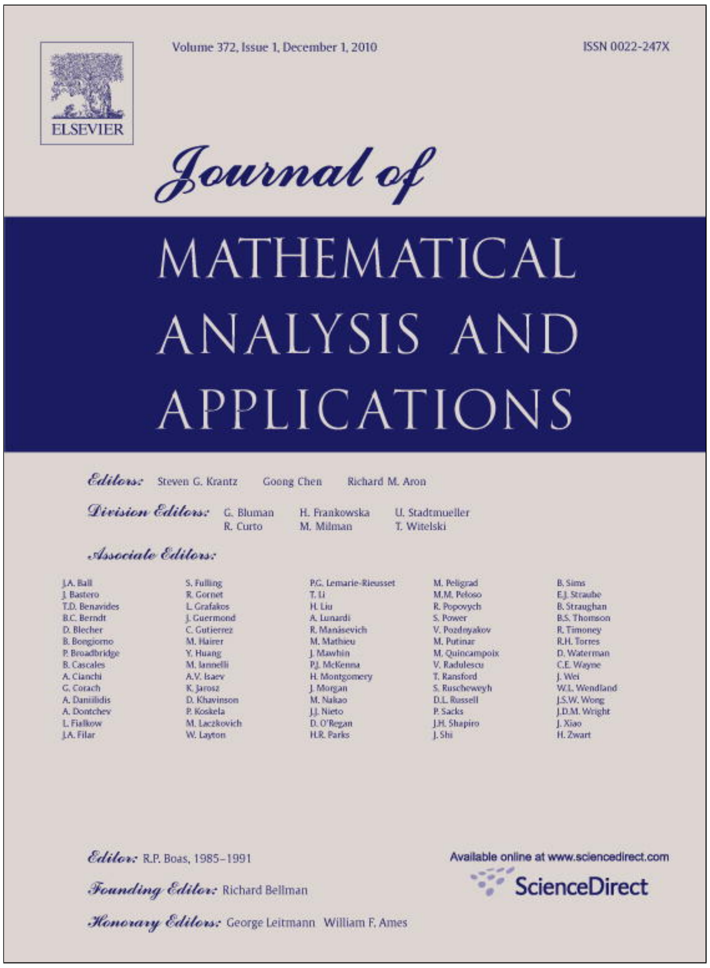
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

J. Math. Anal. Appl. 372 (2010) 181–196

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Generalized semi-inner products with applications to regularized learning ☆

Haizhang Zhang, Jun Zhang *

University of Michigan, Ann Arbor, MI 48109, USA

ARTICLE INFO

Article history:
Received 22 October 2009
Available online 23 May 2010
Submitted by J.A. Ball

Keywords:
Generalized semi-inner products
Duality mappings
Riesz representation theorem
Regularization networks
Representer theorems
Characterization equations

ABSTRACT

We propose a definition of generalized semi-inner products (g.s.i.p.). By relating them to duality mappings from a normed vector space to its dual space, a characterization for all g.s.i.p. satisfying this definition is obtained. We then study the Riesz representation of continuous linear functionals via g.s.i.p. As applications, we establish a representer theorem and characterization equation for the minimizer of a regularized learning from finite or infinite samples in Banach spaces of functions.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Let V be a vector space over $\mathbb{C}$. A *semi-inner product* (s.i.p.) on V is a function $[\cdot,\cdot]$ from $V \times V$ to $\mathbb{C}$ such that for all $\alpha, \beta \in \mathbb{C}$ and $x, y, z \in V$

(1) $[\alpha x + \beta y, z] = \alpha[x, z] + \beta[y, z]$;
(2) $[x, x] > 0$ for $x \neq 0$;
(3) $|[x, y]| \leqslant [x, x]^{1/2}[y, y]^{1/2}$.

It was shown in [21] that if $[\cdot,\cdot]$ is an s.i.p. on V then $\|x\| := [x, x]^{1/2}$ is a norm on V. Conversely, if V is a normed vector space then it has an s.i.p. that induces its norm in this manner.

Fundamental properties and consequences of s.i.p. were explored by Giles [15]. The concept of s.i.p. has been proved useful both theoretically and practically. The application of s.i.p. in the theory of functional analysis was demonstrated, for example, in [10,12,20–22,27–29,34]. They have recently found applications in machine learning. Der and Lee [9] investigated hard margin classification in Banach spaces with the aid of s.i.p. Motivated by the need of developing learning in Banach spaces [3,4,14,16,18,24,36,41,42], we established the theory of reproducing kernel Banach spaces (RKBS) in a recent work [40]. The usage of s.i.p. there made possible a systematic study of standard learning schemes in Banach spaces. In particular, it was a key tool in proving the crucial representer theorem for the following regularized learning from finite

☆ Supported by the National Science Foundation under grant 0631541.
* Corresponding author.
E-mail addresses: haizhang@umich.edu (H. Zhang), junz@umich.edu (J. Zhang).

0022-247X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmaa.2010.04.075

Author's personal copy

samples:

$$\min_{f\in\mathcal{B}}\sum_{j=1}^{n}\left|f(x_j)-y_j\right|^2+\lambda\|f\|_{\mathcal{B}}^2, \tag{1.1}$$

where $\mathcal{B}$ is an RKBS on an input space $\mathcal{Z}$, $\|\cdot\|_{\mathcal{B}}$ denotes its norm, $\{(x_j,y_j)\colon j=1,2,\ldots,n\}\subseteq\mathcal{Z}\times\mathbb{C}$ is a finite set of samples, and λ is a positive regularization parameter. By introducing an s.i.p. $[\cdot,\cdot]$ on $\mathcal{B}$ and using the following differential property of the regularizer $\|\cdot\|_{\mathcal{B}}^2$ in (1.1)

$$\lim_{t\in\mathbb{R},\,t\to 0}\frac{\frac{1}{2}\|f+t\tilde{f}\|_{\mathcal{B}}^2-\frac{1}{2}\|f\|_{\mathcal{B}}^2}{t}=\operatorname{Re}\bigl([\tilde{f},f]\bigr),\quad f,\tilde{f}\in\mathcal{B}, \tag{1.2}$$

it was proved in [40] by a variational method that the unique minimizer f_0 of (1.1) satisfies for some constants $c_j\in\mathbb{C}$, $j=1,2,\ldots,n$ that

$$[g,f_0]=\sum_{j=1}^{n}c_j g(x_j),\quad g\in\mathcal{B}.$$

When $\mathcal{B}$ is a reproducing kernel Hilbert space (RKHS) [2] with a reproducing kernel K, the above fact can be equivalently stated as

$$f_0=\sum_{j=1}^{n}\overline{c_j}K(x_j,\cdot).$$

This result is known as the representer theorem in the machine learning community [1,7,19,30].

Learning schemes more sophisticated than (1.1) often involve a general nondecreasing regularizer, namely, the regularizer $\|\cdot\|_{\mathcal{B}}^2$ in (1.1) is replaced by $g(\|\cdot\|_{\mathcal{B}})$ for some nondecreasing function g from $\mathbb{R}_+:=[0,+\infty)$ to $\mathbb{R}_+$. Considering the importance of the property (1.2), it is desirable to have a similar one for $g(\|\cdot\|_{\mathcal{B}})$ through a generalized s.i.p $[\cdot,\cdot]_\varphi$ on $\mathcal{B}$:

$$\lim_{t\in\mathbb{R},\,t\to 0}\frac{g(\|f+t\tilde{f}\|_{\mathcal{B}})-g(\|f\|_{\mathcal{B}})}{t}=\operatorname{Re}\bigl([\tilde{f},f]_\varphi\bigr),\quad f,\tilde{f}\in\mathcal{B}. \tag{1.3}$$

Here, motivated by the properties (1)–(3) that define an s.i.p., we call a function $[\cdot,\cdot]_\varphi: V\times V\to\mathbb{C}$ a *generalized semi-inner product* (g.s.i.p.) on a vector space V if it satisfies the following three conditions:

(i) Linearity with respect to the first variable:

$$[\alpha x+\beta y,z]_\varphi=\alpha[x,z]_\varphi+\beta[y,z]_\varphi\quad\text{for all }\alpha,\beta\in\mathbb{C}\text{ and }x,y,z\in V; \tag{1.4}$$

(ii) Positivity: $[x,x]_\varphi>0$ for all $x\in V\setminus\{0\}$;

(iii) A generalization of the Cauchy–Schwartz inequality: there holds for some $\varphi,\psi:\mathbb{R}_+\to\mathbb{R}_+$ that

$$\bigl|[x,y]_\varphi\bigr|\leqslant\varphi\bigl([x,x]_\varphi\bigr)\psi\bigl([y,y]_\varphi\bigr),\quad x,y\in V \tag{1.5}$$

and the equality holds when $x=y$.

For notational simplicity, we shall denote for every positive number s by δ_s the power function given as

$$\delta_s(t):=t^s,\quad t\in\mathbb{R}_+.$$

When $\varphi=\delta_{1/p}$ and $\psi=\delta_{1/q}$, where $p,q\in(1,+\infty)$ are a pair of conjugated numbers such that $1/p+1/q=1$, g.s.i.p. satisfying (1.5) were discovered and termed as *semi-inner products of type p* by Nath [26]. The purpose of this paper is to investigate all the generalized s.i.p. and their applications to regularized learning via Eq. (1.3). The organization of the paper and main results to be obtained are summarized below.

Firstly, we shall observe in Section 2 the following relationship that should exist between φ and ψ:

$$\varphi(t)\psi(t)=t,\quad t\in\mathbb{R}_+, \tag{1.6}$$

which justifies the sole subscript φ in a g.s.i.p. $[\cdot,\cdot]_\varphi$. We shall then prove that if $[\cdot,\cdot]_\varphi$ is a g.s.i.p. on a vector space V then $\|x\|:=\varphi([x,x]_\varphi)$ defines a norm on V. Conversely, if φ is surjective onto $\mathbb{R}_+$ then for any normed vector space V there exists a g.s.i.p. on it such that

$$\|x\|_V=\varphi\bigl([x,x]_\varphi\bigr),\quad x\in V. \tag{1.7}$$

Author's personal copy

The latter two results when $\varphi = \delta_{1/2}$ and $\varphi = \delta_{1/p}$, $p \in (1, +\infty)$ were due to Lumer [21] and Nath [26], respectively. It follows immediately from our generalization that the index s in $\varphi = \delta_s$ is not restricted to $(0, 1)$.

We shall also discuss in Section 2 g.s.i.p. with properties desirable for applications. We call a g.s.i.p. $[\cdot,\cdot]_\varphi$ on a normed vector space V *compatible* if (1.7) holds true. We also say that $[\cdot,\cdot]_\varphi$ is *consistent* if $[x,x]_\varphi$ is truly a measurement of the norm of x in V, in other words, if $[x,x]_\varphi$ is a gauge function of $\|x\|_V$. Recall that a function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is a *gauge function* if it is continuous and strictly increasing with $f(0) = 0$ and $\lim_{t\to\infty} f(t) = +\infty$. Let V^* be the space of continuous linear functionals on V. A function $J : V \to V^*$ is called a *duality mapping* (see, for example, [5, page 25]) if there exists a gauge function f on $\mathbb{R}_+$ such that

$$J(x)(x) = \big\|J(x)\big\|_{V^*}\|x\|_V \quad \text{and} \quad \big\|J(x)\big\|_{V^*} = f\big(\|x\|_V\big), \quad x \in V, \tag{1.8}$$

where $J(x)(x)$ denotes the application of the linear functional $J(x)$ on the vector x. We shall show that $[\cdot,\cdot]_\varphi$ is a consistent and compatible g.s.i.p. on a normed vector space V if and only if there exists a duality mapping $J : V \to V^*$ such that

$$[x, y]_\varphi = \frac{J(y)(x)}{\|y\|_V} \quad \text{and} \quad \varphi\big(\big\|J(x)\big\|_{V^*}\big) = \|x\|_V, \quad x, y \in V.$$

Secondly, we shall investigate in Section 3 the Riesz representation of continuous linear functionals via g.s.i.p. Such a representation is crucial for regularized learning where samples are usually considered to be the outcome of the application of continuous linear functionals on a function under sampling. Let φ be a gauge function and $[\cdot,\cdot]_\varphi$ a consistent and compatible g.s.i.p. on a normed vector space V. Introduce the function

$$\gamma(t) := \frac{\varphi^{-1}(t)}{t}, \quad t \in \mathbb{R}_+, \tag{1.9}$$

where we have made the convention throughout the paper that $0/0 := 0$. Saying that for every $u \in V^*$ there exists a vector $y \in V$ such that

$$u(x) = [x, y]_\varphi, \quad x \in V \tag{1.10}$$

is equivalent to requiring the mapping $J_\varphi : V \to V^*$ defined by

$$J_\varphi(x)(y) := [y, x]_\varphi \tag{1.11}$$

to be surjective. Similarly, the representer y in (1.10) is unique for every $u \in V^*$ if and only if J_φ is injective. It will be shown that J_φ is surjective if and only if V is reflexive and γ is surjective onto $\mathbb{R}_+$, and J_φ is injective if V is strictly convex and γ is injective. These two results when $\varphi = \delta_{1/2}$ were established in [12] and extended in [29] to the case when $\varphi = \delta_{1/p}$, $p \in (1, +\infty)$. Assuming that the mapping J_φ is bijective, we shall also prove in Section 3 that for the following g.s.i.p. on V^* defined by

$$[u, v]_\psi := \big[(J_\varphi)^{-1}(v), (J_\varphi)^{-1}(u)\big]_\varphi, \quad u, v \in V^* \tag{1.12}$$

to be consistent and compatible on V^*, it is necessary and sufficient that

$$[x, y]_\varphi = J(y)(x) \quad \text{and} \quad \varphi\big(\big\|J(x)\big\|_{V^*}\|x\|_V\big) = \|x\|_V, \quad x, y \in V,$$

where J is a duality mapping from V to V^*. Induced s.i.p. on V^* were used to yield dual formulations of learning schemes [9,40]. We propose (1.12) for its such future applications.

Finally, we shall fulfill in the last section our main purpose of exploring the application of g.s.i.p. to regularized learning of a target function h in a Banach space V from its finite or infinite samples. Let $(\Omega, \mathcal{F}, \mu)$ be probability measure space and $L^2_\mu(\Omega)$ denote the space of all the $\mathcal{F}$-measurable functions f from Ω to $\mathbb{C}$ such that the norm

$$\|f\|_{L^2_\mu(\Omega)} := \bigg(\int_\Omega \big|f(\omega)\big|^2 \, d\mu(\omega)\bigg)^{1/2}$$

is finite. Suppose that through a prescribed bounded set $\{\nu_\omega \in V^*\colon\ \omega \in \Omega\}$ of sampling functionals in V^*, the following sampled data of the target function h

$$\rho(\omega) := \nu_\omega(h), \quad \omega \in \Omega \tag{1.13}$$

is available. We also assume that $\omega \to \nu_\omega(f) \in L^2_\mu(\Omega)$ for each $f \in V$. By the assumption, $\rho \in L^2_\mu(\Omega)$. We aim at inferring an approximation of h from its sampled data ρ. We shall follow the widely used Tikhonov regularization approach [33]. Specifically, we consider the following minimization problem:

$$\inf\bigg\{\int_\Omega \big|\nu_\omega(f) - \rho(\omega)\big|^2 \, d\mu(\omega) + \lambda g\big(\|f\|_V\big)\colon\ f \in V\bigg\}, \tag{1.14}$$

Author's personal copy

where λ is a positive regularization parameter and g, serving as a regularizer, is a nondecreasing function from $\mathbb{R}_+$ to $\mathbb{R}_+$ with $g(0)=0$. One notes from (1.14) that the use of a probability space accommodates infinite samples and allows the loss for an error to be input-dependent. The problem (1.14) covers the usual regularization learning from finite samples in machine learning. Details are provided in Section 4.

We impose the additional assumptions that V is reflexive, strictly convex, Gâteaux differentiable, g is convex and continuously differentiable, and g' is surjective onto $\mathbb{R}_+$. Under these assumptions, (1.14) has a unique minimizer. Introduce a consistent and compatible g.s.i.p. $[\cdot,\cdot]_\varphi$ on V, where φ is defined by

$$\varphi^{-1}(t) := tg'(t), \quad t \in \mathbb{R}_+. \tag{1.15}$$

Such a g.s.i.p. satisfies (1.3), based on which two main results about the minimization problem (1.14) will be obtained:

(a) Representer theorem: the unique minimizer f_0 of (1.14) satisfies that $J_\varphi(f_0)$ belongs to $\overline{\operatorname{span}}\{\nu_\omega\colon \omega \in \Omega\}$, the closure of the linear span of $\{\nu_\omega\colon \omega \in \Omega\}$.
(b) Characterization equation: $f_0 \neq 0$ is the minimizer of (1.14) if and only if $\lambda J_\varphi(f_0) = \mathcal{T}_{f_0,\Omega}$, while $f_0 = 0$ is the minimizer if and only if $\lambda g'(0) \geqslant \|\mathcal{T}_{0,\Omega}\|_{V^*}$, where we associate each $f \in V$ with a continuous linear functional $\mathcal{T}_{f,\Omega}$ on V set by

$$\mathcal{T}_{f,\Omega}(\tilde f) := 2\int_\Omega \overline{\rho(\omega) - \nu_\omega(f)}\,\nu_\omega(\tilde f)\,d\mu(\omega), \quad \tilde f \in V. \tag{1.16}$$

These results include the classical ones [8,11,19,30–32,35,37,38] for the regularization networks in RKHS as special cases.

2. Generalized semi-inner products

We start with justifying the relationship (1.6). Suppose that $[\cdot,\cdot]_\varphi$ is a g.s.i.p. on a vector space V. Since (1.5) becomes an equality when $x = y$, we get that

$$[x,x]_\varphi = \varphi\big([x,x]_\varphi\big)\psi\big([x,x]_\varphi\big), \quad x \in V. \tag{2.1}$$

By $[x,x]_\varphi > 0$ for $x \neq 0$, we must have

$$\varphi\big([x,x]_\varphi\big) > 0, \quad x \neq 0,$$

which, by Eq. (2.1), implies that

$$\psi\big([x,x]_\varphi\big) = \frac{[x,x]_\varphi}{\varphi([x,x]_\varphi)}, \quad x \neq 0.$$

Furthermore, $[\cdot,\cdot]_\varphi$ being linear with respect to its first variable implies that $[0,0]_\varphi = 0$. This together with (2.1) necessitates that $\varphi(0)\psi(0) = 0$, so we may impose that

$$\varphi(0) = \psi(0) = 0.$$

Since we are only interested in the behavior of φ and ψ on $\{t \in \mathbb{R}_+\colon t = [x,x]_\varphi \text{ for some } x \in V\}$, we may require by the above three equations that

$$\varphi(0) = 0, \qquad \varphi(t) > 0 \quad \text{for } t > 0 \tag{2.2}$$

and

$$\psi(t) = \frac{t}{\varphi(t)}, \quad t \in \mathbb{R}_+. \tag{2.3}$$

According to (2.3), the inequality (1.5) has the equivalent form

$$\big|[x,y]_\varphi\big| \leqslant \varphi\big([x,x]_\varphi\big)\frac{[y,y]_\varphi}{\varphi([y,y]_\varphi)}, \quad x, y \in V. \tag{2.4}$$

We next show that a g.s.i.p. induces a norm.

Theorem 2.1. *Let $[\cdot,\cdot]_\varphi$ be a g.s.i.p. on a vector space V over $\mathbb{C}$. Then $\|x\|_\varphi := \varphi([x,x]_\varphi)$ defines a norm on V.*

Proof. By $[0,0]_\varphi = 0$, $\|0\|_\varphi = \varphi([0,0]_\varphi) = \varphi(0) = 0$. If $x \neq 0$ then $[x,x]_\varphi > 0$. Since $\varphi(t) > 0$ for $t > 0$, $\varphi([x,x]_\varphi) > 0$. Therefore, we obtain that $\|x\|_\varphi \geqslant 0$, $x \in V$ and $\|x\|_\varphi = 0$ if and only if $x = 0$.

Author's personal copy

Let $\alpha \in \mathbb{C} \setminus \{0\}$ and $x \in V \setminus \{0\}$. We proceed by properties (i)–(iii) of g.s.i.p. that

$$[\alpha x, \alpha x]_\varphi = \big|[\alpha x, \alpha x]_\varphi\big| = |\alpha|\big|[x, \alpha x]_\alpha\big| \leqslant |\alpha|\varphi\big([x, x]_\varphi\big)\frac{[\alpha x, \alpha x]_\varphi}{\varphi([\alpha x, \alpha x]_\varphi)},$$

which implies that

$$\varphi\big([\alpha x, \alpha x]_\alpha\big) \leqslant |\alpha|\varphi\big([x, x]_\varphi\big),$$

that is, $\|\alpha x\|_\varphi \leqslant |\alpha|\|x\|_\varphi$. On the other hand,

$$\|x\|_\varphi = \left\|\frac{1}{\alpha}\alpha x\right\|_\varphi \leqslant \frac{1}{|\alpha|}\|\alpha x\|_\varphi,$$

which yields that $\|\alpha x\|_\varphi \geqslant |\alpha|\|x\|_\varphi$. Therefore, $\|\alpha x\|_\varphi = |\alpha|\|x\|_\varphi$ for all $\alpha \in \mathbb{C} \setminus \{0\}$ and $x \in V \setminus \{0\}$. Clearly, the equality also holds when $\alpha = 0$ or $x = 0$.

Finally, we need to show that $\|x + y\|_\varphi \leqslant \|x\|_\varphi + \|y\|_\varphi$, or equivalently,

$$\varphi\big([x + y, x + y]_\varphi\big) \leqslant \varphi\big([x, x]_\varphi\big) + \varphi\big([y, y]_\varphi\big). \tag{2.5}$$

To this end, we proceed by (i) and (iii) that

$$\begin{aligned}
[x + y, x + y]_\varphi &= [x, x + y]_\varphi + [y, x + y]_\varphi \leqslant \big|[x, x + y]_\varphi\big| + \big|[y, x + y]_\varphi\big| \\
&\leqslant \varphi\big([x, x]_\varphi\big)\frac{[x + y, x + y]_\varphi}{\varphi([x + y, x + y]_\varphi)} + \varphi\big([y, y]_\varphi\big)\frac{[x + y, x + y]_\varphi}{\varphi([x + y, x + y]_\varphi)} \\
&= \big(\varphi\big([x, x]_\varphi\big) + \varphi\big([y, y]_\varphi\big)\big)\frac{[x + y, x + y]_\varphi}{\varphi([x + y, x + y]_\varphi)},
\end{aligned}$$

which implies (2.5) immediately. □

Let V be a normed vector space. If $[\cdot,\cdot]_\varphi$ is a g.s.i.p. on V then $\|x\|_\varphi = \varphi([x, x]_\varphi)$ induces a norm on V. It might not be the same as the original one on V. In the rest of the paper, we shall denote the original norm on V by $\|\cdot\|$ and that on V^* by $\|\cdot\|_*$. Recall that $[\cdot,\cdot]_\varphi$ is said to be compatible on V if $\|\cdot\|_\varphi = \|\cdot\|$. Since $\|\cdot\|_\varphi$ is a norm, φ must be surjective onto $\mathbb{R}_+$. This also turns out to be sufficient for the existence of a compatible g.s.i.p. on V.

Theorem 2.2. *If φ is surjective onto $\mathbb{R}_+$ with* (2.2) *then there exists a g.s.i.p. $[\cdot,\cdot]_\varphi$ on V such that $\|\cdot\|_\varphi = \|\cdot\|$.*

Proof. Since φ is surjective, there exists $f : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\varphi\big(f(t)\big) = t, \quad t \in \mathbb{R}_+.$$

For each $x \in V$, by the Hahn–Banach theorem, there exists an $\tilde{x} \in V^*$ such that

$$\tilde{x}(x) = f\big(\|x\|\big)$$

and

$$\|\tilde{x}\|_* = \frac{f(\|x\|)}{\|x\|}.$$

Set

$$[x, y]_\varphi := \tilde{y}(x), \quad x, y \in V.$$

It can be verified that $[\cdot,\cdot]_\varphi$ so defined indeed is a g.s.i.p. on V. Finally, we have for every $x \in V$ that

$$\|x\|_\varphi = \varphi\big([x, x]_\varphi\big) = \varphi\big(f\big(\|x\|\big)\big) = \|x\|,$$

which completes the proof. □

Let V be a nontrivial vector space. By the Hahn–Banach theorem, there exist infinitely many distinct norms on V. As a consequence, by Theorem 2.2, there are infinitely many distinct g.s.i.p. $[\cdot,\cdot]_\varphi$ on V associated with a surjective φ on $\mathbb{R}_+$.

We also remark that Theorems 2.1 and 2.2 can be extended in an effortless manner to n-normed vector spaces [13,25]. The extension for n-semi-inner products of type p has been done, for example, in [23].

Author's personal copy

Another natural consideration about a g.s.i.p. $[\cdot,\cdot]_\varphi$ is that $[x,x]_\varphi$ should be a measurement of the norm of $x \in V$ as $[\cdot,\cdot]_\varphi$ is designed to be a substitute for an inner product. First of all, this implies that $[x,x]_\varphi > [y,y]_\varphi$ if and only if $\|x\| > \|y\|$, which can be equivalently stated as that $[x,x]_\varphi$ is a strictly increasing function of $\|x\|$. Secondly, it would be convenient to require $[x,x]_\varphi$ to be continuous with respect to $\|x\|$ in applications. Thirdly, $[x,x]_\varphi$ should tend to infinity as $\|x\|$ does. Therefore, a g.s.i.p. $[\cdot,\cdot]_\varphi$ on a normed vector space V is consistent if all these three requirements are satisfied. Consistency of a g.s.i.p. imposes some particular conditions on the function φ, which are revealed below.

Proposition 2.3. *Let V be a normed vector space. Then $[\cdot,\cdot]_\varphi$ is consistent on V if and only if φ is a gauge function on $\mathbb{R}_+$ and there exists a positive constant λ such that*

$$[x,x]_\varphi = \varphi^{-1}\big(\lambda\|x\|\big). \tag{2.6}$$

Proof. By definition, $[\cdot,\cdot]_\varphi$ is consistent on V if and only if there exists a gauge function ϕ on $\mathbb{R}_+$ such that

$$[x,x]_\varphi = \phi\big(\|x\|\big), \quad x \in V. \tag{2.7}$$

Thus, if φ is a gauge function on $\mathbb{R}_+$ and (2.6) holds then $[\cdot,\cdot]_\varphi$ is consistent on V. Conversely, suppose that $[\cdot,\cdot]_\varphi$ is consistent on V, that is, (2.7) holds for some gauge function ϕ on $\mathbb{R}_+$. By Theorem 2.1, $\|\cdot\|_\varphi$ is a norm on V. Consequently, we have for all $x \in V$ and $t \geqslant 0$ that

$$\varphi\big([tx,tx]_\varphi\big) = \|tx\|_\varphi = t\|x\|_\varphi = t\varphi\big([x,x]_\varphi\big),$$

which by (2.7) has the following equivalent form

$$\varphi\big(\phi\big(t\|x\|\big)\big) = t\varphi\big(\phi\big(\|x\|\big)\big).$$

By choosing $x \in V$ with $\|x\| = 1$ in the above equation, we get for all $t \geqslant 0$ that

$$\varphi\big(\phi(t)\big) = t\varphi\big(\phi(1)\big).$$

Setting $\lambda := \varphi(\phi(1))$ leads to

$$\varphi\big(\phi(t)\big) = \lambda t, \quad t \in \mathbb{R}_+. \tag{2.8}$$

It follows that φ/λ is the inverse function of ϕ on $\mathbb{R}_+$. Therefore, φ is a gauge function on $\mathbb{R}_+$. Finally, we derive from (2.7) and (2.8) that

$$[x,x]_\varphi = \phi\big(\|x\|\big) = \varphi^{-1}\big(\lambda\|x\|\big), \quad x \in V,$$

which completes the proof. □

Consistent and compatible g.s.i.p. on a normed vector space V are desirable. We get by Proposition 2.3 the following corollary.

Corollary 2.4. *A g.s.i.p. $[\cdot,\cdot]_\varphi$ on a normed vector space V is consistent and compatible if and only if φ is a gauge function on $\mathbb{R}_+$ and*

$$[x,x]_\varphi = \varphi^{-1}\big(\|x\|\big), \quad x \in V. \tag{2.9}$$

It is time to present examples of gauge functions φ and the associated ψ defined by (2.3). We see that if $\varphi = \delta_{1/p}$, $p \in (1,+\infty)$, then $\psi = \delta_{1/q}$. In this case, we obtain an s.i.p. of type p. Theorems 2.1 and 2.2 reveal that the index p is not restricted to $(1,+\infty)$. In fact, one may choose any $s \in (0,+\infty)$ and set

$$\varphi := \delta_{1/s}, \qquad \psi = \delta_{1-1/s}, \quad t > 0.$$

There are many other examples. For instance,

$$\varphi(t) := \ln(1+t), \qquad \psi(t) = \frac{t}{\ln(1+t)}, \quad t \in \mathbb{R}_+$$

and

$$\varphi(t) := e^t - 1, \qquad \psi(t) = \frac{t}{e^t - 1}, \quad t \in \mathbb{R}_+.$$

We end this section with a characterization of consistent and compatible g.s.i.p. on a normed vector space V in terms of duality mappings from V to V^*.

Author's personal copy

Theorem 2.5. *Let V be a normed vector space. Then $[\cdot,\cdot]_\varphi : V \times V \to \mathbb{C}$ is a consistent and compatible g.s.i.p. on V if and only if there exists some duality mapping $J : V \to V^*$ such that*

$$[x, y]_\varphi = \frac{J(y)(x)}{\|y\|}, \quad x, y \in V \tag{2.10}$$

and

$$\varphi\big(\big\|J(x)\big\|_*\big) = \|x\|, \quad x \in V. \tag{2.11}$$

Proof. Let $[\cdot,\cdot]_\varphi$ be a consistent and compatible g.s.i.p. on V. By Corollary 2.4, φ is a gauge function on $\mathbb{R}_+$ and (2.9) holds. We introduce a mapping J from V to the set of linear functionals on V by setting

$$J(x)(y) := \|x\|[y, x]_\varphi, \quad x, y \in V. \tag{2.12}$$

Then (2.10) holds. It remains to show that J is a duality mapping satisfying (2.11). Let $x \in V$. Clearly, $J(x)$ indeed is a linear functional on V. Moreover, we observe by (2.4) and (2.9) that

$$\big|J(x)(y)\big| \leqslant \|x\|\varphi\big([y, y]_\varphi\big)\frac{[x, x]_\varphi}{\varphi([x, x]_\varphi)} = \|x\|\|y\|\frac{\varphi^{-1}(\|x\|)}{\|x\|} = \|y\|\varphi^{-1}\big(\|x\|\big). \tag{2.13}$$

The above equation implies that $J(x) \in V^*$ and $\|J(x)\|_* \leqslant \varphi^{-1}(\|x\|)$. Since (2.13) becomes an equality when $x = y$, $\|J(x)\|_* = \varphi^{-1}(\|x\|)$. Furthermore,

$$J(x)(x) = \|x\|[x, x]_\varphi = \|x\|\varphi^{-1}\big(\|x\|\big) = \big\|J(x)\big\|_*\|x\|.$$

We have hence proved that J is a duality mapping from V to V^* that satisfies (2.11).

Conversely, assume that (2.10) and (2.11) hold true for some duality mapping J from V to V^*. By definition, there exists a gauge function f on $\mathbb{R}_+$ that satisfies (1.8). It follows from (2.11) that $\varphi = f^{-1}$ is a gauge function on $\mathbb{R}_+$ as well. Finally, we observe from (2.10) and (1.8) that

$$[x, x]_\varphi = \frac{J(x)(x)}{\|x\|} = \frac{\|J(x)\|_*\|x\|}{\|x\|} = \big\|J(x)\big\|_* = f\big(\|x\|\big) = \varphi^{-1}\big(\|x\|\big),$$

which, by Corollary 2.4, proves that $[\cdot,\cdot]_\varphi$ is a consistent and compatible g.s.i.p. on V. $\square$

3. The Riesz representation of continuous linear functionals

Let V be a fixed normed vector space, φ a gauge function on $\mathbb{R}_+$, and $[\cdot,\cdot]_\varphi$ a consistent and compatible g.s.i.p. on V. The purpose of this section is to investigate the conditions ensuring that every $u \in V^*$ can be represented via some vector $y \in V$ as

$$u(x) = [x, y]_\varphi, \quad x \in V,$$

and those ensure the uniqueness of the representer y. As mentioned in the introduction section, the task is equivalent to studying the surjectivity and injectivity of the mapping J_φ defined by (1.11).

We begin with several simple observations about J_φ. Firstly, it is clear by the property (2.4) of g.s.i.p. that

$$\big\|J_\varphi(x)\big\|_* = \gamma\big(\|x\|\big), \quad x \in V, \tag{3.1}$$

where γ is defined by (1.9). Secondly, by Theorem 2.5, J_φ can be chosen so that

$$J_\varphi(x) = \frac{\gamma(\|x\|)}{\|x\|} J_{\delta_{1/2}}(x), \quad x \in V. \tag{3.2}$$

This choice enables us to make use of some known properties of $J_{\delta_{1/2}}$. The first one of them, first pointed out by Giles [15], states that $J_{\delta_{1/2}}$ may be chosen to satisfy

$$J_{\delta_{1/2}}(\alpha x) = \bar{\alpha} J_{\delta_{1/2}}(x), \quad \alpha \in \mathbb{C},\ x \in V. \tag{3.3}$$

We shall assume in the rest of the paper properties (3.2) and (3.3). The second one of them addresses the uniqueness of the g.s.i.p. $[\cdot,\cdot]_\varphi$ on V. To explain this, we say that V is *Gâteaux differentiable* if for all $x, y \in V$ with $x \neq 0$

$$\lim_{t\in\mathbb{R},\, t\to 0} \frac{\|x + ty\| - \|x\|}{t} \quad \text{exists}.$$

It was shown in [15] that if V is Gâteaux differentiable then for all $x, y \in V$ with $x \neq 0$

Author's personal copy

$$\lim_{t\in\mathbb{R},\, t\to 0} \frac{\|x+ty\|-\|x\|}{t} = \frac{\mathrm{Re}([y,x]_2)}{\|x\|}.$$

This together with (3.2) yields the following result.

Proposition 3.1. *Let g be a continuously differentiable function on $(0,+\infty)$ such that $g'=\gamma$. Suppose that V is Gâteaux differentiable. Then for all $x, y \in V$ with $x \neq 0$*

$$\lim_{t\in\mathbb{R},\, t\to 0} \frac{g(\|x+ty\|)-g(\|x\|)}{t} = \mathrm{Re}\big([y,x]_\varphi\big).$$

As a consequence, the g.s.i.p. $[\cdot,\cdot]_\varphi$ on V is uniquely given by

$$[x,y]_\varphi = \lim_{t\in\mathbb{R},\, t\to 0} \frac{g(\|y+tx\|)-g(\|y\|)}{t} + i \lim_{t\in\mathbb{R},\, t\to 0} \frac{g(\|iy+tx\|)-g(\|y\|)}{t}, \quad x,y\in V\setminus\{0\}.$$

We next make a trivial observation from (3.1) about the boundedness of J_φ.

Proposition 3.2. *The mapping J_φ is bounded from V to V^* if and only if the function φ^{-1}/δ_2 is bounded. Consequently, if $\varphi=\delta_s$ for some $s\in(0,+\infty)$ then J_φ is bounded if and only if $s=\frac{1}{2}$.*

We then turn to the surjectivity of J_φ. The following lemma was established in [12] and extended to s.i.p. of type p in [29].

Lemma 3.3. *The mapping $J_{\delta_{1/2}}$ is surjective onto V^* if and only if V is reflexive.*

Theorem 3.4. *The mapping J_φ is surjective from V to V^* if and only if γ is surjective onto $\mathbb{R}_+$ and V is reflexive.*

Proof. Suppose that γ is surjective onto $\mathbb{R}_+$ and V is reflexive. Let $u\in V^*$. If $u=0$ then $J_\varphi(0)=u$. Assume that $u\neq 0$. Then by Lemma 3.3, there exists some $x\in V\setminus\{0\}$ such that $J_{\delta_{1/2}}(x)=u$. Since γ is surjective, we may choose some $\alpha>0$ such that

$$\gamma\big(\alpha\|x\|\big)=\|x\|.$$

We then compute by the above choice, (3.2), and (3.3) that

$$J_\varphi(\alpha x) = \frac{\gamma(\alpha\|x\|)}{\|\alpha x\|} J_{\delta_{1/2}}(\alpha x) = \frac{\gamma(\alpha\|x\|)}{\alpha\|x\|}\alpha J_{\delta_{1/2}}(x) = \frac{\gamma(\alpha\|x\|)}{\|x\|} J_{\delta_{1/2}}(x) = J_{\delta_{1/2}}(x) = u.$$

Thus, J_φ is surjective.

Conversely, suppose that J_φ is surjective onto V^*. Then we must have

$$\big\{\big\|J_\varphi(x)\big\|_*\colon x\in V\big\} = \mathbb{R}_+,$$

which and Eq. (3.1) immediately imply that γ is surjective onto $\mathbb{R}_+$. The surjectivity of J_φ also implies that for each $u\in V^*\setminus\{0\}$ there exists some $x\in V$ such that

$$u(y) = J_\varphi(x)(y) = [y,x]_\varphi, \quad y\in V.$$

As a consequence,

$$u\big(x/\|x\|\big) = \frac{[x,x]_\varphi}{\|x\|} = \frac{\varphi^{-1}(\|x\|)}{\|x\|} = \big\|J_\varphi(x)\big\|_* = \|u\|_*.$$

Therefore, each linear functional in V^* assumes its norm on the unit sphere of V. By a celebrated characterization of reflexivity due to James [17], V is reflexive. $\square$

For the purpose of studying the injectivity of J_φ, we shall characterize strict convexity of V in terms of the g.s.i.p. $[\cdot,\cdot]_\varphi$. Such characterizations via s.i.p. can be found in [15,34].

Proposition 3.5. *The normed vector space V is strictly convex if and only if whenever*

$$[x,y]_\varphi = \varphi\big([x,x]_\varphi\big)\frac{[y,y]_\varphi}{\varphi([y,y]_\varphi)}, \quad x,y\neq 0 \tag{3.4}$$

then $y=\alpha x$ for some $\alpha>0$.

Author's personal copy

Proof. Suppose that V is strictly convex and (3.4) holds for some $x, y \neq 0$. Then we get that

$$[x+y, y]_\varphi = [x, y]_\varphi + [y, y]_\varphi = \varphi\big([x,x]_\varphi\big)\frac{[y,y]_\varphi}{\varphi([y,y]_\varphi)} + [y,y]_\varphi = \big(\|x\|+\|y\|\big)\frac{[y,y]_\varphi}{\varphi([y,y]_\varphi)}. \tag{3.5}$$

On the other hand, we have by (2.4) that

$$[x+y, y]_\varphi \leqslant \|x+y\|\frac{[y,y]_\varphi}{\varphi([y,y]_\varphi)}. \tag{3.6}$$

By (3.5) and (3.6), $\|x+y\| \geqslant \|x\|+\|y\|$. Thus, we must have $\|x+y\| = \|x\|+\|y\|$, which implies by the strict convexity of V that $y = \alpha x$ for some $\alpha > 0$.

Conversely, suppose that (3.4) implies $y = \alpha x$ for some $\alpha > 0$. Assume that $\|x+y\| = \|x\|+\|y\|$ for some $x, y \neq 0$. We then proceed that

$$\begin{aligned}
\|x+y\|\frac{[x+y,x+y]_\varphi}{\varphi([x+y,x+y]_\varphi)} &= [x+y,x+y]_\varphi = [x,x+y]_\varphi + [y,x+y]_\varphi \\
&= \mathrm{Re}\big([x,x+y]_\varphi\big) + \mathrm{Re}\big([y,x+y]_\varphi\big) \\
&\leqslant \big|[x,x+y]_\varphi\big| + \big|[y,x+y]_\varphi\big| \\
&\leqslant \|x\|\frac{[x+y,x+y]_\varphi}{\varphi([x+y,x+y]_\varphi)} + \|y\|\frac{[x+y,x+y]_\varphi}{\varphi([x+y,x+y]_\varphi)}.
\end{aligned}$$

Since $\|x+y\| = \|x\|+\|y\|$, all the inequalities must become equalities. Therefore, we obtain that

$$\mathrm{Re}\big([x,x+y]_\varphi\big) = \big|[x,x+y]_\varphi\big| = \|x\|\frac{[x+y,x+y]_\varphi}{\varphi([x+y,x+y]_\varphi)},$$

which implies that

$$[x,x+y]_\varphi = \|x\|\frac{[x+y,x+y]_\varphi}{\varphi([x+y,x+y]_\varphi)}.$$

By the assumption, there exists some $\alpha > 0$ such that $x+y = \alpha x$. Substituting this into $\|x+y\| = \|x\|+\|y\|$ implies that $\alpha > 1$. We hence get that $y = (\alpha-1)x$ and $\alpha - 1 > 0$. Therefore, V is strictly convex. The proof is complete. □

We present another characterization below, which in the case when $\varphi = \delta_{1/2}$ was obtained in [34].

Proposition 3.6. *The space V is strictly convex if and only if whenever $[y,x]_\varphi = 0$ and $y \neq 0$ then $\|x+y\| > \|x\|$.*

Proof. Suppose that V is strictly convex and $[y,x]_\varphi = 0$ with $y \neq 0$. Then $\|x+y\| > \|x\|$ if $x = 0$. Assume that $x \neq 0$. Then $y \neq \alpha x$ for any $\alpha \in \mathbb{R}$. Thus, $x+y \neq \alpha x$ for any $\alpha > 0$. By Proposition 3.5, we get that

$$[x,x]_\varphi = [x+y,x]_\varphi < \|x+y\|\frac{[x,x]_\varphi}{\varphi([x,x]_\varphi)} = \|x+y\|\frac{[x,x]_\varphi}{\|x\|},$$

which implies that $\|x\| < \|x+y\|$.

On the other hand, suppose that whenever $[\tilde y,\tilde x]_\varphi = 0$ with $\tilde y \neq 0$ then $\|\tilde x+\tilde y\| > \|\tilde x\|$. We shall make use of Proposition 3.5 to prove that V is strictly convex. Assume that there holds (3.4) for some $x, y \neq 0$. We want to prove that $y = \alpha x$ for some $\alpha > 0$. Assume to the contrary that this is not true. Then

$$\frac{\|y\|}{\|x\|}x - y \neq 0.$$

We compute by (3.4) that

$$\left[\frac{\|y\|}{\|x\|}x - y, y\right]_\varphi = \frac{\|y\|}{\|x\|}\|x\|\frac{[y,y]_\varphi}{\varphi([y,y]_\varphi)} - \|y\|\frac{[y,y]_\varphi}{\varphi([y,y]_\varphi)} = 0.$$

By the assumption, we reach the contradiction that

$$\|y\| = \left\|\frac{\|y\|}{\|x\|}x - y + y\right\| > \|y\|.$$

The proof is complete. □

With the above characterizations, we shall provide a sufficient condition for J_φ to be injective.

Author's personal copy

Theorem 3.7. *If V is strictly convex and γ is injective then J_φ is injective.*

Proof. Suppose that V is strictly convex and γ is injective. Assume that $J_\varphi(x)=J_\varphi(y)$ for some $x,y\in V$. Note by (3.1) that $J_\varphi(z)=0$ if and only if $z=0$. Thus, we may assume that $x,y\neq 0$. The equality $J_\varphi(x)=J_\varphi(y)$ implies that

$$[z,x]_\varphi=[z,y]_\varphi,\quad z\in V.$$

Choosing $z=x$ in the above equation yields $[x,x]_\varphi=[x,y]_\varphi$. We hence obtain that

$$[x,x]_\varphi=[x,y]_\varphi\leqslant\varphi\big([x,x]_\varphi\big)\frac{[y,y]_\varphi}{\varphi([y,y]_\varphi)}=\|x\|\frac{[y,y]_\varphi}{\|y\|},$$

which implies that

$$\frac{[x,x]_\varphi}{\|x\|}\leqslant\frac{[y,y]_\varphi}{\|y\|}.$$

Similarly, one can show that

$$\frac{[y,y]_\varphi}{\|y\|}\leqslant\frac{[x,x]_\varphi}{\|x\|}.$$

Thus, we obtain that

$$\frac{[y,y]_\varphi}{\|y\|}=\frac{[x,x]_\varphi}{\|x\|}.$$

We deduce from the above relation that

$$\varphi\big([x,x]_\varphi\big)\frac{[y,y]_\varphi}{\varphi([y,y]_\varphi)}=\|x\|\frac{[y,y]_\varphi}{\|y\|}=\|x\|\frac{[x,x]_\varphi}{\|x\|}=[x,x]_\varphi=[x,y]_\varphi.$$

Thus, by Proposition 3.5, $x=\alpha y$ for some $\alpha>0$. We substitute $x=\alpha y$ into $[x,x]_\varphi=[x,y]_\varphi$ to get that

$$\varphi^{-1}\big(\alpha\|y\|\big)=\alpha\varphi^{-1}\big(\|y\|\big),$$

which has the equivalent form that

$$\gamma\big(\alpha\|y\|\big)=\gamma\big(\|y\|\big).$$

The injectivity of γ forces $\alpha=1$. We have hence shown that $x=y$. $\square$

We remark that γ being injective is necessary for J_φ to be injective. To see this, assume that γ is not injective. As a result, there exist $s,t>0$ with $s\neq 1$ such that $\gamma(st)=\gamma(t)$. Take any $x\in V$ with $\|x\|=t$. Then it follows immediately from (3.2) and (3.3) that $J_\varphi(sx)=J_\varphi(x)$ while $sx\neq x$ as $s\neq 1$.

Combining the above remark with Theorems 3.4 and 3.7, we obtain the following result.

Theorem 3.8. *Let V be a strictly convex Banach space and $[\cdot,\cdot]_\varphi$ a consistent and compatible g.s.i.p. on V. Then J_φ is bijective from V to V^* if and only if V is reflexive and the function γ is bijective on $\mathbb{R}_+$.*

In the last part of this section, we aim at inducing a g.s.i.p. on V^* by the existing one on V. It turns out that this can be done when the mapping J_φ is bijective.

Theorem 3.9. *Suppose that V is a reflexive and strictly convex Banach space and γ is bijective on $\mathbb{R}_+$. Then the function $[\cdot,\cdot]_\psi: V^*\times V^*\to\mathbb{C}$ defined by (1.12) is a compatible g.s.i.p. on V^*. It is consistent if and only if γ is a gauge function on $\mathbb{R}_+$.*

Proof. Under the hypotheses, J_φ is bijective by Theorem 3.8. It follows from (1.11) and the definition (1.12) that

$$[u,v]_\psi=u\big((J_\varphi)^{-1}(v)\big),\quad u,v\in V^*,$$

which implies that $[\cdot,\cdot]_\psi$ is linear with respect to its first variable. It is also obvious that $[u,u]_\psi>0$ for all $u\in V^*\setminus\{0\}$. Recalling the definition (2.3) of ψ, we observe for $u=J_\varphi(x)$ and $v=J_\varphi(y)$, $x,y\in V$ that

$$\big|[u,v]_\psi\big|=\big|[y,x]_\varphi\big|\leqslant\|y\|\frac{\varphi^{-1}(\|x\|)}{\|x\|}$$

and

Author's personal copy

$$\psi\big([u,u]_\psi\big)\frac{[v,v]_\psi}{\psi([v,v]_\psi)}=\psi\big([x,x]_\varphi\big)\frac{[y,y]_\varphi}{\psi([y,y]_\varphi)}=\psi\big(\varphi^{-1}\big(\|x\|\big)\big)\frac{\varphi^{-1}(\|y\|)}{\psi(\varphi^{-1}(\|y\|))}=\frac{\varphi^{-1}(\|x\|)}{\|x\|}\|y\|.$$

Therefore, $[\cdot,\cdot]_\psi$ does satisfy

$$\big|[u,v]_\psi\big|\leqslant\psi\big([u,u]_\psi\big)\frac{[v,v]_\psi}{\psi([v,v]_\psi)},\quad u,v\in V^*.$$

We reach the conclusion that $[\cdot,\cdot]_\psi$ defined by (1.12) is a g.s.i.p. on V^*. We also obtain by (3.1) for $u=J_\varphi(x)$, $x\in V$ that

$$\psi\big([u,u]_\psi\big)=\psi\big([x,x]_\varphi\big)=\psi\big(\varphi^{-1}\big(\|x\|\big)\big)=\frac{\varphi^{-1}(\|x\|)}{\|x\|}=\big\|J_\varphi(x)\big\|_*=\|u\|_*. \tag{3.7}$$

Thus, $[\cdot,\cdot]_\psi$ is compatible on V^*. If γ is a gauge function then so is ψ as $\psi=\gamma\circ\varphi$. By (3.7),

$$[u,u]_\psi=\psi^{-1}\big(\|u\|_*\big).$$

By Corollary 2.4, $[\cdot,\cdot]_\psi$ is consistent on V^*. Conversely, if $[\cdot,\cdot]_\psi$ is consistent on V^* then by Proposition 2.3, ψ must be a gauge function on $\mathbb{R}_+$. Consequently, so is γ as $\gamma=\psi\circ\varphi^{-1}$. $\square$

Assuming the hypotheses of Theorem 3.9, we note some dual relationships between V and V^*. Specifically, for all $x,y\in V$ and $u,v\in V^*$,

$$[x,x]_\varphi=\big[J_\varphi(x),J_\varphi(x)\big]_\psi,$$
$$\big\|J_\varphi(x)\big\|_*=\gamma\big(\|x\|\big),\qquad\big\|(J_\varphi)^{-1}(u)\big\|=\gamma^{-1}\big(\|u\|_*\big),$$
$$\big|[x,y]_\varphi\big|\leqslant\varphi\big([x,x]_\varphi\big)\psi\big([y,y]_\varphi\big),\qquad\big|[u,v]_\psi\big|\leqslant\psi\big([u,u]_\psi\big)\varphi\big([v,v]_\psi\big),$$
$$[x,x]_\varphi=\varphi^{-1}\big(\|x\|\big),\qquad[u,u]_\psi=\psi^{-1}\big(\|u\|_*\big).$$

We shall call a g.s.i.p. $[\cdot,\cdot]_\varphi$ on V *natural* if it is consistent and compatible, J_φ is bijective, and the g.s.i.p. on V^* induced by (1.12) is consistent and compatible as well. The following characterization of natural g.s.i.p. is a direct consequence of Corollary 2.4, Theorems 3.8 and 3.9.

Theorem 3.10. *Let V be a reflexive and strictly convex Banach space and $[\cdot,\cdot]_\varphi$ a g.s.i.p. on V. Then $[\cdot,\cdot]_\varphi$ is natural if and only if φ, γ are gauge functions on $\mathbb{R}_+$ and* (2.9) *holds true. Alternatively, $[\cdot,\cdot]_\varphi$ is a natural g.s.i.p. on V if and only if there exists a duality mapping $J:V\to V^*$ such that*

$$[x,y]_\varphi=J(y)(x)\quad\text{and}\quad\varphi\big(\big\|J(x)\big\|_*\|x\|\big)=\|x\|,\quad x,y\in V.$$

We close this section with a table of successively imposed requirements on a g.s.i.p. $[\cdot,\cdot]_\varphi$ on a reflexive and strictly convex Banach space V and the corresponding restrictions on the function φ:

$[\cdot,\cdot]_\varphi$	Necessary and sufficient conditions
exists and is compatible	φ being surjective onto $\mathbb{R}_+$
consistent and compatible	φ being a gauge function on $\mathbb{R}_+$ and (2.9)
natural	$\varphi,\gamma=\varphi^{-1}/t$ being gauge functions on $\mathbb{R}_+$ and (2.9)

4. Applications to regularized learning in Banach spaces

Let V be a Banach space of functions. We shall apply g.s.i.p. to the regularized learning of a target function h in V from its finite or infinite samples. Let us briefly review our strategy which has been introduced at the end of the first section. Suppose that $(\Omega,\mathcal{F},\mu)$ is a probability measure space and $\{\nu_\omega\colon\ \omega\in\Omega\}$ is a bounded subset of V^* of sampling functionals such that for each $f\in V$, $\omega\to\nu_\omega(f)\in L^2_\mu(\Omega)$. Assume that the sampled data ρ of h given by (1.13) is obtained. Introduce an error functional $\mathcal{E}$ from V to $\mathbb{R}_+$ defined by

$$\mathcal{E}(f):=\int_\Omega\big|\nu_\omega(f)-\rho(\omega)\big|^2\,d\mu(\omega)+\lambda g\big(\|f\|\big), \tag{4.1}$$

where λ is a positive regularization parameter and g, serving as a regularizer, is a nondecreasing function from $\mathbb{R}_+$ to $\mathbb{R}_+$ with $g(0)=0$. We consider the following minimization problem:

$$\inf\big\{\mathcal{E}(f)\colon\ f\in V\big\}. \tag{4.2}$$

An approximation of the target function will be taken as the minimizer of (4.2) provided that it uniquely exists.

Author's personal copy

Before proceeding, let us make connections with the familiar regularized learning from finite samples by reproducing kernel methods in machine learning. Assume that V is an RKHS on an input space $\mathcal{Z}$ with the reproducing kernel K. Let $m \in \mathbb{N}$. If

$$\Omega := \{1, 2, \ldots, m\}, \qquad \mathcal{F} := 2^{\Omega}, \qquad \mu\bigl(\{j\}\bigr) := \frac{1}{m}, \quad 1 \leqslant j \leqslant m, \qquad g = \delta_2 \tag{4.3}$$

and there exist $y_j \in \mathbb{C}$ and $x_j \in \mathcal{Z}$ such that

$$\rho(j) := y_j, \qquad \nu_j(f) := f(x_j), \quad f \in V,\ 1 \leqslant j \leqslant m \tag{4.4}$$

then we see that

$$\mathcal{E}(f) = \frac{1}{m}\sum_{j=1}^{m}\bigl|f(x_j) - y_j\bigr|^2 + \lambda\|f\|^2. \tag{4.5}$$

For this choice of error functionals, (4.2) is known as the regularization network and has been extensively studied in the literature of machine learning (see, for example, [8,11,19,30–32,35,37–39]). It has been established in this case that (4.2) has a unique minimizer and a function $f_0 \in V$ is the minimizer if and only if

$$\lambda f_0 = \frac{1}{m}\sum_{j=1}^{m}\bigl(y_j - f_0(x_j)\bigr)K(x_j, \cdot). \tag{4.6}$$

We extended this result to the case when V is an RKBS in [40]. Similarly, we obtained that if V is an RKBS on $\mathcal{Z}$ with the reproducing kernel K and the error functional has the form (4.5) then (4.2) has a unique minimizer and a function $f_0 \in V$ is the minimizer if and only if

$$\lambda J_{\delta_{1/2}}(f_0) = \frac{1}{m}\sum_{j=1}^{m}\overline{y_j - f_0(x_j)}K(\cdot, x_j). \tag{4.7}$$

The purpose of this section is to provide a characterization equation analogous to (4.6) and (4.7) for the minimizer of (4.2) by making use of g.s.i.p. We first discuss the existence and uniqueness of the minimizer.

Proposition 4.1. *If V is a reflexive Banach space and g is continuous, nondecreasing with*

$$\lim_{t\to\infty} g(t) = +\infty \tag{4.8}$$

then there exists a minimizer for (4.2).

Proof. Let β be the infimum (4.2). Then $\beta \leqslant \mathcal{E}(0)$. Since g is nondecreasing with (4.8), there exists a positive number t_0 such that for all $f \in V$ with $\|f\| > t_0$

$$\mathcal{E}(f) \geqslant \lambda g\bigl(\|f\|\bigr) \geqslant \lambda g(t_0) > \beta.$$

Thus, with $S := \{f \in V\colon\ \|f\| \leqslant t_0\}$,

$$\beta = \inf\bigl\{\mathcal{E}(f)\colon\ f \in S\bigr\}.$$

By the above equality, there exists a sequence $f_m \in S$ such that

$$\beta \leqslant \mathcal{E}(f_m) \leqslant \beta + \frac{1}{m}, \quad m \in \mathbb{N}. \tag{4.9}$$

Since V is a reflexive Banach space, S is weakly compact, that is, we may assume that there exists a function $f_0 \in S$ such that

$$\lim_{m\to\infty} u(f_m) = u(f_0) \quad \text{for all } u \in V^*. \tag{4.10}$$

Choosing $u = \nu_\omega$, $\omega \in \Omega$ in the above equation yields that

$$\lim_{m\to\infty} \nu_\omega(f_m) = \nu_\omega(f_0), \quad \omega \in \Omega.$$

By the assumption that $\{\nu_\omega\colon\ \omega \in \Omega\}$ is bounded in V^*,

$$\bigl|\nu_\omega(f_m)\bigr| \leqslant \|f_m\|\|\nu_\omega\|_* \leqslant t_0 \sup\bigl\{\|\nu_\omega\|_*\colon\ \omega \in \Omega\bigr\} < +\infty.$$

Author's personal copy

Thus, we apply the Lebesgue dominated convergence theorem to obtain that

$$\lim_{m\to\infty}\int_\Omega \left|\nu_\omega(f_m)-\rho(\omega)\right|^2 d\mu(\omega)=\int_\Omega \left|\nu_\omega(f_0)-\rho(\omega)\right|^2 d\mu(\omega). \tag{4.11}$$

If $f_0=0$ then it is obvious that $\|f_0\|\leqslant\|f_m\|$, $m\in\mathbb{N}$. If $f_0\neq 0$ then we substitute $u=J_{\delta_{1/2}}(f_0)$ in (4.10) to get for each $\sigma>0$ some $m_0\in\mathbb{N}$ such that for $m>m_0$

$$[f_0,f_0]\leqslant\left|[f_m,f_0]\right|+\sigma\|f_0\|,$$

which implies by the Cauchy–Schwartz inequality of s.i.p. that

$$\|f_0\|^2\leqslant\|f_m\|\|f_0\|+\sigma\|f_0\|.$$

Thus, we have in both cases that for sufficiently large m

$$\|f_0\|\leqslant\|f_m\|+\sigma.$$

Since g is continuous, it is uniformly continuous on $[0,t_0]$. Consequently, for every $\varepsilon>0$, we may choose σ small enough so that for sufficiently large m

$$\lambda g\big(\|f_0\|\big)\leqslant\lambda g\big(\|f_m\|\big)+\varepsilon. \tag{4.12}$$

Combining (4.9), (4.11), and (4.12) proves that

$$\mathcal{E}(f_0)=\beta,$$

which shows that f_0 is a minimizer for (4.2). □

Corollary 4.2. *If V is a reflexive and strictly convex Banach space and g is continuous, convex, and strictly increasing with* (4.8) *then* (4.2) *has a unique minimizer.*

Proof. It suffices to point out that $\mathcal{E}$ is strictly convex on V under the hypotheses. □

For further study of the minimization problem (4.2), we shall assume that g is convex and continuously differentiable with $g'(t)>0$ for $t>0$. Let φ be defined by (1.15). Then φ is a gauge function on $\mathbb{R}_+$. We shall assume that the conditions of Theorem 3.8 are satisfied so that J_φ is bijective from V to V^*. Specifically, we require that V be a reflexive and strictly convex Banach space, and $g'=\gamma$ be bijective on $\mathbb{R}_+$. Our last requirement is that V be Gâteaux differentiable. By Proposition 3.1, there holds for all $f,\tilde f\in V$ with $f\neq 0$ that

$$\lim_{t\in\mathbb{R},\,t\to 0}\frac{g(\|f+t\tilde f\|)-g(\|f\|)}{t}=\mathrm{Re}\big([\tilde f,f]_\varphi\big). \tag{4.13}$$

The next result specified to the case when V is an RKHS and $\mathcal{E}$ is given by (4.5) is known as the representer theorem [7,19,30] in machine learning.

Theorem 4.3. *Let f_0 be the unique minimizer of* (4.2)*. Then $J_\varphi(f_0)\in\overline{\mathrm{span}}\{\nu_\omega:\ \omega\in\Omega\}$.*

Proof. Assume that the result is not true. Then by a geometric consequence of the Hahn–Banach theorem (see, for example, [6, page 111]), there exist a continuous linear functional T on V^* and real number α such that

$$\mathrm{Re}\big(T\big(J_\varphi(f_0)\big)\big)<\alpha\leqslant\mathrm{Re}\big(T(u)\big)\quad\text{for all } u\in\overline{\mathrm{span}}\{\nu_\omega:\ \omega\in\Omega\}. \tag{4.14}$$

Since $\overline{\mathrm{span}}\{\nu_\omega:\ \omega\in\Omega\}$ is a linear space, we must have $T(u)=0$ for all $u\in\overline{\mathrm{span}}\{\nu_\omega:\ \omega\in\Omega\}$. As a consequence, $\alpha\leqslant 0$. By the reflexivity of V, there exists an $f\in V$ such that

$$T(u)=u(f),\quad u\in V^*.$$

We get that for all $t\in\mathbb{R}$

$$\nu_\omega(f_0+tf)=\nu_\omega(f_0)+t\nu_\omega(f)=\nu_\omega(f_0)+tT(\nu_\omega)=\nu_\omega(f_0),$$

which implies that

$$\int_\Omega\left|\nu_\omega(f_0+tf)-\rho(\omega)\right|^2 d\mu(\omega)=\int_\Omega\left|\nu_\omega(f_0)-\rho(\omega)\right|^2 d\mu(\omega). \tag{4.15}$$

Author's personal copy

By Eqs. (4.13) and (4.14),

$$\lim_{t\in\mathbb{R},\, t\to 0}\frac{g(\|f_0+tf\|)-g(\|f_0\|)}{t}=\mathrm{Re}\big([f,f_0]_\varphi\big)=\mathrm{Re}\big(T\big(J_\varphi(f_0)\big)\big)<0.$$

Therefore, by choosing $t\in\mathbb{R}_+$ close enough to 0, we obtain that $g(\|f_0+tg\|)<g(\|f_0\|)$, which together with (4.15) implies that

$$\mathcal{E}(f_0+tf)<\mathcal{E}(f_0),$$

contradicting that f_0 is the minimizer of (4.2). The proof is complete. □

The above theorem provides a representation of the minimizer f_0 in the dual space V^*. In the case when the set Ω consists of finite elements, it states that there exist constants $c_\omega\in\mathbb{C}$, $\omega\in\Omega$ such that

$$J_\varphi(f_0)=\sum_{\omega\in\Omega}c_\omega\nu_\omega.$$

Applying the inverse operator of J_φ to both sides of the above equations yields that

$$f_0=J_\varphi^{-1}\bigg(\sum_{\omega\in\Omega}c_\omega\nu_\omega\bigg),$$

which is a representation of f_0 in the space V.

We are now in a position to establish a characterization equation for the minimizer of (4.2). Recall the linear functional $\mathcal{T}_{f,\Omega}$, $f\in V$ defined by (1.16).

Theorem 4.4. *A function $f_0\neq 0$ is the minimizer of* (4.2) *if and only if*

$$\lambda J_\varphi(f_0)=\mathcal{T}_{f_0,\Omega}. \tag{4.16}$$

The zero function $f_0=0$ is the minimizer of (4.2) *if and only if*

$$\lambda g'(0)\geqslant\|\mathcal{T}_{0,\Omega}\|_*. \tag{4.17}$$

Proof. Suppose that $f_0\neq 0$ is the minimizer of (4.2). Then we have for each $f\in V$ and $t\in\mathbb{R}$ that

$$\mathcal{E}(f_0+tf)\geqslant\mathcal{E}(f_0).$$

Thus, the function

$$\tau_1(t):=\mathcal{E}(f_0+tf),\quad t\in\mathbb{R}$$

attains its minimum at $t=0$. We hence have $\tau_1'(0)=0$, which by Eq. (4.13) has the form

$$2\,\mathrm{Re}\bigg(\int_\Omega\overline{\nu_\omega(f_0)-\rho(\omega)}\,\nu_\omega(f)\,d\mu(\omega)\bigg)+\lambda\,\mathrm{Re}\big([f,f_0]_\varphi\big)=0. \tag{4.18}$$

Since the above equation holds true for all $f\in V$, we obtain that

$$-\mathcal{T}_{f_0,\Omega}(f)+\lambda\big(J_\varphi(f_0)\big)(f)=0,\quad f\in V$$

which is (4.16). Conversely, suppose that (4.16) is true. As a result, there holds (4.18) for all $f\in V$. We get by this fact that for an arbitrary $f\in V$

$$\begin{aligned}\mathcal{E}(f_0+f)-\mathcal{E}(f_0)&=2\,\mathrm{Re}\bigg(\int_\Omega\overline{\nu_\omega(f_0)-\rho(\omega)}\,\nu_\omega(f)\,d\mu(\omega)\bigg)+\int_\Omega\big|\nu_\omega(f)\big|^2\,d\mu(\omega)+\lambda g\big(\|f_0+f\|\big)-\lambda g\big(\|f_0\|\big)\\&=\int_\Omega\big|\nu_\omega(f)\big|^2\,d\mu(\omega)+\lambda\big(g\big(\|f_0+f\|\big)-g\big(\|f_0\|\big)-\mathrm{Re}\big([f,f_0]_\varphi\big)\big).\end{aligned}$$

Therefore, to show that $\mathcal{E}(f_0+f)\geqslant\mathcal{E}(f_0)$, it suffices to show that

$$g\big(\|f_0+f\|\big)-g\big(\|f_0\|\big)-\mathrm{Re}\big([f,f_0]_\varphi\big)\geqslant 0. \tag{4.19}$$

Author's personal copy

For this purpose, we set

$$\tau_2(t) := g\big(\|f_0 + tf\|\big), \quad t \in \mathbb{R}$$

and observe that τ_2 is convex and continuously differentiable at a neighborhood of 0. By (4.13), Eq. (4.19) can be rewritten as

$$\tau_2(1) - \tau_2(0) - \tau_2'(0) \geqslant 0,$$

which is clearly true by the convexity of φ_2. We conclude that the first claim of the theorem is true.

Let us deal with the second one. Suppose that $f_0 = 0$ is the minimizer of (4.2). Then we have for all $f \in V$ and $t \in \mathbb{R}$ that

$$\mathcal{E}(tf) \geqslant \mathcal{E}(0).$$

It follows from the above equation that

$$\lim_{t \to 0^+} \frac{\mathcal{E}(tf) - \mathcal{E}(0)}{t} \geqslant 0.$$

Direct computations show that

$$\lim_{t \to 0^+} \frac{\mathcal{E}(tf) - \mathcal{E}(0)}{t} = \mathrm{Re}\big(\mathcal{T}_{0,\Omega}(f)\big) + \lambda g'(0)\|f\|.$$

Thus, we get for all $f \in V$ that

$$\big|\mathrm{Re}\big(\mathcal{T}_{0,\Omega}(f)\big)\big| \leqslant \lambda g'(0)\|f\|,$$

which immediately implies (4.17). On the other hand, suppose that (4.17) holds true. We verify for an arbitrary $f \in V \setminus \{0\}$ that

$$\mathcal{E}(f) - \mathcal{E}(0) = \int_\Omega \big|\nu_\omega(f)\big|^2 \, d\mu(\omega) - \mathrm{Re}\big(\mathcal{T}_{0,\Omega}(f)\big) + \lambda g\big(\|f\|\big).$$

We use (4.17) to get that

$$\mathcal{E}(f) - \mathcal{E}(0) \geqslant \lambda g\big(\|f\|\big) - \big|\mathcal{T}_{0,\Omega}(f)\big| \geqslant \lambda g\big(\|f\|\big) - \|\mathcal{T}_{0,\Omega}\|_*\|f\| \geqslant \lambda\big(g\big(\|f\|\big) - g'(0)\|f\|\big). \tag{4.20}$$

Since $\tau_3(t) := g(t\|f\|)$, $t \in \mathbb{R}_+$, is convex and continuously differentiable,

$$\tau_3(1) - \tau_3(0) - \tau_3'(0) = g\big(\|f\|\big) - g'(0)\|f\| \geqslant 0. \tag{4.21}$$

By (4.20) and (4.21), $\mathcal{E}(f) \geqslant \mathcal{E}(0)$. Thus, $f_0 = 0$ is the minimizer. The proof is complete. □

We end the paper with the remark that when the error functional is of the form (4.5) and V is an RKHS or RKBS, the characterization equations established in Theorem 4.4 become Eq. (4.6) or (4.7), respectively.

Acknowledgment

The authors express their appreciation for some initial discussions with Dr. Yuesheng Xu.

References

[1] A. Argyriou, C.A. Micchelli, M. Pontil, When is there a representer theorem? Vector versus matrix regularizers, J. Mach. Learn. Res. 10 (2009) 2507–2529.
[2] N. Aronszajn, Theory of reproducing kernels, Trans. Amer. Math. Soc. 68 (1950) 337–404.
[3] K. Bennett, E. Bredensteiner, Duality and geometry in SVM classifier, in: P. Langley (Ed.), Proceeding of the Seventeenth International Conference on Machine Learning, Morgan Kaufmann, San Francisco, 2000, pp. 57–64.
[4] S. Canu, X. Mary, A. Rakotomamonjy, Functional learning through kernel, in: J. Suykens, G. Horvath, S. Basu, C. Micchelli, J. Vandewalle (Eds.), Advances in Learning Theory: Methods, Models and Applications, in: NATO Sci. Ser. III Comput. Syst. Sci., vol. 190, IOS Press, Amsterdam, 2003, pp. 89–110.
[5] I. Ciorănescu, Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems, Kluwer Academic Publishers Group, Dordrecht, 1990.
[6] J.B. Conway, A Course in Functional Analysis, 2nd edition, Springer-Verlag, New York, 1990.
[7] D. Cox, F. O'Sullivan, Asymptotic analysis of penalized likelihood and related estimators, Ann. Statist. 18 (1990) 1676–1695.
[8] F. Cucker, S. Smale, On the mathematical foundations of learning, Bull. Amer. Math. Soc. 39 (2002) 1–49.
[9] R. Der, D. Lee, Large-margin classification in Banach spaces, in: JMLR Workshop and Conference Proceedings 2: AISTATS, 2007 pp. 91–98.
[10] S. Dragomir, Semi-Inner Products and Applications, Nova Science Publishers, Hauppauge, New York, 2004.
[11] T. Evgeniou, M. Pontil, T. Poggio, Regularization networks and support vector machines, Adv. Comput. Math. 13 (2000) 1–50.
[12] G.D. Faulkner, Representation of linear functionals in a Banach space, Rocky Mountain J. Math. 7 (1977) 789–792.

Author's personal copy

[13] S. Gähler, Lineare 2-normierte Räume, Math. Nachr. 28 (1964) 1–43.
[14] C. Gentile, A new approximate maximal margin classification algorithm, J. Mach. Learn. Res. 2 (2001) 213–242.
[15] J.R. Giles, Classes of semi-inner-product spaces, Trans. Amer. Math. Soc. 129 (1967) 436–446.
[16] M. Hein, O. Bousquet, B. Schölkopf, Maximal margin classification for metric spaces, J. Comput. System Sci. 71 (2005) 333–359.
[17] R.C. James, A characterization of reflexivity, Studia Math. 23 (1964) 205–216.
[18] D. Kimber, P.M. Long, On-line learning of smooth functions of a single variable, Theoret. Comput. Sci. 148 (1995) 141–156.
[19] G. Kimeldorf, G. Wahba, Some results on Tchebycheffian spline functions, J. Math. Anal. Appl. 33 (1971) 82–95.
[20] D.O. Koehler, A note on some operator theory in certain semi-inner-product spaces, Proc. Amer. Math. Soc. 30 (1971) 363–366.
[21] G. Lumer, Semi-inner-product spaces, Trans. Amer. Math. Soc. 100 (1961) 29–43.
[22] G. Lumer, On the isometries of reflexive Orlicz spaces, Ann. Inst. Fourier (Grenoble) 13 (1963) 99–109.
[23] R. Malcheski, An n-semi-inner product of characteristic p, Mat. Bilten 25 (2001) 53–66.
[24] C.A. Micchelli, M. Pontil, A function representation for learning in Banach spaces, in: Learning Theory, in: Lecture Notes in Comput. Sci., vol. 3120, Springer-Verlag, Berlin, 2004, pp. 255–269.
[25] A. Misiak, n-Inner product spaces, Math. Nachr. 140 (1989) 299–319.
[26] B. Nath, On a generalization of semi-inner product spaces, Math. J. Okayama Univ. 15 (1971/1972) 1–6.
[27] E. Pap, Functional analysis with K-convergence, in: Convergence Structures 1984, Bechyně, 1984, in: Math. Res., vol. 24, Akademie-Verlag, Berlin, 1985, pp. 245–250.
[28] E. Pap, R. Pavlović, Adjoint theorem on semi-inner product spaces of type (p), Zb. Rad. Prirod.-Mat. Fak. Ser. Mat. 25 (1995) 39–46.
[29] C. Puttamadaiah, H. Gowda, On generalised adjoint abelian operators on Banach spaces, Indian J. Pure Appl. Math. 17 (1986) 919–924.
[30] B. Schölkopf, R. Herbrich, A.J. Smola, A generalized representer theorem, in: Proceeding of the Fourteenth Annual Conference on Computational Learning Theory and the Fifth European Conference on Computational Learning Theory, Springer-Verlag, London, UK, 2001, pp. 416–426.
[31] B. Schölkopf, A.J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, Cambridge, 2002.
[32] J. Shawe-Taylor, N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge University Press, Cambridge, 2004.
[33] A.N. Tikhonov, V.Y. Arsenin, Solutions of Ill-posed Problems, translation editor F. John, V.H. Winston & Sons, New York, 1977 (distributed by Wiley).
[34] E. Torrance, Strictly convex spaces via semi-inner-product space orthogonality, Proc. Amer. Math. Soc. 26 (1970) 108–110.
[35] V.N. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
[36] U. von Luxburg, O. Bousquet, Distance-based classification with Lipschitz functions, J. Mach. Learn. Res. 5 (2004) 669–695.
[37] Y. Xu, H. Zhang, Refinable kernels, J. Mach. Learn. Res. 8 (2007) 2083–2120.
[38] Y. Xu, H. Zhang, Refinement of reproducing kernels, J. Mach. Learn. Res. 10 (2009) 107–140.
[39] H. Zhang, Orthogonality from disjoint support in reproducing kernel Hilbert spaces, J. Math. Anal. Appl. 349 (2009) 201–210.
[40] H. Zhang, Y. Xu, J. Zhang, Reproducing kernel Banach spaces for machine learning, J. Mach. Learn. Res. 10 (2009) 2741–2775.
[41] T. Zhang, On the dual formulation of regularized linear systems with convex risks, Machine Learning 46 (2002) 91–129.
[42] D. Zhou, B. Xiao, H. Zhou, R. Dai, Global geometry of SVM classifiers, Technical Report 30-5-02, Institute of Automation, Chinese, Academy of Sciences, 2002.