Ann. Funct. Anal. 6 (2015), no. 1, 95–108
http://doi.org/10.15352/afa/06-1-8
ISSN: 2008-8752 (electronic)
http://projecteuclid.org/afa

HYPERCIRCLE INEQUALITY FOR PARTIALLY-CORRUPTED DATA

KANNIKA KHOMPURNGSON* AND BORIBOON NOVAPRATEEP

Communicated by J. Soria

Abstract. In recent years, the problem of learning and methods for learning functions have received increasing attention in Machine Learning. This problem is motivated by several applications in which it is required to estimate a function representation from available data. Recently, an extension of hypercircle inequality to data error ($Hide$) was proposed by Kannika Khompurngson and Charles A. Micchelli and the results on this subject have constructed a new learning method. Unfortunately, the material on Hide only applies to circumstances for which all data are known within error. In this paper, our purpose is to extend the hypercircle inequality to circumstances for which data set contains both accurate and inaccurate data.

1. Introduction

In recent years, the problem of learning and methods for learning functions have received increasing attention in Machine Learning. This problem is motivated by several applications in which it is required to estimate a function representation from available data. There are several methods that can be used to determine a function representation from given data [1, 3]. Specifically, the well-known hypercircle inequality, which has a long history in applied mathematics, has been applied to kernel-based learning when data are known exactly. Recently, an extension of hypercircle inequality to data error ($Hide$) was proposed by Kannika Khompurngson and Charles A. Micchelli and the results on this subject have constructed a new learning method. Unfortunately, the material on Hide only applies to circumstances for which all data are known within error. In real−world

Date: Received: Mar. 29, 2013; Accepted: Jul. 18, 2013.

* Corresponding author.

2010 *Mathematics Subject Classification.* Primary 46E22; Secondary 46C07, 74PXX..

Key words and phrases. Hypercircle inequality, Convex optimization and Noise data.

problems, there are several types of data. An example of this includes partially-corrupted data. In this paper, our purpose is to extend the hypercircle inequality to circumstances for which data set contains both accurate and inaccurate data.

Let H be the Hilbert space over the real numbers with inner product $\langle \cdot, \cdot \rangle$. Consequently, we choose a linearly independent finite set $\mathcal{X} = \{x_j : j \in \mathbb{N}_n\}$ in H, where we use $\mathbb{N}_n = \{1, 2, \cdots, n\}$. We shall denote by M the $n-$dimensional subspace of H spanned by the vectors in $\mathcal{X}$. That is, we have that

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

We define the linear operator $Q : H \to \mathbb{R}^n$ as for $x \in H$

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

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

$$Q^T(a) = \sum_{i \in \mathbb{N}_n} a_i x_i. \tag{1.1}$$

Given $d \in \mathbb{R}^n$, the *hyperplane of codimension* n is defined by

$$P(d) := \{x : x \in H, Q(x) = d\}.$$

That is, for all $x \in P(d)$ we have that $\langle x_i, x \rangle = d_i$ for all $i \in \mathbb{N}_n$. For each $d \in \mathbb{R}^n$, it is well-known that there exists a unique vector $x(d) \in M$ such that

$$x(d) := arg \min\{||x|| : x \in H, Q(x) = d\}. \tag{1.2}$$

Clearly, the vector $x(d)$ is the element of $P(d)$ nearest to the origin. The portion of H common to $P(d)$ and to the unit ball $||x|| \leq 1$ is called the *hypercircle* and we use the following notation

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

Therefore, the hypercircle inequality states that:

Let $x(d)$ be the element of $P(d)$ which is nearest the origin and $x_0 \in H$. Then for any $x \in \mathcal{H}(d)$

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

where dist $(x_0, M) := \min\big\{||x_0 - y|| : y \in M\big\}$. *Moreover, if $\mathcal{H}(d) \neq \emptyset$, then there is an $x_\pm(d) \in \mathcal{H}(d)$ for which equality above holds.*

The inequality above ensures the existence of an approximate value which is the vector in the nearest point to the origin in the hyperplane. Moreover, it is independent of the vector x_0. Geometrically speaking, the value of $\langle x(d), x_0 \rangle$ is the best estimator to estimate $\langle x, x_0 \rangle$ when $x \in \mathcal{H}(d)$. It is easily seen that the best estimator $\langle x(d), x_0 \rangle$ is the midpoint of the interval of uncertainty which is defined by $I(x_0, d) := \{\langle x, x_0 \rangle : x \in \mathcal{H}(d)\}$. Indeed, the inequality above ensures that there exists $x_\pm(d) \in \mathcal{H}(d)$ such that $\langle x_+(d), x \rangle = m_+(x_0, d) := \max\{\langle x, x_0 \rangle : x \in \mathcal{H}(d)\}$ and $\langle x_-(d), x \rangle = m_-(x_0, d) := \min\{\langle x, x_0 \rangle : x \in \mathcal{H}(d)\}$ respectively. The detailed proofs appear in [4, 10].

In [7, 8], Kannika Khompurngson and Charles A. Micchelli have extended the hypercircle inequality to data error. First, we assume $E = \{e : e \in \mathbb{R}^n, |e| \le \varepsilon\}$, where $|\cdot| : \mathbb{R}^n \to \mathbb{R}_+$ is some prescribed norm on $\mathbb{R}^n$ and $\varepsilon > 0$. From now on, we follow the notation of [7] and recall the definition of the *hyperellipse* as follows

$$\mathcal{H}(d|E) := \{x : ||x|| \le 1, Q(x) - d \in E\}.$$

In the general observation, we found that

$$\mathcal{H}(d|E) = \bigcup_{e \in E} \mathcal{H}(d + e).$$

That is, each element $x \in \mathcal{H}(d|E)$ verifies the following properties:

$$Q(x) = d + e \quad \text{for some} \quad e \in E$$

and $||x|| \le 1$. Specifically, the uncertainty set defined by

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

fills out a closed and bounded interval in $\mathbb{R}$. We obtain that

$$I(x_0, d|E) := [m_-(x_0, d|E), m_+(x_0, d|E)],$$

where $m_+(x_0, d|E) := \max\big\{\langle x, x_0 \rangle : x \in \mathcal{H}(d|E)\big\}$ as well as $m_-(x_0, d|E) := \min\big\{\langle x, x_0 \rangle : x \in \mathcal{H}(d|E)\big\} = -m_+(x_0, -d|E)$. Furthermore, our result provides that the best estimator to estimate $\langle x, x_0 \rangle$ when $x \in \mathcal{H}(d|E)$ still has the form of linear combination of the vectors in $\mathcal{X}$ but the choice of the coefficients depends on the vector x_0. Therefore, we investigated that the learned function *still* has the form of Representer Theorem (1.1) but the choice of the coefficients is generally different from those obtained from a regularization method, which is the standard method for learning problem. As we said earlier, the best estimator is the midpoint of the interval of uncertainty. We then provided the useful duality formula for the right hand endpoint of the uncertainty interval. The result state as the following.

If $\mathcal{H}(d) \ne \emptyset$, then

$$m_+(x_0, d|E) = \min\big\{||x_0 - Q^T(c)|| + \varepsilon|c|_* + (d, c) : c \in \mathbb{R}^n\big\}, \tag{1.3}$$

where $|\cdot|_* : \mathbb{R}^n \longrightarrow \mathbb{R}_+$ is the conjugate norm of $|\cdot|$ which is used to measure data error and $(\cdot,\cdot)$ is the Euclidean inner product on $\mathbb{R}^n$.

Therefore, the extreme on the right hand side of equation (1.3) is a finite dimensional convex optimization problem. In summary, the midpoint of the uncertainty interval is given by

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

Furthermore, we discussed some results of numerical experiments of learning the value of a function in a reproducing kernel Hilbert space and also compared the midpoint estimator to the regularization estimator. We refer the reader to the papers [7, 8] for more detailed information on the theory and practice of midpoint algorithms for learning the value of a function from inaccurate data.

The paper is organized in the following manner. In section 2, we introduce our notations and show the results of hypercircle inequality for partially-corrupted data. The main result in this section is Theorem 2.9, which establishes that the best estimator still has the form of Representer Theorem. In addition, we provide three important cases of the existence of the minimum of the convex function which is used to obtain the midpoint estimator. In section 3, we specialize the result of section 2 to the case of data error measured with l^p norm and a conclusion appears in section 4.

2. Hypercircle inequality for partially-corrupted data

In this section, we will restrict our attention to the study of hypercircle inequality with partially-corrupted data. We start with $I \subseteq \mathbb{N}_n$ which contains m elements ($m < n$). Consequently, we use the notations $\mathcal{X}_I = \{x_i : i \in I\} \subset \mathcal{X}$ and $\mathcal{X}_J = \{x_i : i \in J\} \subset \mathcal{X}$, where we denote $J = \mathbb{N}_n \backslash I$. For each $e = (e_1, \cdots, e_n) \in \mathbb{R}^n$, we also use the notations $e_I = (e_i : i \in I) \in \mathbb{R}^m$ and $e_J = (e_i : i \in J) \in \mathbb{R}^{n-m}$ respectively. We choose $|||\cdot||| : \mathbb{R}^{n-m} \longrightarrow \mathbb{R}_+$ is a prescribed norm on $\mathbb{R}^{n-m}$ and define $\mathbb{E} = \{e : e \in \mathbb{R}^n : e_I = 0, |||e_J||| \le \varepsilon\}$, where ε is some positive number. For each $d \in \mathbb{R}^n$, we define the *partial hyperellipse* as follows

$$\mathcal{H}(d|\mathbb{E}) := \big\{x : \ x \in H, \ ||x|| \le 1, \ Q(x) - d \in \mathbb{E}\big\}. \tag{2.1}$$

Here, $\mathbb{E}$ contains our a priori information about data error and our notation differs from [7]. For example, if $|||\cdot|||_2 : \mathbb{R}^{n-m} \longrightarrow \mathbb{R}_+$ is the Euclidean norm on $\mathbb{R}^{n-m}$, then $\mathbb{E}$ is chosen to be the set

$$\mathbb{E}_2 = \{e : e \in \mathbb{R}^n : e_I = 0, |||e_J|||_2 \le \varepsilon\}$$

and the *partial hyperellipse* is given by

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

Before we add some relation between (2.1) and our previous work on *Hide*, let us introduce the notations for the linear operator

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

respectively. According to the definition of *hyperellipse* and *hypercircle*, we observe that

$$\mathcal{H}(d|\mathbb{E}) = \mathcal{H}(d_I) \cap \mathcal{H}(d_J|E_J),$$

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

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

where we define $E_J = \{c : c \in \mathbb{R}^{n-m} : |||c||| \le \varepsilon\}$. With this notation, we write E_J instead of E.

We point out that the *partial hyperellipse* is a convex subset of H which is sequentially compact in the weak topology on H. Consequently, we obtain the uncertainty interval

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

is bounded and closed on $\mathbb{R}$. To find the best predictor, we only need to evaluate the two numbers $m_\pm(x_0, d|\mathbb{E})$ and compute the midpoint $m(x_0, d|\mathbb{E})$, where we define $m_+(x_0, d|\mathbb{E}) = \sup\{\langle x, x_0\rangle : x \in \mathcal{H}(d|\mathbb{E})\}$ and $m_-(x_0, d|\mathbb{E}) = \inf\{\langle x, x_0\rangle : x \in \mathcal{H}(d|\mathbb{E})\}$ respectively. According to $\mathcal{X} \subseteq H$, the Gram matrix of the vectors in $\mathcal{X}$ is defined by

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

which is symmetric and positive definite. Again, we follow the notation of (1.2) and point out that for each $e \in \mathbb{E}$ there is a unique element $x(d+e) \in M$ defined as

$$x(d+e) := arg\min\{||x|| : x \in H, Q(x) = d + e\} \tag{2.2}$$

and it is well known that

$$x(d+e) = Q^T\big(G^{-1}(d+e)\big) \in M \text{ and } ||x(d+e)||^2 = (d+e, G^{-1}(d+e))$$

(see [9]). The first Theorem below provides the existence of right and left hand end point of the uncertainty interval.

Theorem 2.1. *If $\mathcal{H}(d|\mathbb{E}) \neq \emptyset$, then there exist $x_\pm \in \mathcal{H}(d|\mathbb{E})$ such that*

$$\langle x_\pm, x_0\rangle = m_\pm(x_0, d|\mathbb{E}).$$

Proof. Our proof start with the observation that $\mathcal{H}(d|\mathbb{E})$ is a sequentially compact subset of H and the function $x \to \langle x_0, x\rangle$ is weakly continuous. Therefore, there exist $x_\pm \in \mathcal{H}(d|\mathbb{E})$ such that $\langle x_\pm, x_0\rangle = m_\pm(x_0, d|\mathbb{E})$. □

Here is another way to obtain the right hand end point of uncertainty interval with different hypotheses. First, we recall the notion of conjugate norm, denoted by $|||\cdot|||_*$ corresponding to a prescribed norm $|||\cdot|||$ on $\mathbb{R}^{n-m}$. The conjugate norm of $|||\cdot|||$ is defined for all $c \in \mathbb{R}^{n-m}$ as

$$|||c|||_* = \max\big\{(c, w) : w \in \mathbb{R}^{n-m}, |||w||| \leq 1\big\},$$

which appears in [6]. Moreover, if $c \neq 0$, then there is a $\hat{c} \in \mathbb{R}^{n-m}$ such that $|||\hat{c}||| = 1$ and $|||c|||_* = (c, \hat{c})$. We also required a useful version of the Von Neumann Minimax Theorem which appears in [2].

Theorem (VN)[2] *Let $f : \mathcal{C} \times \mathcal{U} \to \mathbb{R}$ where $\mathcal{C}$ is a closed convex subset of a Hausdorff topological vector space* X *and $\mathcal{U}$ is a convex subset of a vector space Y. If for any $x \in \mathcal{U}$ the function $c \to f(c, x)$ is convex and lower semi-continuous on $\mathcal{C}$ and for every $c \in \mathcal{C}$ $x \to f(c, x)$ is concave on $\mathcal{U}$ and there is an $\hat{x} \in \mathcal{U}$ such that for all $\lambda \in \mathbb{R}$ the set*

$$\{c : c \in \mathcal{C}, f(c, \hat{x}) \leq \lambda\}$$

is a compact subset of X*, then there is a $c_0 \in \mathcal{C}$ such that*

$$\sup_{x\in\mathcal{U}} f(c_0, x) = \sup_{c\in\mathcal{C}} \inf_{x\in\mathcal{U}} f(c, x).$$

In particular, we have that

$$\min_{c\in\mathcal{C}}\sup_{x\in\mathcal{U}} f(c,x) = \sup_{x\in\mathcal{U}}\inf_{c\in\mathcal{C}} f(c,x).$$

We are almost ready to describe the duality formula for the right hand end point of the uncertainty interval and the result may be proved in much the same way as in the paper [7]. Before doing we prepare the following lemma.

Lemma 2.2. *If $\mathcal{H}(d|\mathbb{E})$ contains more than one point, then there exists $\hat{e} \in \mathbb{E}$ and $|||\hat{e}_{_J}||| < \varepsilon$ such that $x(d+\hat{e}) \in \mathcal{H}(d|\mathbb{E})$, where $x(d+\hat{e}) = Q^T\big(G^{-1}(d+\hat{e})\big)$.*

Proof. By our assumption and (2.2), we obtain that there exists $e \in \mathbb{E}$ such that $x(d+e) \in \mathcal{H}(d|\mathbb{E})$ and

$$||x(d+e)||^2 = (d+e, G^{-1}(d+e)) < 1.$$

Let $\alpha_n \in (0,1)$ and $\alpha_n \to 1$ as $n \to \infty$. We define $e_n = \alpha_n e$ and get that $e_n \to e$ as $n \to \infty$. Consequently, we obtain that

$$(d+e_n, G^{-1}(d+e_n)) \to (d+e, G^{-1}(d+e)) < 1$$

as $n \to \infty$. Thus, there is $\hat{e} = e_n$ and $|\hat{e}_{_J}| < \varepsilon$ for some $n \in \mathbb{N}_n$ such that $(d+\hat{e}, G^{-1}(d+\hat{e})) \leq 1$. Using (2.2) again, we conclude that there is a vector $x(d+\hat{e}) = Q^T(G^{-1}(d+\hat{e})) \in \mathcal{H}(d|\mathbb{E})$. $\square$

Theorem 2.3. *If $\mathcal{H}(d|\mathbb{E})$ contains more than one element, then*

$$m_+(x_0, d|\mathbb{E}) = \min\big\{||x_0 - Q^T(c)|| + \varepsilon|||c_{_J}|||_* + (d,c) : c \in \mathbb{R}^n\big\}, \tag{2.3}$$

where $|||\cdot|||_* : \mathbb{R}^{n-m} \longrightarrow \mathbb{R}_+$ is the conjugate norm of $|||\cdot|||$ which is used to measure data error. Moreover, if either $x_0 \notin M$ or $|||\cdot|||_*$ is strictly convex, then the right hand side of equation (2.3) has a unique solution.

Proof. For any $x \in \mathcal{H}(d|\mathbb{E})$, $c \in \mathbb{R}^n$ and $x_0 \in H$ we have that

$$\begin{aligned} \langle x_0, x\rangle &= \langle x_0 - Q^T(c), x\rangle + (c, Q(x) - d) + (c,d) \\ &= \langle x_0 - Q^T(c), x\rangle + (c_{_J}, Q_{_J}(x) - d_{_J}) + (c,d) \\ &\leq ||x_0 - Q^T(c)|| + \varepsilon|||c_{_J}|||_* + (c,d). \end{aligned}$$

Thus, we conclude that

$$m_+(x_0, d|\mathbb{E}) \leq \inf\big\{||x_0 - Q^T(c)|| + \varepsilon|||c_{_J}|||_* + (c,d) : c \in \mathbb{R}^n\big\}. \tag{2.4}$$

In order to get this inequality, we define the function $f : \mathbb{R}^{n-m} \times \mathcal{H}(d_{_I}) \longrightarrow \mathbb{R}$ by for each $b \in \mathbb{R}^{n-m}$ and $x \in \mathcal{H}(d_{_I})$ as

$$f(b,x) := \langle x_0 - Q_{_J}^T(b), x\rangle + \varepsilon|||b|||_* + (b, d_{_J}).$$

Consequently, we obtain that

$$\inf\big\{||x_0 - Q^T(c)|| + \varepsilon|||c_{_J}|||_* + (c,d) : c \in \mathbb{R}^n\big\} = \inf_{b\in\mathbb{R}^{n-m}}\max_{x\in\mathcal{H}(d_{_I})} f(b,x).$$

This follows by the same method as in [7]. We then identify $\mathcal{C} = \mathbb{R}^{n-m}$ and $\mathcal{U} = \mathcal{H}(d_{_I})$. Clearly, for each $x \in \mathcal{H}(d_{_I})$ the function $b \longrightarrow f(b,\mathcal{U})$ is convex and for $b \in \mathbb{R}^{n-m}$, $x \longrightarrow f(b,x)$ is a linear function of $\mathcal{U}$. Also, it is jointly continuous in $(b,x) \in \mathcal{C} \times \mathcal{U}$. The task is now to show that the set

$$\{b : b \in \mathbb{R}^{n-m}, f(b,\hat{x}) \leq \lambda\}$$

is a compact subset of $\mathbb{R}^n$. By our assumption and lemma 2.2 , there is an $\hat{x} \in B$ and $\hat{e} \in \mathbb{E}$ such that $Q(\hat{x}) = d + \hat{e}$ and $|||\hat{e}_J||| < \varepsilon$. So, we observe that

$$\begin{aligned} f(b,\hat{x}) &= \langle x_0 - Q_J^T(b), \hat{x}\rangle + \varepsilon|||b|||_* + (b, d_J) \\ &= \langle x_0, \hat{x}\rangle + \varepsilon|||b|||_* - (\hat{e}, b) \leq \lambda. \end{aligned}$$

By the Cauchy Schwarz inequality, we obtain that

$$|||b|||_* \leq \frac{\lambda - \langle \hat{x}, x_0\rangle}{\varepsilon - |||\hat{e}|||}.$$

Therefore, $\{b : b \in \mathbb{R}^{n-m}, f(b,\hat{x}) \leq \lambda\}$ is a bounded subset of $\mathbb{R}^{n-m}$. Hence, $\{b : b \in \mathbb{R}^{n-m}, f(b,\hat{x}) \leq \lambda\}$ is a compact subset of $\mathbb{R}^n$. Applying Theorem VN, we conclude that

$$\min\left\{||x_0 - Q^T(c)|| + \varepsilon|||c_J|||_* + (c,d) : c \in \mathbb{R}^n\right\} = \sup_{x \in \mathcal{H}(d_I)} \inf_{b \in \mathbb{R}^{n-m}} f(b,x).$$

To this end, it remains to prove that

$$\inf\left\{\varepsilon|||b|||_* + (b, d_J - Q_J(x)) : b \in \mathbb{R}^{n-m}\right\} = \begin{cases} -\infty, & |||Q_J(x) - d_J||| > \varepsilon \\ 0, & |||Q_J(x) - d_J||| \leq \varepsilon. \end{cases}$$

First, let us assume that $|||Q_J(x) - d_J||| \leq \varepsilon$. Then we see that

$$\begin{aligned} \varepsilon|||b|||_* + (b, d_J - Q_J(x)) &\geq \varepsilon|||b|||_* - |||b|||_* \cdot |||Q_J(x) - d_J||| \\ &= |||b|||_*(\varepsilon - |||Q_J(x) - d_J|||) \geq 0 \end{aligned}$$

and so in this case, the infimum on the left hand side is achieved for $b = 0$. Next we consider in the case that $|||Q_J(x) - d_J||| > \varepsilon$. We choose $\hat{b} \in \mathbb{R}^{n-m} \setminus \{0\}$ such that

$$(\hat{b}, d_J - Q_J(x)) = |||\hat{b}|||_* \cdot |||d_J - Q_J(x)|||.$$

Hence, for all $t > 0$ we have that

$$\begin{aligned} \inf\{\varepsilon|||b|||_* + (b, d_J - Q_J x) : b \in \mathbb{R}^{n-m}\} &\leq \varepsilon||| - t\hat{b}|||_* + (-t\hat{b}, d_J - Q_J x) \\ &= \varepsilon||| - t\hat{b}|||_* - (t\hat{b}, d_J - Q_J x) \\ &\leq \varepsilon||| - t\hat{b}|||_* - ||| - t\hat{b}|||_* \cdot |||d_J - Q_J x||| \\ &= t\Big(\varepsilon - |||d_J - Q_J x|||\Big) \cdot |||\hat{b}|||_* \end{aligned}$$

Letting $t \longrightarrow \infty$, we have that $\inf\{\varepsilon|||b|||_* + (b, d_J - Q_J x) : b \in \mathbb{R}^{n-m}\} = -\infty$. Therefore, we can conclude that

$$\min\left\{||x_0 - Q^T c|| + \varepsilon|||c_J|||_* + (c,d) : c \in \mathbb{R}^n\right\} = \max\left\{\langle x, x_0\rangle : x \in \mathcal{H}(d|\mathbb{E})\right\}.$$

What is left is to show that the right hand side of equation (2.3) has a unique solution. This follows by the same method as in [7]. □

Using the fact that $m_-(x_0, d|\mathbb{E}) = -m_+(x_0, -d|\mathbb{E})$, we obtain the following formula

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

where $m(x_0, d|\mathbb{E})$ is the midpoint of the uncertainty interval. Before we add an important example, let us define the convex function $V : \mathbb{R}^n \longrightarrow \mathbb{R}$ defined for $c \in \mathbb{R}^n$

$$V(c) := ||x_0 - Q^T(c)|| + \varepsilon|||c_J|||_* + (d, c).$$

As the assumption of theorem 2.3, if we drop the assumption $\mathcal{H}(d|\mathbb{E}) \neq \emptyset$, then the convex function V above does not assume its minimum.

Example If $H = \mathbb{R}^3$, $x_1 = (1,0,0), x_2 = (0,1,0)$, $d_1 = 0, d_2 = 1.5,\ x_0 = (0,0,1)$ and $\varepsilon = 0.5$. It is easy to check that $\mathcal{H}(d|\mathbb{E}) \neq \emptyset$ and it contains only one element as follows

$$\mathcal{H}(d|\mathbb{E}) = \Big\{x : ||x|| \leq 1, \langle x, x_1\rangle = 0, |\langle x, x_2\rangle - 1.5| \leq \varepsilon\Big\} = \Big\{(0,1,0)\Big\}$$

and $m_+(x_0, d|\mathbb{E}) = 0$. Next, we claim that $\inf\{\sqrt{c_1^2 + c_2^2 + 1} + 1.5c_2 + 0.5|c_2| : c \in \mathbb{R}^2\} = 0$ and the infimum is not achieved. For any $c = (c_1, c_2) \in \mathbb{R}^2$, we observe that

$$\sqrt{c_2^2 + 1} + 1.5c_2 + 0.5|c_2| \leq \sqrt{c_1^2 + c_2^2 + 1} + 1.5c_2 + 0.5|c_2|.$$

Then, we obtain that

$$\inf\{\sqrt{c_1^2 + c_2^2 + 1} + 1.5c_2 + 0.5|c_2| : c \in \mathbb{R}^2\} = \inf\{\sqrt{c_2^2 + 1} + 1.5c_2 + 0.5|c_2| : c_2 \in \mathbb{R}\}.$$

We observe that the mapping $c \rightarrow \sqrt{c^2 + 1} + 1.5c + 0.5|c|$ is increasing. Next, we claim that

$$\inf\{\sqrt{c^2 + 1} + 1.5c + 0.5|c| : c \in \mathbb{R}\} = 0.$$

This follows from the fact that

$$\lim_{c \rightarrow -\infty} \sqrt{c^2 + 1} + 1.5c + 0.5|c| = 0.$$

Therefore, $\inf\{\sqrt{c_1^2 + c_2^2 + 1} + 1.5c_2 + 0.5|c_2| : c \in \mathbb{R}^2\} = 0$ and the infimum is not achieved.

It is important to pay attention to the special case that the convex function V achieves its minimum at $c^* = 0$.

Theorem 2.4. *If $x_0 \neq 0$, then $0 = arg\ \min\{V(c) : c \in \mathbb{R}^n\}$ if and only if $\dfrac{x_0}{||x_0||} \in \mathcal{H}(d|\mathbb{E})$.*

Proof. We observe that $0 = arg\ \min\{V(c) : c \in \mathbb{R}^n\}$ holds if and only if

$$||x_0|| = V(0) \leq V(c) = ||x_0 - Q^T(c)|| + \varepsilon|||c_J|||_* + (d, c)$$

for all $c \in \mathbb{R}^n$. Since the function V is convex this inequality holds if and only if for all $c \in \mathbb{R}^n$

$$-\varepsilon|||c_J|||_* - (c,d) \leq \inf\{\frac{||x_0 - \lambda Q^T(c)|| - ||x_0||}{\lambda} : \lambda > 0\}$$

which means for all $c \in \mathbb{R}^n$ that

$$-\varepsilon|||c_J|||_* - (c,d) \leq -(\frac{Qx_0}{||x_0||}, c).$$

That is, we have that for all $c \in \mathbb{R}^n$

$$(\frac{Q(x_0)}{||x_0||} - d, c) \leq \varepsilon|||c_J|||_*. \tag{2.5}$$

First, we claim that $Q_I(\frac{x_0}{||x_0||}) = d_I$. We then choose $\hat{c} \in \mathbb{R}^n$ such that $\hat{c}_I = Q_I(\frac{x_0}{||x_0||}) - d_I$ and $\hat{c}_J = 0$. Then we have that

$$\big(Q_I(\frac{x_0}{||x_0||}) - d_I, \hat{c}_I\big) = \big(Q(\frac{x_0}{||x_0||}) - d, \hat{c}\big) \leq \varepsilon|||\hat{c}_J|||_* = 0.$$

We obtain that $Q_I(\frac{x_0}{||x_0||}) = d_I$. This means that $\frac{x_0}{||x_0||} \in \mathcal{H}(d_I)$. According to (2.5), we obtain that for each $c \in \mathbb{R}^n$

$$\big(Q_J(\frac{x_0}{||x_0||}) - d_J, c_J\big) \leq \varepsilon|||c_J|||_*,$$

which is equivalent to saying that $|||Q_J(\frac{x_0}{||x_0||}) - d_J||| \leq \varepsilon$. This means that $\frac{x_0}{||x_0||} \in \mathcal{H}(d|\mathbb{E})$ which completes the proof. □

In addition, we provide the necessary and sufficient condition on $\mathcal{H}(d|\mathbb{E})$ which provides that V achieves its minimum at c^* with $c^*_J = 0$. Let us recall a useful theorem,[7], before providing the proof of the following fact.

Theorem 2.5. *If $\mathcal{H}(d_I)$ contains more than one point, then*

$$m_+(x_0, d) = \min\left\{||x_0 - Q_I^T(a)|| + (a, d_I) : a \in \mathbb{R}^m\right\},$$

where $m_+(x_0, d_I) = \max\{\langle x, x_0\rangle : x \in \mathcal{H}(d_I)\}$. Moreover, the minimum $a^* \in \mathbb{R}^m$ is unique and

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

satisfies

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

Now we are ready to state the result.

Theorem 2.6. *If $x_0 \notin M_I := \left\{Q_I^T(a) : a \in \mathbb{R}^m\right\}$ and $\mathcal{H}(d_I)$ contains more than one point, then $c^* = arg\ \min\{V(c) : c \in \mathbb{R}^n\}$ with $c^*_J = 0$ if and only if* $\dfrac{x_0 - Q_I^T(a^*)}{||x_0 - Q_I^T(a^*)||} \in \mathcal{H}(d|\mathbb{E})$.

Proof. We first prove that $\dfrac{x_0 - Q_I^T(a^*)}{||x_0 - Q_I^T(a^*)||} \in \mathcal{H}(d|\mathbb{E})$. Under the assumption, we observe that $c^* = arg\ \min\{V(c) : c \in \mathbb{R}^n\}$ with $c_J^* = 0$ if and only if

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

which is equivalent to saying that

$$||x_0 - Q_I^T(a^*)|| + (a^*, d_I) \leq V(c)$$

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

$$-\varepsilon|||c_J|||_* - (c_J, d_J) \leq \inf\{\frac{||x_0 - Q_I^T(a^*) - \lambda Q_J^T(c_J)|| - ||x_0 - Q_I^T(a^*)||}{\lambda} : \lambda > 0\},$$

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

$$-\varepsilon|||c_J|||_* - (c_J, d_J) \leq -(\frac{Q_J(x_0 - Q_I^T(a^*))}{||x_0 - Q_I^T(a^*)||}, c_J).$$

That is, we have that

$$\Big(\frac{Q_J(x_0 - Q_I^T(a^*))}{||x_0 - Q_I^T(a^*)||} - d_J, c_J\Big) \leq \varepsilon|||c_J|||_*.$$

Therefore, we have that $\dfrac{x_0 - Q_I^T(a^*)}{||x_0 - Q_I^T(a^*)||} \in \mathcal{H}(d|\mathbb{E})$. Conversely, for each $x \in \mathcal{H}(d|\mathbb{E}) = \mathcal{H}(d_I) \cap \mathcal{H}(d_J|\mathbb{E}_J)$ we observe that

$$\langle x, x_0 \rangle \leq m_+(x_0, d_I).$$

This means, we have that $m_+(x_0, d|\mathbb{E}) \leq m_+(x_0, d_I)$. Since $\dfrac{x_0 - Q_I^T(a_I^*)}{||x_0 - Q_I^T(a_I^*)||} \in \mathcal{H}(d|\mathbb{E})$ and $m_+(x_0, d_I) = ||x_0 - Q_I^T(a_I^*)|| + (a_I^*, d_I)$, we obtain that

$$m_+(x_0, d|\mathbb{E}) = m_+(x_0, d_I) = ||x_0 - Q_I^T(a_I^*)|| + (a_I^*, d_I).$$

From (2.4), it follows that

$$||x_0 - Q_I^T(a^*)|| + (a^*, d_I) = \min\{V(c) : c \in \mathbb{R}^n\}$$

which completes the proof. □

The following result may be proved in much the same way as in theorem 2.6. To this end, let us recall an important fact for the proof, [7].

Theorem 2.7. *If $x_0 \notin M_J := \{Q_J^T(b) : b \in \mathbb{R}^{n-m}\}$ and $\mathcal{H}(d_J|\mathbb{E}_J)$ contains more than one point, then*

$$m_+(x_0, d_J|\mathbb{E}_J) = \min\{||x_0 - Q_J^T b|| + \varepsilon|||b|||_* + (b, d_J) : b \in \mathbb{R}^{n-m}\},$$

where $m_+(x_0, d_J|\mathbb{E}_J) = \max\{\langle x, x_0\rangle : x \in \mathcal{H}((d_J|\mathbb{E}_J)\}$. *Moreover, the minimum* $b^* \in \mathbb{R}^{n-m}$ *is unique and*

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

satisfies

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

.

Theorem 2.8. *If* $x_0 \notin M_J$ *and* $\mathcal{H}(d_J|\mathbb{E}_J)$ *contains more than one point, then*

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

where $c_I^* = 0$ *if and only if*

$$\frac{x_0 - Q_J^T(b^*)}{||x_0 - Q_J^T(b^*)||} \in \mathcal{H}(d|\mathbb{E}).$$

Proof. We begin by proving $\frac{x_0 - Q_J^T(b^*)}{||x_0 - Q_J^T(b^*)||} \in \mathcal{H}(d|\mathbb{E})$. First, we observe that $c^* = arg\ \min\{V(c) : c \in \mathbb{R}^n\}$ with $c_I^* = 0$ if and only if for all $c \in \mathbb{R}^n$

$$\begin{aligned} &||x_0 - Q_J^T(b^*)|| + \varepsilon|||b^*|||_* + (b^*, d_J) \\ &= \ \min\{||x_0 - Q_J^T(b)|| + \varepsilon|||b|||_* + (b, d_J) : b \in \mathbb{R}^{n-m}\} \\ &\leq \ V(c). \end{aligned}$$

For all $c \in \mathbb{R}^n$ with $c_J = b^*$, we obtain that

$$-(c_I, d_I) \leq \inf\{\frac{||x_0 - Q_J^T(b^*) - \lambda Q_I^T c_I|| - ||x_0 - Q_J^T(b^*)||}{\lambda} : \lambda > 0\},$$

which means that

$$-(c_I, d_I) \leq -(\frac{Q_I(x_0 - Q_J^T(b^*))}{||x_0 - Q_J^T(b^*)||}, c_I).$$

That is, we have that

$$(\frac{Q_I(x_0 - Q_J^T(b^*))}{||x_0 - Q_J^T(b^*)||} - d_I, c_I) \leq 0.$$

Therefore, we have that $\frac{Q_I(x_0 - Q_J^T(b^*))}{||x_0 - Q_J^T(b^*)||} \in \mathcal{H}(d|\mathbb{E})$. Conversely, for each $x \in \mathcal{H}(d|\mathbb{E}) = \mathcal{H}(d_I) \cap \mathcal{H}(d_J|\mathbb{E}_J)$ we observe that

$$\langle x, x_0\rangle \leq m_+(x_0, d_J|\mathbb{E}_J).$$

This means, we have that $m_+(x_0, d|\mathbb{E}) \leq m_+(x_0, d_J|\mathbb{E}_J)$. Using (2.4) again, we obtain that

$$||x_0 - Q_J^T b^*|| + \varepsilon|||b^*|||_* + (b^*, d_J) = \min\{V(c) : c \in \mathbb{R}^n\}.$$

□

We add a final result which is the key to construct the best predictor.

Theorem 2.9. *If $\mathcal{H}(d|\mathbb{E}) \neq \emptyset$, then there is an $e_0 \in \mathbb{E}$ such that*

$$\langle x(d+e_0), x_0\rangle = m(x_0, d|\mathbb{E}),$$

where $x(d+e_0) = Q^T\big(G^{-1}(d+e_0\big) \in \mathcal{H}(d|\mathbb{E})$.

Proof. According to theorem 2.1, there exists $x_\pm \in \mathcal{H}(d|\mathbb{E})$ such that

$$\langle x_\pm, x_0\rangle = m_\pm(x_0, d|\mathbb{E}).$$

Consequently, there exists $e_\pm \in \mathbb{E}$ such that $x_\pm \in \mathcal{H}(d+e_\pm)$ and

$$\langle x_\pm, x_0\rangle = \langle x(d+e_\pm), x_0\rangle \pm dist\ (x_0, M)\sqrt{1 - ||x(d+e_\pm)||^2} = m_\pm(x_0, d|\mathbb{E}).$$

We can follow the proof from the paper [7] to obtain the vector $e_0 \in \mathbb{E}$ which is on the line segment joining the vector e_- and e_+. □

Therefore, the Hypercircle inequality for partially-corrupted data error becomes in the following way. If $x_0 \in H$ and $\mathcal{H}(d|\mathbb{E}) \neq \emptyset$, then there is an $e_0 \in \mathbb{E}$ such that for any $x \in \mathcal{H}(d|\mathbb{E})$

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

where $x(d+e_0) = Q^T\big(G^{-1}(d+e_0)\big) \in \mathcal{H}(d|\mathbb{E})$.

3. Hypercircle inequality for partially-corrupted data error measured with l^p norm

In this section, we consider the case that the data error is measured with l^p norm. Firstly, let us denote the notation of partial hyperellipse for data error measured with l^p norm ($1 < p < \infty$) by $\mathbb{E}_p = \big\{e : e \in \mathbb{R}^{n-m}, |||e|||_p \leq \varepsilon\big\}$ where $|||\cdot|||_p$ is the l^p norm on $\mathbb{R}^{n-m}$. For any $d \in \mathbb{R}^n$, the partial hyperellipse is defined by

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

We point out that if $1 < p_1 \leq p_2 < \infty$, then $\mathcal{H}(d|\mathbb{E}_{p_1}) \subseteq \mathcal{H}(d|\mathbb{E}_{p_2})$. To prove this, we observe that for all $x \in \mathcal{H}(d|\mathbb{E}_{p_1})$ it holds that $|||Q_J(x) - d_J|||_{p_1} < \varepsilon$. That is, $|||\dfrac{Q_J(x) - d_J}{\varepsilon}|||_{p_1}^{p_1} < 1$. Since $p_1 \leq p_2$, this inequality implies that $|||\dfrac{Q_J(x) - d_J}{\varepsilon}|||_{p_2}^{p_2} < 1$ which means that $x \in \mathcal{H}(d|\mathbb{E}_{p_2})$.

In this case, we provide the different hypotheses for the unique minimum of the function $V_q : \mathbb{R}^n \to \mathbb{R}$ which is defined by for $c \in \mathbb{R}^n$

$$V_q(c) := ||x_0 - Q^T c|| + \varepsilon|||c_J|||_q + (c, d).$$

Theorem 3.1. *Suppose that $\mathcal{H}(d|\mathbb{E}_p)$ contains more than one point, $x_0 \notin M$, and* $\dfrac{x_0 - Q_I^T(a^*)}{||x_0 - Q_I^T(a^*)||} \notin \mathcal{H}(d|\mathbb{E}_p)$. *Then*

$$m_+(x_0, d) = \min\{||x_0 - Q^T c|| + \varepsilon|||c_J|||_q + (c, d) : c \in \mathbb{R}^n\}, \tag{3.1}$$

where $|||\cdot|||_q : \mathbb{R}^{n-m} \to \mathbb{R}_+$ is the conjugate norm of the l^p-norm which is used to measure data error, $\mathbb{E}_p$. Moreover, the minimum $c^* \in \mathbb{R}^n$ is the unique solution of the nonlinear equation

$$-Q(\frac{x_0 - Q^T c^*}{\| x_0 - Q^T c^* \|}) + \varepsilon w^* + d = 0, \tag{3.2}$$

where w^* is the vector in $\mathbb{R}^n$ with components given by the formula

$$w_i^* = \begin{cases} 0, & i \in I \\ \dfrac{|c_i^*|^{q-1} sgn\ c_i^*}{(|||c_J^*|||_q)^{\frac{q}{p}}}, & i \in J \end{cases}$$

and

$$x_+(d|\mathbb{E}_p) := \frac{x_0 - Q^T c^*}{||x_0 - Q^T c^*||}$$

satisfies

$$x_+(d|\mathbb{E}_p) = arg\ \ \max\{\langle x, x_0\rangle : x \in \mathcal{H}(d|\mathbb{E}_p)\}. \tag{3.3}$$

Proof. According to our hypotheses, the minimum $c^* \in \mathbb{R}^n$ is the unique solution of the function V_q and $c_J^* \neq 0$. Hence, computing the gradient of V_q gives equation (3.2). Therefore, we obtain

$$Q(x_+(d|\mathbb{E}_p)) - d = \varepsilon w^*$$

which confirms that $x_+(d|\mathbb{E}_p) \in \mathcal{H}(d|\mathbb{E}_p)$. Also, we have that $\langle x_0, x_+(d|\mathbb{E}_p)\rangle = \langle x_0 - Q^T c^*, x_+(d|\mathbb{E}_p)\rangle + (Q(x_+(d|\mathbb{E}_p)) - d, c^*) + (c^*, d) = V_q(c^*)$ which proves (3.1), (3.3) and the Theorem. $\square$

We end this section by considering the special case that $\mathcal{X} = \{x_j : j \in \mathbb{N}_n\}$ is an *orthonormal* set of vectors. Consequently, the Gram matrix is the identity matrix and we have the following for any $x(d+e) \in \mathcal{H}(d|\mathbb{E}_p) \cap M$

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

where $x(d_I) \in \mathcal{H}(d_I)$ and $x(d_J + e_J) \in \mathcal{H}(d_J|E_J)$. Moreover, we observe that $\mathcal{H}(d|\mathbb{E}_p) \neq \emptyset$ if and only if

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

For $p = 2$, we recall the formula of the minimum of quadratic polynomial on sphere as presented in [5]. Let $\Lambda = \varepsilon^2 - \sqrt{\sum_{i\in J} d_i^2}$.Then we have the following $\mathcal{H}(d|\mathbb{E}_2) \neq \emptyset$ if and only if

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

where $\mathbb{I} := \{j : d_j = 0, j \in J\}$. Summarizing, we obtain the useful formula for checking when $\mathcal{H}(d|\mathbb{E}_p) \neq \emptyset$. If $\Lambda + \Lambda \sum_{j \notin \mathbb{I}} \frac{d_j^2}{\Lambda - \varepsilon^2} \leq 1 - ||x(d_J)||^2$, then $\mathcal{H}(d|\mathbb{E}_p) \neq \emptyset$ for all $p \geq 2$.

4. Conclusion

In this paper, we presented the existence of learned function from partially-corrupted data which is obtained by the midpoint algorithm. This framework is also specified to the important case of a reproducing kernel Hilbert space. Within the proposed, we provided three important cases of the existence of the minimum of the convex function V which is useful for practice.

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

References

1. R. Akgün, *Approximation by polynomials in rearrangement invariant quasi Banach function spaces*, Banach J. Math. Anal. **6** (2012), no. 2, 113–131.
2. J.P. Aubin, *Mathematical methods of game and economic theory*, Studies in Mathematics and its Application, Volume 7, North Holland, 1982.
3. H.H. Cuenya, F.E Levis and C.N. Rodriguez, *Optimal bondles in normed spaces*, Ann. Funct. Anal. **4** (2013), no. 2, 87–96.
4. P.J. Davis, *Interpolation and approximation*, J Dover Publications, New York, 1975.
5. G.E. Forsythe and G.H. Golub, *On the stationary values of a second-degree polynomial on the unit sphere,* SIAM J. Appl. Math **13** (1965), 1050–1068.
6. R.A. Horn and C.R. Johnson, *Matrix analysis*, Cambridge University Press, 1985.
7. K. Khompurngson and C.A. Micchelli, *Hide*, Jaen J. Approx. **3** (2011), no. 1, 87–115.
8. K. Khompurngson, B. Novaprateep and Y. Lenbury, *Learning the value of a function from inaccurate data*, East-West J. Math. 2010, Special Vol., 128–138.
9. C.A. Micchelli and T.J. Rivlin, *A survey of optimal recovery in: Optimal estimation in approximation theory*, C.A Micchelli, and T. J. Rivlin, eds., Plenum Press, 1977.
10. H.L. Royden, *Real analysis*, Macmillan Publishing Company, 3rd edition, New York, 1988.

[1] Division of Mathematics, School of Science, University of Phayao, THAILAND and Centre of Excellence in Mathematics, PERDO, CHE, THAILAND.
E-mail address: kannika.kh@up.ac.th

[2] Department of Mathematics, Faculty of Science, Mahidol University, THAILAND and Centre of Excellence in Mathematics, PERDO, CHE, THAILAND.
E-mail address: scbnv@mahidol.ac.th