Songklanakarin J. Sci. Technol.
45 (2), 211–218, Mar. – Apr. 2023

http://www.sjst.psu.ac.th

Original Article

Some results on hypercircle inequality for partially corrupted data via orthonormal set

Tanyong Kraiwiradechachai[1] and Kannika Khompurngson[2*]

[1] *School of Mathematics, Faculty of Science, Naressuan University, Mueang, Phitsanulok, 65000 Thailand*

[2] *Division of Mathematics, School of Science, University of Phayao, Mueang, Phayao, 56000 Thailand*

Received: 26 September 2022; Revised: 26 January 2023; Accepted: 20 February 2023

Abstract

In this paper, we briefly review the material on hypercircle inequality for partially corrupted data and its potential for addressing a learning problem in reproducing kernel Hilbert space. The aim of this paper is to present the transformation of its material to orthonormal bases. Specifically, our recent results lead us to improve important results on this subject, which are practically useful.

Keywords: hypercircle inequality, convex optimization, reproducing kernel Hilbert space

1. Introduction and Preliminaries

The basic concept of a learning problem is to find a function representation of given data. The hypercircle method, which is well-known in mathematical physics, has been applied to kernel-based machine learning. Unfortunately, this method is not well adapted to circumstances in which there are known data errors. In 2015, Khompurngson K. and Novaprateep B. extended the hypercircle inequality to circumstances in which there are both accurate and inaccurate known data. Moreover, its potential for a learning problem in reproducing kernel Hilbert space was presented. In this paper, we continue to study this subject and present a transformation of its material to orthonormal bases. Specifically, our recent results lead us to improve important results on this subject, which are practically useful.

Let H be a Hilbert space over the real numbers with inner product $<\bullet,\bullet>$ and $X = \{x_j : j \in \mathbb{N}_n\}$ be a set of *linearly independent* vectors in H, where we denote $\mathbb{N}_n = \{1, 2,\ldots, n\}$.

*Corresponding author
Email address: kannika.kh@up.ac.th

For any $d \in \mathbb{N}^n$, $\mathrm{H}(d) := \{x : \|x\| \le 1, Q(x) = d\}$ is called *hypercircle* where $Q : H \to \mathbb{R}^n$ is a linear operator H onto $\mathbb{R}^n$ given by

$$Qx = (<x, x_j> : j \in \mathbb{N}_n). \qquad (1.1)$$

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

$$Q^T(a) = \sum_{i \in \mathbb{N}_n} a_i, x_j.$$

The Gram matrix of the vectors in X is $G_x = QQ^T = (<x_j, x_l> : j, l \in \mathbb{N}_n)$, a symmetric and positive definite matrix. It is well-known that there exists a unique vector $x(d) \in M$ such that

$$x(d) := \operatorname{argmin}\{\|x\| : x \in H, Q(x) = d\}, \qquad (1.2)$$

where M is the n-dimensional subspace of H spanned by the vectors in X, see Khompurngson and Micchelli (2011). Moreover, the vector satisfies

$$x(d) := Q^T(G_x^{-1}d) \text{ and } \|x(d)\|^2 = (d, G_x^{-1}d),$$

see Davis (1975). Let the vectors in X be orthonormalized according to the Gram-Schmidt process yielding $x^*_1,\dots,x^*_n$. Let M^* be the n-dimensional subspace of H spanned by the vectors $x^*_1,\dots,x^*_n$; see (3). Then the Gram matrix for $x^*_1,\dots,x^*_n$ is the identity matrix. For any $x \in H(d)$, the condition

$$Qx = d \text{ is equivalent to } Rx = d^*$$

where $R : H \rightarrow \mathbb{R}^n$ is a linear operator H onto $\mathbb{R}^n$ as $R(x) = (\langle x, x^*_j\rangle : j \in \mathbb{N}_n)$ and

$$d_1^* = \frac{d_1}{\sqrt{|G(x_1)|}}$$

$$d_k^* = \frac{1}{\sqrt{|G(x_1,\dots,x_{k-1})||G(x_1,\dots,x_k)|}}\begin{bmatrix} \langle x_1,x_1\rangle & \langle x_2,x_1\rangle & \dots & \langle x_k,x_1\rangle \\ \vdots & \vdots & \dots & \vdots \\ \langle x_1,x_{k-1}\rangle & \langle x_2,x_{k-1}\rangle & \dots & \langle x_k,x_{k-1}\rangle \\ d_1 & d_2 & \dots & d_k \end{bmatrix}, k = 2,\dots,n \tag{1.3}$$

Therefore, we represent the coefficients of d^*_k for all $k = 1,\dots,n$ by the following matrix

$$A = \begin{bmatrix} \beta_{11} & 0 & 0 & \cdots & 0 \\ \beta_{21} & \beta_{22} & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ \beta_{n1} & \beta_{n2} & \beta_{n3} & \cdots & \beta_{nn} \end{bmatrix}.$$

Hence, we point out that $\mathrm{H}(d) = \mathrm{H}^*(d^*)$ where $\mathrm{H}^*(d^*) = \{x \in H : \|x\| \le 1, Rx = d^*\}$ and the vector $x(d) = x(d^*) : R^T d^*$ where the adjoint map R^T is given for each $a \in \mathbb{R}^n$ by $R^T(x) = \sum_{j=1}^{n} a_j x_j^*$. In this case, the classical *hypercircle inequality* is stated as follows:

Theorem 1.1 If $x \in \mathrm{H}^*(d^*)$ and $x_0 \in H$ then

$$\left|\left\langle x(d^*), x_0\right\rangle - \left\langle x, x_0\right\rangle\right| \le dist(x_0, M^*)\sqrt{1-(d^*,d^*)}\,. \tag{1.3}$$

where $dist(x_0, M^*) := \min\{\| x_0 - y \| : y \in M^*\} = \sqrt{|G(x_1^*,\dots,x_n^*,x_0)|}$.

Moreover, there is an $x_\pm(d^*) \in \mathrm{H}^*(d^*)$ for which equality holds in (1.3) and the vector $x_\pm(d^*)$ is given by

$$x_\pm(d^*) = \pm\frac{x_0 - R^T a_\pm}{\left\| x_0 - R^T a_\pm \right\|} \tag{1.4}$$

and the vector $a_\pm \in \mathbb{R}^n$ is given by the formula

$$a_\pm := Rx_0 \mp \frac{dist(x_0, M^*)}{\sqrt{1-(d^*,d^*)}} d^*\,. \tag{1.5}$$

Recently, an extension of the hypercircle inequality to *partially - corrupted* data was proposed by Kannika Khompurngson and Boriboon Novaprateep, see Khompurngson and Novaprateep (2015). We start with $I \subseteq \mathbb{N}_n$ which contains m elements ($m < n$). Consequently, we use the notations

$$X_I = \{x_i : i \in I\} \subset X \text{ and } X_J = \{x_i : i \in J\} \subset X\,,$$

where we denote $J = \mathbb{N}_n \setminus I$. Similarly, we use the notations M_I and M_J for the subspaces of H spanned by the vectors in X_I and X_J respectively. For each $e \in \mathbb{R}^n$, we also use the notations $e_I = (e_i : i \in I) \in \mathbb{R}^m$ and $e_J = (e_i : i \in I) \in \mathbb{R}^{n-m}$ respectively. We choose $|||\bullet|||_p : \mathbb{R}^{n-m} \rightarrow \mathbb{R}_+$, i.e. the l^p norm on $\mathbb{R}^{n-m}$, and define

$$E_p = \{e : e \in \mathbb{R}^n: e : e_I = 0, ||| e_J |||_p \leq \varepsilon\},$$

where ε is some positive number and $1 \leq p \leq \infty$. For each $d \in \mathbb{R}^n$, we define the *partial hyperellipse* as follows

$$\mathrm{H}(d \mid E_p) := \{x : x \in H, \| x \| \leq 1, Q(x) - d \in E_p\}, \tag{1.6}$$

where $E_p = \{e : e \in \mathbb{R}^n: e_I = 0, ||| e_J |||_p \leq \varepsilon\}$.

Given $x_0 \in H$, we want to estimate $\langle x, x_0 \rangle$ when $x \in \mathrm{H}(d \mid E_p)$. That is, the dataset contains both accurate and inaccurate data. The best estimator is the *midpoint* of this interval

$$I(x_0, d \mid E_p) := \{\langle x, x_0 \rangle : x \in \mathrm{H}(d \mid E_p)\}.$$

Next let us recall the duality formula for the right-hand end point, $m_+(x_0, d \mid E_p)$, of the uncertainty interval.

Theorem 1.2 If $\mathrm{H}(d \mid E_p)$ contains more than one point and $x_0 \notin M$, then

$$m_+(x_0, d \mid E_p = \min\{\| x_0 - Q^T(c)\| + \varepsilon \, ||| c_J |||_q + (d, c) : c \in \mathbb{R}^n\}. \tag{1.7}$$

Finally, the midpoint is given by

$$m(x_0, d \mid E_p) = \frac{m_+(x_0, d \mid E_p) - m_+(x_0, -d \mid E_p)}{2}.$$

Furthermore, if $X = \{x_j : J \in \mathbb{N}^n\}$ is an *orthonormal* set of vectors (and the Gram matrix is the identity matrix) then for any

$$x(d+e) \in \mathrm{H}(d \mid E_p) \cap M$$

$$x(d+e) = x(d_I) + x(d_J + e_J) \quad \text{and} \quad \| x(d+e) \|^2 = \| x(d_I) \|^2 + \| x(d_J + e_J) \|^2,$$

where $x(d_I) \in \mathrm{H}(d_I)$ and $x(d_J + e_J) \in \mathrm{H}(d_J \mid E_J)$.

Moreover, we observe that $\mathrm{H}(d \mid E_p) \neq \varnothing$ if and only if

$$min\{(d_J + c, d_J + c) : c \in \mathbb{R}^{n-m}, |||c|||_p \leq \varepsilon\} \leq 1 - \| x(d_J) \|^2. \tag{1.8}$$

For $p = 2$, we have the following $\mathrm{H}(d \mid E_2) \neq \varnothing$ if and only if

$$min\{(d_J + c, d_J + c) : c \in \mathbb{R}^{n-m}, |||c|||_2 \leq \varepsilon\} = \Lambda + \Lambda \sum_{j \notin \mathrm{II}} \frac{d_j^2}{\Lambda - \varepsilon^2} \leq 1 - \| x(d_J) \|^2,$$

where $\mathrm{II} := \{j : d_j = 0, j \in J\}$ and $\Lambda = \varepsilon^2 - \sqrt{\sum_{i \in J} d_i^2}$.

Summarizing, we point out that if $\Lambda + \Lambda \sum_{j \notin \mathrm{II}} \frac{d_j^2}{\Lambda - \varepsilon^2} \leq 1 - \| x(d_J) \|^2$ then $\mathrm{H}(d \mid E_2) \neq \varnothing$ for all $p \geq 2$; for more details on theory and proofs, see Forsythe and Golub (1965). After this observation, we continue our study of this subject by presenting the transformation of its material to orthonormal bases in section 2, which includes an example.

2. Main Results

We begin this section by assuming that $X_I = \{x_1, \ldots, x_m\}$ and $X_J = \{x_{m+1}, \ldots, x_n\}$ respectively. Let the vectors in $X_I \cup X_J$ be orthonormalized according to the Gram-Schmidt process yielding $x_1^*, \ldots, x_m^*, x_{m+1}^*, \ldots, x_n^*$. For our purpose, we define

$$E_p(A) := \{Ae : e \in E_p\}$$

and

$$\mathrm{H}^*(d^* \mid E_p(A)) := \{x : \| x \| \leq 1, Rx - d^* \in E_p(A)\}. \tag{2.1}$$

Similarly, we point out that for each $x \in \mathrm{H}(d \mid E_p)$ the condition

$Qx = d + e$ is equivalent to $Rx = A(d + e) = d^* + Ae$.

Alternatively, we point out that for each $x \in \mathrm{H}^*(d^* \mid E_p(A))$

$$R_I(x) = d_I^* \text{ and } R_J(x) - d_J^* = A_J e_J ,$$

where $||| e_J |||_p \leq \varepsilon$ and the n x m matrix A_J is given by

$$A_J = \begin{bmatrix} \beta_{m+1\, m+1} & 0 & 0 & \cdots & 0 \\ \beta_{m+2\, m+1} & \beta_{m+2\, m+2} & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ \beta_{n\, m+1} & \beta_{n\, m+2} & \beta_{n\, m+3} & \cdots & \beta_{nn} \end{bmatrix}. \tag{2.2}$$

In addition, for each $Ae \in E_p(A)$ the vector $x^*(d^* + Ae) \in \mathrm{H}^*(d^* \mid E_p(A)) \bigcap M^*$ can be written in the form $x^*(d^* + Ae) = x^*(d_I^*) + x^*(d_J^* + A_J e_J)$ and

$$\| x^*(d^* + Ae) \|^2 = \| x^*(d_I^*) \|^2 + \| x^*(d_J^* + A_J e_J) \|^2 , \tag{2.3}$$

where the vector $x^*(d_I^*) \in \mathrm{H}^*(d_I^*) := \{x : \| x \| \leq 1, R_I(x) = d_I^*\}$ and
$x^*(d_J^* + A_J e_J) \in \mathrm{H}^*(d_J^* \mid E_p(A_J)) := \{x : \| x \| \leq 1, R_J(x) - d_J^* \in E_p(A_J)\}$,
when $E_p(A_J) = \{A_J c : c \in \mathbb{R}^{n-m} : ||| c |||_p \leq \varepsilon\}$.

Lemma 2.1 $\mathrm{H}^*(d^* \mid E_p(A)) \neq \varnothing$ if and only if

$$\min\left\{\left(A_J^{-1} d_J^* + \varepsilon\xi, A_J^T A_J (A_J^{-1} d_J^* + \varepsilon\xi)\right) : \ ||| \xi |||_p \leq 1\right\} \leq 1 - \| x(d_I^*) \|^2 \Big\}. \tag{2.4}$$

Proof. Let $x \in \mathrm{H}^*(d^* \mid E_p(A))$. Then there is $Ae \in E_p(A)$ such that $x = x^*(d^* + Ae) = x^*(d_I^*) + x^*(d_J^* + A_J e_J) \in M^*$ and $\| x^*(d^* + Ae) \|^2 = \| x^*(d_I^*) \|^2 + \| x^*(d_J^* + A_J e) \|^2 \leq 1$. Next, we observe that

$$\| x^*(d_J^* + A_J e) \|^2 = \left(A_J^{-1} d_J^* + \varepsilon\xi, A_J^T A_J (A_J^{-1} d_J^* + \varepsilon\xi)\right).$$

Hence, $\min\left\{\left(A_J^{-1} d_J^* + \varepsilon\xi, A_J^T A_J (A_J^{-1} d_J^* + \varepsilon\xi)\right) : ||| \xi |||_p \leq 1\right\} \leq 1 - \| x(d_I^*) \|^2 \Big\}$.
Conversely, (2.3) and (2.4) certainly imply that $\mathrm{H}^*(d^* \mid E_p(A)) \neq \varnothing$.

For $p = 2$, we describe the solution of the optimization problem appearing in (2.4) as presented in Forsythe and Golub (1965). We begin with the following definition.

Definition 2.2 Let C be an n x n symmetric matrix and $d \in \mathbb{R}^n$. The *spectrum* of the pair (C, d) is defined to be the set of all real numbers λ for which there exists an $x \in \mathbb{R}^n$ with Euclidean norm one such that

$$C(x - d) = \lambda x. \tag{2.5}$$

Let $0 < \lambda_1 \le \lambda_2 \le ... \le \lambda_{n-m}$ be the eigenvalues of $A_J^T A_J$, and $\{u^j : j \in \mathbb{N}_{n-m}\}$ be the corresponding orthonormal set of eigenvectors, and write the vector $A_J^{-1} d_J^*$ in the form $A_J^{-1} d_J^* = \sum_{j \in \mathbb{N}_{n-m}} \gamma_j u^j$ for some constants $\gamma_j \in \mathbb{R}$ and define the subset II of $\mathbb{N}_{n-m}$ by II := $\{j : \lambda_j \gamma_j = 0\}$.

Theorem 2.3 If Λ is the least value in the spectrum of the pair $\left(\varepsilon^2 A_J^T A_J, \frac{A_J^{-1} d_J^*}{\varepsilon}\right)$

then $\mathrm{H}^*(d^* \mid E_2(A)) \neq \varnothing$ if and only if

$$\Lambda + \Lambda \sum_{j \notin \mathrm{II}} \frac{\lambda_j |\gamma_j|^2}{\Lambda - \varepsilon^2 \lambda_j} \le 1 - (d_I^*, d_I^*) \cdot$$

Proof. This result is proved in much the same way as in the paper by Khompungson and Micchelli (2011), and we refer the reader to the paper by Forsythe and Golub (1965) for proofs of the solution of the optimization problem.

Here is another way of stating (1.7): $\mathrm{H}(d \mid E_p) \neq \varnothing$ if and only if $\Lambda + \Lambda \sum_{j \notin \mathrm{II}} \frac{\lambda_j |\gamma_j|^2}{\Lambda - \varepsilon^2 \lambda_j} \le 1 - (d_I^*, d_I^*)$ for all $p \ge 2$.

Therefore, we establish the new version of theorem 1.2 with a different hypothesis.

Theorem 2.4. Let Λ be the least value in the spectrum of the pair $\left(\varepsilon^2 A_J^T A_J, \frac{A_J^{-1} d_J^*}{\varepsilon}\right)$.

If $\Lambda + \Lambda \sum_{j \notin \mathrm{II}} \frac{\lambda_j |\gamma_j|^2}{\Lambda - \varepsilon^2 \lambda_j} < 1 - (d_I^*, d_I^*)$ then

$$m_+(x_0, d \mid E_p) = \min\{\| x_0 - Q^T(c) \| + \varepsilon \||c_J\||_q + (d, c) : c \in \mathbb{R}^n\}, \tag{2.6}$$

where $\mathrm{II} := \{j : \lambda_j \gamma_j = 0\}$.

To this end, let us specialize the above results to the problem of function estimation in a reproducing kernel Hilbert space (*RKHS*). We let H_K be a *RKHS* of real-valued functions on a set *T*. The real valued function *K*(*t, s*) of *t* and *s* in *T* is called a *reproducing kernel* of *H* if the following property is satisfied for all $t \in T$ and $f \in H$.

$$f(t) = \langle K_t, f \rangle, \tag{2.7}$$

where K_t is the function defined for any $s \in T$ as $K_t(s) = K(t, s)$. Moreover, for any kernel *K* there is unique RKHS with *K* as its reproducing kernel. For our computational experiment we choose the Hardy space of square integrable function on the unit circle with reproducing kernel

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

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

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

see Duren (2000).

Let $T = \{t_j : j \in \mathbb{N}\}$ be distinct points (in 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 we define

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

According to section I, the vectors $\{x_j : j \in \mathbb{N}_n\}$ appearing above are identified with the function $\{Kt_j : j \in \mathbb{N}_n\}$. Therefore, the Gram matrix of the $\{Kt_j : j \in \mathbb{N}_n\}$ is given by

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

We recall the Cauchy determinant identity which states that for any $\{t_j : j \in \mathbb{N}_n\}$, $\{s_j : j \in \mathbb{N}_n\}$

$$\det\left(\frac{1}{1-t_i s_j}\right) i, j \in \mathbb{N}_n = \frac{\prod_{1\le j<i\le n}(t_j - t_i)(s_j - s_i)}{\prod_{i,j\in\mathbb{N}_n}(1-t_i s_j)}, \tag{2.8}$$

see for example. From this formula we obtain that

$$\det G = \frac{\prod_{1\le i\le j\le n}(t_i - t_j)^2}{\prod_{i,j\in\mathbb{N}_n}(1-t_i t_j)}. \tag{2.9}$$

According to equation (2.7), the linear operator $Q : H^2(\Delta) \to \mathbb{R}^n$ is defined for $f \in H^2(\Delta)$ as follows

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

By the Gram-Schmidt process and the formulae (2.8) and (2.9), we obtain the vector K_j^* for any $j \in \mathbb{N}_n$. In particular, the vector K_j^* is given by the formula

$$K_1^* = \sqrt{1-t_1^2}K_{t_1},$$
$$K_k^* = \sqrt{1-t_k^2}\sum_{l=1}^{k}(-1)^{k+l}\frac{\prod_{i\in\mathbb{N}_{k-1}}|1-t_l t_i|}{\prod_{i\in\mathbb{N}_k, i\ne l}|t_l - t_i|}K_{t_l}, \quad (k = 2,3,...,n).$$

For any $d = (d_1, d_2, ..., d_n) \in \mathbb{R}^n$, we obtain that the condition

$$Qf = (f(t_i) : i \in \mathbb{N}_n) = d \text{ is equivalent to } Rf = d^*,$$

where $R : H^2(\Delta) \to \mathbb{R}^n$ is a linear operator $H^2(\Delta)$ onto $\mathbb{R}^n$ as $Rf = (\langle f, K^*_j\rangle : j \in \mathbb{N}_n)$ and

$$d_1^* = \sqrt{1-t_1^2}d_1,$$
$$d_k^* = \sqrt{1-t_k^2}\sum_{l=1}^{k}(-1)^{k+l}\frac{\prod_{i\in\mathbb{N}_{k-1}}|1-t_l t_i|}{\prod_{i\in\mathbb{N}_k, i\ne l}|t_l - t_i|}d_l, \quad (k = 2,3,...,n).$$

For our example, we choose $E := \{e : e \in \mathbb{R}^n, e_I = 0, |e_n| \le \varepsilon\}$ where the set $I = \{1,2,...,n-1\}$.
The *partial hyperellipse* becomes

$$\mathrm{H}(d\,|\,E) := \left\{ f : f \in H^2(\Delta), \|f\| \le 1, Q_I(f) = d_I, |(Q(f) - d)_n| \le \varepsilon \right\}.$$

Clearly, we have only one inaccurate data and for any $f \in H_K$

$$f(t_j) = d_j \text{ for all } j \in \mathbb{N}_{n-1} \text{ and } f(t_n) = d_n + e \text{ where } |e| \le \varepsilon.$$

In this case, the corresponding *partial hyperellipse* $\mathrm{H}(d\,|\,E)$ is given by

$$\mathrm{H}^*(d^* \mid E(A)) := \left\{ f : f \in H^2(\Delta), \| f \| \le 1, R_I(f) = d_I^*, R_n(f) - d_n^* \in A(\varepsilon) \right\}.$$

where $A(\varepsilon) := \{\beta e : e \in \mathbb{R}, |e| \le \varepsilon\}$ and $\beta = \sqrt{\dfrac{\det G(t_1,...,t_{n-1})}{\det G(t_1,...,t_n)}} = \sqrt{1-t_n^2}\,\dfrac{\prod_{i \in \mathbb{N}_{n-1}} |1-t_n t_i|}{\prod_{i \in \mathbb{N}_{n-1}} |t_n - t_i|}$.

Alternatively, we have that $\mathrm{H}^*(d^* \mid E(A)) \neq \varnothing$ if and only if

$$\min\{(d_n^* + \beta e)^2 : |e| \le \varepsilon\} \le 1 - (d_I^*, d_I^*).$$

Moreover, we point out the formula

$$\min\{(d_n^* + \beta e)^2 : |e| \le \varepsilon\} = \begin{cases} 0, & \left|\dfrac{d_n^*}{\beta}\right| \le \varepsilon \\ d_n^* + \beta\varepsilon \dfrac{\hat{e}}{|\hat{e}|}, & \left|\dfrac{d_n^*}{\beta}\right| > \varepsilon \end{cases}$$

where $\hat{e} = -\dfrac{d_n^*}{\beta}$. Therefore, we establish theorem 1.2 in the following way.

Theorem 2.5 Let $T = \{t_j : j \in \mathbb{N}_n\}$ be distinct points (in increasing order) in (-1, 1), $t_0 \in (-1,1)$ and $d = (d_1, d_2 \ldots, d_n) \in \mathbb{R}^n$. Then we have the following:

1. If $\left|\dfrac{d_n^*}{\beta}\right| \le \varepsilon$ and $(d_I^*, d_I^*) < 1$ then

$$m_+(K_{t_0}, d \mid E) = \sum_{i \in \mathbb{N}_{n-1}} d_i^* \left\langle K_{t_i}^*, K_{t_0} \right\rangle + \frac{|B_I(t_0)|}{\sqrt{1-t_0^2}} \sqrt{1-(d_I^*, d_I^*)},$$

where the function B_I is the rational function defined at $t \in \mathbb{C} \setminus \left\{ t_j^{-1} : j \in \mathbb{N}_{n-1} \right\}$ by

$$B_I(t) = \prod_{j \in \mathbb{N}_{n-1}} \frac{t - t_j}{1 - t t_j}.$$

2. If $\left|\dfrac{d_n^*}{\beta}\right| > \varepsilon$ and $d_n^* + \beta\varepsilon \dfrac{\hat{e}}{|\hat{e}|} < 1 - (d_I^*, d_I^*)$ then

$$m_+(K_{t_0}, d \mid E) = \min\left\{ \| K_{t_0} - Q(f) \| + \varepsilon |c_n| + (d, c) : c \in \mathbb{R}^n \right\}.$$

Moreover, if $\dfrac{K_{t_0}}{\sqrt{1-t_0^2}} \notin H(d \mid E)$ then

$$m_+(K_{t_0}, d \mid E) = \sum_{i \in \mathbb{N}_{n-1}} d_i^* \left\langle K_{t_i}^*, K_{t_0} \right\rangle + \mathrm{d}\left\langle K_{t_n}^*, K_{t_0} \right\rangle + \frac{|B_I(t_0)|}{\sqrt{1-t_0^2}} \sqrt{1-(d_I^*, d_I^*) - \mathrm{d}^2},$$

where the value $\mathrm{d} \in \left\{ d_n^* + \beta_{nn}\varepsilon, d_n^* - \beta_{nn}\varepsilon \right\}$.

Proof. See Khompurngson, K., & Nammanee K., (2022).

3. Conclusions

In this paper, we extended the hypercircle inequality to partially corrupted data. That is, we established a new version of theorem 1.2 with a different hypothesis.

In addition, we provided a concrete example for learning problems in reproducing kernel Hilbert space when there is known only one data error. We also provided an explicit solution of a dual problem which is practically useful.

Acknowledgements

The author is supported by the National Research Council of Thailand and University of Phayao under Grant No. RIB64-RIB012.

References

Davis, P. J. (1975). *Interpolation and approximation*. New York, NY: J Dover Publications.

Duren, P. (2000). *Theory of H^P space* (2nd ed.). New York, NY: J Dover Publications.

Forsythe, G. E., & Golub, G. H. (1965). On the stationary values of a second-degree polynomial on the unit sphere. *Journal of the Society for Industrial and Applied Mathematics, 13*(40), 1050–1068.

Khompurngson, K., & Nammanee, K., (2022). A note on hypercircle inequality for data error with l 1 norm. *Journal of Inequalities and Applications, 87*, 1-15.

Khompurngson, K., & Micchelli, C.A. (2011). Hide. *Jaen Journal on Approximation, 3*, 87-115.

Khompurngson, K., & Novaprateep, B. (2015). Hypercircle inequality for partially corrupted data. *Annals of Functional Analysis, 6*, 95-108.