Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH Open Access

# A note on hypercircle inequality for data error with $l^1$ norm

Kannika Khompungson[1*] and Kamonrat Nammanee[1]

*Correspondence:
kannika.kh@up.ac.th
[1]Division of Mathematics, School of Science, University of Phayao, Phayao, 56000, Thailand

**Abstract**

In our previous work, we have extended the hypercircle inequality (HI) to situations where the data error is known. Furthermore, the most recent result is applied to the problem of learning a function value in the reproducing kernel Hilbert space. Specifically, a computational experiment of the method of hypercircle, where the data error is measured with the $l^p$ norm ($1 < p \leq \infty$), is compared to the regularization method, which is a standard method of the learning problem. Despite this breakthrough, there is still a significant aspect of data error measure with the $l^1$ norm to consider in this issue. In this paper, we do not only explore the hypercircle inequality for the data error measured with the $l^1$ norm, but also provide an unexpected application of hypercircle inequality for only one data error to the $l^\infty$ minimization problem, which is a dual problem in this case.

**MSC:** 54A05

**Keywords:** Hypercircle inequality; Convex optimization; Noise data

## 1 Introduction

Many inequalities are known in connection with the approximation problem. In 1947 the hypercircle inequality has been applied to boundary value problems in mathematical physics [11]. In 1959, Golomb and Weinberger demonstrated the relevance of the hypercircle inequality (HI) to a large class of numerical approximation problems [5]. At present the method of hypercircle, which has a long history in applied mathematics, has received attention by mathematicians in several directions [4, 12]. In 2011, Khompurngson and Micchelli [6] described HI and its potential application to kernel-based learning when the data is known exactly and then extended it to situation where there is known data error (Hide). Furthermore, the most recent result is applied to the problem of learning a function value in the reproducing kernel Hilbert space. Specifically, a computational experiment of the method of hypercircle, when data error is measured with the $l^p$ norm ($1 < p \leq \infty$), is compared to the regularization method, which is a standard method of learning problem [6, 8]. We continue our research on this topic by presenting a full analysis of hypercircle inequality for data error (Hide) measured with the $l^\infty$ norm [10]. Despite this breakthrough, there is still a significant aspect of data error measure with the $l^1$ norm to consider in this issue. In this paper, we do not only explore the hypercircle inequality

for data error measured with the $l^1$ norm, but also provide an unexpected application of hypercircle inequality for only one data error to the $l^\infty$ minimization problem, which is a dual problem in this case.

Recently, we are specifically interested in a detailed analysis of the hypercircle inequality for data error (Hide) measured with the $l^\infty$ norm [10]. Given a set of *linearly independent* vectors $X = \{x_j : j \in \mathbb{N}_n\}$ in a real Hilbert space $H$ with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$, where $\mathbb{N}_n = \{1, 2, \ldots, n\}$. The Gram matrix of the vector in $X$ is

$$G = \big(\langle x_i, x_j \rangle : i, j \in \mathbb{N}_n\big).$$

We define the linear operator $L : H \longrightarrow \mathbb{R}^n$ as

$$Lx = \big(\langle x, x_j \rangle : j \in \mathbb{N}_n\big), \quad x \in H.$$

Consequently, the adjoint map $L^T : \mathbb{R}^n \longrightarrow H$ is given as

$$L^T a = \sum_{j \in \mathbb{N}_n} a_j x_j, \quad a \in \mathbb{R}^n.$$

It is well known that for any $d \in \mathbb{R}^n$, there is a unique vector $x(d) \in M$ such that

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

where $M$ is the $n$-dimensional subspace of $H$ spanned by the vectors in $X$; see, for example, [9]. We start with $I \subseteq \mathbb{N}_n$ that contains $m$ elements ($m < n$). For each $e = (e_1, \ldots, e_n) \in \mathbb{R}^n$, we also use the notations $e_I = (e_i : i \in I) \in \mathbb{R}^m$ and $e_J = (e_i : i \in J) \in \mathbb{R}^{n-m}$. We define the set

$$\mathbb{E}_\infty = \big\{e : e \in \mathbb{R}^n : e_I = 0, \|\!|e_J|\!\|_\infty \le \varepsilon\big\},$$

where $\varepsilon$ is some positive number. For each $d \in \mathbb{R}^n$, we define the *partial hyperellipse*

$$\mathcal{H}(d|\mathbb{E}_\infty) := \big\{x : x \in H, \|x\| \le 1, L(x) - d \in \mathbb{E}_\infty\big\}. \tag{2}$$

Given $x_0 \in H$, our main goal here is to estimate $\langle x, x_0 \rangle$ for $x \in \mathcal{H}(d|\mathbb{E}_\infty)$. According to the midpoint algorithm, we define

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

We point out that $I(x_0, d|\mathbb{E}_\infty)$ is a closed bounded subset in $\mathbb{R}$. Therefore we obtain that

$$I(x_0, d|\mathbb{E}_\infty) = \big[m_-(x_0, d|\mathbb{E}_\infty), m_+(x_0, d|\mathbb{E}_\infty)\big],$$

where $m_-(x_0, d|\mathbb{E}_\infty) = \min\{\langle x, x_0 \rangle : x \in \mathcal{H}(d|\mathbb{E}_\infty)\}$ and $m_+(x_0, d|\mathbb{E}_\infty) = \max\{\langle x, x_0 \rangle : x \in \mathcal{H}(d|\mathbb{E}_\infty)\}$. Hence the best estimator is the midpoint of this interval. According to our previous work [10], we give a formula for the right-hand endpoint.

**Theorem 1.1** *If $x_0 \notin M$ and $\mathcal{H}(d|\mathbb{E}_\infty)$ contains more than one point, then*

$$m_+(x_0, d|\mathbb{E}_\infty) = \min\left\{ \left\| x_0 - L^T(c) \right\| + \varepsilon \|\!| c_J |\!\|_1 + (d, c). : c \in \mathbb{R}^n \right\}. \tag{3}$$

Therefore the midpoint of the uncertainty $I(x_0, d|\mathbb{E}_\infty)$ is given by

$$m(x_0, d|\mathbb{E}_\infty) = \frac{m_+(x_0, d|\mathbb{E}_\infty) - m_+(x_0, -d|\mathbb{E}_\infty)}{2}. \tag{4}$$

Furthermore, we describe every solution to the error bound problem (3) that is required to find the uncertainty interval midpoint. Specifically, the result is applied to a problem with learning the value of a function in the Hardy space of square-integrable functions on the unit circle, which has a well-known reproducing kernel. These formulas allow us to give explicitly the right-hand endpoint $m_+(x_0, d|\mathbb{E}_\infty)$ when only the data error is known. We conjecture that the results of this case appropriately extend to the case of data error measured with the $l^1$ norm, which is our motivation to study this subject.

The paper is organized as follows. In Sect. 2, we provide basic concepts of the particular case of hypercircle inequality for only one data error. Specifically, we provide an explicit solution of a dual problem, which we need for main results. In Sect. 3, we solve the problem of hypercircle inequality for data error measured with the $l^1$ norm. The main result in this section is Theorem 3.3, which establishes the solution for the $l^\infty$ minimization problem, which is a dual problem in this case. Finally, we provide an example of a learning problem in the Hardy space of square-integrable functions on the unit circle and report on numerical experiments of the proposed methods.

## 2 Hypercircle inequality for only one data error

In this section, we describe *HI* for only one data error and its potential relevance to kernel-based learning. Given a set $I \subseteq \mathbb{N}_n$ that contains $n-1$ elements, we assume that $\mathbf{j} \notin I$. For each $e = (e_1, \ldots, e_n) \in \mathbb{R}^n$, we also use the notation $e_I = (e_i : i \in I) \in \mathbb{R}^m$. For each $d \in \mathbb{R}^n$, we define the *partial hyperellipse*

$$\mathcal{H}(d, \varepsilon) := \left\{ x : x \in H, \|x\| \le 1, L_I(x) = d_I, \left| \langle x, x_{\mathbf{j}} \rangle - d_{\mathbf{j}} \right| \le \varepsilon \right\}. \tag{5}$$

Let $x_0 \in H$. Our purpose here is to find the best estimator for $\langle x, x_0 \rangle$ knowing that $\|x\| \le 1$,

$$\langle x, x_i \rangle = d_i \quad \text{for all } j \in \mathbb{N}_{n-1} \quad \text{and} \quad \langle x, x_{\mathbf{j}} \rangle = d_{\mathbf{j}} + e, \quad \text{where } |e| \le \varepsilon.$$

According to our previous work [7], we point out that $\mathcal{H}(d, \varepsilon)$ is weakly sequentially compact in the weak topology on $H$. It follows that $I(x_0, d, \varepsilon) := \{\langle x, x_0 \rangle : x \in \mathcal{H}(d, \varepsilon)\}$ fills out a closed bounded interval in $\mathbb{R}$. Clearly, the midpoint of the uncertainty interval is the best estimator for $\langle x, x_0 \rangle$ when $x \in \mathcal{H}(d, \varepsilon)$. Therefore the hypercircle inequality for partially corrupted data becomes as follows.

**Theorem 2.1** *If $x_0 \in H$ and $\mathcal{H}(d, \varepsilon) \neq \emptyset$, then there is $e_0 \in \mathbb{R}$ such that $|e_0| \le \varepsilon$ and for any $x \in \mathcal{H}(d, \varepsilon)$,*

$$\left| \langle x(d + e_0), x_0 \rangle - \langle x, x_0 \rangle \right| \le \frac{1}{2}(m_+(x_0, d, \varepsilon) + m_-(x_0, d, \varepsilon),$$

*where* $x(d+e_0) = Q^T(G^{-1}(d+e_0)) \in \mathcal{H}(d,\varepsilon)$,

$$m_+(x_0,d,\varepsilon) := \max\{\langle x,x_0\rangle : x \in \mathcal{H}(d,\varepsilon)\}, \tag{6}$$

*and*

$$m_-(x_0,d,\varepsilon) := \min\{\langle x,x_0\rangle : x \in \mathcal{H}(d,\varepsilon)\}. \tag{7}$$

For the particular case $\varepsilon = 0$, let us provide an explicit HI bound and a hypercircle inequality as follows.

**Theorem 2.2** *If* $x \in \mathcal{H}(d)$ *and* $x_0 \in H$, *then*

$$\left|\langle x(d),x_0\rangle - \langle x,x_0\rangle\right| \le \operatorname{dist}(x_0,M)\sqrt{1-\|x(d)\|^2}, \tag{8}$$

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

The inequality above guarantees the presence of an approximation value, which is the vector in the closest point of a hyperplane to the origin. Moreover, it is independent of the vector $x_0$. For the detailed proofs, see [2].

A more complete right-hand endpoint of the uncertainty interval may be obtained by the following results. To this end, we define the function $V : \mathbb{R}^n \to \mathbb{R}$ for each $c \in \mathbb{R}^n$ by

$$V(c) := \|x_0 - L^T(c)\| + \varepsilon|c_j| + (d,c).$$

**Theorem 2.3** *If* $x_0 \notin M$ *and* $\mathcal{H}(d,\varepsilon)$ *contains more than one element, then*

$$m_+(x_0,d,\varepsilon) = \min\{V(c) : c \in \mathbb{R}^n\}, \tag{9}$$

*and the right-hand side of equation* (9) *has a unique solution.*

*Proof* See [7]. □

To state the midpoint of the uncertainty interval, we point out the following fact. We begin with the left-hand side of the interval

$$-m_+(x_0,-d,\varepsilon) = m_-(x_0,d,\varepsilon) := \min\{\langle x,x_0\rangle : x \in \mathcal{H}(d,\varepsilon)\}.$$

The midpoint is given by

$$m(x_0,d,\varepsilon) = \frac{m_+(x_0,d,\varepsilon) - m_+(x_0,-d,\varepsilon)}{2}. \tag{10}$$

In the remainder of this section, we provide an explicit solution to (9).

**Theorem 2.4** *If* $x_0 \notin M$ *and* $\mathcal{H}(d,\varepsilon)$ *contain more than one element, then we have*:

1. $\frac{x_0}{\|x_0\|} \in \mathcal{H}(d,\varepsilon)$ *if and only if* $m_+(x_0,d,\varepsilon) = \|x_0\|$,

2. $x_+(d_I) \in \mathcal{H}(d,\varepsilon)$ *if and only if*

$$m_+(x_0,d,\varepsilon) = \langle x(d_I), x_0\rangle + \operatorname{dist}(x_0, M_I)\sqrt{1 - \|x(d_I)\|^2},$$

*where the vector* $x_+(d_I) := \arg\max\{\langle x, x_0\rangle : x \in \mathcal{H}(d_I)\}$,

3. $\frac{x_0}{\|x_0\|}, x_+(d_I) \notin \mathcal{H}(d,\varepsilon)$ *if and only if*

$$m_+(x_0,d,\varepsilon) = \max\{\langle x_+(d+\varepsilon\mathbf{e}), x_0\rangle, \langle x_+(d-\varepsilon\mathbf{e}), x_0\rangle\},$$

*where the vector* $\mathbf{e} \in \mathbb{R}^n$ *with* $\mathbf{e}_I = 0$ *and* $|\mathbf{e_j}| = 1$.

*Proof* According to our hypotheses, the minimum $c^* \in \mathbb{R}^n$ is the unique solution of the right-hand side of equation (9).

(1) The proof directly follows from [7], that is, we can state that if $x_0 \notin M$, then

$$0 = \arg\min\{V(c) : c \in \mathbb{R}^n\}$$

if and only if $\frac{x_0}{\|x_0\|} \in \mathcal{H}(d,\varepsilon)$.

(2) Again from [7] it follows that $c^* = \arg\min\{V(c) : c \in \mathbb{R}^n\}$ with $c^*_{\mathbf{j}} = 0$ if and only if

$$x_+(d_I) \in \mathcal{H}(d,\varepsilon).$$

By the hypercircle inequality and (8) we obtain that

$$\langle x_+(d_I), x_0\rangle = \langle x(d_I), x_0\rangle + \operatorname{dist}(x_0, M_I)\sqrt{1 - \|x(d_I)\|^2}.$$

(3) Under our hypotheses and [7], the minimum $c^* \in \mathbb{R}^n$ is the unique solution of the function $V$, and $c^*_{\mathbf{j}} \neq 0$. Computing the gradient of $V$ yields

$$-L\left(\frac{x_0 - L^T c^*}{\|x_0 - L^T c^*\|}\right) + \varepsilon \operatorname{sgn}(c^*_n)\mathbf{e} + d = 0, \tag{11}$$

which confirms that

$$L_I(x_+(d,\varepsilon)) = d_I \text{ and } \langle x_+(d,\varepsilon), x_{\mathbf{j}}\rangle = d_{\mathbf{j}} + \operatorname{sgn}(c^*)\varepsilon,$$

where the vector $x_+(d,\varepsilon)$ is given by

$$x_+(d,\varepsilon) := \frac{x_0 - L^T c^*}{\|x_0 - L^T c^*\|} \in \mathcal{H}(d+\varepsilon\mathbf{e}) \cup \mathcal{H}(d-\varepsilon\mathbf{e}).$$

Therefore we obtain that

$$m_+(x_0,d,\varepsilon) = \max\{\langle x_+(d+\varepsilon\mathbf{e}), x_0\rangle, \langle x_+(d-\varepsilon\mathbf{e}), x_0\rangle\}. \qquad \square$$

We end this section by discussing a concrete example of the hypercircle inequality for only one data error for function estimation in a reproducing kernel Hilbert space. Specifically, we report on a new numerical experiment in a reproducing kernel Hilbert space by

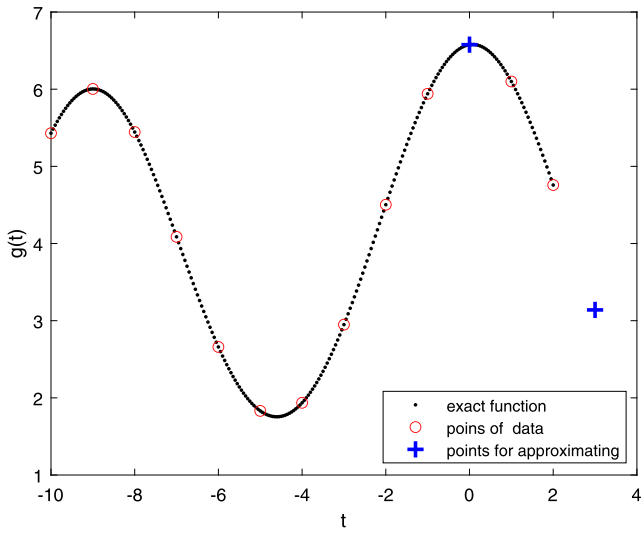


**Figure 1** Exact function

using the available material from HI and our recent results. A real-valued function $K(t,s)$ of $t$ and $s$ in $\mathcal{T}$ is called a *reproducing kernel* of $H$ if the following property is satisfied for all $t \in \mathcal{T}$ and $f \in H$:

$$f(t) = \langle K_t, f \rangle, \tag{12}$$

where $K_t$ is the function defined for $s \in \mathcal{T}$ as $K_t(s) = K(t,s)$. Moreover, for any kernel $K$, there is unique RKHS with $K$ as its reproducing kernel [1]. In our example, we choose the Gaussian kernel on $\mathbb{R}$, that is,

$$K(s,t) = e^{-\frac{(s-t)^2}{10}}, \quad s,t \in \mathbb{R}.$$

The computational steps are organized in the following way. Let $T = \{t_j : j \in \mathbb{N}_n\}$ be points of increasing order in $\mathbb{R}$. Consequently, we have a finite set of linearly independent elements $\{K_{t_j} : j \in \mathbb{N}_n\}$ in $H$, where

$$K_{t_j}(t) := e^{-\frac{(t_j-t)^2}{10}}, \quad j \in \mathbb{N}_n, t \in \mathbb{R}.$$

Thus the vectors $\{x_j : j \in \mathbb{N}_n\}$ appearing above are identified with the function $\{K_{t_j} : j \in \mathbb{N}_n\}$ Therefore the Gram matrix of $\{K_{t_j} : j \in \mathbb{N}_n\}$ is given by

$$G(t_1, \ldots, t_n) := \big(K(t_i, t_j) : i,j \in \mathbb{N}_n\big).$$

In our experiment, we choose the exact function

$$g(t) = -0.15K_{0.5}(t) + 0.05K_{0.85}(t) - 0.25K_{-0.5}(t)$$

and compute the vector $d = \{g(t_j) : j \in \mathbb{N}_{12}\}$ as shown in Fig. 1.

Given $t_0 = 3$, we want to estimate $f(3) = \langle K_{t_0}, f\rangle$ knowing that $\|f\|_K \leq \rho$ and $f(t_i) = \langle K_{t_i}, f\rangle = d_i$ for all $i \in \mathbb{N}_{12}$. In addition, we assume that there is one missing data, that is, we assume that $g(0)$ is missing. Therefore we proximate $f(0)$ by $f_{d_I}(0) = 6.5768973$, which is obtained from the hypercircle inequality, Theorem 2.2, whereas the exact value $g(0) = 6.576978$. Next, we wish to estimate $f(3) = \langle f, K_{t_0}\rangle$ knowing that $f(t_j) = d_j$ for all $j \in \mathbb{N}_{12}$ and $f(0) = 6.5768973 + e$, where $|e| \leq \varepsilon$. Clearly, our data set contains both accurate and inaccurate data. Specifically, there is only one data error in this case. By Theorem 2.4 we easily see that $f_{d_I} \in \mathcal{H}(d, \varepsilon)$. Thus the best value to estimate $f(3)$ is $f_{d_I}(3) = 3.137912$ knowing that $f(t_j) = d_j$ for all $j \in \mathbb{N}_{12}$ and $f(0) = 6.5768973 + e$, where $|e| \leq \varepsilon$. The exact value is $g(3) = 3.1395855$.

## 3 Hypercircle inequality for data error measured with $l^1$ norm

In the previous section, we have provided basic concepts of the particular case of hypercircle inequality for only one data error. For our purpose, we restrict our attention to the study of hypercircle inequality for partially corrupted data with the $l^1$ norm. We start with $I \subseteq \mathbb{N}_n$ that contains $m$ elements ($m < n$). For each $e = (e_1, \ldots, e_n) \in \mathbb{R}^n$, we also use the notations $e_I = (e_i : i \in I) \in \mathbb{R}^m$ and $e_J = (e_i : i \in J) \in \mathbb{R}^{n-m}$. We define $\mathbb{E}_1 = \{e : e \in \mathbb{R}^n : e_I = 0, \||e_J|\|_1 \leq \varepsilon\}$, where $\varepsilon$ is some positive number. For each $d \in \mathbb{R}^n$, we define the *partial hyperellipse*

$$\mathcal{H}(d|\mathbb{E}_1) := \big\{x : x \in H, \|x\| \leq 1, L(x) - d \in \mathbb{E}_1\big\}. \tag{13}$$

As we said earlier, it follows that $\mathcal{H}(d|\mathbb{E}_1)$ is weakly sequentially compact in the weak topology on $H$ and $I(x_0, d|\mathbb{E}_1) := \{\langle x, x_0\rangle : x \in \mathcal{H}(d|\mathbb{E}_1)\}$ is a closed bounded interval in $\mathbb{R}$. Again, the midpoint of the uncertainty interval is the best estimator for $\langle x, x_0\rangle$ when $x \in \mathcal{H}(d|\mathbb{E}_1)$. Therefore the midpoint of the uncertainty $I(x_0, d|\mathbb{E}_1)$ is given by

$$m(x_0, d|\mathbb{E}_1) = \frac{m_+(x_0, d|\mathbb{E}_1) - m_+(x_0, -d|\mathbb{E}_1)}{2}. \tag{14}$$

We easily see that the data set contains both accurate and inaccurate data. In the same manner, we provide the duality formula to obtain the right-hand endpoint of the uncertainty interval $I(x_0, d|\mathbb{E}_1)$. To this end, let us define the convex function $\mathbb{V} : \mathbb{R}^n \longrightarrow \mathbb{R}$ by

$$\mathbb{V}(c) := \big\|x_0 - L^T(c)\big\| + \varepsilon\|\!|c_{\mathbf{J}}|\!\|_\infty + (d, c), \quad c \in \mathbb{R}^n. \tag{15}$$

**Theorem 3.1** *If $x_0 \notin M$ and $\mathcal{H}(d|\mathbb{E}_1)$ contains more than one point, then*

$$m_+(x_0, d|\mathbb{E}_1) = \min\big\{\mathbb{V}(c) : c \in \mathbb{R}^n\big\}, \tag{16}$$

*and the right-hand side of equation (16) has a unique solution. Moreover, $x_+(d_I) \in \mathcal{H}(d|\mathbb{E}_1)$ if and only if $c_j^* = 0$ where, $c^* = \arg\min\{\mathbb{V}(c) : c \in \mathbb{R}^n\}$.*

*Proof* See [10]. □

We begin our main result of this section by providing a useful observation. To this end, let us introduce the following notations. For each $j \in \mathbf{J}$, define the function $\mathbb{V}_j : \mathbb{R}^n \to \mathbb{R}$ by

$$\mathbb{V}_j(c) := \big\|x_0 - L^T(c)\big\| + \varepsilon|c_j| + (c, d), \quad c \in \mathbb{R}^n. \tag{17}$$

Clearly, we see that the duality formula (17) corresponds to the hyperellipse with only one data error

$$\mathcal{H}_j(d,\varepsilon) = \{x : x \in B, L_{\mathbb{N}_n\setminus\{j\}}(x) = d_{\mathbb{N}_n\setminus\{j\}}, |\langle x, x_j\rangle - d_j| < \varepsilon\}. \tag{18}$$

By Theorem 2.4, if $x_0 \notin M$ and $\mathcal{H}_j(d,\varepsilon)$ contain more than one element, then there is unique $a^* \in \mathbb{R}^n$ such that

$$\mathbb{V}_j(a^*) = \min\{\mathbb{V}_j(a) : a \in \mathbb{R}^n\}.$$

We can now state the first result.

**Theorem 3.2** *If $x_0 \notin M, \mathcal{H}(d|\mathbb{E}_1), \mathcal{H}_j(\mathbf{d},\varepsilon)$ contains more than one point and there exists $a^* = \arg\min\{\mathbb{V}_j(a) : a \in \mathbb{R}^{m+1}\}$ with $\|a^*\|_\infty = |a_j^*|$, then*

$$\min\{\mathbb{V}(c) : c \in \mathbb{R}^n\} = \min\{\mathbb{V}_j(a) : a \in \mathbb{R}^n\}. \tag{19}$$

*Proof* For each $c \in \mathbb{R}^n$, we observe that

$$\begin{aligned}\mathbb{V}_j(c) &= \|x_0 - L^T(c)\| + \varepsilon|c_j| + (c,d)\\ &\le \|x_0 - L^T(c)\| + \varepsilon|||c_{\mathbf{J}}|||_\infty + (c,d)\\ &= \mathbb{V}(c),\end{aligned}$$

which means that $\min\{\mathbb{V}_j(c) : c \in \mathbb{R}^n\} \le \min\{\mathbb{V}(c) : c \in \mathbb{R}^n\}$.

According to our assumption, we obtain that

$$\mathbb{V}_j(a^*) = \|x_0 - L^T(a^*)\| + \varepsilon|a_j^*| + (a^*,d) \tag{20}$$

$$= \|x_0 - L^T(a^*)\| + \varepsilon\|a_{\mathbf{J}}^*\|_\infty + (a^*,d), \tag{21}$$

which completes the proof. □

To study the general case, let us introduce the following notations. We first denote the set

$$\Lambda_\infty = \{\lambda : \lambda \in \mathbb{R}^n, \lambda_{\mathbf{I}} = 0, |||\lambda_{\mathbf{J}}|||_\infty \le 1\}.$$

For each $\lambda \in \Lambda_\infty$, we denote the set of linearly independent vectors

$$X(\lambda^j) = \{x_i : i \in \mathbb{N}_m\} \cup \{\mathbf{x}(\lambda^j)\}$$

in $H$, where the vector

$$\mathbf{x}(\lambda^j) = x_j + \sum_{i\in\mathbf{J}\setminus\{j\}} \lambda_i x_i.$$

Consequently, we denote by $M(X(\lambda^j))$ the $(m+1)$-dimensional linear subspace of $H$ spanned by the vectors in $X(\lambda^j)$. From now on, we denote by $G(X(\lambda^j))$ the Gram matrix of the vectors in $X_j(\lambda_{\mathbf{J}})$, which is symmetric and positive definite. The vector $\mathbf{d}(\lambda^j) \in \mathbb{R}^{m+1}$ has the components

$$\mathbf{d}\big(\lambda^j\big)_i = d_i \quad \text{for } i \in I \text{ and } \mathbf{d}\big(\lambda^j\big)_{m+1} = d_j + \sum_{i \in \mathbf{J}\setminus\{j\}} \lambda_i d_i.$$

Therefore we obtain the following *partial hyperellipse* with constant $\mathbf{d}(\lambda^j)$:

$$\mathcal{H}\big(\mathbf{d}\big(\lambda^j\big), \varepsilon\big) = \big\{x : x \in B, L_I(x) = d_I, \big|\big\langle x, \mathbf{x}\big(\lambda^j\big)\big\rangle - \mathbf{d}\big(\lambda^j\big)_{m+1}\big| < \varepsilon\big\}. \tag{22}$$

Next, this partial hyperellipse with only one data error as (22) corresponds to a duality formula for the right-hand endpoint of uncertainty interval, $m_+(x_0, \mathbf{d}(\lambda^j), \varepsilon)$, as shown the following way. For all $j \in \mathbf{J}$ and $\lambda \in \Lambda_\infty$, we define the function $\mathbb{V}_j(\cdot|\lambda) : \mathbb{R}^{m+1} \to \mathbb{R}$ by

$$\mathbb{V}_j(c|\lambda) := \big\|x_0 - L_I^T(c_I) - c_{m+1}\big(\mathbf{x}\big(\lambda^j\big)\big)\big\| + \varepsilon|c_{m+1}| + \big(c, \mathbf{d}\big(\lambda^j\big)\big), \quad c \in \mathbb{R}^{m+1}. \tag{23}$$

**Theorem 3.3** *If $x_0 \notin M$, $\frac{x_0}{\|x_0\|} \notin \mathcal{H}(d|\mathbb{E}_1)$, and $\mathcal{H}(d|\mathbb{E}_1)$ contains more than one point, then there are $\hat{\lambda} \in \Lambda_\infty$ and $j \in \mathbf{J}$ such that*

$$\min\big\{\mathbb{V}(c) : c \in \mathbb{R}^n\big\} = \min\big\{\mathbb{V}_{\mathbf{j}}(c|\hat{\lambda}) : c \in \mathbb{R}^{m+1}\big\}. \tag{24}$$

*Proof* According to our assumption, we can conclude that the right-hand side of equation (24) has a unique solution. Since $\frac{x_0}{\|x_0\|} \notin \mathcal{H}(d|\mathbf{E}_1)$, the vector $c^* \neq 0$. We then assume that $\||c^*_{\mathbf{j}}|\|_\infty = |c^*_{\mathbf{j}}|$ for some $\mathbf{j} \in \mathbf{J}$. Alternatively, we find that there is $\hat{\lambda} \in \Lambda_\infty$ such that $c^*_i = \hat{\lambda}_i c^*_{\mathbf{j}}$ for all $i \in \mathbf{J}\setminus\{\mathbf{j}\}$. Therefore we obtain that

$$\begin{aligned}\min\big\{\mathbb{V}(c) : c \in \mathbb{R}^n\big\} &= \big\|x_0 - L_I^T\big(c_I^*\big) - c^*_{\mathbf{j}}\big(\mathbf{x}\big(\hat{\lambda}^j\big)\big)\big\| + \varepsilon\big|c^*_{\mathbf{j}}\big| + \big(c^*, \mathbf{d}\big(\hat{\lambda}^j\big)\big) \\ &= \min\big\{\mathbb{V}_{\mathbf{j}}(a|\hat{\lambda}) : a \in \mathbb{R}^{m+1}\big\}. \end{aligned} \tag{25}$$

□

Computing the gradient of $\mathbb{V}_{\mathbf{j}}(\cdot|\hat{\lambda})$, the minimum $c^*_{I\cup\{\mathbf{j}\}} = a^* \in \mathbb{R}^{m+1}$ is a unique solution of the nonlinear equations

$$L_{\mathbf{I}}\big(x_+\big(\mathbf{d}\big(\hat{\lambda}^j\big), \varepsilon\big)\big) = d_I,$$

and

$$\big\langle x_+\big(\mathbf{d}\big(\hat{\lambda}^j\big), \varepsilon\big), \mathbf{x}\big(\lambda^j\big)\big\rangle - \mathbf{d}\big(\lambda^j\big)_{m+1} = \operatorname{sgn}\big(c^\star_{\mathbf{j}}\big)\varepsilon,$$

where the vector $x_+(\mathbf{d}(\hat{\lambda}^j), \varepsilon)$ is given by

$$x_+\big(\mathbf{d}\big(\hat{\lambda}^j\big), \varepsilon\big) := \frac{x_0 - L_I^T(a_I^*) - a^*_{\mathbf{j}}(\mathbf{x}(\hat{\lambda}^j))}{\|x_0 - L_I^T(a_I^*) - a^*_{\mathbf{j}}(\mathbf{x}(\hat{\lambda}^j))\|} \in \mathcal{H}\big(\mathbf{d}\big(\hat{\lambda}^j\big), \varepsilon\big),$$

and the partial hyperellipse with the constant $\mathbf{d}(\hat{\lambda}^j)$ is

$$\mathcal{H}\big(\mathbf{d}(\hat{\lambda}^j),\varepsilon\big) = \big\{x : x \in B, L_I(x) = d_I, \big|\langle x, \mathbf{x}(\hat{\lambda}^j)\rangle - \mathbf{d}(\hat{\lambda}^j)_{m+1}\big| < \varepsilon\big\}.$$

To this end, let us introduce the following set: For each $\lambda \in \Lambda_\infty$, we define

$$W(\lambda) := \min\big\{m_i(\lambda) : i \in \mathbf{J}\big\},$$

where $m_i(\lambda) = \min\{\mathbb{V}_i(c|\lambda) : c \in \mathbb{R}^{m+1}\}$.

**Theorem 3.4** *If $\mathcal{H}(d|\mathbb{E}_1)$ contains more than one point, then*

$$m_+(x_0, d|\mathbb{E}_1) = \min\big\{W(\lambda) : \lambda \in \Lambda_\infty\big\} \tag{26}$$

*Proof* For each $\lambda \in \Lambda_\infty$, we see that

$$\min\big\{\mathbb{V}(c) : c \in \mathbb{R}^n\big\} \le \min\big\{\mathbb{V}_i(a|\lambda) : a \in \mathbb{R}^{m+1}\big\}$$

for all $i \in \mathbf{J}$, that is, for each $\lambda \in \Lambda_\infty$,

$$\min\big\{\mathbb{V}(c) : c \in \mathbb{R}^n\big\} \le W(\lambda).$$

Consequently, we obtain that

$$\min\big\{\mathbb{V}(c) : c \in \mathbb{R}^n\big\} \le \inf\big\{W(\lambda) : \lambda \in \Lambda_\infty\big\}.$$

According to Theorem 3.2, there are $\hat{\lambda} \in \Lambda_\infty$ and $\mathbf{j} \in \mathbf{J}$ such that

$$\min\big\{\mathbb{V}(c) : c \in \mathbb{R}^n\big\} = \min\big\{V_{\mathbf{j}}(c|\hat{\lambda}) : c \in \mathbb{R}^{m+1}\big\}.$$

Therefore we can conclude that

$$\min\big\{\mathbb{V}(c) : c \in \mathbb{R}^n\big\} = \min\big\{W(\lambda) : \lambda \in \Lambda_\infty\big\}. \qquad \square$$

We end this section by extending these results to estimate optimally any number of features. Let us define the function $W : \mathcal{H}(d|\mathbb{E}_1) \to \mathbb{R}^k$ defined for $x \in \mathcal{H}(d|\mathbb{E}_1)$ by

$$Wx = \big(\langle x, x_{-k+j}\rangle : j \in \mathbb{N}_k\big). \tag{27}$$

In the case of estimating a single feature, the uncertainty set is an interval. For multiple features, the uncertainty set is a bounded set in a finite-dimensional space. Consequently, the corresponding uncertainty set is given as

$$U(d|\mathbb{E}_1) := \big\{Wx : x \in \mathcal{H}(d|\mathbb{E}_1)\big\}. \tag{28}$$

It is easy to check that $U(d|\mathbb{E}_1)$ is a convex compact subset of $\mathbb{R}^k$. To get the best estimator, we need to find the center and radius of $U(d|\mathbb{E}_1)$. We recall the Chebyshev radius and

center. For this purpose, we choose the $l^\infty$ norm $\|\|\cdot\|\|_\infty$ on $\mathbb{R}^k$ and define the radius of $U(d|\mathbb{E}_1)$ as

$$r_\infty\big(U(d|\mathbb{E}_1)\big) := \inf_{y\in\mathbb{R}^k} \sup_{u\in U(d|\mathbb{E}_1)} \|\|u-y\|\|_\infty.$$

We denote its center as $m_\infty \in \mathbb{R}^k$. In the theorem below, we will show that the $l^\infty$ center of the set $U(d|\mathbb{E})$ is given by the vector

$$m_\infty := \big(m(x_{-k+j}, d|\mathbb{E}_1) : j \in \mathbb{N}_k\big),$$

where $m(x_{-k+j}, d|\mathbb{E}_1)$ is the center of the interval $I(x_{-k+j}, d|\mathbb{E}_1)$ for all $j \in \mathbb{N}_k$.

**Theorem 3.5** *If $\mathcal{H}(d|\mathbb{E}_1) \neq \emptyset$, then the $l^\infty$ center of the uncertainty set is $m_\infty = (m(x_{-k+j}, d|\mathbb{E}_1) : j \in \mathbb{N}_k)$, and its radius is given by*

$$r_\infty\big(U(d|\mathbb{E}_1)\big) = \max\big\{r\big(I(x_{-k+j}, d|\mathbb{E}_1)\big) : j \in \mathbb{N}_k\big\}.$$

*Proof* This follows by the same method as in [6]. □

To this end, we present some results of a numerical experiment on estimating multiple features of a vector in the partial hyperellipse $\mathcal{H}(d|\mathbb{E}_1)$. For our computational experiments, we choose the Hardy space of square-integrable functions on the unit circle with reproducing kernel

$$K(z,\zeta) = \frac{1}{1-\bar{\zeta}z}, \zeta, \quad z \in \Delta,$$

where $\Delta := \{z : |z| \leq 1\}$, [3]. Specifically, let $H^2(\Delta)$ be the set of all functions analytic in the unit disc $\Delta$ with norm

$$\|f\| = \sup_{0<r<1}\left(\frac{1}{2\pi}\int_0^{2\pi} \big|f\big(re^{i\theta}\big)\big|^2 d\theta\right)^{\frac{1}{2}}.$$

Specifically, let $T = \{t_j : j \in \mathbb{N}_n\}$ be points of increasing order in $(-1, 1)$. Consequently, we have a finite set of linearly independent elements $\{K_{t_j} : j \in \mathbb{N}_n\}$ in $H$, where

$$K_{t_j}(t) := \frac{1}{1-t_j t}, j \in \mathbb{N}_n, \quad t \in \Delta.$$

Thus the vectors $\{x_j : j \in \mathbb{N}_n\}$ appearing above are identified with the functions $\{K_{t_j} : j \in \mathbb{N}_n\}$. Therefore the Gram matrix of $\{K_{t_j} : j \in \mathbb{N}_n\}$ is given by

$$G(t_1, \ldots, t_n) := \big(K(t_i, t_j) : i, j \in \mathbb{N}_n\big).$$

Next, we recall the Cauchy determinant defined for $\{t_j : j \in \mathbb{N}_n\}$ and $\{s_j : j \in \mathbb{N}_n\}$ as

$$\det\left(\frac{1}{1-t_i s_j}\right)_{i,j\in\mathbb{N}_n} = \frac{\prod_{1\leq j<i\leq n}(t_j - t_i)(s_j - s_i)}{\prod_{i,j\in\mathbb{N}_n}(1-t_i s_j)}; \tag{29}$$

see, for example, [2]. From this formula we obtain that

$$\det G(t_1,\dots,t_n) = \frac{\prod_{1\le i<j\le n}(t_i - t_j)^2}{\prod_{i,j\in\mathbb{N}_n}(1 - t_i t_j)}. \tag{30}$$

In our case, for any $t_0 \in (-1,1)$ and $t_0 \notin T := \{t_j : j \in \mathbb{N}_n\}$, we obtain that

$$\operatorname{dist}\big(K_{t_0}, \operatorname{span}\{K_{t_j} : j \in \mathbb{N}_n\}\big) = \frac{|B(t_0)|}{\sqrt{1 - t_0^2}}, \tag{31}$$

where $B$ is the rational function defined for $t \in \mathbb{C} \setminus \{t_j^{-1} : j \in \mathbb{N}_n\}$ by

$$B(t) := \prod_{j\in\mathbb{N}_n} \frac{t - t_j}{1 - t t_j}, \tag{32}$$

and the vector $x_0$ appearing previously is identified with the function $K_{t_0}$. We organize the computational steps as follows. We choose a finite set of linear independent elements $\{K_{t_j} : j \in \mathbb{N}_6\}$ in $H$ with

$$t_1 = -0.9, \qquad t_2 = -0.6, \qquad t_3 = -0.3, \qquad t_4 = 0.3, \qquad t_5 = 0.6 \quad \text{and} \quad t_6 = 0.9.$$

We choose the exact function

$$g(t) = -0.15K_{0.5}(t) + 0.05K_{0.85}(t) - 0.25K_{-0.5}(t)$$

and compute the vector $d = \{g(t_j) : j \in \mathbb{N}_6\}$. By the definition of (12) the linear operator $L : H^2(\Delta) \longrightarrow \mathbb{R}^5$ is defined for $f \in H^2(\Delta)$ as follows:

$$Lf := \big(f(t_i) : i \in \mathbb{N}_6\big) = \big(\langle f, K_{t_i}\rangle : i \in \mathbb{N}_6\big).$$

In our experiment, we choose

$$t_{-2} = -0.4, \qquad t_{-1} = 0, \qquad t_0 = 0.4,$$

and we wish to estimate

$$Wf = \big(\langle f, K_{t_{-3+j}}\rangle = f(t_{-3+j}) : j \in \mathbb{N}_3\big)$$

when we known that

$$f(t_j) = d_j \quad \text{for all } j \in \mathbb{N}_6 \setminus \{3,4\} \quad \text{and} \quad \big|f(t_3) - d_3\big| + \big|f(t_4) - d_4\big| \le \varepsilon = 0.1.$$

According to Theorem 3.2, the functions $\mathbb{V}_1$ and $\mathbb{V}_2$ become

$$\begin{aligned}\mathbb{V}_1(c|\lambda) = {} & \big\|K_{t_0} - (c_1K_{t_1} + c_2K_{t_2} + c_3K_{t_5} + c_4K_{t_6}) - c_5(K_{t_3} + \lambda K_{t_4})\big\| \\ & + \varepsilon|c_5| + (c_I, d_I) + c_5(d_3 + \lambda d_4)\end{aligned}$$

**Table 1** Optimal value

| $\varepsilon$ | $m(t_{-3+j}, d\|\mathbb{E}_1)$ | $g(t_{-3+j})$ |
|---|---|---|
| −0.4 | −0.32936 | −0.3617 |
| 0 | −0.349967 | −0.35 |
| 0.4 | −0.32468 | −0.3585 |

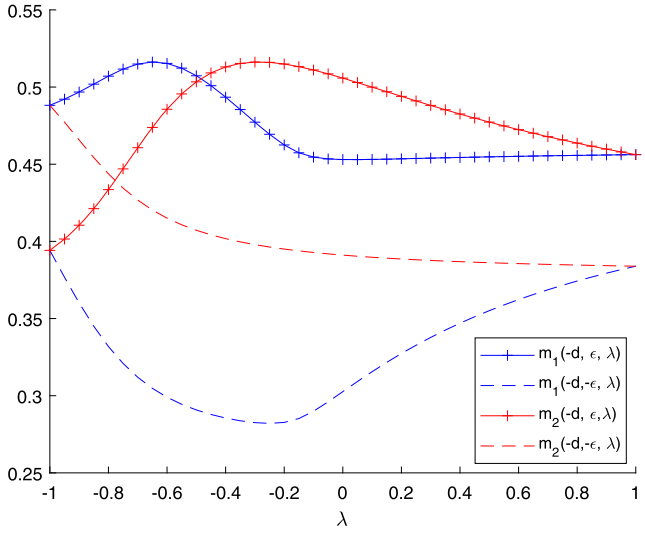


**Figure 2** $m_i(-0.4, \lambda)$ for $-d$

and

$$\mathbb{V}_2(c|\lambda) = \left\|K_{t_0} - (c_1K_{t_1} + c_2K_{t_2} + c_3K_{t_5} + c_4K_{t_6}) - c_5(K_{t_4} + \lambda K_{t_3})\right\| + \varepsilon|c_5| + (c_I, d_I) + c_5(d_4 + \lambda d_3).$$

In this computation, we found that $f^+_{d_I} \notin \mathcal{H}(d|\mathbb{E}_1)$. To obtain the minimum of $W$, we must compare the values of $m_1(t_{-3+j}, \lambda)$ and $m_2(t_{-3+j}, \lambda)$, where

$$m_1(t_{-3+j}, \lambda) = \max\left\{f^+_{d(\lambda^1)+\varepsilon\mathbf{e}}(t_{-3+j}), f^+_{d(\lambda^1)-\varepsilon\mathbf{e}}(t_{-3+j})\right\}$$

and

$$m_2(t_{-3+j}, \lambda) = \max\left\{f^+_{d(\lambda^2)+\varepsilon\mathbf{e}}(t_{-3+j}), f^+_{d(\lambda^2)-\varepsilon\mathbf{e}}(t_{-3+j})\right\},$$

that is,

$$m_j(t_{-3+j}, \lambda) = \max\left\{f^+_{d(\lambda^j)\pm\varepsilon\mathbf{e}}(t_{-3+j})\right\} = \max\left\{f_{d(\lambda^j)\pm\varepsilon\mathbf{e}}(t_{-3+j}) + \operatorname{dist}\left(K_{t_{-3+j}}, M\left(\lambda^j\right)\right)\sqrt{1 - \|f_{d(\lambda^j\pm\varepsilon\mathbf{e})}\|^2}\right\}.$$

To obtain the right-hand endpoint, we need to find the minimum of $W$ defined for $\lambda \in [-1, 1]$. As explained earlier, the midpoint algorithm requires us to find numerically the minimum of the function $\mathbb{V}$ for $d$ and $-d$, that is, we compute $v_\pm := \min\{\mathbb{V}(c, \pm d) : c \in \mathbb{R}^n\}$,

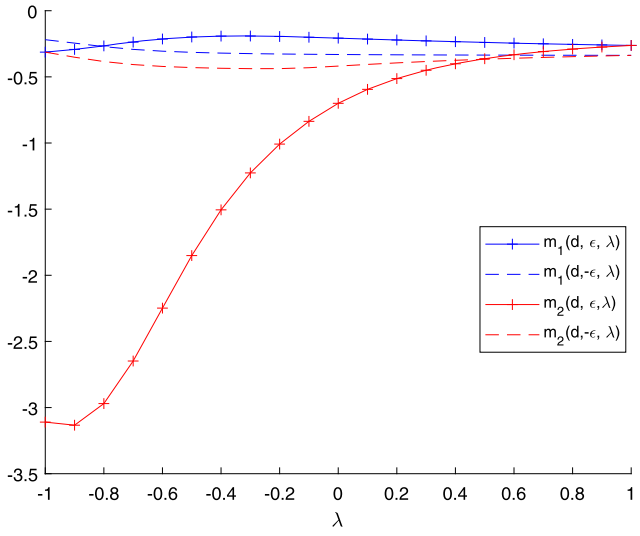


**Figure 3** $m_i(0.4, \lambda)$ for $d$

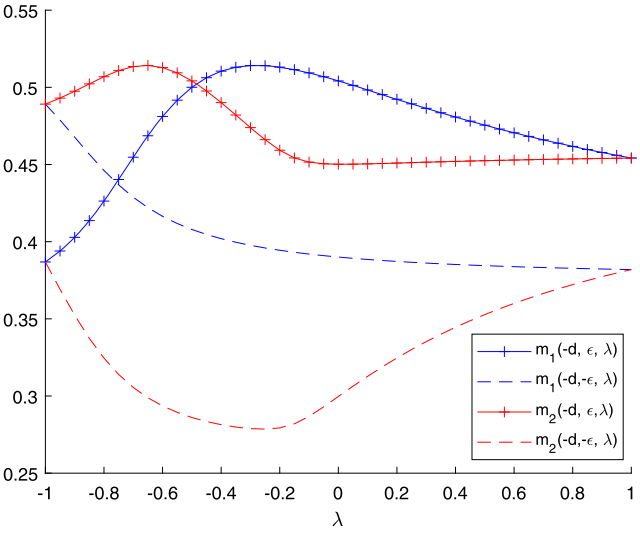


**Figure 4** $m_i(0.4, \lambda)$ for $-d$

and then our midpoint estimator is given by $\frac{\nu_+ - \nu_-}{2}$. The result of this computation is shown in Table 1.

Furthermore, we see that $m_+(t_{-2}, d|\mathbb{E}_1) = \min\{f^+_{d(\lambda^2)+\varepsilon\mathbf{e}}(t_{-2}) : \lambda \in [-1, 1]\}$. To obtain $m_+(t_{-2}, -d|\mathbb{E}_1)$, we then plot $m_1$ and $m_2$ as functions of $\lambda$ for $t_{-2} = -0.4$, as shown in Fig. 2.

Similarly, we find that

$$m_+(t_{-1}, d|\mathbb{E}_1) = \min\left\{f^+_{d(\lambda^2)+\varepsilon\mathbf{e}}(t_{-1}) : \lambda \in [-1, 1]\right\}$$

and

$$m_+(t_{-1}, -d|\mathbb{E}_1) = \min\left\{f^+_{-d(\lambda^1)+\varepsilon\mathbf{e}}(t_{-1}) : \lambda \in [-1,1]\right\}.$$

For the case $t_0 = 0.4$, we plot $m_1$ and $m_2$ as functions of $\lambda$ for $d$ and $-d$ as shown in Figs. 3 and 4, respectively.

## 4 Conclusions

In this paper, we described an unexpected application of hypercircle inequality for only one data error to the $l^\infty$ minimization problem (16). In two different circumstances, we applied what we have learned from recent results to the problem of learning the value of a function in RKHS, which can be beneficial in practice.

**Acknowledgements**
The authors would like to thank the referee for his/her comments and suggestions on the manuscript. This work was supported by PMU (B05F630094) and University of Phayao, Thailand.

**Funding**
PMU (B05F630094) and University of Phayao, Thailand.

**Availability of data and materials**
Not applicable.

## Declarations

**Competing interests**
The authors declare that they have no competing interests.

**Authors' contributions**
KK developed the theoretical part and performed the analytic calculations and numerical simulations. Both authors contributed to the final version of the manuscript. Both authors read and approved the final manuscript.

## Publisher's Note



Received: 12 December 2021 Accepted: 13 June 2022 Published online: 25 June 2022

**References**

1. Aronszajn, N.: Theory of reproducing kernels. Trans. Am. Math. Soc. **68**, 337–404 (1950)
2. Davis, P.J.: Interpolation and Approximation. Dover, New York (1975)
3. Duren, P.: Theory of $H^p$ Space, 2nd edn. Dover, New York (2000)
4. Garcia, A.G., Portal, A.: Hypercircle inequalities and sampling theory. Appl. Anal. **82**, 1111–1125 (2003)
5. Golomb, M., Weinberger, H.F.: Optimal approximation and error bounds, on numerical approximation. In: Langer, R.E. (ed.) Proceedings of a Symposium, Madison, April 21–23, 1958, pp. 117–190. The University of Wisconsin Press, Madison (1959). Publication No. 1 of the Mathematics Research Center, U.S. Army, the University of Wisconsin
6. Khompurngson, K., Micchelli Hide, C.A.: Jaen J. Approx. **3**, 87–115 (2011)
7. Khompurngson, K., Novaprateep, B.: Hypercircle inequality for partially-corrupted data. Ann. Funct. Anal. **6**, 95–108 (2015)
8. Khompurngson, K., Novaprateep, B., Lenbury, Y.: Learning the value of a function from inaccurate data. East-West J. Math. Special Vol., 128–138 (2010)
9. Micchelli, C.A., Rivlin, T.J.: A survey of optimal recovery. In: Micchelli, C.A., Rivlin, T.J. (eds.) Optimal Estimation in Approximation Theory. Plenum, New York (1977)
10. Nammanee, K., Khompurngson, K.: Hypercircle inequality for data error measured with $l^\infty$ norm. Jaen J. Approx. **11**, 151–167 (2019)
11. Synge, J.L.: The method of the hypercircle in function-space for boundary-value problems. Proc. R. Soc. Lond. Ser. A, Math. Phys. Sci. **1027**, 447–467 (1947)
12. Valentin, R.A.: The use of the hypercircle inequality in deriving a class of numerical approximation rules for analytic functions. Math. Compet. **22**, 110–117 (1968)