma|\frac{\varphi^s(\sigma)}{\sigma^{\mu_0 s}}\sigma^{\mu_0 s}\right\} \\
&\le \max\left\{\gamma\varphi^s(\lambda), \frac{1}{c^s}\varphi^s(\lambda)\gamma^{1-\mu_0 s/v_0}\gamma_0^{\mu_0 s/v_0}\lambda^{(\alpha-1)\mu_0 s}\right\} \\
&= \max\left\{\gamma, \frac{1}{c^s}\gamma^{1-\mu_0 s/v_0}\gamma_0^{\mu_0 s/v_0}\right\}\varphi^s(\lambda)\lambda^{(\alpha-1)\mu_0 s} \doteq \alpha_s\varphi^s(\lambda)\lambda^{(\alpha-1)\mu_0 s}.
\end{aligned} \tag{2.6}$$

In order to prove the third inequality, let $\tau = \min\{2\nu_0/(2\nu_0-1), \nu_0/\mu_0\}$ and $\tau(1-1/2\nu_0) = (1/\mu_0)\min\{\mu_0, \nu_0 - 1/2\}$, by (2.2), we have

$$\begin{aligned}
\sup_{0<\sigma\le\kappa^2} |1-g_\lambda(\sigma)\sigma|^{1-1/2\nu_0}\varphi(\sigma) &= \sup_{0<\sigma\le\kappa^2}\left[|1-g_\lambda(\sigma)\sigma|\varphi^\tau(\sigma)\right]^{1-1/2\nu_0}(\varphi(\sigma))^{1-\tau(1-1/2\nu_0)} \\
&\le \left(\varphi\left(\kappa^2\right)\right)^{1-\tau(1-1/2\nu_0)}\alpha_\tau^{1-1/2\nu_0}\lambda^{\min\{\mu_0,\ \nu_0-1/2\}(\alpha-1)} \\
&\quad\times(\varphi(\lambda))^{\min\{(2\nu_0-1)/2\mu_0,\ 1\}}.
\end{aligned} \tag{2.7}$$

Thus,

$$\begin{aligned}
\sup_{0<\sigma\le\kappa^2} |1-g_\lambda(\sigma)\sigma|\varphi(\sigma)\sqrt{\sigma} &= \sup_{0<\sigma\le\kappa^2}\left[|1-g_\lambda(\sigma)\sigma|\sigma^{\nu_0}\right]^{1/2\nu_0}\times|1-g_\lambda(\sigma)\sigma|^{1-1/2\nu_0}\varphi(\sigma) \\
&\le \beta_1\lambda^{\min\{\mu_0,\nu_0-1/2\}(\alpha-1)+\alpha/2}(\varphi(\lambda))^{\min\{(2\nu_0-1)/2\mu_0,\ 1\}},
\end{aligned} \tag{2.8}$$

where β_1 is a constant only dependent on $\nu_0, \mu_0, \gamma, \gamma_0, c, \varphi(\kappa^2)$.

If $0 < \mu_0 \le 1/2$, we have

$$\begin{aligned}
\sup_{0<\sigma\le\kappa^2}\left|g_\lambda(\sigma)\sigma^{1/2}\varphi(\sigma)\right| &= \max\left\{\sup_{0<\sigma<\lambda}\left|g_\lambda(\sigma)\sigma^{1/2}\varphi(\sigma)\right|, \sup_{\lambda\le\sigma\le\kappa^2}\left|g_\lambda(\sigma)\sigma^{\mu_0+1/2}\frac{\varphi(\sigma)}{\sigma^{\mu_0}}\right|\right\} \\
&\le \max\left\{\sup_{0<\sigma<\lambda}|g_\lambda(\sigma)\sigma|^{1/2}|g_\lambda(\sigma)|^{1/2}\varphi(\lambda), \frac{\varphi(\lambda)}{c\lambda^{\mu_0}}D^{\mu_0+1/2}B^{1/2-\mu_0}\lambda^{-(1/2-\mu_0)\alpha}\right\} \\
&\le \max\left\{\sqrt{BD}, c^{-1}D^{\mu_0+1/2}B^{1/2-\mu_0}\right\}\varphi(\lambda)\lambda^{\alpha(\mu_0-1/2)-\mu_0}.
\end{aligned} \tag{2.9}$$

Similarly computation shows that, for $\mu_0 \ge 1/2$,

$$\sup_{0<\sigma\le\kappa^2}\left|g_\lambda(\sigma)\sigma^{1/2}\varphi(\sigma)\right| \le \max\left\{\sqrt{BD}, c^{-1}D\kappa^{2\mu_0-1}\right\}\varphi(\lambda)\lambda^{-\mu_0}. \tag{2.10}$$

Thus, the last inequality holds, and we complete the proof. $\square$

By taking $s = 1/2$ in (2.1), we have

$$\sup_{0<\sigma\le\kappa^2} \left|1 - g_\lambda(\sigma)\sigma\right|\sigma^{1/2} \le \gamma^{1-1/2v_0}\gamma_0^{1/2v_0}\lambda^{\alpha/2}. \tag{2.11}$$

The estimates of operator norm mainly adopt the following classical argument in operator theory. Argument: let A be a positive operator in a Hilbert space, for $f \in C[0, \|A\|]$, then $f(A)$ is self-adjoint by [17, Proposition 4.4.7] and $\sigma(f(A)) = \{f(t) : t \in \sigma(A)\}$ by [17, Theorem 4.4.8] where $\sigma(A)$ is the spectral set of A. Consequently, $\|f(A)\| \le \|f\|_\infty$.

The following probability inequality concerning random variables with values in a Hilbert space is proved in [18].

Lemma 2.2. *Let H be a Hilbert space and ξ a random variable on (Z, ρ) with values in H. Assume $\|\xi\| \le \widetilde{M} < \infty$ almost surely. Denote $\sigma^2(\xi) = E(\|\xi\|^2)$. Let $\{z_i\}_{i=1}^m$ be independent random drawers of ρ. For any $0 < \delta < 1$, with confidence $1 - \delta$, there holds*

$$\left\|\frac{1}{m}\sum_{i=1}^m [\xi(z_i) - E(\xi)]\right\| \le \frac{2\widetilde{M}\log(2/\delta)}{m} + \sqrt{\frac{2\sigma^2(\xi)\log(2/\delta)}{m}}. \tag{2.12}$$

Let $\mathrm{HS}(\mathscr{H}_K)$ be the class of all the Hilbert Schmidt operators on $\mathscr{H}_K$. It forms a Hilbert space with inner product

$$\langle T, S\rangle_{\mathrm{HS}} := \sum_{i=1}^\infty \langle T\varphi_i, S\varphi_i\rangle_K, \tag{2.13}$$

where φ_i is an orthonormal basis of $\mathscr{H}_K$ and this definition does not depend on the choice of the basis. The integral operator L_K, as an operator on $\mathscr{H}_K$, belongs to $\mathrm{HS}(\mathscr{H}_K)$ and $\|L_K\|_{\mathrm{HS}} \le \kappa^2$ (see [9]). By Lemma 2.2, we can estimate the following operator norms.

Lemma 2.3. *Let $\mathbf{x} = \{x_i\}_{i=1}^m$ be a sample set i.i.d drawn from (X, ρ_X). With confidence $1 - \delta$, we have*

$$\left\|L_K - S_{\mathbf{x}}^T S_{\mathbf{x}}\right\| \le \kappa^2\left(\frac{2\log(2/\delta)}{m} + \sqrt{\frac{2\log(2/\delta)}{m}}\right). \tag{2.14}$$

Proof. Observe that $S_{\mathbf{x}}^T S_{\mathbf{x}} = (1/m)\sum_{i=1}^m K_{x_i}\langle\cdot, K_{x_i}\rangle_K$. Denote $S_{\mathbf{x}}^T S_{\mathbf{x}} = (1/m)\sum_{i=1}^m \xi(x_i)$. Here ξ is the random variable on (X, ρ_X) given by $\xi(x) = K_x\langle\cdot, K_x\rangle_K$.

Consider

$$\langle\xi(x), \xi(x)\rangle_{\mathrm{HS}} = \sum_{i=1}^\infty \left\langle K_x\langle\varphi_i, K_x\rangle_K, K_x\langle\varphi_i, K_x\rangle_K\right\rangle_K = \sum_{i=1}^\infty \langle\varphi_i, K_x\rangle_K^2 K(x, x) \le \kappa^4. \tag{2.15}$$

For $x \in X$ and $f \in \mathscr{H}_K$, the reproducing property insures that

$$\xi(x)(f) = K_x\langle f, K_x\rangle_K = f(x)K_x. \tag{2.16}$$

Hence, $E(\xi) = L_K$, and thereby

$$\left(L_K - S_{\mathbf{x}}^T S_{\mathbf{x}}\right) = E\xi - \frac{1}{m}\sum_{i=1}^{m}\xi(x_i). \tag{2.17}$$

According to (2.15), there holds $\sigma^2(\xi) = E\|\xi\|_{\mathrm{HS}}^2 \leq \kappa^4$. Inequality (2.14) then follows from (2.12) and the fact that $\|L_K - S_{\mathbf{x}}^T S_{\mathbf{x}}\| \leq \|L_K - S_{\mathbf{x}}^T S_{\mathbf{x}}\|_{\mathrm{HS}}$. □

Lemma 2.4. *Under the assumption of Lemma 2.1. Let* $\mathbf{z} = \{z_i\}_{i=1}^m$ *be a sample set i.i.d drawn from* (Z, ρ). *With confidence* $1 - \delta$, *we have*

$$\begin{aligned}\left\|S_{\mathbf{x}}^T y - T_{\mathbf{x}} f_\lambda\right\|_K \leq & 2\kappa\left(M + \kappa\beta_2 R\varphi(\lambda)\lambda^{-\mu_0+\min\{\alpha(\mu_0-1/2),\,0\}}\right)\frac{\log(2/\delta)}{m} + \beta_1 R\lambda^{\min\{\mu_0,\nu_0-1/2\}(\alpha-1)+\alpha/2} \\ & \times \left(\varphi(\lambda)\right)^{\min\{(2\nu_0-1)/2\mu_0,\,1\}} + \kappa\left(\alpha_1\lambda^{(\alpha-1)\mu_0}\varphi(\lambda)R + c_\rho\right)\sqrt{\frac{2\log(2/\delta)}{m}}.\end{aligned} \tag{2.18}$$

Proof. Define $\varsigma = (f_\lambda(x) - y)K_x$, so ς is a random variable from Z to $\mathscr{H}_K$. Combing the reproducing property with Cauchy-Schwartz inequality, we get

$$\|f_\lambda\|_\infty = \sup_{x\in X}\left|\langle f_\lambda, K_x\rangle_K\right| \leq \kappa\|f_\lambda\|_K. \tag{2.19}$$

Since $L_K^{1/2}$ is an isometric isomorphism from $(\overline{\mathscr{H}_K}, \|\cdot\|_{\rho_X})$ onto $(\mathscr{H}_K, \|\cdot\|_K)$ (see [16]), we obtain

$$\begin{aligned}\|f_\lambda\|_K &= \left\|g_\lambda(L_K)L_K^{1/2}\varphi(L_K)L_K^{1/2}h_0\right\|_K \\ &\leq \left\|g_\lambda(L_K)L_K^{1/2}\varphi(L_K)\right\| \times \|h_0\|_{\rho_X} \\ &\leq \sup_{0<t\leq\kappa^2}\left|g_\lambda(t)t^{1/2}\varphi(t)\right| \times R \\ &\leq \beta_2\varphi(\lambda)\lambda^{\min\{\alpha(\mu_0-1/2),\,0\}-\mu_0}R,\end{aligned} \tag{2.20}$$

where the last inequality follows from (2.4).

By $|y| \leq M$ almost surely, there holds

$$\|\varsigma\|_K^2 = \langle (f_\lambda(x) - y)K_x, (f_\lambda(x) - y)K_x\rangle_K \leq \kappa^2\left(M + \kappa\beta_2\varphi(\lambda)\lambda^{\min\{\alpha(\mu_0-1/2),\,0\}-\mu_0}R\right)^2. \tag{2.21}$$

By (2.3) and $L_K f_\rho = L_K f_{\mathcal{H}}^+$, we get

$$\begin{aligned}
\|E\varsigma\|_K &= \left\|L_K\left(f_\rho - f_\lambda\right)\right\|_K = \left\|L_K\left(f_{\mathcal{H}}^+ - f_\lambda\right)\right\|_K \\
&= \left\|L_K\varphi(L_K)\left(I - g_\lambda(L_K)L_K\right)h_0\right\|_K \\
&\le \left\|\left(I - g_\lambda(L_K)L_K\right)\varphi(L_K)L_K^{1/2}\right\| \times \left\|L_K^{1/2}h_0\right\|_K \\
&\le \beta_1 R\lambda^{\min\{\mu_0,\, \nu_0-1/2\}(\alpha-1)+\alpha/2}\left(\varphi(\lambda)\right)^{\min\{(2\nu_0-1)/2\mu_0,\, 1\}}, \\
E\|\varsigma\|_K^2 &= E\left(y - f_\lambda(x)\right)^2 K(x,x) \\
&\le \kappa^2 E\left(y - f_\lambda(x)\right)^2 \\
&\le \kappa^2\left[\int_Z \left(y - f_\rho(x)\right)^2 d\rho + \int_X \left(f_\rho(x) - f_{\mathcal{H}}^+(x)\right)^2 d\rho_X + \int_X \left(f_{\mathcal{H}}^+(x) - f_\lambda(x)\right)^2 d\rho_X\right] \\
&\le \kappa^2\left(\alpha_1^2\lambda^{2(\alpha-1)\mu_0}\varphi^2(\lambda)R^2 + \int_Z \left(y - f_\rho(x)\right)^2 d\rho + \left\|f_\rho - f_{\mathcal{H}}^+\right\|_{\rho_X}^2\right),
\end{aligned} \tag{2.22}$$

where, in the last step, we used the result of Proposition 3.1 in Section 3. For simplicity, we write c_ρ^2 for $\int_Z (y - f_\rho(x))^2 d\rho + \|f_\rho - f_{\mathcal{H}}^+\|_{\rho_X}^2$. Applying Lemma 2.2, there holds

$$\begin{aligned}
\left\|\frac{1}{m}\sum_{i=1}^{m}\left[\varsigma(z_i) - E\varsigma\right]\right\|_K &\le 2\kappa\left(M + \kappa\beta_2 R\varphi(\lambda)\lambda^{-\mu_0+\min\{\alpha(\mu_0-1/2),\, 0\}}\right)\frac{\log(2/\delta)}{m} \\
&\quad + \kappa\left(\alpha_1\lambda^{(\alpha-1)\mu_0}\varphi(\lambda)R + c_\rho\right)\sqrt{\frac{2\log(2/\delta)}{m}}.
\end{aligned} \tag{2.23}$$

Then, we can use the following inequality to get the desired error bound,

$$\left\|S_{\mathbf{x}}^T y - T_{\mathbf{x}} f_\lambda\right\|_K \le \left\|\frac{1}{m}\sum_{i=1}^{m}\left[\varsigma(z_i) - E\varsigma\right]\right\|_K + \|E\varsigma\|_K. \tag{2.24}$$

This completes the proof of Lemma 2.4. □

3. Error Analysis

Proposition 3.1. *Let φ be an index function with $\mu_0 > 0$ covering φ and $\nu_0 > \max\{1/2, \mu_0\}$, so under the assumptions of* (1.21)*, there holds $\|f_\lambda - f_{\mathcal{H}}^+\|_{\rho_X} \le \alpha_1\lambda^{(\alpha-1)\mu_0}\varphi(\lambda)R$.*

Proof. From the definition of f_λ and $f_{\mathcal{H}}^+$, we have

$$f_\lambda - f_{\mathcal{H}}^+ = g_\lambda(L_K)L_K f_{\mathcal{H}}^+ - f_{\mathcal{H}}^+ = \left(g_\lambda(L_K)L_K - I\right)\varphi(L_K)h_0. \tag{3.1}$$

So the following error estimation holds

$$\begin{aligned}\|f_\lambda - f_{\mathscr{H}}^{+}\|_{\rho_X} &\le \|(g_\lambda(L_K)L_K - I)\varphi(L_K)\| \times \|h_0\|_{\rho_X}\\ &\le \sup_{0<\sigma\le\kappa^2} \left|(g_\lambda(\sigma)\sigma - 1)\varphi(\sigma)\right| \times \|h_0\|_{\rho_X}\\ &\le \alpha_1 \lambda^{(\alpha-1)\mu_0}\varphi(\lambda)R,\end{aligned} \tag{3.2}$$

where the last inequality follows from (2.2). □

Let us focus on the estimation of sample error.

Consider

$$\begin{aligned}\left\|f_z^\lambda - f_\lambda\right\|_{\rho_X} &= \left\|L_K^{1/2}(f_z^\lambda - f_\lambda)\right\|_K \le \left\|L_K^{1/2}(g_\lambda(T_{\mathbf{x}})T_{\mathbf{x}} - I)f_\lambda - (g_\lambda(T_{\mathbf{x}})T_{\mathbf{x}} - I)L_K^{1/2}f_\lambda\right\|_K\\ &\quad + \left\|(g_\lambda(T_{\mathbf{x}})T_{\mathbf{x}} - I)g_\lambda(L_K)L_K\varphi(L_K)L_K^{1/2}h_0\right\|_K + \left\|L_K^{1/2}g_\lambda(T_{\mathbf{x}})\left(S_{\mathbf{x}}^T y - T_{\mathbf{x}}f_\lambda\right)\right\|_K\\ &:= \|I_1\|_K + \|I_2\|_K + \|I_3\|_K.\end{aligned} \tag{3.3}$$

The idea is to separately bound each term in $\mathscr{H}_K$. We start dealing with the first term of (3.3).

Consider

$$\begin{aligned}I_1 &= \left(L_K^{1/2} - T_{\mathbf{x}}^{1/2}\right)(g_\lambda(T_{\mathbf{x}})T_{\mathbf{x}} - I)(\varphi(L_K) - \varphi(T_{\mathbf{x}}))g_\lambda(L_K)L_K^{1/2}L_K^{1/2}h_0\\ &\quad + \left(L_K^{1/2} - T_{\mathbf{x}}^{1/2}\right)(g_\lambda(T_{\mathbf{x}})T_{\mathbf{x}} - I)\varphi(T_{\mathbf{x}})g_\lambda(L_K)\mathrm{L}_K^{1/2}L_K^{1/2}h_0\\ &\quad + (g_\lambda(T_{\mathbf{x}})T_{\mathbf{x}} - I)\left(T_{\mathbf{x}}^{1/2} - L_K^{1/2}\right)\varphi(L_K)g_\lambda(L_K)L_K^{1/2}L_K^{1/2}h_0\\ &= J_1 + J_2 + J_3.\end{aligned} \tag{3.4}$$

According to (1.4) and (1.5), we derive the following bound:

$$\begin{aligned}\left\|g_\lambda(L_K)L_K^{1/2}\right\| &\le \sup_{0<\sigma\le\kappa^2}\left|g_\lambda(\sigma)\sigma^{1/2}\right| = \sup_{0<\sigma\le\kappa^2}\sqrt{|g_\lambda(\sigma)\sigma|} \times \sqrt{|g_\lambda(\sigma)|}\\ &\le \lambda^{-\alpha/2}\sqrt{DB}.\end{aligned} \tag{3.5}$$

Now, we are in the position to bound (3.4).

Suppose that $m \geq 2\log(4/\delta)$, then

$$\begin{aligned}
&\kappa^2\left(\frac{2\log(4/\delta)}{m}+\sqrt{\frac{2\log(4/\delta)}{m}}\right) \leq 2\kappa^2\sqrt{\frac{2\log(4/\delta)}{m}} := \zeta,\\
&2\kappa\left(M+\kappa\beta_2 R\varphi(\lambda)\lambda^{-\mu_0+\min\{\alpha(\mu_0-1/2),\,0\}}\right)\frac{\log(4/\delta)}{m}+\kappa\left(\alpha_1\lambda^{(\alpha-1)\mu_0}\varphi(\lambda)R+c_\rho\right)\sqrt{\frac{2\log(4/\delta)}{m}}\\
&\quad+\beta_1 R\lambda^{\min\{\mu_0,\nu_0-1/2\}(\alpha-1)+\alpha/2}(\varphi(\lambda))^{\min\{(2\nu_0-1)/2\mu_0,1\}}\\
&\quad\leq 2\kappa^2\beta_2 R\varphi(\lambda)\lambda^{-\mu_0+\min\{\alpha(\mu_0-1/2),\,0\}}\frac{\log(4/\delta)}{m}+\kappa\left(M+\alpha_1 R\varphi(\lambda)\lambda^{(\alpha-1)\mu_0}+c_\rho\right)\sqrt{\frac{2\log(4/\delta)}{m}}\\
&\quad+\beta_1 R\lambda^{\min\{\mu_0,\nu_0-1/2\}(\alpha-1)+\alpha/2}(\varphi(\lambda))^{\min\{(2\nu_0-1)/2\mu_0,1\}}\\
&\quad\leq C_0\Bigg(\varphi(\lambda)\lambda^{-\mu_0+\min\{\alpha(\mu_0-1/2),\,0\}}m^{-1}\log\frac{4}{\delta}+\left(1+\varphi(\lambda)\lambda^{(\alpha-1)\mu_0}\right)m^{-1/2}\sqrt{\log\frac{4}{\delta}}\\
&\quad+\lambda^{\min\{\mu_0,\nu_0-1/2\}(\alpha-1)+\alpha/2}(\varphi(\lambda))^{\min\{(2\nu_0-1)/2\mu_0,1\}}\Bigg)\\
&\quad:= C_0\eta.
\end{aligned} \tag{3.6}$$

By Lemmas 2.3 and 2.4, with confidence $1-\delta$, the following inequalities hold simultaneously:

$$\left\|L_K - S_{\mathbf{x}}^T S_{\mathbf{x}}\right\| \leq \zeta, \qquad \left\|S_{\mathbf{x}}^T y - T_{\mathbf{x}} f_\lambda\right\|_K \leq C_0\eta. \tag{3.7}$$

Combing (1.6), (3.5) together with the operator monotonicity property of $\varphi(t)$ and $t^{1/2}$, we obtain

$$\begin{aligned}
\|J_1\|_K &\leq \left\|L_K^{1/2}-T_{\mathbf{x}}^{1/2}\right\| \times \|g_\lambda(T_{\mathbf{x}})T_{\mathbf{x}}-I\| \times \|\varphi(L_K)-\varphi(T_{\mathbf{x}})\|\\
&\quad\times \left\|g_\lambda(L_K)L_K^{1/2}\right\| \times \left\|L_K^{1/2}h_0\right\|_K\\
&\leq c_\varphi\gamma\sqrt{DB}R\lambda^{-\alpha/2}\|L_K-T_{\mathbf{x}}\|^{1/2}\times\varphi(\|L_K-T_{\mathbf{x}}\|)\\
&\leq c_\varphi\gamma\sqrt{DB}R\lambda^{-\alpha/2}\zeta^{1/2}\varphi(\zeta).
\end{aligned} \tag{3.8}$$

By Lemma 2.1 and (3.5),

$$\begin{aligned}
\|J_2\|_K &\leq \left\|L_K^{1/2}-T_{\mathbf{x}}^{1/2}\right\| \times \|(g_\lambda(T_{\mathbf{x}})T_{\mathbf{x}}-I)\varphi(T_{\mathbf{x}})\| \times \left\|g_\lambda(L_K)L_K^{1/2}\right\| \times \left\|L_K^{1/2}h_0\right\|_K\\
&\leq R\sqrt{DB}\|L_K-T_{\mathbf{x}}\|^{1/2}\times\alpha_1\lambda^{(\alpha-1)\mu_0-\alpha/2}\varphi(\lambda)\\
&\leq \alpha_1 R\sqrt{DB}\zeta^{1/2}\varphi(\lambda)\lambda^{(\alpha-1)\mu_0-\alpha/2}.
\end{aligned} \tag{3.9}$$

For the purpose of bounding $\|J_3\|_K$, we rewritten J_3 as the following form:

$$\begin{aligned} J_3 = {} & \left(g_\lambda(T_{\mathbf{x}})T_{\mathbf{x}} - I\right)T_{\mathbf{x}}^{1/2}\left(\varphi(L_K) - \varphi(T_{\mathbf{x}})\right)g_\lambda(L_K)L_K^{1/2}L_K^{1/2}h_0 \\ & + \left(g_\lambda(T_{\mathbf{x}})T_{\mathbf{x}} - I\right)T_{\mathbf{x}}^{1/2}\varphi(T_{\mathbf{x}})g_\lambda(L_K)L_K^{1/2}L_K^{1/2}h_0 \\ & - \left(g_\lambda(T_{\mathbf{x}})T_{\mathbf{x}} - I\right)\left(\varphi(L_K) - \varphi(T_{\mathbf{x}})\right)g_\lambda(L_K)L_K L_K^{1/2}h_0 \\ & - \left(g_\lambda(T_{\mathbf{x}})T_{\mathbf{x}} - I\right)\varphi(T_{\mathbf{x}})g_\lambda(L_K)L_K L_K^{1/2}h_0. \end{aligned} \tag{3.10}$$

In the same way, we have that

$$\begin{aligned} \|J_3\|_K \leq {} & \gamma^{1-1/2v_0}\gamma_0^{1/2v_0}R\sqrt{DB}\varphi(\zeta) + \beta_1\lambda^{\min\{\mu_0,\nu_0-1/2\}(\alpha-1)}\left(\varphi(\lambda)\right)^{\min\{(2\nu_0-1)/2\mu_0,1\}}\sqrt{DB}R \\ & + \gamma RD\varphi(\zeta) + RD\alpha_1\varphi(\lambda)\lambda^{(\alpha-1)\mu_0}. \end{aligned} \tag{3.11}$$

Thus, we can get the bound for $\|I_1\|_K$ by combining (3.8), (3.9), and (3.11). What left is to estimate $\|I_2\|_K$ and $\|I_3\|_K$, we can employ the same way used in the estimation of $\|I_1\|_K$.

Consider

$$\begin{aligned} \|I_2\|_K \leq {} & \left\|\left(g_\lambda(T_{\mathbf{x}})T_{\mathbf{x}} - I\right)\left(\varphi(L_K) - \varphi(T_{\mathbf{x}})\right)g_\lambda(L_K)L_K L_K^{1/2}h_0\right\|_K \\ & + \left\|\left(g_\lambda(T_{\mathbf{x}})T_{\mathbf{x}} - I\right)\varphi(T_{\mathbf{x}})g_\lambda(L_K)L_K L_K^{1/2}h_0\right\|_K \\ \leq {} & \gamma RD\varphi(\zeta) + \alpha_1 RD\varphi(\lambda)\lambda^{(\alpha-1)\mu_0}, \end{aligned} \tag{3.12}$$

$$\begin{aligned} \|I_3\|_K \leq {} & \left\|\left(L_K^{1/2} - T_{\mathbf{x}}^{1/2}\right)g_\lambda(T_{\mathbf{x}})\left(S_{\mathbf{x}}^T y - T_{\mathbf{x}}f_\lambda\right)\right\|_K + \left\|T_{\mathbf{x}}^{1/2}g_\lambda(T_{\mathbf{x}})\left(S_{\mathbf{x}}^T y - T_{\mathbf{x}}f_\lambda\right)\right\|_K \\ \leq {} & C_0 B\lambda^{-\alpha}\zeta^{1/2}\eta + C_0\sqrt{DB}\eta\lambda^{-\alpha/2}. \end{aligned} \tag{3.13}$$

Lastly, combining (3.8) to (3.13) with Proposition 3.1, we have Theorem 1.5 holds.

4. Learning Rates

Significance of this paper lies in two facts; firstly, we generalize the definition of regularization and enrich the content of spectral regularization algorithms; secondly, analysis of this paper is able to undertake on the very general prior condition (1.21). Thus, our results can be applied to many different kinds of regularization, such as regularized least square learning, coefficient regularization learning, and (accelerate) landweber iteration and spectral cutoff. In this section, we will choose a suitable index function and apply Theorem 1.5 to some specific algorithms mentioned in Section 1.

4.1. Least Square Regularization

In this case, the regularization is $g_\lambda(\sigma) = 1/(\sigma + \lambda)$, $\lambda \in (0,1]$ with $B = D = \gamma = \gamma_0 = \alpha = 1$. The qualification of this algorithm is $\nu_0 = 1$. Suppose $\varphi(t) = t^r$ with $0 < r \leq 1$, that means $f_{\mathcal{H}}^+ = L_K^r h_0$, $h_0 \in L_{\rho_X}^2(X)$. Thus, we have that $\mu_0 = r$ covering $\varphi(t)$.

Using the result of Theorem 1.5, we obtain the following corollary.

Corollary 4.1. *Under the assumptions of Theorem 1.5, we have the following.*

(i) *For $0 < r \leq 1/2$, with confidence $1-\delta$, there holds*

$$\left\| f_z^{\lambda} - f_{\mathcal{H}}^{+} \right\|_{\rho_X} \leq O\left(\left(\lambda^{r} + m^{-r/2} + \lambda^{-1} m^{-3/4} \right) \left(1 + \lambda^{-1/2} m^{-1/4} \right) \left(\log \frac{4}{\delta} \right)^{5/4} \right). \tag{4.1}$$

By taking $\lambda = m^{-1/2}$, we have the following learning rate:

$$\left\| f_z^{\lambda} - f_{\mathcal{H}}^{+} \right\|_{\rho_X} \leq O\left(m^{-r/2} \left(\log \frac{4}{\delta} \right)^{5/4} \right). \tag{4.2}$$

(ii) *For $1/2 \leq r < 1$, with confidence $1-\delta$, there holds*

$$\left\| f_z^{\lambda} - f_{\mathcal{H}}^{+} \right\|_{\rho_X} \leq O\left(\left(\lambda^{1/2} + \lambda^{-1/2} m^{-1/2} + \lambda^{-1} m^{-3/4} + m^{-1/4} \right) \left(\log \frac{4}{\delta} \right)^{5/4} \right). \tag{4.3}$$

By taking $\lambda = m^{-1/2}$, we have the following learning rate:

$$\left\| f_z^{\lambda} - f_{\mathcal{H}}^{+} \right\|_{\rho_X} \leq O\left(m^{-1/4} \left(\log \frac{4}{\delta} \right)^{5/4} \right). \tag{4.4}$$

4.2. Coefficient Regularization with the l^2 Norm

In this case, the regularization is $g_\lambda(\sigma) = \sigma/(\sigma^2 + \lambda)$, $\lambda \in (0,1]$ with $B = D = \gamma = \gamma_0 = 1$, $\alpha = 1/2$. The qualification is $\nu_0 = 2$. We also consider the index function $\varphi(t) = t^r$ with $0 < r \leq 1$ and $\mu_0 = r$.

Corollary 4.2. *Under the assumptions of Theorem 1.5, we have the following.*

(i) *For $0 < r \leq 1/2$, with confidence $1-\delta$, there holds*

$$\left\| f_z^{\lambda} - f_{\mathcal{H}}^{+} \right\|_{\rho_X} \leq O\left(\left(1 + \lambda^{-1/4} m^{-1/4} \right) \left(\lambda^{r/2} + m^{-r/2} + m^{-1/2} \lambda^{-1/4} + m^{-1} \lambda^{(1/2)(r-1)} \right) \left(\log \frac{4}{\delta} \right)^{5/4} \right). \tag{4.5}$$

By taking $\lambda = m^{-1}$, we have the following learning rate:

$$\left\| f_z^{\lambda} - f_{\mathcal{H}}^{+} \right\|_{\rho_X} \leq O\left(m^{-r/2} \left(\log \frac{4}{\delta} \right)^{5/4} \right). \tag{4.6}$$

(ii) *For $1/2 \le r \le 1$, with confidence $1-\delta$, there holds*

$$\left\| f_z^\lambda - f_{\mathcal{H}}^+ \right\|_{\rho_X} \le O\left(\left(1+\lambda^{-1/4} m^{-1/4}\right)\left(\lambda^{r/2} + m^{-r/2} + m^{-1/2}\lambda^{-1/4}\right)\left(\log\frac{4}{\delta}\right)^{5/4} \right). \quad (4.7)$$

By taking $\lambda = m^{-2/(2r+1)}$, we have the following learning rate:

$$\left\| f_z^\lambda - f_{\mathcal{H}}^+ \right\|_{\rho_X} \le O\left(m^{-r/(2r+1)} \left(\log\frac{4}{\delta}\right)^{5/4} \right). \quad (4.8)$$

For coefficient regularization, the learning rates derived by Theorem 1.5 are almost the same, see Corollary 5.2 in [2]. For least square regularization, the learning rates in Corollary 4.1 are weak, the analysis in [8] by integral operator method gives learning rate $O(m^{-3r/4(1+r)})$ for $0 < r \le 1/2$; leave one out analysis in [11] gives the rate $O(m^{-r/(1+2r)})$.

Our analysis is influenced by both the prior condition and the regularization. Under the weaker prior condition (1.21), some techniques for error analysis in [1] are inapplicable; we take more complicated error decomposition and refined analysis to estimate error bounds and learning rates.

Acknowledgment

This work is supported by the Natural Science Foundation of China (Grant no. 11071276).

References

[1] F. Bauer, S. Pereverzev, and L. Rosasco, "On regularization algorithms in learning theory," *Journal of Complexity*, vol. 23, no. 1, pp. 52–72, 2007.
[2] H. Sun and Q. Wu, "Least square regression with indefinite kernels and coefficient regularization," *Applied and Computational Harmonic Analysis*, vol. 30, no. 1, pp. 96–109, 2011.
[3] L. Lo Gerfo, L. Rosasco, F. Odone, E. De Vito, and A. Verri, "Spectral algorithms for supervised learning," *Neural Computation*, vol. 20, no. 7, pp. 1873–1897, 2008.
[4] F. Cucker and S. Smale, "On the mathematical foundations of learning," *Bulletin of the American Mathematical Society*, vol. 39, no. 1, pp. 1–49, 2002.
[5] T. Evgeniou, M. Pontil, and T. Poggio, "Regularization networks and support vector machines," *Advances in Computational Mathematics*, vol. 13, no. 1, pp. 1–50, 2000.
[6] T. Poggio and S. Smale, "The mathematics of learning: dealing with data," *Notices of the American Mathematical Society*, vol. 50, no. 5, pp. 537–544, 2003.
[7] S. Smale and D.-X. Zhou, "Learning theory estimates via integral operators and their approximations," *Constructive Approximation*, vol. 26, no. 2, pp. 153–172, 2007.
[8] H. Sun and Q. Wu, "A note on application of integral operator in learning theory," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 416–421, 2009.
[9] H. Sun and Q. Wu, "Regularized least square regression with dependent samples," *Advances in Computational Mathematics*, vol. 32, no. 2, pp. 175–189, 2010.
[10] Q. Wu, Y. Ying, and D.-X. Zhou, "Learning rates of least-square regularized regression," *Foundations of Computational Mathematics*, vol. 6, no. 2, pp. 171–192, 2006.
[11] T. Zhang, "Leave-one-out bounds for kernel methods," *Neural Computation*, vol. 15, pp. 1397–1437, 2003.
[12] V. N. Vapnik, *Statistical Learning Theory*, A Wiley-Interscience, New York, NY, USA, 1998.
[13] Q. Guo and H. W. Sun, "Asymptotic convergence of coefficient regularization based on weakly depended samples," *Journal of Jinan University*, vol. 24, no. 1, pp. 99–103, 2010.

[14] Q. Wu and D.-X. Zhou, "Learning with sample dependent hypothesis spaces," *Computers & Mathematics with Applications*, vol. 56, no. 11, pp. 2896–2907, 2008.

[15] Y. Yao, L. Rosasco, and A. Caponnetto, "On early stopping in gradient descent learning," *Constructive Approximation*, vol. 26, no. 2, pp. 289–315, 2007.

[16] H. Sun and Q. Wu, "Application of integral operator for regularized least-square regression," *Mathematical and Computer Modelling*, vol. 49, no. 1-2, pp. 276–285, 2009.

[17] R. V. Kadison and J. R. Ringrose, *Fundamentals of the Theory of Operator Algebras*, vol. 1 of *Elementary Theory*, Academic Press, San Diego, Calif, USA, 1983.

[18] I. Pinelis, "Optimum bounds for the distributions of martingales in Banach spaces," *The Annals of Probability*, vol. 22, no. 4, pp. 1679–1706, 1994.