Some Remarks on Hypercircle Inequality for Inaccurate Data and Its Application

K. Khompurngson, B. Novaprateep and Y. Lenbury

Abstract—In a recent paper, we extended the hypercircle inequality for data error and applied our results to the problem of learning the value of a function from inaccurate data in the reproducing kernel Hilbert space. In the present paper we continue to present some other recent results on this subject within this approach.

Keywords—hypercircle inequality, reproducing kernel Hilbert space, convex optimization and noise data.

I. Introduction

MOST of the previous studies on learning problem has focused on finding the best function representation from data. There are several methods that can be used to determine a learned function which best describes given data [2], [5], [6], [12]. Specifically, the well-known hypercircle inequality has been applied to kernel- based learning when data is known exactly [3], [4], [11]. Therefore, our previous work has extended it to the circumstance for which data is known within error [9]. In this paper, we continue to present some other recent results on this subject. The first objective is to consider the case that data error is measured with different error tolerance. Specifically, we consider the case that data error is measured with square loss. Moreover, we provide two importance cases of the existence of the minimum of the convex function which is used to obtain the best predictor. The second objective is to report on further computational experiment of learning the value of a function from partial corruption data in the reproducing kernel Hilbert space .

Specifically, the theory of reproducing kernel Hilbert space (RKHS) has recently emerged as a powerful framework for the learning problem. A reproducing kernel Hilbert space is a Hilbert space of functions with special properties [1]. It plays an important role in approximation and regularization theory as it allows us to write in a simple way the solution of learning from empirical data problem. However, the choice of kernel is critical to the success of many learning algorithms but it is typically left to the user.

Given an input set $\mathcal{T}$, we assume H to be a reproducing kernel Hilbert space over the real numbers (RKHS). We recall that an RKHS is a Hilbert space of real-valued functions everywhere defined on $\mathcal{T}$. Corresponding to Hilbert space H is a reproducing kernel $K : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ such that for all $t \in \mathcal{T}$ and $t \in H$

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

The Aronszajn's theory of reproducing kernel Hilbert spaces states that a function $K : \mathcal{T} \times \mathcal{T} \longrightarrow \mathbb{R}$ is a reproducing kernel if it is symmetric, that is $K(s,t) = K(t,s)$, and positive definite:

$$\sum_{i,j=1}^{n} a_j a_i K(t_j, t_i) \geq 0$$

for any $n \in \mathbb{N}$ and the choice of inputs $T = \{t_j : j \in \mathbb{N}_n\} \subseteq \mathcal{T}$ and $a = (a_1, ..., a_n) \in \mathbb{R}^n$ where we use the notation $\mathbb{N}_n = \{1, 2, ..., n\}$. This useful theorem allows us to specify a hypothesis space by $choosing\ K$.

Let $d = (d_j : j \in \mathbb{N}_n) \in \mathbb{R}^n$ be an inaccurate representation of $f(t_j)$ where $f : \mathcal{T} \to \mathbb{R}$ is a functional representation in H. Given $t_0 \in \mathcal{T}$, we want to estimate $f(t_0)$ knowing that

$$||f||_K \leq 1$$

and the data error $e := If - d$ is measured with some norm on $\mathbb{R}^n$. where we define

$$If := (f(t_i) = \langle f, K_{t_i}\rangle : i \in \mathbb{N}_n).$$

As we state earlier, there are several methods that can be used to determine a function which best describes given data. Specifically, hypercircle inequality (Hi) has been applied to kernel-based machine leaning. Unfortunately, Hi has only applied to circumstances for which data is known exactly. Recently, Kannika Khompurngson and Charles A. Micchelli have extended it to inaccurate data and constructed a new learning method [8] [9] [13] . In fact, the method is described with a abstract Hilbert space setting. This framework is also specific to the practically important case of reproducing kernel Hilbert space.

According to the midpoint algorithm in our previous work, the best estimator to learn $f(t_0)$ is the midpoint of the uncertainty interval as follows

$$I(t_0, d) = \{f(t_0) : ||f||_K \leq 1, ||If - d|| \leq \varepsilon\}.$$

In addition, we showed that the best estimator still has the form of a linear combination of the functions $K(t_j, \cdot), t_i \in T$.

K. Khompurngson is with Division of Mathematics, School of Science, University of Phayao, THAILAND and Centre of Excellence in Mathematics, PERDO, CHE, THAILAND e-mail: kannika.kh@up.ac.th.

B. Novaprateep and Y. Lenbury are with Deparment of Mathematics, Faculty of Science, Mahidol University, THAILAND and Centre of Excellence in Mathematics, PERDO, CHE, THAILAND e-mail: scbnv@mahidol.ac.th.

That is, we have that

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

for some real vector $c = (c_j : j \in \mathbb{N}_n)$.

This paper is organized as follows. In Section II, we restrict our attention to the study of hypercircle inequality for data error when it has different empirical data error. Specifically, we consider the case that data error is measured with square loss and it has different empirical data error. Moreover, we provide two importance cases of the existence of the minimum of the convex function which is used to obtain the right hand endpoint of the uncertainty interval and provide possible iteration method to solve for such function. In Section III, we briefly review the recent result on hypercircle inequality for partial corruption data and we discuss some numerical experiment on learning the value of a function from partial corruption data in the reproducing kernel Hilbert space which will appear in Section IV.

II. Hypercircle Inequality for Data Error with Different Error Tolerance

Let H be a Hilbert space over the real numbers with inner product $\langle \cdot, \cdot \rangle$. We choose a finite set of *linearly independent* elements $\mathcal{X} = \{x_j : j \in \mathbb{N}_n\}$ in H. Consequently, let M be the $n-$dimensional subspace of H spanned by the vectors in $\mathcal{X}$. That is, we have that

$$M := \{\sum_{i \in \mathbb{N}_n} a_i x_i : a \in \mathbb{R}^n\}.$$

Let $Q : H \to \mathbb{R}^n$ be a linear operator H onto $\mathbb{R}^n$ which is defined for any $x \in H$ as

$$Q(x) = \big(\langle x, x_j \rangle : j \in \mathbb{N}_n\big)$$

Since $\{x_j : j \in \mathbb{N}_n\}$ is linearly independent, we obtain that Q is onto $\mathbb{R}^n$. That is, for all $d \in \mathbb{R}^n$ there is an $x(d) \in H$ such that

$$Q(x(d)) = d.$$

Consequently, the adjoint map $Q^T : \mathbb{R}^n \to H$ is given by

$$Q^T(a) = \sum_{j \in \mathbb{N}_n} a_i x_i$$

Therefore, the Gram matrix of the vector in $\mathcal{X}$ is given by

$$\begin{aligned} G &= \big(\langle x_j, x_l \rangle : j, l \in \mathbb{N}_n\big) \\ &= \begin{bmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle & ... & \langle x_1, x_n \rangle \\ \langle x_2, x_1 \rangle & \langle x_2, x_2 \rangle & ... & \langle x_2, x_n \rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle x_n, x_1 \rangle & \langle x_n, x_2 \rangle & ... & \langle x_n, x_n \rangle \end{bmatrix} \\ &= QQ^T \end{aligned}$$

We remark that G is symmetric and positive definite and then we obtain the vector $x(d)$ as

$$x(d) := Q^T\big(G^{-1}d\big).$$

Before we review the results of Hypercircle inequality for data error with different error tolerance, let us introduce the following notations. We start with $I \subseteq \mathbb{N}_n$ which contains m elements ($m < n$). We define $|\cdot|_2 : \mathbb{R}^m \longrightarrow \mathbb{R}_+$ and $|||\cdot|||_2 : \mathbb{R}^{n-m} \longrightarrow \mathbb{R}_+$ as Euclidean norms on $\mathbb{R}^m$ and $\mathbb{R}^{n-m}$ respectively.

For each $e = (e_1, ..., e_n) \in \mathbb{R}^n$, we use the notations

$$e_I = (e_i : i \in I) \text{ and } e_J = (e_i : i \in J).$$

where we denote $J = \mathbb{N}_n \backslash I$. We assume that

$$E_\infty = \{e : e \in \mathbb{R}^n, |e|_\infty \leq 1\}$$

where we use $|\cdot|_\infty$ as following

$$|e|_\infty = \max\big\{\frac{1}{\varepsilon}|e_I|_2, \frac{1}{\varepsilon'}|||e_J|||_2\big\}$$

where $\varepsilon, \varepsilon' > 0$. In this case, we define the *hyperellipse* in the following way.

$$\mathcal{H}(d|E_\infty) = \{x : x \in B, |Q(x) - d|_\infty \leq 1\} \tag{1}$$

where B is a unite ball in H. That is, for each $x \in \mathcal{H}(d|\mathbb{E})$

$$|(Q(x) - d)_I|_2 \leq \varepsilon \text{ and } |||(Q(x) - d)_J|||_2 \leq \varepsilon'.$$

In the special case that $\varepsilon = 0$, we know that the hyperellipse becomes to *hypercircle* [3] as shown below

$$\mathcal{H}(d) = \{x : x \in B, Qx = d\}.$$

First, we point out that $\mathcal{H}(d|E_\infty) \neq \emptyset$ if

$$||x(d)||^2 = (d, G^{-1}d) \leq 1.$$

Before we add some relations, let us introduce the notations for the linear operators:

$$Q_I(x) := \big(\langle x_j, x \rangle : j \in I\big) \in \mathbb{R}^m$$

and

$$Q_J(x) := \big(\langle x_j, x \rangle : j \in J\big) \in \mathbb{R}^{n-m}.$$

According to (1), we observe that

$$\mathcal{H}(d|E_\infty) = \mathcal{H}(d_I|E_I) \cap \mathcal{H}(d_J|E_J) \tag{2}$$

where we denote

$$\mathcal{H}(d_I|E_I) = \big\{x : x \in B, |Q_I(x) - d_I|_2 \leq \varepsilon\big\}$$

and

$$\mathcal{H}(d_J|E_J) = \big\{x : x \in B, |||Q_J(x) - d_J|||_2 \leq \varepsilon'\big\}.$$

Alternatively, we consider here the following point of view. Given $x_0 \in H$, we want to estimate $\langle x, x_0 \rangle$ when it is known that $x \in \mathcal{H}(d|E_\infty)$. We point out that $\mathcal{H}(d|E_\infty)$ is a convex subset of H which is sequentially compact in the weak topology on H [14].

Consequently, we obtain that the uncertainty interval

$$I(x_0, d|E_\infty) := \{\langle x, x_0\rangle : x \in \mathcal{H}(d|E_\infty)\}$$

is a bounded and closed interval in $\mathbb{R}$. Consequently, we define

$$m_+(x_0, d|E_\infty) = \sup\{\langle x, x_0\rangle : x \in \mathcal{H}(d|E_\infty)\}$$

and

$$m_-(x_0, d|E_\infty) = \inf\{\langle x, x_0\rangle : x \in \mathcal{H}(d|E_\infty)\}$$

respectively. Clearly, the best estimator is the midpoint of the uncertainty interval and we use the following notation $m(x_0, d|E_\infty)$.

Before we provide the duality formula for the right hand endpoint, let us recall the conjugate norm of $|\cdot|$ which is defined for all $c \in \mathbb{R}^n$ as

$$|c|_* = \max_{\substack{w\in\mathbb{R}^n \\ |w|\le 1}} (c, w).$$

These facts can be found in [7].

Moreover, if $c \neq 0$ then there is a $\hat{c} \in \mathbb{R}^n$ such that $|\hat{c}| = 1$ and $|c|_* = (c, \hat{c})$. Therefore, the conjugate norm $|\cdot|_\infty$ is given, for each $e \in \mathbb{R}^n$, by

$$|e|_1 = \varepsilon|e_I|_2 + \varepsilon'|||e_J|||_2.$$

Theorem 1: If $\mathcal{H}(d|E_\infty)$ contains more than one point then

$$m_+(x_0, d|E_\infty) = \min_{c\in\mathbb{R}^n} V_2(c)$$

where the function

$$V_2(c) := ||x_0 - Q^T c|| + \varepsilon|c_I|_2 + \varepsilon'|||c_J|||_2 + (d, c)$$

for all $c \in \mathbb{R}^n$.

Proof. see [9]

In additional, we provide the necessary and sufficient condition on $\mathcal{H}(d|E_\infty)$ which provide that V_2 achieves its minimum at c^* with $c^*_I \neq 0$ and $c^*_J \neq 0$. Let us recall a useful theorem [9] before providing the proof of the following facts.

Theorem 2: If $\mathcal{H}(d_I|E_I)$ contains more than one point and

$$x_0 \notin M_I := \{Q_I^T(a) : a \in \mathbb{R}^m\}$$

then

$m_+(x_0, d_I|E_I) =$

$$\min\{||x_0 - Q_I^T a|| + \varepsilon|a|_2 + (a, d_I) : a \in \mathbb{R}^m\}.$$

where we use the notation

$$m_+(x_0, d_I|E_I) = \max\{\langle x, x_0\rangle : x \in \mathcal{H}(d_I|E_I)\}.$$

Moreover, the minimum $a^* \in \mathbb{R}^m$ is unique and

$$x_+(d_I|E_I) := \frac{x_0 - Q_I^T(a^*)}{||x_0 - Q_I^T(a^*)||}$$

satisfies

$$x_+(d_I|E_I) := arg\min\{\langle x, x_0\rangle : x \in \mathcal{H}(d_I|E_I)\}.$$

Proof. see [9]

The following theorem shows the necessary and sufficient condition on $\mathcal{H}(d|E_\infty)$ which provides that V_2 achieves its minimum at c^* with $c^*_J = 0$.

Theorem 3: If $x_0 \notin M_I$ and $\mathcal{H}(d_I|E_I)$ contains more than one point then

$$c^* = arg\ \min\{V_2(c) : c \in \mathbb{R}^n\} \text{ with } c^*_J = 0$$

if and only if

$$\frac{x_0 - Q_I^T(a^*)}{||x_0 - Q_I^T(a^*)||} \in \mathcal{H}(d|E_\infty)$$

Proof. We begin by proving $\dfrac{x_0 - Q_I^T(a^*)}{||x_0 - Q_I^T(a^*)||} \in \mathcal{H}(d|E_\infty)$.

First, we observe that

$$c^* = arg\ \min\{V_2(c) : c \in \mathbb{R}^n\} \text{ with } c^*_J = 0$$

if and only if for all $c \in \mathbb{R}^n$

$$\begin{aligned} & ||x_0 - Q_I^T(a^*)|| + +\varepsilon|a^*|_* + (a^*, d_I) \\ = \ & \min\{||x_0 - Q_I^T(a)|| + \varepsilon|a|_* + (a, d_I) : a \in \mathbb{R}^m\} \\ \le \ & V_2(c). \end{aligned}$$

Since the function $\mathbb{V}$ is a convex this inequality holds if and only if for all $c \in \mathbb{R}^n$ with $c_I = a^*$ we obtain that

$$-\varepsilon'|||c_J|||_* - (c_J, d_J) \le$$

$$\inf\left\{\frac{||x_0 - Q_I^T(a^*) - \lambda Q_J^T(c_J)|| - ||x_0 - Q_I^T(a^*)||}{\lambda} : \lambda > 0\right\}$$

which means for all $c_J \in \mathbb{R}^{n-m}$ that

$$-\varepsilon'|||c_J|||_* - (c_J, d_J) \le -\Big(\frac{Q_J(x_0 - Q_I^T(a^*))}{||x_0 - Q_I^T(a^*)||}, c_J\Big).$$

That is , we have that

$$\Big(\frac{Q_J(x_0 - Q_I^T(a^*))}{||x_0 - Q_I^T(a^*)||} - d_J, c_J\Big) \le \varepsilon'|||c_J|||_*.$$

Therefore, we have that

$$\frac{x_0 - Q_I^T(a^*)}{||x_0 - Q_I^T(a^*)||} \in \mathcal{H}(d|E_\infty).$$

Conversely, for each $x \in \mathcal{H}(d|E_\infty) = \mathcal{H}(d_I|E_I) \cap \mathcal{H}(d_J|E_J)$ we observe that

$$\langle x, x_0\rangle \le m_+(x_0, d_I|E_I).$$

This means we have that

$$m_+(x_0, d|E_\infty) \le m_+(x_0, d_I|E_I).$$

Since $\dfrac{x_0 - Q_I^T(a_I^*)}{||x_0 - Q_I^T(a_I^*)||} \in \mathcal{H}(d|E_\infty)$ and

$$m_+(x_0, d_I|E_I) = ||x_0 - Q_I^T(a^*)|| + \varepsilon|a^*|_* + (a^*, d_I),$$

we obtain that

$$m_+(x_0, d|E_\infty) = ||x_0 - Q_I^T(a^*)|| + \varepsilon|a^*|_2 + (a^*, d_I)$$

which completes the proof. □

Similarly, we also have the following theorem which provides the different hypothesis which ensures that V_2 achieves its minimum at c^* with $c_I^* = 0$.

Theorem 4: If $\mathcal{H}(d_J|E_J)$ contains more than one point and

$$x_0 \notin M_J := \left\{Q_J^T(b) : b \in \mathbb{R}^{n-m}\right\}$$

then

$$c^* = arg\ \min\{V_2(c) : c \in \mathbb{R}^n\} \text{ with } c_I^* = 0$$

if and only if

$$\frac{x_0 - Q_J^T(b^*)}{||x_0 - Q_J^T(b^*)||} \in \mathcal{H}(d|E_\infty)$$

where the vector $b^* \in \mathbb{R}^{n-m}$ is the unique minimum of the following function

$$b \to ||x_0 - Q_J^T(b)|| + \varepsilon|||b|||_2 + (b, d_I)$$

and

$$x_+(d_J|E_J) := \frac{x_0 - Q_J^T(b^*)}{||x_0 - Q_I^T(b^*)||}$$

satisfies $x_+(d_J|E_J) := arg\min\{\langle x, x_0\rangle : x \in \mathcal{H}(d_J|E_J)\}$.

Proof. This follows by the same method as in the proof of Theorem 3.

The following theorem is the main result of this section.

Theorem 5: If $\mathcal{H}(d|E_\infty)$ contains more than one point, $x_0 \notin M$ and

$$\frac{x_0 - Q_I^T(a^*)}{||x_0 - Q_I^T(a^*)||}, \frac{x_0 - Q_J^T(b^*)}{||x_0 - Q_J^T(b^*)||} \notin \mathcal{H}(d|E_\infty)$$

then

$$m_+(x_0, d) = \min\{V_2(c) : c \in \mathbb{R}^n\}. \tag{3}$$

Moreover, the minimum $c^* \in \mathbb{R}^n$ is the unique solution of the nonlinear equation

$$-Q\left(\frac{x_0 - Q^T c^*}{||x_0 - Q^T c^*||}\right) + w + d = 0 \tag{4}$$

where the vector $w \in \mathbb{R}^n$ and

$$w = \begin{cases} \dfrac{\varepsilon c_I^*}{|c_I^*|_2}, & \text{if } i \in I \\ \\ \dfrac{\varepsilon' c_J^*}{|||c_J^*|||_2}, & \text{if } i \in J. \end{cases} \tag{5}$$

Proof. Let c^* be the unique minimum of V_2. Our hypothesis guarantees that $c_I^* \neq 0$ and $c_J^* \neq 0$ and $x_0 \neq Q^T c^*$. Hence, computing the gradient of V_2 gives equation (4). □

In summary, to obtain the best predictor in hyperellipse $\mathcal{H}(d|E_\infty)$ for the $\langle x, x_0\rangle$ requires the solution of two nonlinear optimization problems in (3). That is, we obtain that

$$m(x_0, d|E_\infty) = \frac{m_+(x_0, d|E_\infty) - m_+(x_0, -d|E_\infty)}{2}.$$

A possible iterative method to solve equation (4) proceeds in the following manner. We introduce the matrix D which is an $n \times n$ diagonal matrix and we define the elements on the diagonal by

$$d_{ii} = \begin{cases} \dfrac{\varepsilon}{\rho_I}, & \text{if } i \in I \\ \\ \dfrac{\varepsilon'}{\rho_J}, & \text{if } i \in J. \end{cases} \tag{6}$$

where $\rho_I := |c_I^*|_2$ and $\rho_J := |||c_J^*|||_2$ and rewrite the equation (4) in the equivalent form

$$c^* = (G + \tau D)^{-1}(Qx_0 - \tau d).$$

where $\tau = ||x_0 - Q^T c^*||$.

We choose an initial vector $c^0 \neq 0$ and then successively define $c^k, k \in \mathbb{N}$, by the formula

$$c^{k+1} = (G + \tau^k D^k)^{-1}(Qx_0 - \tau^k d) \tag{7}$$

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 the diagonal by

$$d_{ii}^k = \begin{cases} \dfrac{\varepsilon}{\rho_I^k}, & \text{if } i \in I \\ \\ \dfrac{\varepsilon'}{\rho_J^k}, & \text{if } i \in J. \end{cases} \tag{8}$$

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

III. Hypercircle Inequality for Partial Corruption Data with Square Loss

As we have described above, Hide has only applied to circumstance for which all data are inaccurate data. In the real situation, there are several types of data and an example of this is the partial corruption data. Therefore, we have extended the hypercircle inequality for partial corruption data in a recent work [10].

In this section, we briefly review hypercircle inequality for partial corruption data when data error is measured with square loss and it has different error tolerance. We start with $J_1, J_2 \subseteq J$ which contains m_1, m_2 elements $(m_1, m_2 < n - m)$ and $m_1 + m_2 = n - m$. We define $|\cdot|_2 : \mathbb{R}^{m_1} \longrightarrow \mathbb{R}_+$ and $|||\cdot|||_2 : \mathbb{R}^{m_2} \longrightarrow \mathbb{R}_+$ as Euclidean norms on $\mathbb{R}^{m_1}$ and $\mathbb{R}^{m_2}$, respectively. For each $e = (e_1, ..., e_n) \in \mathbb{R}^n$, we use the notations

$$e_{J_1} = (e_i : i \in J_1) \text{ and } e_{J_2} = (e_i : i \in J_2).$$

We assume that

$$\mathbb{E}_\infty = \{e : e \in \mathbb{R}^n, e_I = 0, |e|_\infty \leq 1\}$$

where we define $|\cdot|_\infty : \mathbb{R}^{n-m} \longrightarrow \mathbb{R}_+$ as follows.

$$|e|_\infty = \max\big\{\frac{1}{\varepsilon}|e_{J_1}|_2, \frac{1}{\varepsilon'}|||e_{J_2}|||_2\big\}.$$

where $\varepsilon, \varepsilon' > 0$.

Similarly, let us introduce the notations for the linear operators.

$$Q_{J_1}(x) := \big(\langle x_j, x\rangle : j \in J_1\big) \in \mathbb{R}^{m_1}$$

and

$$Q_{J_2}(x) := \big(\langle x_j, x\rangle : j \in J_2\big) \in \mathbb{R}^{m_2}.$$

For each $d \in \mathbb{R}^n$, we define the $partial\ \ hyperellipse$ as follows.

$$\mathcal{H}(d|\mathbb{E}_\infty) := \big\{x : \ x \in H, \ ||x|| \le 1, \ Q(x) - d \in \mathbb{E}_\infty\big\} \quad (9)$$

That is, for each $x \in \mathcal{H}(d|\mathbb{E}_\infty)$ we have the following information from data

$$(Q(x) - d))_I = 0$$

$$|(Q(x) - d)_{J_1}|_2 \le \varepsilon \text{ and } |||(Q(x) - d)_{J_2}|||_2 \le \varepsilon'.$$

Clearly, our data set contains both accurate and inaccurate data and there is known different error tolerance of data error.

According to the definition of $hyperellipses$ and $hypercircles$, we observe that

$$\mathcal{H}(d|\mathbb{E}_\infty) = \mathcal{H}(d_I) \cap \mathcal{H}(d_J|E_\infty) \quad (10)$$

where we denote the $hypercircle$ with the constant d_I as

$$\mathcal{H}(d_I) = \big\{x : ||x|| \le 1, Q_I(x) = d_I\big\}$$

and the $hyperellipse$ with the constant d_J as

$$\mathcal{H}(d_J|E_\infty) = \big\{x : ||x|| \le 1, Q_J(x) - d_J \in E_\infty\big\}$$

where $E_\infty = \{e : e \in \mathbb{R}^{n-m}, |e|_\infty \le 1\}$ and we use $|\cdot|_\infty : \mathbb{R}^{n-m} \to \mathbb{R}^+$ as following

$$|e|_\infty = \max\big\{\frac{1}{\varepsilon}|e_{J_1}|_2, \frac{1}{\varepsilon'}|||e_{J_2}|||_2\big\}.$$

Indeed, we obtain that

$$\mathcal{H}(d|\mathbb{E}_\infty) = \mathcal{H}(d_I) \cap \mathcal{H}(d_I|E_{J_1}) \cap \mathcal{H}(d_J|E_{J_2})$$

where we denote the the $hyperellipse$ with the constant d_{J_1} and d_{J_2} as

$$\mathcal{H}(d_{J_1}|E_{J_1}) = \big\{x : ||x|| \le 1, |Q_{J_1}(x) - d_{J_1}| \le \varepsilon\big\}.$$

and

$$\mathcal{H}(d_{J_2}|E_{J_1}) = \big\{x : ||x|| \le 1, |||Q_{J_2}(x) - d_{J_2}||| \le \varepsilon'\big\}$$

respectively.

Next, let us add some remarks when $\mathcal{H}(d|\mathbb{E}_\infty) \neq \emptyset$. According to $Qx(d) = d$, we obtain that if

$$||x(d)||^2 = (d, G^{-1}d) \le 1$$

then $\mathcal{H}(d|\mathbb{E}_\infty) \neq \emptyset$.

Given $x_0 \in H$, we want to estimate $\langle x, x_0\rangle$ when it is known that $x \in \mathcal{H}(d|\mathbb{E}_\infty)$. That is, our data set contains both accurate and inaccurate data. Again, we point out that the $partial\ \ hyperellipse$ is convex subset of H which is sequentially compact in the weak topology on H. Consequently, we obtain that the uncertainty interval

$$I(x_0, d|\mathbb{E}_\infty) := \big\{\langle x, x_0\rangle : x \in \mathcal{H}(d|\mathbb{E}_\infty)\big\}$$

is a bounded and closed interval on $\mathbb{R}$. Therefore, we use the following the notation for the right and left hand endpoints

$$m_+(x_0, d|\mathbb{E}_\infty) = \sup\{\langle x, x_0\rangle : x \in \mathcal{H}(d|\mathbb{E}_\infty)\}$$

and

$$m_-(x_0, d|\mathbb{E}_\infty) = \inf\{\langle x, x_0\rangle : x \in \mathcal{H}(d|\mathbb{E}_\infty)\},$$

respectively. According to the midpoint algorithm again, the midpoint, $m(x_0, d|\mathbb{E}_\infty)$, of the uncertainty interval is the best estimator.

According to the previous section, we only need to evaluate the two numbers $m_\pm(x_0, d|\mathbb{E}_\infty)$ and compute the midpoint $m(x_0, d|\mathbb{E}_\infty)$. Therefore, let us provide the following facts before we show the duality formula for the right hand endpoint. We found that the conjugate norm $|\cdot|_\infty : \mathbb{R}^{n-m} \longrightarrow \mathbb{R}_+$ is also given for each $e \in \mathbb{R}^{n-m}$ by

$$|e|_1 = \varepsilon|e_{J_1}|_2 + \varepsilon'|||e_{J_2}|||_2.$$

Theorem 6: If $\mathcal{H}(d|\mathbb{E}_\infty)$ contains more than one point then

$$m_+(x_0, d|\mathbb{E}_\infty) = \min_{c \in \mathbb{R}^n} \mathbb{V}_2(c)$$

where the function

$$\mathbb{V}_2(c) := ||x_0 - Q^T c|| + \varepsilon|c_{J_1}|_2 + \varepsilon'|||c_{J_2}|||_2 + (d, c)$$

for all $c \in \mathbb{R}^n$.
Proof. see [9].

In this section, we also provide the necessary and sufficient condition on $\mathcal{H}(d|\mathbb{E}_\infty)$ which provides that $\mathbb{V}_2$ achieves its minimum at c^* with $c^*_{J_1} \neq 0$ and $c^*_{J_2} \neq 0$.

To this end, let us introduce the following vectors

$$x_+(d_J|E_{J_1}) := \frac{x_0 - Q^T_{J_1}(a^*)}{||x_0 - Q^T_{J_1}(a^*)||}$$

and

$$x_+(d_J|E_{J_2}) := \frac{x_0 - Q^T_{J_2}(b^*)}{||x_0 - Q^T_{J_2}(b^*)||}.$$

These vectors satisfies

$$x_+(d_{J_1}|E_{J_1}) := arg\min\{\langle x, x_0\rangle : x \in \mathcal{H}(d_{J_1}|E_{J_1})\}$$

and

$$x_+(d_{J_2}|E_{J_2}) := arg\min\{\langle x, x_0\rangle : x \in \mathcal{H}(d_{J_2}|E_{J_2})\}.$$

Now we are ready to state the theorem.

Theorem 7: If $\mathcal{H}(d|\mathbb{E}_\infty)$ contains more than one point, $x_0 \notin M$ and

$$\frac{x_0 - Q_I^T(a^*)}{||x_0 - Q_I^T(a^*)||}, \frac{x_0 - Q_J^T(b^*)}{||x_0 - Q_J^T(b^*)||} \notin \mathcal{H}(d|\mathbb{E}_\infty)$$

then

$$m_+(x_0, d) = \min\{\mathbb{V}_2(c) : c \in \mathbb{R}^n\}. \tag{11}$$

Moreover, the minimum $c^* \in \mathbb{R}^n$ is the unique solution of the nonlinear equation

$$-Q(\frac{x_0 - Q^T c^*}{||x_0 - Q^T c^*||}) + w + d = 0 \tag{12}$$

where the vector $w \in \mathbb{R}^n$ and

$$w = \begin{cases} 0, & \text{if } i \in I \\ \frac{\varepsilon c^*_{J_1}}{|c^*_{J_1}|_2}, & \text{if } i \in J_1 \\ \frac{\varepsilon' c^*_{J_2}}{|||c^*_{J_2}|||_2}, & \text{if } i \in J_2. \end{cases} \tag{13}$$

Proof. This follows by the same method as in the proof of Theorem 5. □

A possible iterative method to solve equation (12) proceeds in the following manner. We introduce the matrix D which is an $n \times n$ diagonal matrix and we define the elements on diagonal by

$$d_{ii} = \begin{cases} 0, & \text{if } i \in I \\ \frac{\varepsilon}{\rho_{J_1}}, & \text{if } i \in J_1 \\ \frac{\varepsilon'}{\rho_{J_2}}, & \text{if } i \in J_2. \end{cases} \tag{14}$$

where $\rho_{J_1} := |c^*_{J_1}|_2$ and $\rho_{J_2} := |||c^*_{J_2}|||_2$ and rewrite the equation (12) in the equivalent form

$$c^* = (G + \tau D)^{-1}(Qx_0 - \tau d).$$

where $\tau = ||x_0 - Q^T c^*||$.

We choose an initial vector $c^0 \neq 0$ and then successively define $c^k, k \in \mathbb{N}$, by the formula

$$c^{k+1} = (G + \tau^k D^k)^{-1}(Qx_0 - \tau^k d) \tag{15}$$

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 the diagonal by

$$d^k_{ii} = \begin{cases} 0, & \text{if } i \in I \\ \frac{\varepsilon}{\rho^k_{J_1}}, & \text{if } i \in J_1 \\ \frac{\varepsilon'}{\rho^k_{J_2}}, & \text{if } i \in J_2. \end{cases} \tag{16}$$

where $\rho^k_{J_1} := |c^k_{J_1}|_2$ and $\rho^k_{J_2} := |||c^k_{J_2}|||_2$.

IV. Numerical Experiments

In this Section, we shall report some results from a numerical experiment on learning the value of a function in $RKHS$ from partial corruption data by the midpoint algorithm.

The computation below is organized in the following way. Let H be the RKHS which contains the real-valued function on some input space $\mathcal{T}$. Given $t_0 \in \mathcal{T}$ and $\{(t_j, d_j) : j \in \mathbb{N}_n\} \subseteq \mathcal{T} \times \mathbb{R}$, we want to estimate $f(t_0)$ knowing that

$$||f||_K \leq \delta \text{ and } f \in \mathcal{H}(d|\mathbb{E}_\infty).$$

Therefore, the vectors $\{x_j : j \in \mathbb{N}_n\}$ appearing in Section 2 are identified with the functions

$$\{K_{t_j} : j \in \mathbb{N}_n\}$$

where $K_{t_j}(t) := K(t, t_j); t \in \mathcal{T}$ and the vector x_0 with the function K_{t_0}.

Therefore, the linear operator $Q : H \to \mathbb{R}^n$ becomes for any $f \in H$,

$$Q(f) = \big(f(t_j) : j \in \mathbb{N}_n\big)$$

Consequently, the adjoint map $Q^T : \mathbb{R}^n \to H$ is given by

$$Q^T(a) = \sum_{j \in \mathbb{N}_n} a_i K(t, t_j)$$

The Gram matrix of the function $\{K_{t_j} : j \in \mathbb{N}_n\}$ is given as

$$\begin{aligned} G &= \big(K(t_l, t_j) : j, l \in \mathbb{N}_n\big) \\ &= \begin{bmatrix} K(t_1,t_1) & K(t_1,t_2) & ... & K(t_1,t_n) \\ K(t_2,t_1) & K(t_2,t_2) & ... & K(t_2,t_n) \\ \vdots & \vdots & \vdots & \vdots \\ K(t_n,t_1) & K(t_n,t_2) & ... & K(t_n,t_n) \end{bmatrix} \end{aligned}$$

We choose a exact function $g \in H$ and randomly choose a vector e representing the "noise" and define

$$d = g(t_j) + e.$$

Next, we choose the I which is the subset of $\mathbb{N}_n$. Consequently, we get that $J = \mathbb{N}_n \setminus I$. We then choose $J_1, J_2 \subseteq J$ which contains m_1, m_2 elements ($m_1, m_2 < n - m$) and $m_1 + m_2 = n - m$. Therefore, we obtain the *partial hyperellipse* as in the following

$$\mathcal{H}(d|\delta\mathbb{E}_\infty) := \big\{f : \ f \in H, \ ||f||_K \leq \delta, \ Q(f) - d \in \mathbb{E}_\infty\big\}$$

where δ is any positive number. That is, for each $f \in \mathcal{H}(d|\mathbb{E}_\infty)$ we have the following information from data

$$f(t_j) - d_j = 0 \text{ for all } j \in I$$

$$|(f(t_j) - d)_{J_1}|_2 \leq \varepsilon \text{ and } |||(f(t_j) - d)_{J_2}|||_2 \leq \varepsilon'.$$

First, we need to evaluate the value of δ such that $\mathcal{H}(d|\delta\mathbb{E}_\infty) \neq \emptyset$. We then consider the norm of the vector $x(d)$

$$||x(d)||^2 = (d, G^{-1}d)$$

where the vector

$$x(d) : f_d(t) = \sum_{i=1}^{n} a_i K(t, t_i)$$

and $a = G^{-1}d$. Therefore we obtain that if we choose the value of $\delta \geq \sqrt{||x(d)||}$ then $\mathcal{H}(d|\mathbb{E}_\infty) \neq \emptyset$.

With no effort at all, the observations we made so far extend to the case that the unit ball B is replaced by δB. Consequently, the duality formula becomes

$$\mathbb{V}_2(c) = \delta\sqrt{K(t_0,t_0) - 2\sum_{j\in\mathbb{N}_n} c_j K(t_0,t_j) + \sum_{i,j\in\mathbb{N}_n} c_i c_j K(t_i,t_j)}$$
$$+\varepsilon\sqrt{\sum_{j\in J_1} c_j^2} + \varepsilon'\sqrt{\sum_{j\in J_2} c_j^2} + \sum_{j\in\mathbb{N}_n} c_j d_j.$$

According to the previous section, we only need to evaluate the two numbers $m_+(x_0, \pm d|\delta\mathbb{E}_\infty)$ and compute the midpoint

$$m(x_0, d|\delta\mathbb{E}_\infty) = \frac{m_+(x_0, d|\delta\mathbb{E}_\infty) - m_+(x_0, -d|\delta\mathbb{E}_\infty)}{2}.$$

For the computation of $\mathbb{V}_2$ we use the program fminunc in the optimization toolbox of Matlab 7.3.0.

The algorithm use to find the value of a function by using the the midpoint estimator has shown below.

The Algorithm

<ins>**Step 1**</ins> : Set $d = g(t_j) + e$ and choose t_0

<ins>**Step 1**</ins> : Calcuate $||f_d||^2 = (d, G^{-1}d)$
Choose $\delta \geq ||f_d|| = \sqrt{(d, G^{-1}d)}$.

<ins>**Step 4**</ins> : Find $m_+(t_0, \pm d|\delta\mathbb{E}_\infty)$, we use the program fminunc in the optimization toolbox of MATLAB 7.3.0.

$$m_+(t_0, \pm d|\delta\mathbb{E}_\infty) = \min_{c\in\mathbb{R}^n} \delta||K_{t_0} - Q^T c|| + \varepsilon|c_{J_1}|_2 + \varepsilon'|||c_{J_2}|||_2 \pm (c, d).$$

<ins>**Step 5**</ins> : Calculate $m(t_0, d|\delta\mathbb{E}_\infty)$ by the formula

$$m(t_0, d|\delta\mathbb{E}_\infty) = \frac{1}{2}\Big(m_+(t_0, d|\delta\mathbb{E}_\infty) - m_+(t_0, -d|\delta\mathbb{E}_\infty)\Big)$$

Example We choose the gaussian kernel on $\mathbb{R}$. That is,

$$K(t,s) := e^{-\frac{(t-s)^2}{10}}$$

where $t, s \in \mathbb{R}$.

and the exact function

$$g(t) := 4e^{-\frac{(t-7.5)^2}{10}} + 2e^{-\frac{(t-2.5)^2}{10}} - 0.5e^{-\frac{(t+2.5)^2}{10}} + 5e^{-\frac{(t+7.5)^2}{10}}.$$

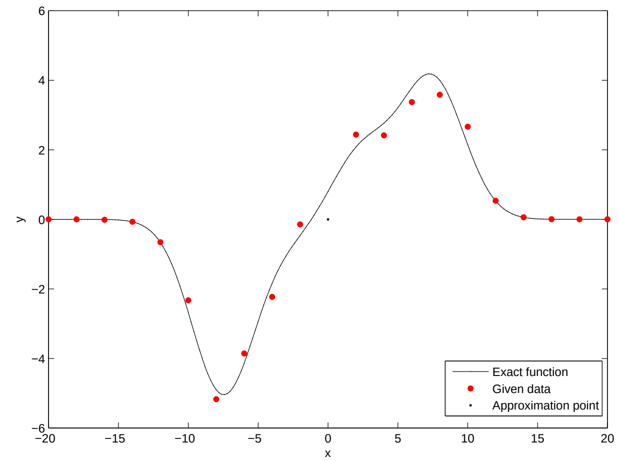

Fig. 1. Graph of the exact function for Gaussian kernel on $\mathbb{R}$ and given points which is known within different error tolerance.

We choose $t_0 = 0$ and generate a training set of twenty points $\{(t_j, d_j) : j \in \mathbb{N}_n\} \subseteq \mathbb{R}\times\mathbb{R}$ obtained by sampling g with noise. That is, we define

$$d = g(t_j) + e.$$

We want to estimate the value of $f(0)$ when we know information from data as following

$$f(t_j) - d_j = 0 \text{ for all } j \in \{6, 7, ..., 15\}$$
$$|(f(t_j) - d)_{J_1}|_2 \leq 0.3 \text{ and } |||(f(t_j) - d)_{J_2}|||_2 \leq 0.5.$$

where $J_1 = \{1, 2, ..., 5\}$ and $J_2 = \{16, 17, ..., 20\}$.

First, we need to evaluate the value of δ such that $\mathcal{H}(d|\mathbb{E}_\infty) \neq \emptyset$. We then consider the norm of the vector $x(d)$

$$f_d(t) = \sum_{i=1}^{n} a_i e^{-\frac{(t-t_i)^2}{10}}$$

and $a = G^{-1}d$. Therefore we obtain that if we choose the value of $\delta \geq 7.6408$ then $\mathcal{H}(d|\delta\mathbb{E}_\infty) \neq \emptyset$.

Consequently, the duality formula to obtain the right and left hand endpoinst become

$$\mathbb{V}_2^{\pm}(c) = \delta\sqrt{1 - 2\sum_{j\in\mathbb{N}_{20}} c_j e^{-\frac{(t_j)^2}{10}} + \sum_{i,j\in\mathbb{N}_{20}} c_i c_j e^{-\frac{(t_i-t_j)^2}{10}}}$$

$$+ 0.3\sqrt{\sum_{j\in J_1} c_j^2} + 0.5\sqrt{\sum_{j\in J_2} c_j^2} \pm \sum_{j\in\mathbb{N}_{20}} c_j d_j.$$

That is, we compute

$$v_{\pm} := \min\{\mathbb{V}_2^{\pm}(c) : c \in \mathbb{R}^n\}$$

and then our midpoint estimator is given by $\frac{v_+ - v_-}{2}$. The results of the computation is indicated in Figure 2 while the exact value $g(0) = 0.7993$.

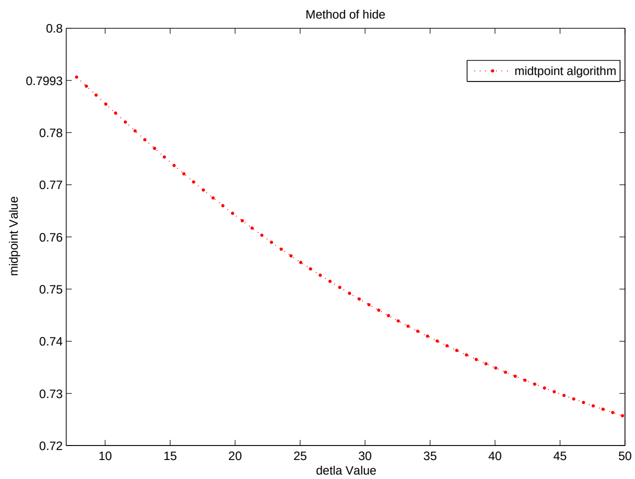

Fig. 2. Midpoint estimator from Gaussian kernel on $\mathbb{R}$

V. Conclusion

In this paper we continue our study on Hide. Specifically, we considered the case that data error is measured with square loss and has different empirical data error. Moreover, we provided two importance cases of the existence of the minimum of the convex functions and also provided possible iteration method to solve for function in equation (4) which is useful for practical. In Section III, we also provided the importance case of the existence of the minimum of the convex functions which is used to obtain the right hand endpoint for the case of partial corruption data. In addition, we reported some numerical experiment on learning the value of a function from partial corruption data in the reproducing kernel Hilbert space.

Acknowledgements:

This research is supported by the Centre of Excellence in Mathematics, the Commission on Higher Education, Thailand.

References

[1] N. Aronszajn, Theory of reproducing kernels, *Transactions American Mathematical Society,* 686, 1950, pp.337–404.

[2] O.H. Choon, L.C. Hoong, T.S. Huey, A functional approximation comparison between neural networks and polynomial regressio, *WSEAS Transactions on Mathematics*, vol. 6, 2008, pp.353-363.

[3] P. J. Davis, Interpolation and Approximation, *Dover Publications,Inc., New York*, 1975.

[4] M. Golomb, H. F. Weinberger, Optimal approximation and error bounds, *In R. E. Langer, editor, The University of Wisconsin Press*. 1959, pp.117-190.

[5] G. Gnecco, M. Sanguineti, Weight-decay regularization in Reproducing Kernel Hilbert Spaces by variable-basis schemes, *WSEAS Transactions on Mathematics*. vot.8, 2009, pp.1109-2769.

[6] U. Hamarik, T. Raus, Choice of the regularization parameter in ill-posed problems with rough estimate of the noise level of data, *WSEAS Transactions on Mathematics*. vol.4, 2005, pp.76-81.

[7] R. A. Horn, C. R. Johnson, Matrix Analysis, *Cambridge University Press*, 1985.

[8] K. Khompurngson, The Hypercircle Inequality for Inaccurate data, *PhD thesis, Mahidol University, Bangkok,* Thailand.

[9] K. Khompurngson, C. A. Micchelli, Hide, *Jean Journal on Approximation,* Vol. 3, 2011, pp.87-115

[10] K. Khompurngson, B. Novaprateep, Hypercircle inequality for partial corruption data, Banach Journal of Mathematical Analysis (submitted).

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

[12] M.R.Norazan , A.H.Habshah Midi, M. R. Imon, Estimating Regression Coefficients using Weighted Bootstrap with Probability, *WSEAS Transactions on Mathematics,* vol. 8, 2009, pp.362-371

[13] B. Novaprateep, K. Khompurngson and D. Poltem, Learning the value of a function by using Hypercircle inequlity for data error, *International Journal of Mathematics and Computers in Simulation,* issue 1, Vol.5, 2011.

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