Thai Journal of Mathematics
Volume 14 (2016) Number 2 : 517–526
http://thaijmath.in.cmu.ac.th
ISSN 1686-0209

Alternative Approximation Method for Learning Multiple Feature

Kannika Khompurngson[†,1] **and Suthep Suantai**[‡]

[†]Division of Mathematics, School of Science University of Phayao,
Phayao 56000, Thailand
e-mail : kannika.kh@up.ac.th

[‡]Department of Mathematics, Faculty of Science, Chiang Mai
University, Chiang Mai 50200, Thailand
e-mail : suthep.s@cmu.ac.th

Abstract : The theory of reproducing kernel Hilbert space (RKHS) has recently appeared as a powerful framework for the learning problem. The principal goal of the learning problem is to determine a functional which best describes given data. Recently, we have extended the hypercircle inequality to data error in two ways: First, we have extended it to circumstance for which all data is known within error. Second, we have extended it to partially-corrupted data. That is, data set contains both accurate and inaccurate data. In this paper, we report on further computational experiments by using the material from both previous work.

Keywords : hypercircle inequality; convex optimization; reproducing kernel Hilbert space.
2010 Mathematics Subject Classification : 46E22; 46C07.

1 Introduction

The theory of reproducing kernel Hilbert space (RKHS) has recently appeared as a powerful framework for the learning problem. The principal goal of the learning problem is to determine a functional which best describes given data. Specifically, Hypercircle inequality (Hi) has been applied to kernel-based machine leaning

[1]Corresponding author.

[1]. Unfortunately, the material on Hi only applies to circumstance for which data is accurate data. Recently, we have extended the hypercircle inequality to data error in two ways: First, we have extended it to circumstance for which all data is known within error [2, 3]. Second, we have extended it to partially-corrupted data [4]. That is, data set contains both accurate and inaccurate data. In this paper, we report on further computational experiments by using the available material from from both previous work.

Let $T = \{t_j : j \in \mathbb{N}_n\} \subseteq \mathcal{T}$ be finite set of points where $\mathcal{T}$ is prescribed and we use the notation $\mathbb{N}_n = \{1, 2, ..., n\}$. Let $d = (d_j : j \in \mathbb{N}_n) \in \mathbb{R}^n$ be an *inaccurate* data representation of $f(t_j)$ where $f : \mathcal{T} \longrightarrow \mathbb{R}$ is the functional in the a hypothesis space H which is assumed to be a reproducing kernel Hilbert space over the real numbers (RKHS). That is, for all $t \in \mathcal{T}$ and $f \in H$, the reproducing kernel $K : \mathcal{T} \times \mathcal{T} \longrightarrow \mathbb{R}$ is defined by

$$f(t) = \langle K(t, \cdot), f \rangle.$$

The Aronszajn-Moore theorem, [5], states that K is a reproducing kernel for some RKHS if and only if for any inputs $T = \{t_j : j \in \mathbb{N}_n\}$ the $n \times n$ matrix $(K(t_i, t_j) : i, j \in \mathbb{N}_n)$ is positive semi-definite.

Given $t_{-2}, t_{-1}, t_0 \in \mathcal{T}$, we want to estimate $f(t_{-2}), f(t_{-1})$ and $f(t_0)$ knowing that $\| f \|_K \leq 1$ and the data error vector $e := (d_j - f(t_j) : j \in \mathbb{N}_n)$ has the Euclidean norm $| \cdot |_2 \leq \varepsilon$ where ε is any positive number. In our case, we still consider the method of regularization for learning the values $f(t_{-2}), f(t_{-1})$ and $f(t_0)$. That is, we choose the function which minimizes from the functional R_ρ defined for $f \in H$ as

$$R_\rho(f) := | d - If |_2^2 + \rho \| f \|_K^2 \tag{1.1}$$

where $If := (f(t_j) : j \in \mathbb{N}_n)$ and ρ is any positive number. The Representer Theorem say that the learned function has the form

$$f_\rho(t) = \sum_{j \in \mathbb{N}_n} c_j K(t_j, t), \quad t \in \mathcal{T} \tag{1.2}$$

for some real vector $c = (c_j : j \in \mathbb{N}_n)$ which is determined by minimizing (1.1) over all functions of the form (1.2); see for example [6, 7] We then choose $f_\rho(t_{-2}), f_\rho(t_{-1}), f_\rho(t_0)$ as regularization estimators. Our goal is to estimate the values of $f(t_{-2}), f(t_{-1})$ and $f(t_0)$ by using the material from Hypercircle inequality for data error and compare the result to the regularization estimator.

The remainder of this paper is presented as follows. In section II, we briefly describe the extension of hypercircle inequality to inaccurate data and partially-corrupted data respectively. Section III contains some results of numerical experiments, discussion of learning the value of a function in a RKHS by our proposed midpoint algorithm.

2 Preliminaries

In this section we briefly review the extension of hypercircle inequality to inaccurate data and partially-corrupted data respectively.

2.1 Hypercircle Inequality for Data Error

Let H be the Hilbert space over the real numbers with inner product $\langle \cdot, \cdot \rangle$. Given a finite set of *linearly independent* vectors $\mathcal{X} = \{x_j : j \in \mathbb{N}_n\}$ in H, we define $M := \left\{\sum_{j\in\mathbb{N}_n} a_j x_j : (a_j : j \in \mathbb{N}_n) \in \mathbb{R}^n\right\}$ which is the $n-$dimensional linear subspace of H spanned by the vectors in $\mathcal{X}$. Let $Q : H \longrightarrow \mathbb{R}^n$ be a linear operator H onto $\mathbb{R}^n$ which is defined for $x \in H$ as

$$Qx = (\langle x, x_j \rangle : j \in \mathbb{N}_n). \tag{2.1}$$

Consequently, the adjoint map $Q^T : \mathbb{R}^n \longrightarrow H$ is given at $a = (a_j : j \in \mathbb{N}_n) \in \mathbb{R}^n$ as

$$Q^T a = \sum_{j\in\mathbb{N}_n} a_j x_j \tag{2.2}$$

and the Gram matrix of the vectors in $\mathcal{X}$ is

$$G = QQ^T = (\langle x_j, x_l \rangle : j, l \in \mathbb{N}_n). \tag{2.3}$$

Moreover, G is positive definite matrix. For each $d \in \mathbb{R}^n$, it is well-known that there exist a unique vector $x(d) \in M$ such that

$$x(d) := Q^T(G^{-1}d) := arg\min\{||x|| : x \in H, Qx = d\} \tag{2.4}$$

and $||x(d)|| = \sqrt{(d, G^{-1}d)}$. Consequently, let us recall the definition of *hypercircle* as follows

$$\mathcal{H}(d) := \{x : ||x|| \leq 1, Q(x) = d\}.$$

We remark that $\mathcal{H}(d) \neq \emptyset$ if and only if $||x(d)|| = \sqrt{(d, G^{-1}d)} \leq 1$. If $H \neq M$ then $\mathcal{H}(d)$ consists of exactly one point if and only if $||x(d)|| = 1$. In particular, the *Hypercircle inequality* is given by the following.

Theorem 2.1. *If $x(d) \in \mathcal{H}(d)$ and $x_0 \in H$ then for any $x \in \mathcal{H}(d)$*

$$|\langle x(d), x_0 \rangle - \langle x, x_0 \rangle| \leq \text{dist}(x_0, M)\sqrt{1 - ||x(d)||^2},$$

where dist $(x_0, M) := \min\{||x_0 - y|| : y \in M\}$.

Alternatively, the value of $\langle x(d), x_0 \rangle$ is the best estimator to estimate $\langle x, x_0 \rangle$ when $x \in \mathcal{H}(d)$. Geometrically speaking, the best estimator $\langle x(d), x_0 \rangle$ is the mid-point of the interval of uncertainty which is defined by $I(x_0, d) := \{\langle x, x_0 \rangle : x \in \mathcal{H}(d)\}$.

Now let us review basic fact about Hypercircle inequality for data error and discuss what we need for section 3. We begin with $E_2 = \{e : e \in \mathbb{R}^n, |e|_2 \leq \varepsilon\}$, where $|\cdot|_2 : \mathbb{R}^n \to \mathbb{R}_+$ is the Euclidean norm on $\mathbb{R}^n$ and $\varepsilon > 0$. For each $d \in \mathbb{R}^n$, the definition of the *hyperellipse* is given by

$$\mathcal{H}(d|E_2) := \{x : x \in B, Q(x) - d \in E_2\}.$$

Therfore, the Hypercircle inequality for data error becomes in the following way.

Theorem 2.2. *If $x_0 \in H$ and $\mathcal{H}(d|E_2) \neq \emptyset$ then there is a $e_0 \in E_2$ such that for any $x \in \mathcal{H}(d|E_2)$*

$$|\langle x(d+e_0), x_0\rangle - \langle x, x_0\rangle| \leq \frac{1}{2}\Big(m_+(x_0, d|E_2) - m_-(x_0, d|E_2)\Big),$$

where $x(d+e_0) = Q^T\big(G^{-1}(d+e_0)\big) \in \mathcal{H}(d|E_2)$, $m_+(x_0, d|E_2) := \max\big\{\langle x, x_0\rangle : x \in \mathcal{H}(d|E_2)\big\}$ and $m_-(x_0, d|E_2) := \min\big\{\langle x, x_0\rangle : x \in \mathcal{H}(d|E_2)\big\} = -m_+(x_0, -d|E_2)$ respectively.

Similarly, the best estimator to estimate $\langle x, x_0\rangle$ is the midpoint of the following uncertainty interval $I(x_0, d|E_2) := \{\langle x, x_0\rangle : x \in \mathcal{H}(d|E_2)\}$. To find the best predictor, we provided the useful duality formula for the right hand endpoint of the uncertainty interval. We then define the convex function $V : \mathbb{R}^n \longrightarrow \mathbb{R}$ defined for $c \in \mathbb{R}^n$

$$V(c) := ||x_0 - Q^T(c)|| + \varepsilon|c|_2 + (d, c),$$

The result state as the following.

Theorem 2.3. *If $\mathcal{H}(d) \neq \emptyset$ then*

$$m_+(x_0, d|E) = \min\big\{V(c) : c \in \mathbb{R}^n\big\}. \tag{2.5}$$

Moreover, $0 = arg\min\{V(c) : c \in \mathbb{R}^n\}$ if and only if $\dfrac{x_0}{||x_0||} \in \mathcal{H}(d|E)$.

The detailed proof will appear in [2, 8, 9]. Alternatively, to find the best predictor, we only need evaluate the two numbers $m_+(x_0, \pm d|E)$ and then compute $\frac{1}{2}(m_+(x_0, d|E_2) - m_+(x_0, -d|E_2))$. We provided a possible iterative method to solve the minimum vector c^* proceeds in the following manner, [3]. Let us introduce two positive constants given $\rho_k := |c^k|_2$ and $\tau_k := ||x_0 - Q^T c^k||$. Next, we choose an initial vector $c^0 \neq 0$ and then successively define $c^k, k \in \mathbb{N}$, by the formula

$$c^{k+1} = (\rho_k G + \varepsilon\tau_k I)^{-1}(\rho_k Q x_0 - \rho_k \tau_k d). \tag{2.6}$$

Our computation experience indicates that this iteration converges if the vector Qx_0 and d are linearly independent in $\mathbb{R}^n$. However, this has not been proved.

2.2 Hypercircle Inequality for Partially-Corrupted Data

Let $I \subseteq \mathbb{N}_n$ which contains m elements ($m < n$). Consequently, we use the notations $\mathcal{X}_I = \{x_i : i \in I\} \subset \mathcal{X}$ and $\mathcal{X}_J = \{x_i : i \in J\} \subset \mathcal{X}$, where we denote $J = \mathbb{N}_n \backslash I$. For each $e = (e_1, ..., e_n) \in \mathbb{R}^n$, we also use the notations $e_I = (e_i : i \in I) \in \mathbb{R}^m$ and $e_J = (e_i : i \in J) \in \mathbb{R}^{n-m}$ respectively. We choose $||| \cdot |||_2 : \mathbb{R}^{n-m} \longrightarrow \mathbb{R}_+$ is the Euclidean norm on $\mathbb{R}^{n-m}$ and define $\mathbb{E}_2 = \{e : e \in \mathbb{R}^n : e_I = 0, |||e_J||| \leq \varepsilon\}$, where ε is some positive number. For each $d \in \mathbb{R}^n$, we define the *partial hyperellipse* as follows

$$\mathcal{H}(d|\mathbb{E}_2) := \big\{x : \ x \in H, \ ||x|| \leq 1, \ Q(x) - d \in \mathbb{E}_2\big\}. \tag{2.7}$$

In this case, we also provided the existence of the best estimator which *still* has the form of linear combination of vectors in $\mathcal{X}$ and the results follows by the same method as in [2]. Again, we provided the useful duality formula for the right hand endpoint of the uncertainty interval and the midpoint is given by $\frac{1}{2}\Big(m_+(x_0, d|\mathbb{E}) - m_+(x_0, -d|\mathbb{E})\Big)$. The result state as the following.

Theorem 2.4. *If $\mathcal{H}(d|\mathbb{E}_2)$ contains more than one element then*

$$m_+(x_0, d|\mathbb{E}_2) = \min\big\{||x_0 - Q^T(c)|| + \varepsilon|||c_J|||_2 + (d, c) : c \in \mathbb{R}^n\big\}, \tag{2.8}$$

Moreover, we provided the necessary and sufficient condition on $\mathcal{H}(d|\mathbb{E}_2)$ such that the minimum vector c^* achieves with $c^*_J = 0$ which is useful for practice. To this end, let us define the convex function $\mathbb{V} : \mathbb{R}^n \longrightarrow \mathbb{R}$ defined for $c \in \mathbb{R}^n$

$$\mathbb{V}(c) := ||x_0 - Q^T(c)|| + \varepsilon|||c_J|||_2 + (d, c).$$

Theorem 2.5. *If $x_0 \notin M_I := \big\{Q^T_I(a) : a \in \mathbb{R}^m\big\}$ and $\mathcal{H}(d_I)$ contain more than one point then $c^* = arg\ \min\{\mathbb{V}(c) : c \in \mathbb{R}^n\}$ with $c^*_J = 0$ if and only if*

$$\frac{x_0 - Q^T_I(a^*)}{||x_0 - Q^T_I(a^*)||} \in \mathcal{H}(d|\mathbb{E})$$

where the vector

$$\frac{x_0 - Q^T_I(a^*)}{||x_0 - Q^T_I(a^*)||} := arg \min\{\langle x, x_0\rangle : x \in \mathcal{H}(d_I)\}.$$

The detailed proof will appear in [4]. As we already said, we going to apply the results in this section in learning problem. We then provide a possible iterative method to solve the minimum vector c^* on the right hand side in equation (2.8) when data error is measured with square loss. We choose an initial vector $c^0 \neq 0$ with $c^0_J \neq 0$ and then successively define $c^k, k \in \mathbb{N}$, by the formula

$$c^{k+1} = (G + \tau^k D^k)^{-1}(Qx_0 - \tau^k d) \tag{2.9}$$

where $\tau^k := ||x_0 - Q^T c^k||$ and the matrix D^k is an $n \times n$ diagonal matrix and we define the elements on diagonal by

$$d_{ii}^k = \begin{cases} 0, & \text{if } i \in I \\ \dfrac{\varepsilon}{\rho_J^k}, & \text{if } i \in J \end{cases} \tag{2.10}$$

where $\rho_J^k := |||c_J^k|||_2$.

3 Computation

In this section, we shall apply the material from previous section to the problem of learning the values of a function in a reproducing kernel Hilbert space(RKHS). They have an origin in the theory of reproducing kernel in the classical paper of Golomb and Weinberger [5, 10]. Specifically, we choose the gaussian kernel on $\mathbb{R}^2$, that is

$$K(x,y) := e^{-|x-y|_2^2}, \quad x, y \in \mathbb{R}^2.$$

In our example, we choose the value of $T = \{t_j : j \in \mathbb{N}_{20}\}$ on a ellipse curve surrounding the origin. Consequently, we have a finite set of linearly independent elements $\{K_{t_j} : j \in \mathbb{N}_{20}\}$ in H where $K_{t_j}(t) := K(t_j, t), j \in \mathbb{N}_n$ and $t \in \mathbb{R}^2$. Therefore, the vectors $\{x_j : j \in \mathbb{N}_{20}\}$ appearing in previous section are identified with the function $\{K_{t_j} : j \in \mathbb{N}_{20}\}$. These vectors determine a linear operator $Q : H \longrightarrow \mathbb{R}^n$ defined for $f \in H$ as

$$Qf = \Big(\langle f, K_{t_j}\rangle = f(t_j) : j \in \mathbb{N}_{20}\Big).$$

Moreover, the Gram matrix of the $\{K_{t_j} : j \in \mathbb{N}_{20}\}$ is given by

$$G := (K(t_i, t_j) : i, j \in \mathbb{N}_{20}).$$

Next, we choose the exact function is given by the formula

$$g(t) = 3.5K_{(1,1)}(t) + 1.75K_{(1,-1)}(t) + 3.25K_{(-1,-1)}(t) - 3.5K_{(-1,1)}(t), \quad t \in \mathbb{R}^2$$

and compute the vector $d := (g(t_j) : j \in \mathbb{N}_{20})$. Given $t_{-2}, t_{-1}, t_0 \in \mathbb{R}^2$, we want to estimate $f(t_{-2}), f(t_{-1}), f(t_0)$ knowing that $||f||_K \leq \delta$ and the data error vector $e := (d_j - f(t_j)) : j \in \mathbb{N}_{20}$ has Euclidean norm $\leq \varepsilon$ and δ are prescribed. However, with no effort at all the observations we made so far extend to the case that the unit ball B is replaced by δB where δ is any positive number. In addition, we shall compare the midpoint estimator discussed in previous section to the regularization estimator which is the stand method for learning problem. The computational steps are organized in the following steps:

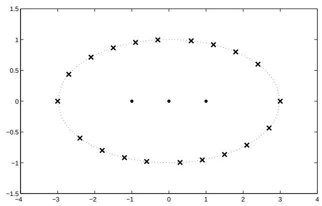

Figure 1. Point on a ellipse curve and estimating points.

Step 1. Given $\rho > 0$, we define the quadratic functional R_ρ for $f \in H$ as

$$R_\rho(f) := \sum_{i=1}^{n} \big(d_i - f(t_i)\big)^2 + \rho ||f||_K^2.$$

The Representer Theorem say that the unique function with minimizes R_ρ has the form

$$f_\rho(t) := \sum_{j \in \mathbb{N}_n} c_\rho(j) K(t_{t_j}, t), \quad t \in \mathbb{R}^2$$

and $c_\rho = (G + \rho I)^{-1} d$. Next, we introduce two functions of ρ given by the formula

$$\varepsilon(\rho) := |d - Q f_\rho|_2, \quad \delta(\rho) := ||f_\rho||_K \tag{3.1}$$

and choose $f_\rho(t_{-2}), f_\rho(t_{-1}), f_\rho(t_0)$ as regularization estimators.

Step 2. To compare midpoint estimator and estimate $f(t_0)$, we need to identify an *hyperellipse* which contains f_ρ. To this end, we define the hyperellipse as follows $\mathcal{H}(d|\delta(\rho)E) := \{f : ||f||_K \leq \delta(\rho), Q(f) - d \in E\}$ where $E = \{e : e \in \mathbb{R}^n, |e| \leq \varepsilon(\rho)\}$. Clearly, the regularization estimator f_ρ can be view as an element in hyperelipse

Since hyperllipse $\mathcal{H}(d|\delta(\rho)E)$ consists one point our strategy compare the regularization and midpoint estimator must consider a bigger hyperllipse. Consequently, we then compute both the regularization estimator and midpoint estimator corresponding to this hyperellipse $\mathcal{H}(d|\delta E)$ where $\delta = 3\delta(\rho)$ and compare to the true value of g at t_0.

As explained earlier, the midpoint algorithm requires us to find numerically the minimum of the function in (2.5) for d and $-d$. then our midpoint estimator is given by $m(t_0) = \frac{m_+(t_0, d|\delta E) - m_+(t_0, -d|\delta E)}{2}$. For the computation of $m_+(t_0, \pm d|\delta E)$ we use the program fminunc in the optimization toolbox of Matlab 7.3.0 and also for comparison sake we use the iteration scheme (2.6) with an arbitrary chosen non zero initial vector.

Step 3. To estimate $f(t_{-1})$, we define the *partial hyperellipse* as follows

$$\mathcal{H}(\mathrm{d}|\delta\mathbb{E}) := \{f : ||f||_K \leq \delta, f(t_0) = f_\rho(t_0), Q(f) - d \in E\}.$$

Clearly, the regularization estimator f_ρ can be view as an element in $\mathcal{H}(\mathrm{d}|\delta\mathbb{E})$. Moreover, we define the quadratic functional as $x \in H$

$$S_\rho(f) := \sum_{i=1}^{n} \big(d_i - f(t_i)\big)^2 + \big(f_\rho(t_0) - f(t_0)\big)^2 + \rho||f||_K^2.$$

It is easy to check that the function which minimize this functional over H is $f_\rho(t)$. That is, the regularization estimator is also given by $f_\rho(t_{-1})$.

To obtain the midpoint, the algorithm requires us to find numerically the minimum of the function in (2.8) for d and $-$d where the vector $\mathrm{d} := (\mathrm{d}_j : j \in \mathbb{N}_{20} \cup \{0\})$, $\mathrm{d}_0 = f_\rho(t_0)$ and $\mathrm{d}_j = d_j$ for $j \in \mathbb{N}_{20}$. Our midpoint estimator is given by $m(t_{-1}) = \frac{m_+(t_{-1},\mathrm{d}|\delta\mathbb{E}) - m_+(t_{-1},-\mathrm{d}|\delta\mathbb{E})}{2}$. For the computation of $m_+(t_{-1}, \pm\mathrm{d}|\delta\mathbb{E})$ we use the program fminunc in the optimization toolbox of Matlab 7.3.0 and also for comparison sake (2.9) with an arbitrary chosen non zero initial vector.

Step 4. To estimate $f(t_{-2})$, we again define *partial hyperellipse* as follows

$$\mathcal{H}(\mathbf{d}|\delta\mathbb{E}') := \{f : ||f||_K \le \delta, f(t_{-1}) = f_\rho(t_{-1}), f(t_0) = f_\rho(t_0), Q(f) - d \in E\}.$$

Clearly, the regularization estimator f_ρ can be view as an element in $\mathcal{H}(\mathbf{d}|\delta\mathbb{E}')$ and the regularization estimator is also given by $f_\rho(t_{-2})$ which follows by the same method as in step 3. Our midpoint estimator becomes as the following $m(t_{-2}) = \frac{m_+(t_{-2},\mathbf{d}|\delta\mathbb{E}') - m_+(t_{-2},-\mathbf{d}|\delta\mathbb{E}')}{2}$ where the vector $\mathbf{d} := (\mathbf{d}_j : j \in \mathbb{N}_{20} \cup \{0, -1\})$ where $\mathbf{d}_{-1} = f_\rho(t_{-1}), \mathbf{d}_0 = f_\rho(t_0)$ and $\mathbf{d}_j = d_j$ for $j \in \mathbb{N}_{20}$.

The results of the computation are displayed below and at least for small values of the regularization parameter, but away from zero, the midpoint algorithm is better than regularization.

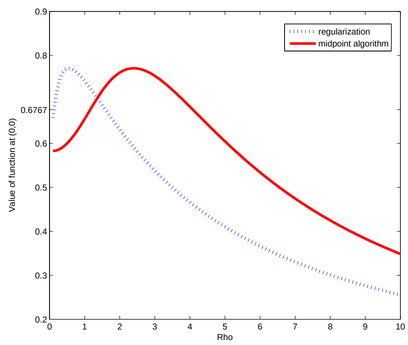

Figure 2. The result of regularization and midpoint algorithm at the point t_0 and exact value is $g(t_0) = 0.6767$

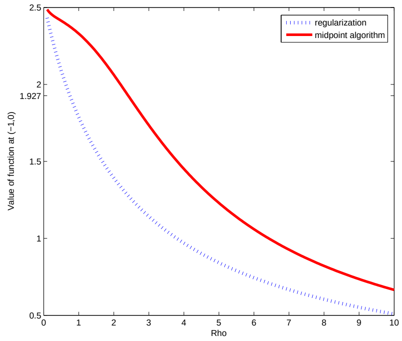

Figure 3.The result of regularization and midpoint algorithm at the point t_{-1} and exact value is $g(t_{-1}) = 1.927$

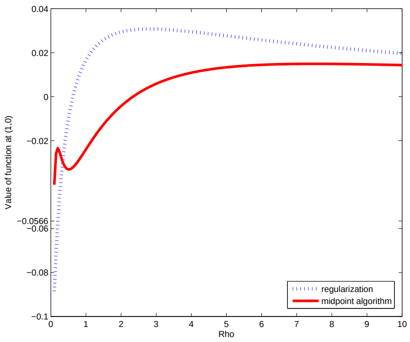

Figure 4.The result of regularization and midpoint algorithm at the point t_{-2} and exact value is $g(t_{-2}) = -0.0566$

Acknowledgements : The author is supported by the Thailand Research Fund, the Commission on Higher Education and University of Phayao under Grant TRG5680030.

References

[1] C.A Micchelli, T.J Rivlin, A survey of optimal recovery, Optimal Estimation in Approximation Theory, Plenum Press (1977), 1-53

[2] K. Khompurngson, C.A. Micchelli, Hide, Jaen Journal on Approximation, 3 (2011) 87-115

[3] K Khompurngson, B Novaprateep, Y. Lenbury, Learning the value of a function from inaccurate data, East-Weast Journal of Mathematics Special 5, (2011)

[4] K. Khompurngson, B. Novaprateep, Hypercircle inequlaity for partially-corrupted data, Annals of Functional Analysis, Ann. Funct. Anal. 6 (1) (2015) 95-108.

[5] N. Aronszajn, Theory of reproducing kernels, Trans. Amer. Math. Soc. 68 (1950) 337-404

[6] T. Evergniou, T. Poggio, M. Pontil, Regularization networks and support vector machines, Advances in Computational Mathematics, 13 (2000) 1-50.

[7] B. Scholkopf, A.J. Smola, Learning with Kernels, The MIT Press, Cambridge, MA, USA, 2002.

[8] P.J. Davis, Interpolation and Approximation, Dover Publications, New York, 1975.

[9] H.L. Royden, Real Analysis, Macmillan Publishing Company, 3rd edition, New York, 1988.

[10] M. Golomb, H.F. Weinberger, Optimal Approximation and Error Bounds, In R. E. Langer, editor, The University of Wisconsin Press, 1959.

(Received 9 August 2015)
(Accepted 12 July 2016)