Journal of Mathematical Research with Applications
Jul., 2017, Vol. 37, No. 4, pp. 496–504
DOI:10.3770/j.issn:2095-2651.2017.04.011
Http://jmre.dlut.edu.cn

An ℓ^1 Regularized Method for Numerical Differentiation Using Empirical Eigenfunctions

Junbin LI*, **Renhong WANG**, **Min XU**
School of Mathematical Sciences, Dalian University of Technology, Liaoning 116024*, P. R. China*

Abstract We propose an ℓ^1 regularized method for numerical differentiation using empirical eigenfunctions. Compared with traditional methods for numerical differentiation, the output of our method can be considered directly as the derivative of the underlying function. Moreover, our method could produce sparse representations with respect to empirical eigenfunctions. Numerical results show that our method is quite effective.

Keywords numerical differentiation; empirical eigenfunctions; ℓ^1 regularization; mercer kernel

MR(2010) Subject Classification 65D15; 65F22

1. Introduction

Numerical differentiation is a problem to determine the derivatives of a function from the values on scattered points. It plays an important role in scientific research and application, such as solving Volterra integral equation [1], image processing [2], option pricing models [3] and identification [4]. The main difficulty of numerical differentiation is that it is an ill-posed problem, which means, the small error of measurement may cause huge error in the computed derivatives [5]. Several methods for numerical differentiation have been proposed in the literature, including difference methods [6] and interpolation methods [7]. In particular, some researchers proposed to use Tikhonov regularization for numerical differentiation problems, which has been shown quite effective [8–10].

Note that most regularization methods for numerical differentiation consist of estimating a function from the given data and then computing derivatives of the function. However, in many practical applications, the thing we need to obtain is the derivative of the underlying function not the underlying function itself [3,4]. Thus, a natural approach to computing derivatives would be to directly estimate the derivatives. In this paper, we propose an algorithm for numerical differentiation in the framework of statistical learning theory. More specifically, we study an ℓ^1 regularized algorithm for numerical differentiation using empirical eigenfunctions. The key

Received April 25, 2017; Accepted May 25, 2017
Supported by the National Nature Science Foundation of China (Grant Nos. 11301052; 11301045; 11271060; 11601064; 11671068), the Fundamental Research Funds for the Central Universities (Grant No. DUT16LK33) and the Fundamental Research of Civil Aircraft (Grant No. MJ-F-2012-04).
* Corresponding author
E-mail address: junbin@mail.dlut.edu.cn (Junbin LI)

advantage of the algorithm is that its output could be considered directly as the derivative of the underlying function. Moreover, the algorithm produces sparse representations with respect to empirical eigenfunctions, without assuming sparsity in terms of any basis or system.

The remainder of this paper is organized as follows. In Section 2, we first review some basic facts in statistical learning theory and then present our main algorithm. In Section 3, we present an approach for computing explicitly the empirical eigenfunctions. In Section 4, we establish the representer theorem of the algorithm. To illustrate the effectiveness of the algorithm, we provide several numerical examples in Section 5. Finally, some concluding remarks are given in Section 6.

2. Formulation of the method

To present our main algorithm, let us first describe the basic setting of statistical learning theory.

Let X be the input space and $Y \subset \mathbb{R}$ the output space. Assume that ρ is a Borel probability measure on $Z = X \times Y$. Let ρ_X be the marginal distribution on X and $\rho(\cdot|x)$ the conditional distribution on Y at given x. Let f_ρ be the function defined by

$$f_\rho(x) = \int_Y y \mathrm{d}\rho(y|x), \quad x \in X.$$

Given a sample $\mathbf{z} = \{(x_i, y_i)\}_{i=1}^m$ drawn independently and identically according to ρ, we are interested in estimating the derivative of f_ρ. More precisely, we want to find a function $f_{\mathbf{z}} : X \to \mathbb{R}$ that can be used as an approximation of the derivative of f_ρ.

Before proceeding further, we need to introduce some notions related to kernels [11,12]. A Mercer kernel on X is defined to be a symmetric continuous function $K : X \times X \to \mathbb{R}$ such that for any finite subset $\{x_i\}_{i=1}^m$ of X, the $m \times m$ matrix $\mathbf{K}$ whose (i, j) entry is $K(x_i, x_j)$ is positive semi-definite. Let $\mathrm{span}\{K_x : x \in X\}$ denote the space spanned by the set $\{K_x = K(\cdot, x) : x \in X\}$. We define an inner product in the space $\mathrm{span}\{K_x : x \in X\}$ as follows:

$$\left\langle \sum_{i=1} \alpha_i K_{x_i}, \sum_{j=1} \beta_j K_{t_j} \right\rangle_K = \sum_{i=1} \sum_{j=1} \alpha_i \beta_j K(x_i, t_j).$$

The reproducing kernel Hilbert space $\mathcal{H}_K$ associated with K is defined to be the completion of $\mathrm{span}\{K_x : x \in X\}$ under the norm $\|\cdot\|_K$ induced by the inner product $\langle \cdot, \cdot \rangle_K$. The reproducing property in $\mathcal{H}_K$ takes the form $f(x) = \langle f, K_x \rangle_K$ for all $x \in X$ and $f \in \mathcal{H}_K$. Let $\kappa = \sup_{x,y \in X} |K(x, y)|$. Then it follows from the reproducing property that

$$\|f\|_\infty \le \kappa \|f\|_K, \quad \forall f \in \mathcal{H}_K.$$

Taylor's expansion of a function $g(u)$ about the point x gives us, for $u \approx x$, $g(u) \approx g(x) + g'(x)(u - x)$. Thus the empirical error incurred by the function f on the sample points $x = x_i,\ u = x_j$ can be measured by

$$(g(u) - g(x) - g'(x)(u - x))^2 = (y_i - y_j + f(x_i)(x_j - x_i))^2.$$

The restriction $u \approx x$ could be enforced by the weight $\omega_{i,j} = \omega_{i,j}^{(s)} > 0$ corresponding to (x_i, x_j) with the requirement that $\omega_{i,j}^{(s)} \to 0$ as $|x_i - x_j|/s \to \infty$. One possible choice of weights is given by a Gaussian with variance $s > 0$. Let ω be the function on $\mathbb{R}$ given by $\omega(x) = \frac{1}{s^4} e^{-\frac{x^2}{2s^2}}$. Then this choice of weights is $\omega_{i,j} = \omega_{i,j}^{(s)} = \omega(x_i - x_j)$. The following regularized algorithm for numerical differentiation was proposed in [13]:

$$\min_{f \in \mathcal{H}_K} \left\{ \frac{1}{m^2} \sum_{i,j=1}^{m} \omega(x_i - x_j)(y_i - y_j - f(x_i)(x_i - x_j))^2 + \gamma \|f\|_K^2 \right\}.$$

In this paper, we shall modify the algorithm (1) by using an ℓ^1 regularizer. Note that the ℓ^1 regularizer plays a key role in producing sparse approximations. This phenomenon has been observed in LASSO [14] and compressed sensing [15], under the assumption that the approximated function has a sparse representation with respect to some basis.

Let $L_{K,s}$ denote the operator defined by

$$L_{K,s}(f) = \int_X \int_X \omega(x-u) K_x(u-x)^2 f(x) \mathrm{d}\rho_X(x) \mathrm{d}\rho_X(u), \quad f \in \mathcal{H}_K.$$

The operator $L_{K,s}$ is compact, positive, and self-adjoint [13]. Therefore it has at most countably many eigenvalues, and all of these eigenvalues are nonnegative. One can arrange these eigenvalues $\{\lambda_l\}$ (with multiplicities) as a nonincreasing sequence tending to 0 and take an associated sequence of eigenfunctions $\{\phi_l\}$ to be an orthonormal basis of $\mathcal{H}_K$. Let $\mathbf{x}$ denote the unlabeled part of the samples $\mathbf{z} = \{(x_i, y_i)\}_{i=1}^m$, i.e., $\mathbf{x} = \{x_i\}_{i=1}^m$. We consider another operator $L_{K,s}^{\mathbf{x}}$ defined on $\mathcal{H}_K$ as follows:

$$L_{K,s}^{\mathbf{x}}(f) = \frac{1}{m(m-1)} \sum_{i,j=1}^{m} \omega(x_i - x_j) K_{x_i}(x_j - x_i)^2 f(x_i), \quad f \in \mathcal{H}_K. \tag{3}$$

It is easy to show that $\mathbb{E}_{\mathbf{x}}(L_{K,s}^{\mathbf{x}} f) = L_{K,s} f$, which means $\mathbb{E}_{\mathbf{x}}(L_{K,s}^{\mathbf{x}}) = L_{K,s}$. As a result, $L_{K,s}^{\mathbf{x}}$ can be viewed as an empirical version of the operator $L_{K,s}$ with respect to $\mathbf{x}$. The operator $L_{K,s}^{\mathbf{x}}$ is self-adjoint, positive. Its eigensystem, called an empirical eigensystem, is denoted by $\{(\lambda_l^{\mathbf{x}}, \phi_l^{\mathbf{x}})\}$, where the eigenvalues $\{\lambda_l^{\mathbf{x}}\}$ are arranged in nonincreasing order. We notice here two important facts: for one thing, all the empirical eigenfunctions $\{\phi_l^{\mathbf{x}}\}$ form an orthonormal basis of $\mathcal{H}_K$; for another, at most m eigenvalues are nonzero, i.e., $\lambda_l^{\mathbf{x}} = 0$ whenever $l > m$.

Based on the first m empirical eigenfunctions $\{\phi_l^{\mathbf{x}}\}_{l=1}^m$, we are now in a position to present our main algorithm as follows:

$$c_\gamma^{\mathbf{z}} = \arg\min_{c \in \mathbb{R}^m} \left\{ \frac{1}{m^2} \sum_{i,j=1}^{m} \omega(x_i - x_j) \left(y_i - y_j - \sum_{l=1}^{m} c_l \phi_l^{\mathbf{x}}(x_i) \right) (x_j - x_i) \right)^2 + \gamma \|c\|_1 \right\}. \tag{4}$$

The output function of algorithm (4) is

$$f_\gamma^{\mathbf{z}} = \sum_{l=1}^{m} c_{\gamma,l}^{\mathbf{z}} \phi_l^{\mathbf{x}},$$

which is expected to approximate the derivative of the underlying target function f_ρ. Next we shall focus on the computations of empirical eigenpairs, the representer theorem (i.e., the explicit solution to problem (4), and the sparsity of coefficients in the representation $f_\gamma^{\mathbf{z}} = \sum_{l=1}^{m} c_{\gamma,l}^{\mathbf{z}} \phi_l^{\mathbf{x}}$.

3. Computations of empirical eigenpairs

We shall establish in this section an approach for computing explicitly the empirical eigenpairs $\{(\lambda_l^{\mathbf{x}}, \phi_l^{\mathbf{x}})\}$. To present our method, some notations and definitions are needed. Recall that $\mathbf{K}$ denotes the $m \times m$ matrix whose (i,j) entry is $K(x_i, x_j)$. For $1 \le i \le m$, define $b_i = \sum_{j=1}^{m} \omega(x_i - x_j)(x_j - x_i)^2$, $d_i = \sqrt{b_i}$. Let $\mathbf{B} = \mathrm{diag}\{b_1, b_2, \ldots, b_m\}, \mathbf{D} = \mathrm{diag}\{d_1, d_2, \ldots, d_m\}$, and $\mathbf{A} = \mathbf{DKD}$. Denote $\mathrm{rank}(\mathbf{A})$ and $\mathrm{rank}(L_{K,s}^{\mathbf{x}})$ to be the ranks of the matrix $\mathbf{A}$ and the operator $L_{K,s}^{\mathbf{x}}$, respectively.

In the following theorem, we shall express the empirical eigenpairs of the operator $L_{K,s}^{\mathbf{x}}$ in terms of the eigenpairs of a matrix.

Theorem 3.1 *Let $d = \mathrm{rank}(\mathbf{A})$. Denote all eigenvalues of $\mathbf{A}$ as $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_d > \lambda_{d+1} = \cdots = \lambda_m = 0$, and the corresponding orthonormal eigenvectors as $u_1, u_2, \ldots, u_m$. Then $\mathrm{rank}(L_{K,s}^{\mathbf{x}}) = \mathrm{rank}(\mathbf{A})$, and the empirical eigenpairs $\{(\lambda_l^{\mathbf{x}}, \phi_l^{\mathbf{x}})\}_{l=1}^{d}$ of $L_{K,s}^{\mathbf{x}}$ can be computed in terms of the eigenpairs of $\mathbf{A}$ as follows:*

$$\lambda_l^{\mathbf{x}} = \frac{\lambda_l}{m(m-1)}, \quad \phi_l^{\mathbf{x}} = \frac{1}{\sqrt{\lambda_l}} \sum_{j=1}^{m} d_j (u_l)_j K_{x_j}.$$

Proof By the definitions of $L_{K,s}^{\mathbf{x}}$ and $\phi_l^{\mathbf{x}}$, we have

$$\begin{aligned}
L_{K,s}^{\mathbf{x}}(\phi_l^{\mathbf{x}}) &= \frac{1}{m(m-1)\sqrt{\lambda_l}} \sum_{j=1}^{m} \sum_{i,p=1}^{m} \omega(x_i - x_p) K_{x_i} (x_p - x_i)^2 d_j (u_l)_j K(x_i, x_j) \\
&= \frac{1}{m(m-1)\sqrt{\lambda_l}} \sum_{i=1}^{m} K_{x_i} \sum_{j=1}^{m} \left(\sum_{p=1}^{m} \omega(x_i - x_p)(x_p - x_i)^2 \right) K(x_i, x_j) d_j (u_l)_j \\
&= \frac{1}{m(m-1)\sqrt{\lambda_l}} \sum_{i=1}^{m} K_{x_i} \sum_{j=1}^{m} b_i K(x_i, x_j) d_j (u_l)_j \\
&= \frac{1}{m(m-1)\sqrt{\lambda_l}} \sum_{i=1}^{m} K_{x_i} d_i \sum_{j=1}^{m} d_i K(x_i, x_j) d_j (u_l)_j \\
&= \frac{1}{m(m-1)\sqrt{\lambda_l}} \sum_{i=1}^{m} K_{x_i} d_i \lambda_l (u_l)_i \\
&= \frac{\lambda_l}{m(m-1)} \frac{1}{\sqrt{\lambda_l}} \sum_{i=1}^{m} d_i (u_l)_i K_{x_i} = \lambda_l^{\mathbf{x}} \phi_l^{\mathbf{x}}
\end{aligned}$$

and

$$\begin{aligned}
\langle \phi_p^{\mathbf{x}}, \phi_q^{\mathbf{x}} \rangle_K &= \frac{1}{\sqrt{\lambda_p \lambda_q}} \sum_{i,j=1}^{m} \langle d_i (u_p)_i K_{x_i}, d_j (u_q)_j K_{x_j} \rangle_K \\
&= \frac{1}{\sqrt{\lambda_p \lambda_q}} \sum_{i,j=1}^{m} d_i (u_p)_i K(x_i, x_j) d_j (u_q)_j \\
&= \frac{1}{\sqrt{\lambda_p \lambda_q}} (Du_p)^T K (Du_q) = \frac{1}{\sqrt{\lambda_p \lambda_q}} u_p^T D^T K D u_q
\end{aligned}$$

$$= \frac{1}{\sqrt{\lambda_p \lambda_q}} u_p^T \mathbf{A} u_q = \sqrt{\frac{\lambda_q}{\lambda_p}} u_p^T u_q = \delta_{p,q}.$$

Therefore, the numbers $\{\lambda_l^{\mathbf{x}}\}_{l=1}^d$ are eigenvalues of $L_{K,s}^{\mathbf{x}}$ with corresponding orthonormal eigenfunctions $\{\phi_l^{\mathbf{x}}\}_{l=1}^d$, and $\mathrm{rank}(L_{K,s}^{\mathbf{x}}) \geq \mathrm{rank}(\mathbf{A})$.

On the other hand, let $t = \mathrm{rank}(L_{K,s}^{\mathbf{x}})$. Then, for $1 \leq l \leq t$, it follows from $L_{K,s}^{\mathbf{x}}(\phi_l^{\mathbf{x}}) = \lambda_l^{\mathbf{x}} \phi_l^{\mathbf{x}}$ that

$$\frac{1}{m(m-1)} \sum_{i,j=1}^{m} \omega(x_i - x_j) K(x_i, x_p)(x_j - x_i)^2 \phi_l^{\mathbf{x}}(x_i)$$
$$= \frac{1}{m(m-1)} \sum_{i=1}^{m} K(x_i, x_p) b_i \phi_l^{\mathbf{x}}(x_i) = \lambda_l^{\mathbf{x}} \phi_l^{\mathbf{x}}(x_p), \quad 1 \leq p \leq m.$$

Let $\phi_l^{\mathbf{x}} |_{\mathbf{x}} = (\phi_l^{\mathbf{x}}(x_1), \ldots, \phi_l^{\mathbf{x}}(x_m))^T$. Then

$$\frac{1}{m(m-1)} \mathbf{K}\mathbf{B} \phi_l^{\mathbf{x}} |_{\mathbf{x}} = \lambda_l^{\mathbf{x}} \phi_l^{\mathbf{x}} |_{\mathbf{x}},$$
$$\frac{1}{m(m-1)} \mathbf{D}\mathbf{K}\mathbf{D}^2 \phi_l^{\mathbf{x}} |_{\mathbf{x}} = \lambda_l^{\mathbf{x}} \mathbf{D} \phi_l^{\mathbf{x}} |_{\mathbf{x}},$$
$$\frac{1}{m(m-1)} \mathbf{A}\mathbf{D} \phi_l^{\mathbf{x}} |_{\mathbf{x}} = \lambda_l^{\mathbf{x}} \mathbf{D} \phi_l^{\mathbf{x}} |_{\mathbf{x}}.$$

Now, for $1 \leq p, q \leq m$, we have

$$\begin{aligned}
\delta_{p,q} \lambda_p^{\mathbf{x}} &= \langle L_{K,s}^{\mathbf{x}}(\phi_p^{\mathbf{x}}), \phi_q^{\mathbf{x}} \rangle_{\mathbf{K}} \\
&= \frac{1}{m(m-1)} \langle \sum_{i,j=1}^{m} \omega(x_i - x_j) K_{x_i} (x_j - x_i)^2 \phi_p^{\mathbf{x}}, \phi_q^{\mathbf{x}} \rangle_{\mathbf{K}} \\
&= \frac{1}{m(m-1)} \sum_{i,j=1}^{m} \omega(x_i - x_j)(x_j - x_i)^2 \phi_p^{\mathbf{x}}(x_i) \phi_q^{\mathbf{x}}(x_i) \\
&= \frac{1}{m(m-1)} \sum_{i=1}^{m} \phi_p^{\mathbf{x}}(x_i) \phi_q^{\mathbf{x}}(x_i) \sum_{j=1}^{m} \omega(x_i - x_j)(x_j - x_i)^2 \\
&= \frac{1}{m(m-1)} \sum_{i=1}^{m} \phi_p^{\mathbf{x}}(x_i) b_i \phi_q^{\mathbf{x}}(x_i) = \frac{1}{m(m-1)} \sum_{i=1}^{m} (d_i \phi_p^{\mathbf{x}}(x_i))(d_i \phi_q^{\mathbf{x}}(x_i)) \\
&= \frac{1}{m(m-1)} \langle \mathbf{D} \phi_p^{\mathbf{x}} |_{\mathbf{x}}, \mathbf{D} \phi_q^{\mathbf{x}} |_{\mathbf{x}} \rangle.
\end{aligned}$$

It follows that for $1 \leq l \leq t$, $\mathbf{D}\phi_l^{\mathbf{x}} |_{\mathbf{x}}$ are the orthonormal eigenvector system of $\mathbf{A}$, and $\mathrm{rank}(\mathbf{A}) \geq \mathrm{rank}(L_{K,s}^{\mathbf{x}})$. The proof of the theorem is now completed. □

Remark 3.2 According to the proof of Theorem 3.1, we know that the eigenfunctions $\{\phi_l^{\mathbf{x}}\}_{l=1}^d$ satisfy the following two properties:

- $\frac{1}{m(m-1)} \sum_{i,j=1}^{m} \omega(x_i - x_j)(x_j - x_i)^2 \phi_p^{\mathbf{x}}(x_i) \phi_q^{\mathbf{x}}(x_i) = \delta_{p,q} \lambda_p^{\mathbf{x}}$.
- If $\lambda_l^{\mathbf{x}} = 0$, then $\phi_l^{\mathbf{x}}(x_i)(x_j - x_i) = 0$.

4. Representer Theorem

The following theorem provides the solution to problem (4) explicitly.

Theorem 4.1 *For $1 \le l \le m$, denote*

$$S_l^{\mathbf{z}} = \begin{cases} \frac{1}{m^2\lambda_l^{\mathbf{x}}} \sum_{i,j=1}^m \omega^{(s)}(x_i - x_j)(y_i - y_j)\phi_l^{\mathbf{x}}(x_i)(x_j - x_i), & \text{if } \lambda_l^{\mathbf{x}} > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Then the solution to problem (4) *is given by*

$$c_{\gamma,l}^{\mathbf{z}} = \begin{cases} 0, & \text{if } 2\lambda_l^{\mathbf{x}}|S_l^{\mathbf{z}}| \le \gamma, \\ -\frac{m}{(m-1)}\left(S_l^{\mathbf{z}} - \frac{\gamma}{2\lambda_l^{\mathbf{x}}}\right), & \text{if } 2\lambda_l^{\mathbf{x}}|S_l^{\mathbf{z}}| > \gamma \text{ and } S_l^{\mathbf{z}} > \frac{\gamma}{2\lambda_l^{\mathbf{x}}}, \\ -\frac{m}{(m-1)}\left(S_l^{\mathbf{z}} + \frac{\gamma}{2\lambda_l^{\mathbf{x}}}\right), & \text{if } 2\lambda_l^{\mathbf{x}}|S_l^{\mathbf{z}}| > \gamma \text{ and } S_l^{\mathbf{z}} < -\frac{\gamma}{2\lambda_l^{\mathbf{x}}}. \end{cases} \tag{5}$$

Proof Let $\omega_{i,j} = \omega(x_i - x_j)$. By using Remark 3.2, we can reduce the empirical error part in algorithm (4) as follows:

$$\begin{aligned}
&\frac{1}{m^2}\sum_{i,j=1}^m \omega_{i,j}\left(y_i - y_j + \sum_{l=1}^m c_l\phi_l^{\mathbf{x}}(x_i) - (x_j - x_i)\right)^2 \\
&= \frac{1}{m^2}\sum_{i,j=1}^m \omega_{i,j}\left(\sum_{l=1}^m c_l\phi_l^{\mathbf{x}}(x_i)(x_j - x_i)\right)^2 + 2(y_i - y_j)\sum_{l=1}^m c_l\phi_l^{\mathbf{x}}(x_i)(x_j - x_i) + (y_i - y_j)^2 \\
&= \frac{1}{m^2}\sum_{i,j=1}^m \omega_{i,j}\sum_{l=1}^m c_l\phi_l^{\mathbf{x}}(x_i)(x_j - x_i)^2\sum_{l=1}^m c_l\phi_l^{\mathbf{x}}(x_i) + \frac{1}{m^2}\sum_{i,j=1}^m \omega_{i,j}(y_i - y_j)^2 + \\
&\quad \frac{2}{m^2}\sum_{i,j=1}^m \omega_{i,j}(y_i - y_j)\sum_{l=1}^m c_l\phi_l^{\mathbf{x}}(x_i)(x_j - x_i) \\
&\quad \frac{1}{m^2}\sum_{p,q=1}^m c_p c_q\sum_{i,j=1}^m \omega_{i,j}\phi_p^{\mathbf{x}}(x_i)(x_j - x_i)^2\phi_q^{\mathbf{x}}(x_i) + \frac{1}{m^2}\sum_{i,j=1}^m \omega_{i,j}(y_i - y_j)^2 + \\
&\quad \frac{2}{m^2}\sum_{l=1}^m c_l\sum_{i,j=1}^m \omega_{i,j}(y_i - y_j)\phi_l^{\mathbf{x}}(x_i)(x_j - x_i) \\
&= \frac{1}{m^2}\sum_{l=1}^m c_l^2\lambda_l^{\mathbf{x}}m(m-1) + \frac{1}{m^2}\sum_{i,j=1}^m \omega_{i,j}(y_i - y_j)^2 + \\
&\quad \frac{2}{m^2}\sum_{l=1}^m c_l\sum_{i,j=1}^m \omega_{i,j}(y_i - y_j)\phi_l^{\mathbf{x}}(x_i)(x_j - x_i) \\
&= \frac{m-1}{m}\sum_{l=1}^m c_l^2\lambda_l^{\mathbf{x}} + 2\sum_{l=1}^m \lambda_l^{\mathbf{x}}S_l^{\mathbf{z}}c_l + \frac{1}{m^2}\sum_{i,j=1}^m \omega_{i,j}(y_i - y_j)^2.
\end{aligned}$$

We now have an equivalent form of the algorithm as

$$c_\gamma^{\mathbf{z}} = \arg\min_{c\in\mathbb{R}^m}\sum_{l=1}^m \frac{m-1}{m}\lambda_l^{\mathbf{x}}(c_l + \frac{m}{(m-1)}S_l^{\mathbf{z}})^2 + \gamma|c_l| \ .$$

It is easy to see that $c_{\gamma,l}^{\mathbf{z}} = 0$ when $\lambda_l^{\mathbf{x}} = 0$. When $\lambda_l^{\mathbf{x}} > 0$, the components $c_{\gamma,l}^{\mathbf{z}}$ can be found by solving the following optimization problem

$$c_{\gamma,l}^{\mathbf{z}} = \arg\min_{c\in\mathbb{R}} \ (c + \frac{m}{(m-1)}S_l^{\mathbf{z}})^2 + \frac{m}{(m-1)}\frac{\gamma}{\lambda_l^{\mathbf{x}}}|c| \ ,$$

which has the solution given by (5). This proves the theorem. □

5. Numerical examples

We present three numerical examples to illustrate the approximating performance for the numerical differentiation. We consider the following functions

$$f_1(x) = x^2 * \exp(-x^2/4), \tag{6}$$

$$f_2(x) = \sin(x) * \exp(-x^2/8), \tag{7}$$

$$f_3(x) = x^2 * \cos(x)/8, \tag{8}$$

$$f_4(x) = x * \sin(x). \tag{9}$$

To estimate the computational error, we choose N test points $\{t_i\}_{i=0}^N$ on the interval $[-4, 4]$ and then compute the errors by using the following two formulae:

$$E_1(f) = \frac{1}{N}\sum_{i=0}^{N} |(f_\gamma^{\mathbf{z}}(t_i) - f'(t_i)|,$$

$$E_2(f) = \sqrt{\frac{1}{N}\sum_{i=0}^{N} (f_\gamma^{\mathbf{z}}(t_i) - f'(t_i))^2}.$$

In the experiments, the points $\{x_i\}_{i=0}^{20}$ are uniformly distributed over $[-4, 4]$, i.e., $x_i = -4 + 0.4i$ $(0 \le i \le 20)$. The parameters s and γ are chosen as 0.1 and 0.001, respectively. The resulting numerical results are shown in Figures 1 and 2. Moreover, the errors are listed in Table 1. From these figures, it could be observed that the function $f_\gamma^{\mathbf{z}}$ matches the derivative function f'_ρ well. Meanwhile, the sparse properties could be explicitly found from the number of non-zero coefficients in Table 1.

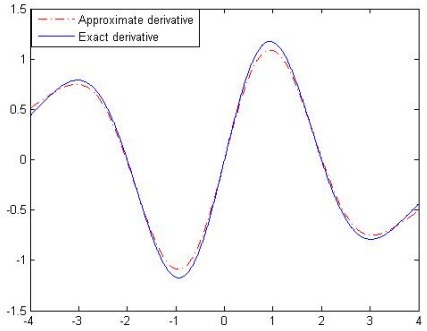
Function f	$E_1(f)$	$E_2(f)$	rate of non-zero coefficients
$f_1(x)$	0.0412	0.0473	10/21
$f_2(x)$	0.0222	0.0257	10/21
$f_3(x)$	0.0489	0.0868	10/21
$f_4(x)$	0.0301	0.0349	10/21

Table 1 Errors

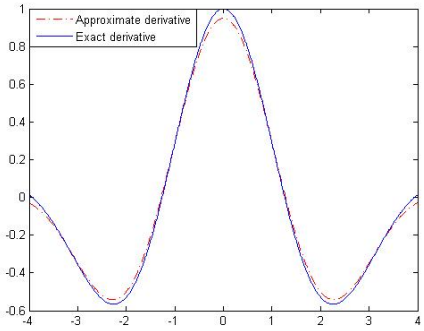
6. Discussion

In this paper, we study a method for numerical differentiation in the framework of statistical learning theory. Based on empirical eigenfunctions, we propose an ℓ^1 regularized algorithm. We present an approach for computing explicitly the empirical eigenfunctions and establish the representer theorem of the algorithm. Compared with traditional methods for numerical differentiation, the output of our method could be considered directly as the derivative of the

underlying function. Moreover, the algorithm could produce sparse representations with respect to empirical eigenfunctions, without assuming sparsity in terms of any basis or system. Finally, this work leaves several open issues for further study. For example, it is interesting to extend our method to the estimation of gradient in high dimensional space.

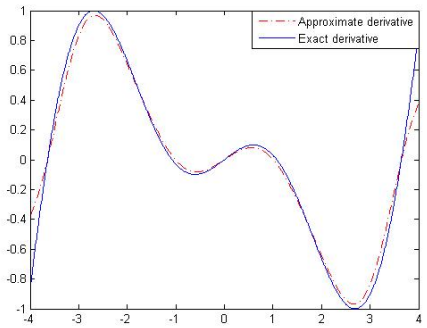


(a)

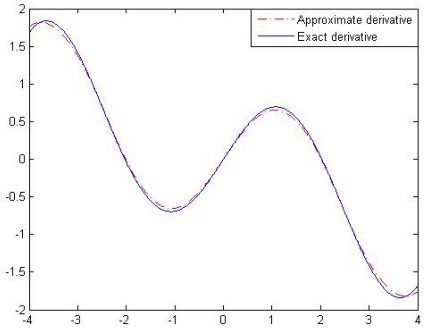


(b)

Figure 1 (a) Approximate derivative of $f_1(x)$ (b) Approximate derivative of $f_2(x)$



(a)



(b)

Figure 2 (a) Approximate derivative of $f_3(x)$ (b) Approximate derivative of $f_4(x)$

Acknowledgements The authors are indebted to the anonymous reviewers for their careful comments and constructive suggestions.

References

[1] Jinquan CHENG, Y. C. HON, Yanbo WANG. *A Numerical Method for the Discontinuous Solutions of Abel Integral Equations*. Amer. Math. Soc., Providence, RI, 2004.
[2] S. R. DEANS. *The Radon Transform and Some of Its Applications*. Courier Corporation, 2007.
[3] E. G. HAUG. *The Complete Guide to Option Pricing Formulas*. McGraw-Hill Companies, 2007.
[4] M. HANKE, O. SCHERZER. *Error analysis of an equation error method for the identification of the diffusion coefficient in a quasi-linear parabolic differential equation*. SIAM J. Appl. Math., 1999, **59**(3): 1012–1027.
[5] A. N. TIKHONOV, V. Y. ARSENIN. *Solutions of Ill-Posed Problems*. Washington, DC: Winston, 1977.
[6] R. S. ANDERSSEN, M. HEGLAND. *For numerical differentiation, dimensionality can be a blessing!*. Math. Comp., 1999, **68**(227): 1121–1141.
[7] T. J. RIVLIN. *Optimally stable Lagrangian numerical differentiation*. SIAM J. Numer. Anal., 1975, **12**(5): 712–725.
[8] J. CULLUM. *Numerical differentiation and regularization*. SIAM J. Numer. Anal., 1971, **8**(2): 254–265.

[9] Shuai LU, S. PEREVERZEV. *Numerical differentiation from a viewpoint of regularization theory*. Math. Comp., 2006, **75**(256): 1853–1870.
[10] Ting WEI, Y. C. HON, Yanbo WANG. *Reconstruction of numerical derivatives from scattered noisy data*. Inverse Problems, 2005, **21**(2): 657–672.
[11] N. ARONSZAJN. *Theory of reproducing kernels*. Trans. Amer. Math. Soc., 1950, **68**(3): 337–404.
[12] F. CUKER, Dingxuan ZHOU. *Learning Theory: An Approximation Theory Viewpoint*. Cambridge University Press, Cambridge, 2007.
[13] S. MUKHERJEE, Dingxuan ZHOU. *Learning coordinate covariances via gradients*. J. Mach. Learn. Res., 2006, **7**: 519–549.
[14] E. J. CANDÈS, J. ROMBERG, T. TAO. *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*. IEEE Trans. Inform. Theory, 2006, **52**(2): 489–509.
[15] Hongyan WANG, Quanwu XIAO, Dingxuan ZHOU. *An approximation theory approach to learning with ℓ^1 regularization*. J. Approx. Theory, 2013, **167**: 240–258.