Radial Basis Function Regularization for Linear Inverse Problems with Random Noise

Carlos Valencia and Ming Yuan

H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology

(February 20, 2011)

Abstract

In this paper, we study the statistical properties of method of regularization with radial basis functions in the context of linear inverse problems. Radial basis function regularization is widely used in machine learning because of its demonstrated effectiveness in numerous applications and computational advantages. From a statistical viewpoint, one of the main advantages of radial basis function regularization in general and Gaussian radial basis function regularization in particular is their ability to adapt to varying degree of smoothness in a direct problem. We show here that similar approaches for inverse problems not only share such adaptivity to the smoothness of the signal but also can accommodate different degrees of ill-posedness. These results render further theoretical support to the superior performance observed empirically for radial basis function regularization.

Keywords: Inverse problem, minimax rate of convergence, radial basis function, regularization.

1 Introduction

Radial basis function regularization is one of the most popular tools in machine learning (see, e.g., Girosi, Jones, and Poggio (1993); Smola, Schölkopf, and Müller (1998); Wahba (1999); Evgeniou, Pontil, and Poggio (2000); Lin and Brown (2004); Zhang, Genton and Liu (2004); Lin and Yuan (2006); Shi and Yu (2006)). Let $\Phi(x) = \phi(\|x\|)$ for vector $x \in \mathbb{R}^d$ be a radial basis function where $\phi : [0, +\infty) \to \mathbb{R}$ is a univariate function. Typical examples include $\phi(r) = r^{2m}\log(r)$ (thin plate spline), $\phi(r) = e^{-\varrho r^2/2}$ (Gaussian), and $\phi(r) = (c^2 + r^2)^{1/2}$ (multiquadrics) among others. When $K_\Phi(x, y) = \Phi(x - y)$ is (conditionally) positive definite in that for any $n \in \mathbb{Z}$ and any distinct $x_1, ..., x_n \in \mathbb{R}^d$,

$$\sum_{j=1}^{n}\sum_{k=1}^{n} a_j a_k K(x_j, x_k) > 0,$$

Φ can be identified with a reproducing kernel Hilbert space (Aronszajn (1950)), denoted by $\mathcal{H}_\Phi$. The squared norm in $\mathcal{H}_\Phi$ can be written as

$$J(f) = (2\pi)^{-d/2} \int_{R^d} |\tilde{f}(\omega)|^2 / \tilde{\Phi}(\omega) d\omega$$

for any function $f \in \mathcal{H}_\Phi$, where $\tilde{f}$ stands for the Fourier transform of f, that is,

$$\tilde{f}(\omega) = (2\pi)^{-d/2} \int_{R^d} f(x) e^{-ix^T\omega} dx.$$

The method of regularization with a radial basis function estimates a functional parameter by the solution to

$$\min_{f \in \mathcal{H}_\Phi} \{L(f, \text{data}) + \lambda J(f)\},$$

where L is the empirical loss, often taken to be the negative log-likelihood. The tuning parameter $\lambda > 0$ controls the trade-off between minimizing the empirical loss and obtaining a smooth solution.

Consider in particular estimating a periodic function $f_0 : [-\pi, \pi] \to \mathbb{R}$ based on noisy observations of Af where A is a bounded linear operator, i.e.,

$$dY(t) = (Af_0)(t)dt + \epsilon dW(t), \qquad t \in [-\pi, \pi]. \tag{1}$$

Here $\epsilon > 0$ is the noise level and $W(t)$ is a standard Brownian motion on $[-\pi, \pi]$. The white noise model (1) connects to a number of common statistical problems in the light of results on its equivalence to nonparametric regression (Brown and Low, 1996), density estimation (Nussbaum, 1996), spectral density estimation (Golubev and Nussbaum, 1998), and nonparametric generalized regression (Grama and Nussbaum, 1997). The radial basis function regularization in this case gives the following estimate of f_0:

$$\hat{f}_\lambda = \underset{f \in \mathcal{H}_\Phi}{\arg\min} \left\{ \|Y - Af\|^2_{\mathcal{L}_2} + \lambda J(f) \right\}.$$

Lin and Brown (2004) and Lin and Yuan (2006) recently studied the statistical properties of $\hat{f}_\lambda$ in a special case when A is the identity operator. They found that when f is a member of any finite-order Sobolev spaces, the method of regularization with many radial basis functions is rate optimal when the tuning parameter is appropriately chosen, which partially explains the success of such methods in this particular setting. Of course in many applications, A is not an identity operator but rather a general compact operator. Problems of this type can be found in almost all areas of science and engineering (see, e.g., Chalmond (2008); Kaipo and Somersalo (2004); Ramm (2009)). These problems, commonly referred to as inverse problems, are often ill-posed and therefore, fundamentally more difficult than the case when A is the identity, often referred to as direct problems (see, e.g., Cavalier (2008)). In this paper, we study the statistical properties of radial basis function regularization estimator $\hat{f}_\lambda$ in this setting.

Similar to direct problems, the difficulty in estimating f_0 in an inverse problem is determined by the complexity of the functional class it belongs to. Differing from direct problems, in an inverse problem, the difficulty of estimating f_0 also depends on the degree of ill-posedness of the linear operator A. We consider a variety of combinations of functional classes and linear operators and show that for many common choices of radial basis functions, $\hat{f}_\lambda$ is rate optimal whenever λ is appropriately tuned. Our results suggest that the superior statistical properties established earlier for the direct problems continue to hold in the inverse problems and therefore further make clear why the radial basis function regularization is so effective in a wider range of applications.

The rest of this article is organized as follows. In the next section, we describe in more details the parameter spaces and the ill-posedness of the problem. We study in Section 3 the statistical properties of radial basis function regularization. All proofs are relegated to Section 6. Section 4 reports results from numerical experiments to illustrate the implications of our theoretical development.

2 Radial Basis Function Regularization in Linear Inverse Problems

The white noise model (1) can be expressed in terms of the corresponding Fourier coefficients and leads to a sequence model that is often more amenable to statistical analysis (see, e.g., Johnstone, 1998).

2.1 Sequence model via singular value decomposition

Let A^* be the adjoint operator of A. Because of the compactness of A, A^*A admits spectral decomposition

$$A^*Af = \sum_{k=1}^{\infty} b_k^2 \langle f, \varphi_k \rangle_{\mathcal{L}_2} \varphi_k \tag{2}$$

for any square integrable periodic function f, where the eigenfunctions $\{\varphi_1, \varphi_2, \ldots\}$ constitute an orthornormal basis of $\mathcal{L}_2$, the collection of square integrable periodic functions, and the eigenvalues $\{b_1^2, b_2^2, \ldots\}$ are arranged in a non-increasing order without loss of generality. Denote by ψ_k the normalized image of φ_k, that is, $A\varphi_k = b_k \psi_k$. It is easy to show that

$$A^*\psi_k = b_k \varphi_k.$$

From the singular value decomposition, we can convert the linear inverse problem (1) into a sequence model. More specifically,

$$y_k := \langle Y, \psi_k \rangle_{\mathcal{L}_2} = \langle Af_0, \psi_k \rangle_{\mathcal{L}_2} + \langle \epsilon W, \psi_k \rangle_{\mathcal{L}_2} = b_k \langle f_0, \varphi_k \rangle_{\mathcal{L}_2} + \epsilon \langle W, \psi_k \rangle_{\mathcal{L}_2} =: b_k \theta_k + \epsilon \xi_k$$

for $k = 1, 2, \ldots$.

Unlike the direct problem where all singular values are one, in an inverse problem, $b_k \to 0$ as $k \to \infty$. The vanishing singular values poses challenges in inverting the linear operator A and makes the problem ill-posed. As a result, the estimation of f_0 becomes fundamentally more difficult for an inverse problem than for a direct problem. The rate of decay of $\{b_k : k \geq 1\}$ quantifies the ill-posedness. Typically, an inverse problem is called mildly ill-posed if $b_k \sim k^{-\beta}$ and severe ill-posed if $b_k \sim \exp(-\beta k)$ for some parameter $\beta > 0$ often referred to as the degree of ill-posedness. Hereafter, $a_k \sim b_k$ means that both a_k/b_k and b_k/a_k are bounded away from zero.

2.2 Parameter spaces

In addition to the ill-posedness, the difficulty of estimating f_0 in (1) is also determined by the parameter space for the functional parameter. It is often convenient to describe the parameters space using the Fourier coefficient with respect to the basis $\{\varphi_k : k \geq 1\}$. Typically, f_0 belongs to the functional class corresponding to an ellipsoid Θ in the space of Fourier coefficients $\{\theta_k : k \geq 1\}$:

$$\Theta = \left\{ (\theta_k : k \geq 1) : \sum_{k\geq 1} a_k^2 \theta_k^2 \leq Q \right\}, \tag{3}$$

for a non-deceasing sequence $0 \leq a_1 \leq a_2 \leq \ldots$ such that $a_k \to \infty$ as $k \to \infty$, and a positive constant Q.

It is instructive to consider the case when $\{\varphi_k : k \geq 1\}$ is the usual trigonometric basis, that is, $\varphi_1(t) = (2\pi)^{-1/2}$, $\varphi_{2l}(t) = \pi^{-1/2}\sin(lt)$ and $\varphi_{2l+1}(t) = \pi^{-1/2}\cos(lt)$ for $l \geq 1$. In this case, the usual Sobolev spaces are perhaps the most popular examples of Θ. Let $\mathcal{S}^m(Q)$ be the mth order Sobolev space of periodic functions on $[-\pi, \pi]$, that is,

$$\mathcal{S}^m(Q) = \left\{ f \in \mathcal{L}_2 : f \text{ is } 2\pi\text{−periodic, and } \int_{-\pi}^{\pi} f^2 + (f^{(m)})^2 \leq Q \right\}.$$

Simple calculation shows that $\mathcal{S}^m(Q)$ can also be equivalently expressed as

$$\mathcal{S}^m(Q) = \left\{ f \in \mathcal{L}_2 : f = \sum_{k\geq 1} \theta_k \varphi_k, \sum_{k\geq 1} a_k^2 \theta_k^2 \leq Q, a_1 = 1, a_{2l} = a_{2l+1} = k^{2m} + 1 \right\}.$$

In the same spirit, analytic functions or sometimes referred to as infinit-order Sobolev space can be described as

$$\mathcal{S}^\infty(\alpha; Q) = \left\{ f \in \mathcal{L}_2 : f = \sum_{k\geq 1} \theta_k \varphi_k, \sum_{k\geq 1} a_k^2 \theta_k^2 \leq Q, a_1 = 1, a_{2l} = a_{2l+1} = e^{\alpha l} \right\}.$$

See Johnstone (1998) for details.

Appealing to this connection, in what follows, we shall write

$$\Theta^\alpha(Q) = \left\{ (\theta_k : k \geq 1) : \sum_{k\geq 1} a_k^2 \theta_k^2 \leq Q, a_1 = 1, a_{2l} = a_{2l+1} = k^\alpha + 1 \right\}$$

as Sobolev type of spaces of order α; and

$$\Theta^\infty(\alpha; Q) = \left\{ (\theta_k : k \geq 1) : \sum_{k\geq 1} a_k^2 \theta_k^2 \leq Q, a_1 = 1, a_{2l} = a_{2l+1} = e^{\alpha k} \right\}$$

to represent spaces similar to $\mathcal{S}^\infty$.

2.3 Radial basis function regularization

We now describe the radial basis functions and the reproducing kernel Hilbert spaces they induce. Because we focus here on periodic functions, it is natural to consider periodized radial basis functions

$$\Phi_0(r) = \sum_{k\in\mathbb{Z}} \Phi(r - 2\pi k),$$

where Φ is a radial basis function. See Smola, Schölkopf and Müller (1998), Lin and Brown (2004) among others for further discussion of periodized radial basis functions and their applications in machine learning. As shown in Lin and Yuan (2006), Φ_0 (or equivalently K_{Φ_0}) is positive definite so long as Φ is positive definite and furthermore the norm of $\mathcal{H}_{\Phi_0}$ can be given by

$$\|f\|^2_{\mathcal{H}_{\Phi_0}} = \sum_{k\geq 1} \gamma_k \theta_k^2,$$

where θ_ks are the Fourier coefficients of f, and $\gamma_1 = (2\pi)^{-1/2}\{\tilde{\Phi}(0)\}^{-1}$, $\gamma_{2l} = \gamma_{2l+1} = (2\pi)^{-1/2}\{\tilde{\Phi}(l)\}^{-1}$, $l = 1, 2, \ldots$. When $\{\varphi_k : k \geq 1\}$ is taken to be the classical trigonometric basis, the method of regularization with radial basis function Φ_0 can be equivalently expressed in terms of the sequence of Fourier coefficients:

$$\hat{f}_\lambda = \underset{f=\sum_{k\geq 0}\theta_k\varphi_k \in \mathcal{H}_{\Phi_0}}{\arg\min} \left\{ \sum_{k\geq 1} (y_k - b_k\theta_k)^2 + \lambda \sum_{k\geq 1} \gamma_k \theta_k^2 \right\}.$$

Consider, for example, the periodic Gaussian kernel

$$G_0(r) = \sum_{k\in\mathbb{Z}} G(r - 2\pi k),$$

where

$$G(r) = \frac{1}{\sqrt{2\pi\varrho^2}} \exp\left(-\frac{r^2}{2\varrho^2}\right)$$

for some parameter $\varrho > 0$. Simple calculation yields that $\gamma_{2l} = \gamma_{2l+1} = e^{l^2\varrho^2/2}$. Other popular examples include periodic multiquadratics and Wendland kernels (Wendland (1998)) that corresponds to $\gamma_{2l} = \gamma_{2l+1} = e^{l\varrho}$ and $\gamma_{2l} = \gamma_{2l+1} = k^{\varrho}$ respectively. There are also other common choices of radial basis functions for which γ_k behaves similarly to these three examples. See Buhlmann (2003) for further details.

3 Main Results

Following the discussion before, we shall focus on the following sequence model hereafter:

$$y_k = b_k\theta_k + \epsilon\xi_k, \qquad k = 1, 2, \ldots. \tag{4}$$

The inverse problem under investigation is either mildly or severely ill-posed, that is, $b_k \sim k^{-\beta}$ or $b_k \sim e^{-\beta k}$ respectively. We shall also consider Sobolev type of parameter spaces, that is, $(\theta_k : k \geq 1) \in \Theta^\alpha$ for some $\alpha > 1/2$ or $\Theta^\infty(\alpha, Q)$. Our primary interest is to evaluate the statistical performance of radial basis function regularization:

$$(\hat{\theta}_{k\lambda} : k \geq 1) = \underset{(\eta_k : k\geq 1)}{\arg\min} \left\{ \sum_{k\geq 1} (y_k - b_k\eta_k)^2 + \lambda \sum_{k\geq 1} \gamma_k \eta_k^2 \right\}. \tag{5}$$

In particular, we consider three different types of radial basis functions: (1) $\gamma_k \sim e^{\gamma k^2}$ for some $\gamma > 0$ with periodic Gaussian kernel as a typical example; (2) $\gamma_k \sim e^{\gamma k}$ with periodic multiquadrics kernel as a typical example; and (3) $\gamma_k \sim k^\gamma$ with periodic Wendland kernel or the usual spline kernels (see, e.g., Wahba (1990)) as typical examples.

We begin with Gaussian type of kernel, that is, $\gamma_k \sim e^{\gamma k^2}$ for some $\gamma > 0$.

Theorem 1 *Assume that $\gamma_k \sim e^{\gamma k^2}$ for some $\gamma > 0$.*

(a) (Mildly ill-posed with Sobolev spaces) If $b_k \sim k^{-\beta}$ and

$$\lambda \sim \exp\left(-\epsilon^{\frac{4}{2\alpha+2\beta+1}}\right),$$

then

$$\sup_{(\theta_k : k\geq 1) \in \Theta^\alpha(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 \sim \epsilon^{\frac{4\alpha}{2\alpha+2\beta+1}}.$$

(b) (Mildly ill-posed with analytic functions) If $b_k \sim k^{-\beta}$ and

$$\lambda \sim \exp\left(-\frac{\gamma}{4\alpha^2}\left(\log\frac{1}{\epsilon^2}\right)^2\right),$$

then

$$\sup_{(\theta_k : k\geq 1) \in \Theta^\infty(\alpha, Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 \sim \epsilon^2\left(\log\frac{1}{\epsilon^2}\right)^{2\beta+1}.$$

(c) (Severely ill-posed with Sobolev spaces) If $b_k \sim k^{-\beta}$ and

$$\lambda \sim \exp\left(-\left(\log\frac{1}{\epsilon^2}\right)^2\right),$$

then

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\alpha(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda}-\theta_k\right)^2 \sim \left(\log\frac{1}{\epsilon}\right)^{-2\alpha}.$$

(d) (Severely ill-posed with analytic functions) If $b_k \sim e^{-\beta k}$ and

$$\lambda \sim \exp\left(-\frac{\gamma}{(2\alpha+2\beta)^2}\left(\log\frac{1}{\epsilon^2}\right)^2\right),$$

then

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\infty(\alpha,Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda}-\theta_k\right)^2 \sim \epsilon^{\frac{2\alpha}{\alpha+\beta}}.$$

We note that all the rates obtained in Theorem 1 are minimax optimal (see, e.g., Cavalier (2008)). In other words, when the tuning parameter λ is appropriately chosen, Gaussian radial basis function regularization is rate optimal for all combinations of ill-posedness as well as parameter spaces. This result, together with similar results for direct problems (Lin and Brown (2004)), partly explain its success in numerous applications.

Next we consider the case with a multiquadrics type of kernel.

Theorem 2 *Assume that $\gamma_k \sim e^{\gamma k}$ for some $\gamma > 0$.*

(a) (Mildly ill-posed with Sobolev spaces) If $b_k \sim k^{-\beta}$ and

$$\lambda \sim \exp\left(-\epsilon^{-\frac{2}{2\alpha+2\beta+1}}\right),$$

then

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\alpha(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda}-\theta_k\right)^2 \sim \epsilon^{\frac{4\alpha}{2\alpha+2\beta+1}}.$$

(b) (Mildly ill-posed with analytic functions) If $b_k \sim k^{-\beta}$, then

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\infty(\alpha,Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda}-\theta_k\right)^2 \sim \epsilon^2\left(\log\frac{1}{\epsilon^2}\right)^{2\beta+1}.$$

provided that

$$\lambda \sim \begin{cases} \epsilon^{\frac{\gamma}{\alpha}} & \gamma > \alpha - 2\beta \\ \epsilon & \gamma \leq \alpha - 2\beta \end{cases}.$$

(c) *(Severely ill-posed with Sobolev spaces) If* $b_k \sim k^{-\beta}$ *and* $\lambda \sim \epsilon^2$*, then*

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\alpha(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 \sim \left(\log \frac{1}{\epsilon}\right)^{-2\alpha}.$$

(d) *(Severely ill-posed with analytic functions) Suppose that* $b_k \sim e^{-\beta k}$*. If* $\gamma > \alpha - 2\beta$ *and*

$$\lambda \sim \epsilon^{-\frac{\beta+\gamma}{\alpha+\beta}},$$

then

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\infty(\alpha,Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 \sim \epsilon^{\frac{2\alpha}{\alpha+\beta}}.$$

If $\gamma \leq \alpha - 2\beta$*, then the best achievable rate is*

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\infty(\alpha,Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 \sim \epsilon^{\frac{4\beta+2\gamma}{3\beta+\gamma}},$$

and it is attained when

$$\lambda \sim \epsilon^{\frac{2\beta+\gamma}{3\beta+\gamma}}.$$

From Theorem 2, regularization with multiquadrics type of kernel is also rate optimal for finite-order Sobolev spaces. For analytic functions, however, its behavior is more complex. When the inverse problem is mildly ill-posed, it can still achieve the optimal rate but different tuning parameters are needed to attain the optimal rate depending on whether γ is larger than $\alpha - 2\beta$. However, for severely ill-posed problems, the minimax optimal rate can only be achieved when $\gamma > \alpha - 2\beta$. The transition point $\alpha - 2\beta$ is somewhat surprising. Observe that $\mathcal{H}_{\Phi_0} \subseteq \mathcal{S}^\infty(\alpha, Q)$ if $\gamma \geq \alpha$ and $\mathcal{S}^\infty(\alpha, Q) \subset \mathcal{H}_{\Phi_0}$ otherwise. Thus Theorem 2 essentially states that regularization with multiquadrics type of kernel is always rate optimal if the reproducing kernel Hilbert space induced by the radial basis function is smaller than the parameter space. But even when the parameter space is larger than the induced space, that is, $\gamma < \alpha$, it is still capable of achieving the minimax optimal rate so long as $\gamma > \alpha - 2\beta$.

Now consider the Wendland/spline type of kernel.

Theorem 3 *Assume that* $\gamma_k \sim k^\gamma$ *for some* $\gamma > 1/2$*.*

(a) *(Mildly ill-posed with Sobolev spaces) Suppose that* $b_k \sim k^{-\beta}$*. If* $\gamma > \alpha - 2\beta$ *and*

$$\lambda \sim \epsilon^{\frac{4\alpha}{2\alpha+2\beta+1}},$$

then

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\alpha(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 \sim \epsilon^{\frac{4\alpha}{2\alpha+2\beta+1}}.$$

If $\gamma \leq \alpha - 2\beta$, *the best achivable rate is*

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\alpha(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 \sim \epsilon^{\frac{2(4\beta+2\gamma)}{6\beta+2\gamma+1}},$$

and it is attained when

$$\lambda \sim \epsilon^{\frac{4\beta+2\gamma}{6\beta+2\gamma+1}}.$$

(b) *(Mildly ill-posed with analytic functions) Suppose the* $b_k \sim k^{-\beta}$. *If* $\gamma > \alpha - 2\beta$ *and*

$$\lambda \sim \epsilon^2 \left(\log \frac{1}{\epsilon}\right)^{-2\beta-\gamma},$$

then

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\alpha(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 \sim \epsilon^2 \left(\log \frac{1}{\epsilon}\right)^{2\beta+1}.$$

If $\gamma \leq \alpha - 2\beta$, *the best achievable rate is*

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\alpha(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 \sim \epsilon^{\frac{2(4\beta+2\gamma)}{6\beta+2\gamma+1}},$$

and it is attained when

$$\lambda \sim \epsilon^{\frac{4\beta+2\gamma}{6\beta+2\gamma+1}}.$$

(c) *(Severely ill-posed with Sobolev spaces) If* $b_k \sim e^{-\beta k}$

$$\lambda \sim \epsilon^2$$

then

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\alpha(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 \sim \left(\log \frac{1}{\epsilon}\right)^{-2\alpha}$$

(d) *(Severely ill-posed with analytic functions) Suppose* $b_k \sim e^{-\beta k}$. *If* $\gamma > \alpha - 2\beta$, *then the achievable rate is*

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\alpha(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 \sim \epsilon^{\frac{2\beta}{\alpha+2\beta}},$$

and it is attained when

$$\lambda \sim \epsilon^{\frac{4\beta}{\alpha+2\beta}}.$$

When $\gamma \le \alpha - 2\beta$, the best achievable rate is

$$\sup_{(\theta_k:k\ge 1)\in\Theta^\alpha(Q)} \sum_{k\ge 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 \sim \epsilon^{\frac{4}{3}},$$

and it is attained when $\lambda \sim \epsilon^{\frac{2}{3}}$.

Theorem 3 indicates that the method of regularization with Wendland or spline type of kernel is also capable of attaining the minimax optimal rate but only so if γ is sufficiently large, or equivalently, the reproducing kernel Hilbert space $\mathcal{H}_{\Phi_0}$ is sufficiently small.

Our main results are summarized in Table 1.

4 Numerical Experiments

To illustrate the performance of the radial basis function regularization estimates, we carried out some numerical experiments. The main purpose is to demonstrate the actual convergence rates when the noise level ϵ goes to zero.

All the simulations are made in the domain of the coefficients for the trigonometric basis $\{\psi_k : k \ge 1\}$ of $L^2[-\pi, \pi]$, implying that all the parameters are generated as sequences in ℓ^2. We consider in particular, two functions $f = \sum_{k\ge 1} \theta_k \psi_k$ where

$$\theta_k = k^{-2} \qquad \text{or} \qquad \exp(-2k),$$

representing Sobolev type or analytic type of functions respectively. We also consider two operators A corresponding to mildly or severly ill-posed situations respectively:

$$b_k = k^{-2} \qquad \text{or} \qquad \exp(-2k).$$

We also chose $\gamma = 2$ for all three types of kernels under consideration to ensure that $\gamma > \alpha - 2\beta$ in each possible sccenario.

To understand the asymptotic behavior of the regularized estimator, we consider a set of values for the noise level as $\epsilon = j/100$ for $j = 1, 2, \cdots, 15$. In each case we estimate the parameter $(\theta_k : k \ge 1)$ using (6) and calculate the integrated squared error by $\|\hat{\theta}_\lambda - \theta\|_{\ell^2}$. We performed 100 replications for each setting to obtain a fair approximation of the expected risk. As usual in nonparametric estimators, the tuning parameter should be selected in order to minimize the risk. To do so, in each setting we calculate $(\hat{\theta}_k : k \ge 1)$ for each $\lambda \in \Lambda$,

Kernel Type		Θ^α		Θ^∞	
		Mildly Ill-posed	Severely Ill-posed	Mildly Ill-posed	Severely Ill-posed
Gaussian	λ	$\exp\left(-\epsilon^{-\frac{4}{2\alpha+2\beta+1}}\right)$	$\exp\left(-\left(\log\frac{1}{\epsilon^2}\right)^2\right)$	$\exp\left(-\frac{\gamma}{4\alpha^2}\left(\log\frac{1}{\epsilon^2}\right)^2\right)$	$\exp\left(-\frac{\gamma}{(2\beta+2\alpha)^2}\left(\log\frac{1}{\epsilon^2}\right)^2\right)$
	$\mathcal{R}(\lambda)$	$\epsilon^{\frac{4\alpha}{2\alpha+2\beta+1}}$	$\left(\log\frac{1}{\epsilon}\right)^{-2\alpha}$	$\epsilon^2\left(\log\frac{1}{\epsilon^2}\right)^{2\beta+1}$	$\epsilon^{\frac{2\alpha}{\alpha+\beta}}$
Multiquadrics	λ	$\exp\left(-\epsilon^{-\frac{2}{2\alpha+1}}\right)$	ϵ^2	$\begin{cases}\epsilon^{\frac{\gamma}{\alpha}} & \text{if } \gamma > \alpha-2\beta \\ \epsilon & \text{if } \gamma \le \alpha-2\beta\end{cases}$	$\begin{cases}\epsilon^{-\frac{\beta+\gamma}{\alpha+\beta}} & \text{if } \gamma > \alpha-2\beta \\ \epsilon^{\frac{2\beta+\gamma}{3\beta+\gamma}} & \text{if } \gamma \le \alpha-2\beta\end{cases}$
	$\mathcal{R}(\lambda)$	$\epsilon^{\frac{4\alpha}{2\alpha+2\beta+1}}$	$\left(\log\frac{1}{\epsilon}\right)^{-2\alpha}$	$\epsilon^2\left(\log\frac{1}{\epsilon^2}\right)^{2\beta+1}$	$\begin{cases}\epsilon^{\frac{2\alpha}{\alpha+\beta}} & \text{if } \gamma > \alpha-2\beta \\ \epsilon^{\frac{4\beta+2\gamma}{3\beta+\gamma}} & \text{if } \gamma \le \alpha-2\beta\end{cases}$
Wendland	λ	$\begin{cases}\epsilon^{\frac{4\beta+2\gamma}{2\alpha+1\beta+1}} & \text{if } \gamma > \alpha-2\beta \\ \epsilon^{\frac{4\beta+2\gamma}{6\beta+2\gamma+1}} & \text{if } \gamma \le \alpha-2\beta\end{cases}$	ϵ^2	$\begin{cases}\left(\log\frac{1}{\epsilon}\right)^{-2\beta-\gamma} & \text{if } \gamma > \alpha-2\beta \\ \epsilon^{\frac{4\beta+2\gamma}{6\beta+2\gamma+1}} & \text{if } \gamma \le \alpha-2\beta\end{cases}$	$\begin{cases}\epsilon^{\frac{4\beta}{\alpha+2\beta}} & \text{if } \gamma > \alpha-2\beta \\ \epsilon^{\frac{2}{3}} & \text{if } \gamma \le \alpha-2\beta\end{cases}$
or Spline	$\mathcal{R}(\lambda)$	$\begin{cases}\epsilon^{\frac{4\alpha}{2\alpha+2\beta+1}} & \text{if } \gamma > \alpha-2\beta \\ \epsilon^{\frac{4\gamma+8\beta}{6\beta+2\gamma+1}} & \text{if } \gamma \le \alpha-2\beta\end{cases}$	$\left(\log\frac{1}{\epsilon}\right)^{-2\alpha}$	$\begin{cases}\epsilon^2\left(\log\frac{1}{\epsilon}\right)^{2\beta+1} & \text{if } \gamma > \alpha-2\beta \\ \epsilon^{\frac{8\beta+4\gamma}{6\beta+2\gamma+1}} & \text{if } \gamma \le \alpha-2\beta\end{cases}$	$\begin{cases}\epsilon^{\frac{2\alpha}{\alpha+2\beta}} & \text{if } \gamma > \alpha-2\beta \\ \epsilon^{\frac{4}{3}} & \text{if } \gamma \le \alpha-2\beta\end{cases}$

Table 1: We list here, for different combinations of parameter spaces and radial basis functions, the best achievable convergence rates of $\mathcal{R}(\lambda) = \sup_{(\theta_k:k\ge1)\in\Theta^\alpha(Q)}\sum_{k\ge1}\mathbb{E}\left(\hat{\theta}_{k\lambda}-\theta_k\right)^2$ and the order of the tuning parameter λ needed to attain the rate. α reflects the smoothness of the parameter space, β determines the ill-posedness of the inverse problem and γ depends on the choice of the radial basis function.

where $\Lambda = \{\lambda_i : \lambda_i = \exp(-i/5), i = 1, \cdots, 100\}$, and select the estimator $\hat{\theta}_{\lambda^*}$ such that the risk $\|\hat{\theta}_\lambda - \theta\|_{\ell^2}$ is minimized.

The results are presented in Figure 1. In each plot, we include also the minimax optimal rate adjusted by a constant. As can be seen, the simulated rates have a similar decay as the theoretical minimax counterparts, indicating the estimate is rate optimal.

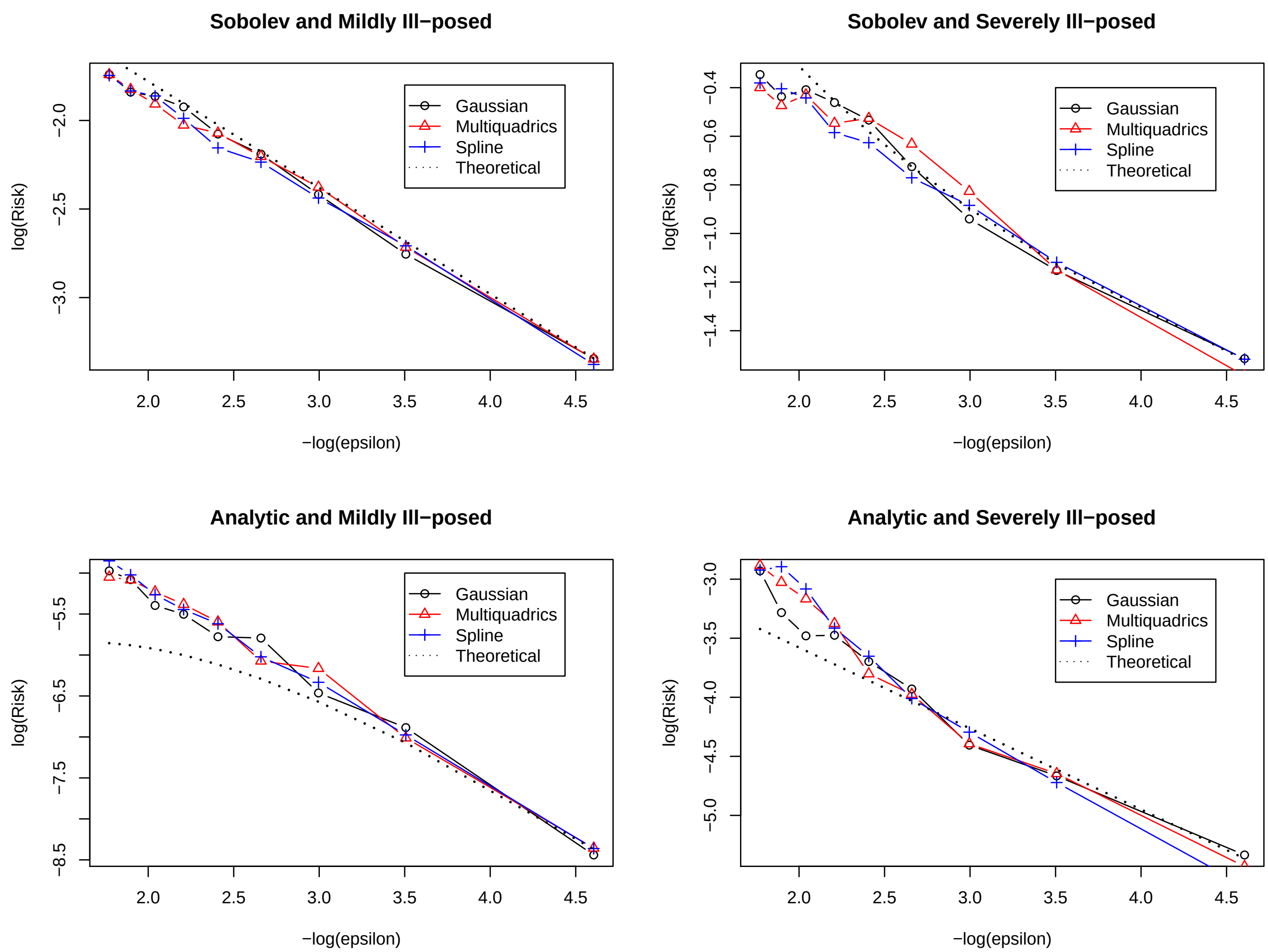

Figure 1: Comparison of the risk of radial basis function regularization. The results are averaged over 100 replications.

5 Risk Analysis of Radial Basis Function Regularization

We now set out to establish the results presented in the previous section. Recall that the regularization estimator $(\hat{\theta}_{k\lambda} : k \geq 1)$ is defined as

$$(\hat{\theta}_{k\lambda} : k \geq 1) = \underset{(\eta_k : k\geq 1)}{\arg\min} \left\{ \sum_{k\geq 1} (y_k - b_k\eta_k)^2 + \lambda \sum_{k\geq 1} \gamma_k \eta_k^2 \right\}.$$

It can be written explicitly as

$$\hat{\theta}_{k\lambda} = \frac{b_k}{b_k^2 + \lambda\gamma_k^{-1}} y_k \, , k = 1, 2, \cdots . \tag{6}$$

In particular, here we consider

$$b_k \sim \begin{cases} k^{-\beta} & \text{Mildly ill} - \text{posed} \\ \exp(-\beta k) & \text{Severely ill} - \text{posed} \end{cases} .$$

Furthermore, the true Fourier coefficients $(\theta_k : k \geq 1)$ are assumed to be in an ellipsiod

$$\Theta(Q) = \left\{ (\theta_k : k \geq 1) : \sum_{k\geq 1} a_k^2 \theta_k^2 \leq Q \right\},$$

where

$$a_k \sim \begin{cases} k^{\alpha} & \Theta = \Theta^{\alpha}(Q) \\ \exp(\alpha k) & \Theta = \Theta^{\infty}(\alpha; Q) \end{cases} .$$

Observe that the risk of the radial basis function regularization estimator $(\hat{\theta}_{k\lambda} : k \geq 1)$ can be decomposed as the sum of the squared bias and the variance:

$$\sum_{k\geq 1} \mathbb{E} \left(\hat{\theta}_{k\lambda} - \theta_k \right)^2 = \sum_{k\geq 1} \left(\mathbb{E}\hat{\theta}_{k\lambda} - \theta_k \right)^2 + \sum_{k\geq 1} \mathrm{Var} \left(\hat{\theta}_{k\lambda} \right) =: B_\theta^2 \left(\hat{\theta}_\lambda \right) + \mathrm{Var}_\theta \left(\hat{\theta}_\lambda \right). \tag{7}$$

By (6), we can further write

$$B_\theta^2 \left(\hat{\theta}_\lambda \right) = \sum_{k\geq 1} \frac{\lambda^2 \gamma_k^{-2} \theta_k^2}{\left(b_k^2 + \lambda\gamma_k^{-1} \right)^2}$$

and

$$\mathrm{Var}_\theta \left(\hat{\theta}_\lambda \right) = \epsilon^2 \sum_{k\geq 1} \frac{b_k^2}{\left(b_k^2 + \lambda\gamma_k^{-1} \right)^2}.$$

The squared bias and variance can be further bounded as follows:

$$B_\theta^2\left(\hat{\theta}_\lambda\right) \le \max_k \left\{ \frac{\lambda^2\gamma_k^{-2}a_k^{-2}}{\left(b_k^2+\lambda\gamma_k^{-1}\right)^2} \right\} \left(\sum_{k\ge1} a_k^2\theta_k^2\right) \le \max_k \left(\frac{\lambda^2\gamma_k^{-2}a_k^{-2}}{b_k^4+\lambda^2\gamma_k^{-2}}\right)\left(\sum_{k\ge1} a_k^2\theta_k^2\right), \tag{8}$$

and

$$\mathrm{Var}_\theta\left(\hat{\theta}_\lambda\right) \le \epsilon^2 \sum_{k\ge1} \frac{b_k^2\gamma_k^2}{b_k^4\gamma_k^2+\lambda^2}. \tag{9}$$

5.1 Proof of Theorem 1

We begin with the case when $\gamma_k \sim e^{\gamma k^2}$ for some $\gamma > 0$.

5.1.1 Mildly ill-posed with Sobolev spaces

In this case, $b_k \sim k^{-\beta}$ and $a_k \sim k^{\alpha}$. From (8)

$$\sup_{\theta\in\Theta^\alpha(Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) \le C\lambda^2\left(\min_{x\ge1}\left\{x^{2\alpha-4\beta}\exp(-2\gamma x^2)+\lambda^2x^{2\alpha}\right\}\right)^{-1}. \tag{10}$$

Hereafter we use C as a generic positive constant which may take different values at each appearance. By the first order condition, the minimum on the right hand side is achieved at the root of

$$\left(\frac{2\alpha-4\beta}{x}-4\gamma x\right)x^{2\alpha-4\beta}\exp(-2\gamma x^2)+\frac{2\alpha\lambda^2}{x}x^{2\alpha}=0,$$

implying that

$$\sup_{\theta\in\Theta^\alpha(Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) \le C\left(-\log\lambda\right)^{-\alpha}. \tag{11}$$

Now consider $\mathrm{Var}_\theta\left(\hat{\theta}_\lambda\right)$. From (9)

$$\mathrm{Var}_\theta\left(\hat{\theta}_\lambda\right) \le \epsilon^2\sum_{k\ge1}\frac{k^{-2\beta}\exp(-2\gamma k^2)}{k^{-4\beta}\exp(-2\gamma k^2)+\lambda^2} \approx \epsilon^2\int_1^\infty \frac{x^{-2\beta}\exp(-2\gamma x^2)}{x^{-4\beta}\exp(-2\gamma x^2)+\lambda^2}dx.$$

The integral on the rightmost hand side can be bounded by

$$\int_1^\infty \frac{1}{x^{-2\beta}+\lambda^2x^{2\beta}\exp(2\gamma x^2)}dx \le \int_1^{x_0} x^{2\beta}dx + \int_{x_0}^\infty \lambda^{-2}x^{-2\beta}\exp(-2\gamma x^2)dx,$$

where x_0 is the positive root of

$$x^{-2\beta} = \lambda^2x^{2\beta}\exp(2\gamma x^2),$$

which is of the order $(-\gamma^{-1}\log\lambda)^{\frac{1}{2}}$. Because

$$\int_{x_0}^{\infty}\lambda^{-2}x^{-2\beta}\exp(-2\gamma x^2)dx = o\left(x_0^{2\beta}\right),$$

for small values of λ, we have

$$\sum_{k\geq 1}\mathrm{Var}\left(\hat{\theta}_{k\lambda}\right) = O\left(\epsilon^2\left(\log\frac{1}{\lambda}\right)^{2\alpha+\frac{1}{2}}\right) \tag{12}$$

as $\lambda \to 0$.

Combining (11) and (12), we have

$$\sum_{k\geq 1}\mathbb{E}\left(\hat{\theta}_{k\lambda}-\theta_k\right)^2 = O\left(\left(\log\frac{1}{\lambda}\right)^{-\alpha} + \epsilon^2\left(\log\frac{1}{\lambda}\right)^{\beta+1/2}\right)$$

as $\epsilon \to 0$. Taking

$$\lambda = O\left(\exp\left(-\epsilon^{\frac{4}{2\alpha+2\beta+1}}\right)\right)$$

yields

$$\sup_{(\theta_k:k\geq 1)\in\Theta^{\alpha}(Q)}\sum_{k\geq 1}\mathbb{E}\left(\hat{\theta}_{k\lambda}-\theta_k\right)^2 = O\left(\exp\left(-\epsilon^{\frac{4\alpha}{2\alpha+2\beta+1}}\right)\right),$$

as $\epsilon \to 0$.

5.1.2 Mildly ill-posed with analytic function

For this case, $b_k \sim k^{-\beta}$ and $a_k \sim \exp(\alpha k)$. First observe that the variance $\mathrm{Var}_{\theta}\left(\hat{\theta}_{\lambda}\right)$ does not change with the parameter space and can still be bounded as in (12). On the other hand, from (8),

$$\sup_{(\theta_k:k\geq 1)\in\Theta^{\infty}(\alpha,Q)} B_{\theta}^2\left(\hat{\theta}_{\lambda}\right) \leq C\lambda^2\left(\min_{x\geq 1}\left\{x^{-4\beta}\exp(2\alpha x-2\gamma x^2)+\lambda^2\exp(2\alpha x)\right\}\right)^{-1},$$

and following the first order condition for the minimization on the right hand side, we have

$$\sup_{(\theta_k:k\geq 1)\in\Theta^{\infty}(\alpha,Q)} B_{\theta}^2\left(\hat{\theta}_{\lambda}\right) = O\left(\exp\left[-2\alpha\left(-\frac{1}{\gamma}\log\lambda\right)^{1/2}\right]\right), \tag{13}$$

as $\lambda \to 0$. Summing up, we have

$$\sum_{k\geq 1}\mathbb{E}\left(\hat{\theta}_{k\lambda}-\theta_k\right)^2 = O\left(\exp\left[-2\alpha\left(-\frac{1}{\gamma}\log\lambda\right)^{1/2}\right] + \epsilon^2\left(\log\frac{1}{\lambda}\right)^{\beta+1/2}\right), \tag{14}$$

as $\epsilon \to 0$. Consequently, if λ takes the optimal value

$$\lambda = O\left(\exp\left(-\frac{\gamma}{2\alpha^2}\left(\log\frac{1}{\epsilon^2}\right)^2\right)\right),$$

the risk is minimax rate optimal, i.e.,

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\infty(\alpha,Q)}\sum_{k\geq 1}\mathbb{E}\left(\hat{\theta}_{k\lambda}-\theta_k\right)^2 = O\left(\epsilon^2\left(\log\frac{1}{\epsilon}\right)^{2\beta+1}\right),$$

as $\epsilon \to 0$.

5.1.3 Severely ill-posed with Sobolev spaces

In this case, $b_k \sim \exp(-\beta k)$ and $a_k \sim k^\alpha$. Inequality (8) implies

$$\sup_{\theta\in\Theta^\alpha(Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) \leq C\lambda^2\left(\min_{x\geq 1}\left\{x^{2\alpha}\exp(-4\beta x - 2\gamma x^2) + \lambda^2 x^{2\alpha}\right\}\right)^{-1},$$

where, after minimizing the function inside the brackets, we get

$$\sup_{\theta\in\Theta^\alpha(Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) = O\left((-\log\lambda)^{-\alpha}\right) \;\; as\ \lambda \to 0. \tag{15}$$

The variance $\mathrm{Var}_\theta\left(\hat{\theta}_\lambda\right)$ can be bounded using (9). In particular,

$$\begin{aligned}\mathrm{Var}_\theta\left(\hat{\theta}_\lambda\right) &\leq \epsilon^2\sum_{k\geq 1}\frac{\exp(-2\beta k - 2\gamma k^2)}{\exp(-4\beta k - 2\gamma k^2)+\lambda^2}\\ &\approx \epsilon^2\int_1^\infty\frac{\exp(-2\beta x - 2\gamma x^2)dx}{\exp(-4\beta x - 2\gamma x^2)+\lambda^2}\\ &\leq \epsilon^2\epsilon^2\left(\int_1^{x_0}\exp(2\beta x)dx + \int_{x_0}^\infty \lambda^{-2}\exp(-2\beta x - 2\gamma x^2)dx\right),\end{aligned}$$

where x_0 is the positive root of

$$\exp(-2\beta x) = \lambda^2\exp(2\beta x + 2\gamma x^2).$$

It can be easily derived that

$$x_0 = O\left(\left(-\gamma^{-1}\log\lambda\right)^{1/2}\right)$$

as $\lambda \to 0$. Observing that

$$\int_{x_0}^\infty \lambda^{-2}\exp(-2\beta x - 2\gamma x^2)dx = o\left(\exp(2\beta x_0)\right),$$

we have

$$\mathrm{Var}_\theta\left(\hat{\theta}_\lambda\right) = O\left(\epsilon^2 \exp\left(2\beta\left(-\gamma^{-1}\log\lambda\right)^{1/2}\right)\right) \tag{16}$$

as $\epsilon \to 0$. Combining (15) and (16), we have

$$\sum_{k\geq 1}\mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left((-\log\lambda)^{-\alpha} + \epsilon^2\exp\left(2\beta\left(-\gamma^{-1}\log\lambda\right)^{1/2}\right)\right) \tag{17}$$

as $\epsilon \to 0$, attaining the minimax optimal rate of convergence

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\alpha(Q)}\sum_{k\geq 1}\mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\left(\log\frac{1}{\epsilon}\right)^{-2\alpha}\right),$$

when

$$\lambda = O\left(\exp\left(-\left(\log\frac{1}{\epsilon^2}\right)^2\right)\right)$$

as $\epsilon \to 0$.

5.1.4 Severely ill-posed with Analytic functions

For this case, $b_k \sim \exp(-\beta k)$ and $a_k \sim \exp(\alpha k)$. Following similar arguments as before, from Inequality (8)

$$\begin{aligned}\sup_{(\theta_k:k\geq 1)\in\Theta^\infty(\alpha,Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) &\leq C\lambda^2\left(\min_{x\geq 1}\left\{\exp[(2\alpha-4\beta)x - 2\gamma x^2] + \lambda^2\exp(2\alpha x)\right\}\right)^{-1}\\ &= O\left(\exp\left[-2\alpha\left(-\gamma^{-1}\log\lambda\right)^{1/2}\right]\right)\end{aligned}$$

as λ goes to 0. On the other hand, the variance can still be bounded by (16). Hence,

$$\sum_{k\geq 1}\mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\exp\left[-2\alpha\left(-\gamma^{-1}\log\lambda\right)^{1/2}\right] + \epsilon^2\exp\left(2\beta\left(\frac{1}{\gamma}\log\frac{1}{\lambda}\right)^{1/2}\right)\right) \tag{18}$$

as $\epsilon \to 0$, which implies that if

$$\lambda = O\left(\exp\left(-\frac{\gamma}{(2\alpha+2\beta)^2}\left(\log\frac{1}{\epsilon^2}\right)^2\right)\right),$$

the radial basis function regularization achieves the optimal rate of convergence

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\infty(\alpha,Q)}\sum_{k\geq 1}\mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\epsilon^{\frac{4\alpha}{2\alpha+2\beta}}\right)$$

as $\epsilon \to 0$.

5.2 Proof of Theorem 2

We now consider the case when $\gamma_k \sim e^{\gamma k}$.

5.2.1 Mildly ill-posed with Sobolev spaces

Observe that $b_k \sim k^{-\beta}$ and $a_k \sim k^{\alpha}$. From (8),

$$\sup_{\theta \in \Theta^\alpha(Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) \leq C\lambda^2 \left(\min_{x\geq 1}\{x^{\alpha-2\beta}\exp(-\gamma x) + \lambda x^\alpha\}\right)^{-2}.$$

By the first order condition, the minimun is acieved when x is the root of the equation

$$\left(x^{-1}(\alpha - 2\beta) - \gamma\right) x^{\alpha-2\beta}\exp(-\gamma x) + \alpha\lambda x^{\alpha-1} = 0,$$

whose solution, after simple algebraic manipulations, implies

$$\sup_{\theta \in \Theta^\alpha(Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) = O\left((-\log\lambda)^{-2\alpha}\right) \tag{19}$$

as λ becomes small. On the other hand, in the light of (9),

$$\mathrm{Var}_\theta\left(\hat{\theta}_\lambda\right) \leq \epsilon^2 \sum_{k=1}^{\infty} \frac{k^{-2\beta}\exp(-2\gamma k)}{k^{-4\beta}\exp(-2\gamma k) + \lambda^2} \approx \epsilon^2 \int_1^\infty \frac{dx}{x^{-2\beta} + \lambda^2 x^{2\beta}\exp(2\gamma x)},$$

where the integral on the right hand side can be bounded by

$$\int_1^{x_0} x^{2\beta} dx + \int_{x_0}^\infty \lambda^{-2} x^{-2\beta}\exp(-2\gamma x) dx,$$

and x_0 is the positive root of

$$x^{-2\beta} = \lambda^2 x^{2\beta}\exp(2\gamma x).$$

It is easy to check that

$$x_0 = O\left(-\gamma^{-1}\log\lambda\right)$$

as $\lambda \to 0$. Observing that

$$\int_{x_0}^\infty \lambda^{-2} x^{-2\beta}\exp(-2\gamma x) dx = o\left(x_0^{2\beta}\right),$$

we have

$$\mathrm{Var}_\theta\left(\hat{\theta}_\lambda\right) = O\left(\epsilon^2 (-\log\lambda)^{2\beta+1}\right) \tag{20}$$

as $\epsilon \to 0$. Combining (19) and (20), we have

$$\sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left((-\log\lambda)^{-2\alpha} + \epsilon^2(-\log\lambda)^{2\beta+1}\right), \tag{21}$$

which implies that if

$$\lambda = O\left(\exp\left(-\epsilon^{-\frac{2}{2\alpha+2\beta+1}}\right)\right),$$

it achieves the minimax optimal rate of convergence

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\alpha(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\epsilon^{\frac{4\alpha}{2\alpha+2\beta+1}}\right)$$

as $\epsilon \to 0$.

5.2.2 Mildly ill-posed with analytic function

In this case, $b_k \sim k^{-\beta}$ and $a_k \sim \exp(\alpha k)$. From (8),

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\infty(\alpha,Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) \leq C\lambda^2\left(\min_{x\geq 1}\{x^{-4\beta}\exp(2\alpha x - 2\gamma x) + \lambda^2\exp(2\alpha x)\}\right)^{-1}.$$

By the first order condition, the minimum on the right hand side is attained at the root of

$$\left(-4\beta x^{-1} + 2\alpha - 2\gamma\right)x^{-4\beta}\exp(-2\gamma x) + 2\alpha\lambda^2 = 0.$$

Thus,

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\infty(\alpha,Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) = \begin{cases} O\left(\exp\left(\frac{2\alpha}{\gamma}\log(\lambda)\right)\right) & \text{if } \gamma > \alpha - 2\beta \\ O\left(\lambda^2\right) & \text{if } \gamma \leq \alpha - 2\beta \end{cases}. \tag{22}$$

Combining (22) and the bound on the variance given in (20), we have

(a) if $\gamma > \alpha - 2\beta$,

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\infty(\alpha,Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\exp\left(\frac{2\alpha}{\gamma}\log(\lambda)\right) + \epsilon^2(-\log\lambda)^{2\beta+1}\right);$$

(b) if $\gamma \leq \alpha - 2\beta$.

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\infty(\alpha,Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\lambda^2 + \epsilon^2(-\log\lambda)^{2\beta+1}\right).$$

Taking

$$\lambda = \begin{cases} O\left(\epsilon^{\frac{\gamma}{\alpha}}\right) & \text{if } \gamma > \alpha - 2\beta \\ O(\epsilon) & \text{if } \gamma \leq \alpha - 2\beta \end{cases},$$

the radial basis function regularization achieves the minimax optimal rate of

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\infty(\alpha,Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\epsilon^2\left(\log\frac{1}{\epsilon}\right)^{2\beta+1}\right),$$

as $\epsilon \to 0$.

5.2.3 Severely ill-posed with Sobolev spaces

Observe that $b_k \sim \exp(-\beta k)$ and $a_k \sim k^\alpha$. From (8),

$$\begin{aligned} \sup_{(\theta_k:k\geq 1)\in\Theta^\alpha(Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) &\leq C\lambda^2\left(\min_{x\geq 1}\{x^\alpha \exp(-2\beta x - \gamma x) + \lambda x^\alpha\}\right)^{-2} \\ &= O\left((-\log\lambda)^{-2\alpha}\right) \end{aligned}$$

as λ goes to 0. To bound the variance, note that from (9),

$$\sum_{k\geq 1} \mathrm{Var}\left(\hat{\theta}_{k\lambda}\right) \approx \epsilon^2 \int_1^\infty \frac{dx}{\exp(-2\beta x) + \lambda^2 \exp(2\beta x + 2\gamma x)},$$

where the integral on the right side can be bounded by

$$\int_1^{x_0} \exp(2\beta x)dx + \int_{x_0}^\infty \lambda^{-2}\exp(-2\beta x - 2\gamma x)dx,$$

and x_0 is the positive root of

$$\exp(-2\beta x) = \lambda^2 \exp(2\beta x + 2\gamma x).$$

It can be easily derived that

$$x_0 = O\left(-(2\beta+\gamma)^{-1}\log\lambda\right)$$

as λ goes to 0. Using

$$\int_{x_0}^\infty \lambda^{-2}\exp(-2\beta x - 2\gamma x)dx = o\left(\exp(2\beta x_0)\right),$$

we conclude

$$\sum_{k\geq 1} \mathrm{Var}\left(\hat{\theta}_{k\lambda}\right) = O\left(\epsilon^2 \exp\left(-\frac{2\beta}{2\beta+\gamma}\log\lambda\right)\right). \tag{23}$$

In summary,

$$\sup_{(\theta_k:k\geq 1)\in\Theta^{\alpha}(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda}-\theta_k\right)^2 = O\left(\left(\log\frac{1}{\lambda}\right)^{-2\alpha} + \epsilon^2\exp\left(\frac{-2\beta}{2\beta+\gamma}\log\lambda\right)\right),$$

which attains the minimax optimal rate of

$$\sup_{(\theta_k:k\geq 1)\in\Theta^{\alpha}(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda}-\theta_k\right)^2 = O\left(\left(\log\frac{1}{\epsilon}\right)^{-2\alpha}\right)$$

if $\lambda = O\left(\epsilon^2\right)$ as $\epsilon \to 0$.

5.2.4 Severely ill-posed with Analytic functions

For this case, $b_k \sim \exp(-\beta k)$ and $a_k \sim \exp(\alpha k)$. Similar to (8),

$$\sup_{(\theta_k:k\geq 1)\in\Theta^{\infty}(\alpha,Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) \leq C\lambda^2\left(\min_{x\geq 1}\{\exp(2\alpha x-4\beta x-2\gamma x)+\lambda^2\exp(2\alpha x)\}\right)^{-1}.$$

By the first order condition, the minimum on the right hand side is achieved at the root of

$$(2\alpha-4\beta+2\gamma)\exp(-4\beta x-2\gamma x)+2\alpha\lambda^2=0.$$

if and only if $\alpha < \gamma + 2\beta$. Otherwise, it is achieved at one. Thus, for small values of λ,

$$\sup_{(\theta_k:k\geq 1)\in\Theta^{\infty}(\alpha,Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) = \begin{cases} O\left(\exp\left(\frac{2\alpha}{2\beta+\gamma}\log\lambda\right)\right) & \text{if } \gamma > \alpha-2\beta \\ O\left(\lambda^2\right) & \text{if } \gamma \leq \alpha - 2\beta \end{cases}. \tag{24}$$

Similarly, from (23), as ϵ goes to 0,

$$\sum_{k\geq 1}\mathrm{Var}\left(\hat{\theta}_{k\lambda}\right) = O\left(\epsilon^2\exp\left(-\frac{2\beta}{2\beta+\gamma}\log\lambda\right)\right).$$

(a) if $\gamma > \alpha - 2\beta$,

$$\sum_{k\geq 1}\mathbb{E}\left(\hat{\theta}_{k\lambda}-\theta_k\right)^2 = O\left(\exp\left(\frac{2\alpha}{2\beta+\gamma}\log\lambda\right)+\epsilon^2\exp\left(-\frac{2\beta}{2\beta+\gamma}\log\lambda\right)\right), \tag{25}$$

which attains the optimal rate of convergence

$$\sup_{(\theta_k:k\geq 1)\in\Theta^{\infty}(\alpha,Q)} \sum_{k\geq 1}\mathbb{E}\left(\hat{\theta}_{k\lambda}-\theta_k\right)^2 = O\left(\epsilon^{\frac{2\alpha}{\alpha+\beta}}\right),$$

when

$$\lambda = O\left(\epsilon^{-\frac{\beta+\gamma}{\alpha+\beta}}\right),$$

as $\epsilon \to 0$.

(b) if $\gamma \leq \alpha - 2\beta$, following a similar argument as before, we note that

$$\sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\lambda^2 + \epsilon^2 \exp\left(-\frac{2\beta}{2\beta+\gamma}\log\lambda\right)\right). \tag{26}$$

The best achievable rate is

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\infty(\alpha,Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\epsilon^{\frac{4\beta+2\gamma}{3\beta+\gamma}}\right) \text{ as } \epsilon \to 0,$$

and it is attained when

$$\lambda = O\left(\epsilon^{\frac{2\beta+\gamma}{3\beta+\gamma}}\right)$$

as $\epsilon \to 0$. It is clear then that in this case the optimal minimax rate is not attained.

5.3 Proof of Theorem 3

In this setting $\gamma_k \sim k^\gamma$.

5.3.1 Mildly ill-posed with Sobolev spaces

Observe that $b_k \sim k^{-\beta}$ and $a_k \sim k^\alpha$. Similar to before,

$$\sup_{\theta\in\Theta^\alpha(Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) \leq C\lambda^2\left(\min_{x\geq 1}\{x^{\alpha-2\beta-\gamma} + \lambda x^\alpha\}\right)^{-2},$$

where it is easy to see that if $\gamma \leq \alpha - 2\beta$, the function inside the brackets is strictly increasing on $x \geq 1$. By the first order condition, for small values of λ,

$$\sup_{\theta\in\Theta^\alpha(Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) = \begin{cases} O\left(\lambda^{\frac{2\alpha}{2\beta+\gamma}}\right) & \text{if } \gamma > \alpha - 2\beta \\ O\left(\lambda^2\right) & \text{if } \gamma \leq \alpha - 2\beta \end{cases}. \tag{27}$$

Similarly, from inequality (9),

$$\mathrm{Var}_\theta\left(\hat{\theta}_\lambda\right) \leq \epsilon^2 \sum_{k\geq 1} \frac{k^{-2\beta-2\gamma}}{k^{-4\beta-2\gamma} + \lambda^2} \approx \epsilon^2 \int_1^\infty \frac{dx}{x^{-2\beta} + \lambda^2 x^{2\beta+2\gamma}}.$$

The integral on the right side can be bounded by

$$\int_1^{x_0} x^{2\beta}dx + \int_{x_0}^\infty \lambda^{-2}x^{-2\beta-2\gamma}dx,$$

where x_0 is the positive root of

$$x^{-2\beta} - \lambda^2 x^{2\beta+2\gamma} = 0,$$

i.e., as λ goes to 0,

$$x_0 = O\left(\lambda^{-\frac{1}{2\beta+\gamma}}\right).$$

Observe that

$$\int_{x_0}^{\infty} \lambda^{-2} x^{-2\beta-2\gamma} dx = o\left(x_0^{2\beta}\right).$$

Thus, for small values of λ,

$$\mathrm{Var}_\theta\left(\hat{\theta}_\lambda\right) = O\left(\epsilon^2 \lambda^{-\frac{2\beta+1}{2\beta+\gamma}}\right). \tag{28}$$

Combining (27) and (28),

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\alpha(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = \begin{cases} O\left(\lambda^{\frac{2\alpha}{2\beta+\gamma}} + \epsilon^2 \lambda^{-\frac{1+2\beta}{2\beta+\gamma}}\right) & \text{if } \gamma > \alpha - 2\beta \\ O\left(\lambda^2 + \epsilon^2 \lambda^{-\frac{1+2\beta}{2\beta+\gamma}}\right) & \text{if } \gamma \leq \alpha - 2\beta \end{cases},$$

implying that

(a) if $\gamma > \alpha - 2\beta$, the estimator achieves the optimal rate in the minimax sense, that is

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\alpha(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\epsilon^{\frac{4\alpha}{2\alpha+2\beta+1}}\right),$$

provided that

$$\lambda = O\left(\epsilon^{\frac{4\beta+2\gamma}{2\alpha+2\beta+1}}\right)$$

as $\epsilon \to 0$;

(b) if $\gamma \leq \alpha - 2\beta$, the best achievable rate is

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\alpha(Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\epsilon^{\frac{2(4\beta+2\gamma)}{6\beta+2\gamma+1}}\right),$$

and it is attained when

$$\lambda = O\left(\epsilon^{\frac{4\beta+2\gamma}{6\beta+2\gamma+1}}\right)$$

as $\epsilon \to 0$.

5.3.2 Mildly ill-posed with analytic function

Similar to before,

$$\sup_{(\theta_k:k\geq 1)\in\Theta^\infty(\alpha,Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) \leq C\lambda^2 \left(\min_{x\geq 1}\{x^{-4\beta-2\gamma}\exp(2\alpha x) + \lambda^2\exp(2\alpha x)\}\right)^{-1}.$$

Then, by the first order condition, the minimum on the right hand side is achieved at one if $\gamma \leq \alpha - 2\beta$ and at the root of

$$\left(-\left(4\beta + 2\gamma\right) x^{-1} + 2\alpha\right) x^{-4\beta-2\gamma} + 2\alpha\lambda^2 = 0$$

otherwise. Thus, for small values of λ,

$$\sup_{(\theta_k : k\geq 1)\in\Theta^\infty(\alpha,Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) = \begin{cases} O\left(\exp\left(-2\alpha\lambda^{-\frac{1}{2\beta+\gamma}}\right)\right) & \text{if } \gamma > \alpha - 2\beta \\ O\left(\lambda^2\right) & \text{if } \gamma \leq \alpha - 2\beta \end{cases}.$$

Together with (28), this implies that

$$\sup_{(\theta_k : k\geq 1)\in\Theta^\infty(\alpha,Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = \begin{cases} O\left(\exp\left(-2\alpha\lambda^{-\frac{1}{2\beta+\gamma}}\right) + \epsilon^2\lambda^{-\frac{1+2\beta}{\gamma+2\beta}}\right) & \text{if } \gamma > \alpha - 2\beta \\ O\left(\lambda^2 + \epsilon^2\lambda^{-\frac{1+2\beta}{\gamma+2\beta}}\right) & \text{if } \gamma \leq \alpha - 2\beta \end{cases}. \tag{29}$$

Thus,

(a) if $\gamma > \alpha - 2\beta$, the estimator is optimal in the minimax sense, that is

$$\sup_{(\theta_k : k\geq 1)\in\Theta^\infty(\alpha,Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\epsilon^2\left(\log\frac{1}{\epsilon}\right)^{2\beta+1}\right)$$

when

$$\lambda = O\left(\left(\frac{1}{2\alpha}\log\frac{1}{\epsilon^2}\right)^{-2\beta-\gamma}\right)$$

as ϵ goes to 0.

(b) if $\gamma \leq \alpha - 2\beta$, the best achievable rate is

$$\sup_{(\theta_k : k\geq 1)\in\Theta^\infty(\alpha,Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\epsilon^{\frac{2(4\beta+2\gamma)}{6\beta+2\gamma+1}}\right),$$

and it is attained when

$$\lambda = O\left(\epsilon^{\frac{4\beta+2\gamma}{6\beta+2\gamma+1}}\right)$$

as $\epsilon \to 0$.

5.3.3 Severely ill-posed with Sobolev spaces

Observe that

$$\sup_{\theta\in\Theta^\alpha(Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) \le C\lambda^2\left(\min_{x\ge 1}\{x^{\alpha-\gamma}\exp(-2\beta x)+\lambda x^\alpha\}\right)^{-2} = O\left((-\log\lambda)^{-2\alpha}\right)$$

as $\lambda\to 0$, and

$$\begin{aligned}\sum_{k\ge 1}\mathrm{Var}\left(\hat{\theta}_{k\lambda}\right) &\approx \epsilon^2\int_1^\infty \frac{dx}{\exp(-2\beta x)+\lambda^2\exp(2\beta x)x^{2\gamma}} \\ &\le \epsilon^2\left(\int_1^{x_0}\exp(2\beta x)dx+\int_{x_0}^\infty \lambda^{-2}\exp(2\beta x)x^{2\gamma}dx\right),\end{aligned}$$

where x_0 is the root of

$$\exp(-2\beta x)=\lambda^2\exp(2\beta x)x^{2\gamma},$$

and therefore

$$\mathrm{Var}_\theta\left(\hat{\theta}_\lambda\right)=O\left(\lambda^{-1}\epsilon^2\right) \tag{30}$$

as $\epsilon\to 0$. Thus,

$$\sup_{(\theta_k:k\ge 1)\in\Theta^\alpha(Q)}\sum_{k\ge 1}\mathbb{E}\left(\hat{\theta}_{k\lambda}-\theta_k\right)^2 = O\left((-\log\lambda)^{-2\alpha}+\frac{\epsilon^2}{\lambda}\right).$$

It is minimax rate optimal, i.e.,

$$\sup_{(\theta_k:k\ge 1)\in\Theta^\alpha(Q)}\sum_{k\ge 1}\mathbb{E}\left(\hat{\theta}_{k\lambda}-\theta_k\right)^2 = O\left(\left(\log\frac{1}{\epsilon}\right)^{-2\alpha}\right),$$

if

$$\lambda = O\left(\epsilon^2\right)$$

as $\epsilon\to 0$.

5.3.4 Severely ill-posed with Analytic functions

In this case, $b_k\sim\exp(-\beta k)$ and $a_k\sim\exp(\alpha k)$, and therefore,

$$\sup_{(\theta_k:k\ge 1)\in\Theta^\infty(\alpha,Q)} B_\theta^2\left(\hat{\theta}_\lambda\right)\le C\lambda^2\left(\min_{x\ge 1}\{\exp(2\alpha-4\beta x)x^{-2\gamma}+\lambda^2\exp(2\alpha x)\}\right)^{-1}.$$

By the first order condition, we conclude that

$$\sup_{(\theta_k:k\ge 1)\in\Theta^\infty(\alpha,Q)} B_\theta^2\left(\hat{\theta}_\lambda\right) = \begin{cases} O\left(\exp\left(\frac{\alpha}{2\beta}\log\lambda\right)\right) & \text{if } \gamma>\alpha-2\beta \\ O\left(\lambda^2\right) & \text{if } \gamma\le\alpha-2\beta\end{cases} \tag{31}$$

as $\lambda\to 0$. Combining (30) and (31), we have

(a) if $\gamma \geq \alpha - 2\beta$,

$$\sup_{(\theta_k : k\geq 1)\in\Theta^\infty(\alpha,Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\exp\left(\frac{\alpha}{2\beta}\log\lambda\right) + \lambda^{-1}\epsilon^2\right). \tag{32}$$

The best achievable rate is

$$\sup_{(\theta_k : k\geq 1)\in\Theta^\infty(\alpha,Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\epsilon^{\frac{2\alpha}{\alpha+2\beta}}\right),$$

and it can be attained when

$$\lambda = O\left(\epsilon^{\frac{4\beta}{\alpha+2\beta}}\right)$$

as $\epsilon \to 0$;

(b) if $\gamma < \alpha - 2\beta$,

$$\sup_{(\theta_k : k\geq 1)\in\Theta^\infty(\alpha,Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\lambda^2 + \lambda^{-1}\epsilon^2\right).$$

The best achievable rate is

$$\sup_{(\theta_k : k\geq 1)\in\Theta^\infty(\alpha,Q)} \sum_{k\geq 1} \mathbb{E}\left(\hat{\theta}_{k\lambda} - \theta_k\right)^2 = O\left(\epsilon^{\frac{4}{3}}\right)$$

and it is attained when

$$\lambda = O\left(\epsilon^{\frac{4}{3}}\right)$$

as $\epsilon \to 0$.

References

[1] Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, **68** (3), 337-404.

[2] Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Annals of Statististics*, **24**, 2384-2398.

[3] Buhmann, M. D. (2003). *Radial Basis Functions: Theory and Implementations*. Cambridge: University Press.

[4] Cavalier, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems*, **24**, 1-19.

[5] Chalmond, B. (2008). *Modeling and Inverse Problems in Image Analysis*. New York: Springer.

[6] Evgeniou, T., Pontil, M., and Poggio,T. (2000). Statistical Learning Theory: A Primer. *International Journal of Computer Vision*, **38** (1), 9-13.

[7] Girosi, F., Jones, M. and Poggio, T. (1993). Priors, stabilizers and basis functions: From regularization to radial, tensor and additive splines. *Artificial Intelligence memo 1430*, MIT, Artificial Intelligence Laboratory.

[8] Golubev, G. and Nussbaum, M. (1998). Asymptotic equivalence of spectral density and regression estimation. Technical report, Weierstrass Institute for Applied Analysis and Stochastics, Berlin.

[9] Grama, I. and Nussbaum, M. (1997). Asymptotic equivalence for nonparametric generalized linear models. Technical report, Weierstrass Institute for Applied Analysis and Stochastics, Berlin.

[10] Johnstone, I. (1998). *Function Estimation and Gaussian Sequence Models*. Unpublished manuscript.

[11] Kaipo, J. and Somersalo, E. (2004). *Statistical and Computational Inverse Problems*. New York: Springer.

[12] Lin, Y. and Brown, L. (2004). Statistical properties of the method of regularization with periodic Gaussian reproducing kernel, *Annals of Statistics*, **32**, 1723-1743.

[13] Lin, Y. and Yuan, M. (2006). Convergence rates of compactly supported radial basis function regularization. *Statistica Sinica*, **16**, 425-439.

[14] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Annals of Statististics*, **24**, 2399-2430.

[15] Ramm, A. (2009). *Inverse Problems: Mathematical and Analytical Techniques with Applications to Engineering*. New York: Springer.

[16] Shi, T. and Yu, B. (2005). Binning in Gaussian Kernel Regularization. *Statistica Sinica* **16**, 541-567.

[17] Smola, A., Schölkopf, B. and Müller, K.R. (1998). The connection between regularization operators and support vector kernels. *Neural Networks*, **11**, 637-649.

[18] Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia: SIAM.

[19] Wahba, G. (1999). Support vector machines, reproducing kernel Hilbert spaces and the randomized GACV. In B. Schölkopf, C. Burges & A. Smola, eds, *Advances in Kernel Methods Support Vector Learning*. Cambridge: MIT Press, 69-88

[20] Wendland, H. (1998). Error estimates for interpolation by compactly supported radial basis functions of minimal degree. *Journal of Approximation Theory*, **93**, 258-272.

[21] Zhang, H., Genton, M. and Liu, P. (2004). Compactly supported radial basis function kernels. Institute of Statistics Mimeo Series 2570, North Carolina State University.