Journal of Machine Learning Research 25 (2024) 1-45 Submitted 5/23; Revised 12/23; Published 2/24

Sparse Representer Theorems for Learning in Reproducing Kernel Banach Spaces

Rui Wang rwang11@jlu.edu.cn
School of Mathematics
Jilin University
Changchun, 130012, P. R. China

Yuesheng Xu y1xu@odu.edu
Mingsong Yan myan007@odu.edu
Department of Mathematics and Statistics
Old Dominion University
Norfolk, VA 23529, USA

Editor: Lorenzo Rosasco

Abstract

Sparsity of a learning solution is a desirable feature in machine learning. Certain reproducing kernel Banach spaces (RKBSs) are appropriate hypothesis spaces for sparse learning methods. The goal of this paper is to understand what kind of RKBSs can promote sparsity for learning solutions. We consider two typical learning models in an RKBS: the minimum norm interpolation (MNI) problem and the regularization problem. We first establish an explicit representer theorem for solutions of these problems, which represents the extreme points of the solution set by a linear combination of the extreme points of the subdifferential set, of the norm function, which is *data-dependent*. We then propose sufficient conditions on the RKBS that can transform the explicit representation of the solutions to a sparse kernel representation having fewer terms than the number of the observed data. Under the proposed sufficient conditions, we investigate the role of the regularization parameter on sparsity of the regularized solutions. We further show that two specific RKBSs, the sequence space $\ell_1(\mathbb{N})$ and the measure space, can have sparse representer theorems for both MNI and regularization models.

Keywords: sparse representer theorem, reproducing kernel Banach space, minimum norm interpolation, regularization, sparse learning

1. Introduction

The goal of this paper is to study a class of RKBSs that can promote sparsity for learning solutions in the spaces. In order to alleviate the computational burden brought by big data, developing sparse learning methods is the future of machine learning (Hoefler et al. (2021)). Reproducing kernel Banach spaces (RKBSs), spaces of functions on which the point-evaluation functionals are continuous, were introduced in Zhang et al. (2009) as potential appropriate hypothesis spaces for sparse learning methods. Some of these spaces have sparsity promoting norms which lead to sparse representations for learning solutions under suitable bases. Following Xu (2023), we are interested in a class of RKBSs, each of which has an adjoint RKBS. Such an RKBS provides a reproducing kernel for representing not only the point evaluation functionals but also learning solutions in the spaces.

We first clarify the notion of sparsity. Intuitively, a vector or a sequence is said to be sparse if most of its components are zero. We say that a function in an RKBS has a sparse representation under the kernel sessions (a kernel with one of its two variables evaluated at given points) if the coefficient vector of the representation is sparse. It is well-known from the celebrated representer theorem (deBoor and Lynch (1966); Argyriou et al. (2009); Cox and O'Sullivan (1990); Kimeldorf and Wahba (1970); Schölkopf et al. (2001)) that when we learn a target function in a reproducing kernel Hilbert space (RKHS) from n function values, a solution of the regularization problem is a linear combination of the n kernel sessions. Often large amount of data are used to learn a target function and the learning solution is used in prediction or other decision-making procedures repeatedly. As a result, a dense learning solution will lead to large computational costs. As we have known, RKHSs do not lead to sparse learning solutions, see for example, Xu (2023). We then appeal to RKBSs as hypothesis spaces for learning methods and hope that some of them can offer sparse representations for their learning solutions under the kernel sessions, which have terms significantly fewer than the number of the given data points.

We consider two typical learning models in an RKBS: the minimum norm interpolation (MNI) problem and the regularization problem. Representer theorems for the solutions of these two models in RKBSs have received considerable attention in the literature (Huang et al. (2021); Unser (2021); Unser et al. (2016); Wang and Xu (2021); Xu and Ye (2019); Zhang et al. (2009); Zhang and Zhang (2012)). In particular, a systematic study of the representer theorems for a solution of the MNI problem and the regularization problem in a Banach space was conducted by Wang and Xu (2021). The resulting representer theorem stated that the solutions lie in a subdifferential set of the norm function evaluated at a finite linear combination of given functionals. On the other hand, an explicit representer theorem for a variational problem in a Banach space was proved in Boyer et al. (2019); Bredies and Carioni (2020) in which the extreme points of the solution set of the variational problem were expressed as a finite linear combination of extreme points of the unit ball in the Banach space. The representer theorem of this kind is *data-independent*, as the searching area for extreme points claimed to express the solution is always the unit ball no matter what the given data are.

The road-map for establishing the sparse representer theorem for the solutions of the MNI problem and the regularization problem in an RKBS may be described as follows. By combining the advantages of the two representer theorems in Wang and Xu (2021) and Boyer et al. (2019), we first put forward an explicit solution representation for the MNI problem in a general Banach space, which is assumed to have a pre-dual space. The new explicit representer theorem allows us to represent the extreme points of the solution set as a linear combination of the extreme points of a subdifferential set of a pre-dual norm, evaluated at a linear combination of the functionals determined by given data. As a result, unlike the representer theorems presented in Boyer et al. (2019); Bredies and Carioni (2020), which are data-independent, the new representer theorem is *data-dependent.* Moreover, we prove that the extreme point set of the subdifferential set is a subset of the extreme point set of the unit ball and its cardinality is finite for certain specific Banach spaces, while the extreme point set of the unit ball includes infinitely many elements. By making use of the explicit representer theorem, we propose a sufficient condition on an RKBS so that the solution of the MNI problem in the RKBS has a sparse kernel representation. The establishment of a sparse kernel representation for the learning solution requires addressing two issues. The first one is to represent the learning solution by the kernel sessions. Observing from the explicit representer theorem, it is intuitive to require that the elements in the subdifferential set, the building blocks of the solution set, coincide with the kernel sessions. This is the first assumption imposed on the RKBS. The second issue is to ensure that the kernel representation has fewer terms than the number of the given data points. To this end, we impose an additional assumption on the RKBS by requiring its norm to be equivalent to the ℓ_1 norm which is well-known to promote sparsity. Under these two assumptions, we install kernel representations for the extreme points of the solution set of the MNI problem. The number of the kernel sessions emerging in the representation is bounded above by the rank of a matrix determined by the observed data. Under a mild condition, the rank of the matrix may be less than the number of the observed data. This leads to a sparse representer theorem for the solutions of the MNI problem in the RKBS. We then convert the resulting sparse representer theorem for the MNI problem to that for the regularization problem through a relation between the solutions of the two problems pointed out in Micchelli and Pontil (2004); Wang and Xu (2021). Unlike the MNI problem, the regularization problem involves a regularization parameter. Under the assumptions on the RKBS, we reveal that the regularization parameter can further promote the sparsity level of the solution. Moreover, we show that the sequence space $\ell_1(\mathbb{N})$ and the space of functions constructed by the measure space satisfy the imposed assumptions. In this way, the sparse representer theorems for the MNI problem and the regularization problem in these two RKBSs are established, showing that they indeed can promote sparsity in kernel representations for learning functions in these two spaces.

Banach spaces were recently employed to understand neural networks and to promote sparsity. A sparse technique was successfully applied in Rosset et al. (2007) to learning with regularization for a feature space of infinite (possibly non-countable) dimension. Neural networks of a single hidden layer with infinitely many neurons as functions by integral representation with a variational norm were studied in Bach (2017). Based on the variational framework of L-splines developed in Unser et al. (2017) and the representer theorem established in Bredies and Carioni (2020), Parhi and Nowak (2021) obtained a representer theorem expressing neural networks of a single hidden layer as solutions of a variation problem with the TV regularization in the Radon domain. The representer theorem for single-output neural networks of one hidden layer was extended in Shenouda et al. (2023) to multi-output networks by considering vector-valued variation spaces. RKBSs were employed in Bartolucci et al. (2023) to study neural networks with one hidden layer. It was shown in Spek et al. (2023) that the Barron spaces were a class of integral RKBSs and their dual spaces as RKBSs were studied. Results developed in Rosset et al. (2007) were found useful in understanding neural networks in RKBSs.

We organize this paper in six sections and two appendices. In Section 2, we review the framework of RKBSs and describe the MNI and regularization problems to be considered in this paper. Also, we define precisely the notion of sparsity of a learning solution in an RKBS. Moreover, we present several new observations for RKBSs, including the closure and weak* closure of the space of point evaluation functionals and the unique reproducing kernel of a RKBS whose δ-dual space is isometrically isomorphic to a Banach space of functions. In Section 3, we first establish a representer theorem for the solutions of the MNI problem in a general Banach space that is assumed to have a pre-dual space. We then impose two assumptions on the RKBSs, which ensure that the spaces have the sparsity promoting property. Under the assumptions, we convert a representation of the solutions to a sparse kernel representation. In Section 4, we translate the sparse representer theorem established in Section 3 for the MNI problem to the regularization problem via a connection between the solutions of these two problems. We also study the effect of the regularization parameter on the sparsity of the regularized solutions and obtain choices of the regularization parameter for sparse solutions. In Section 5, we specialize the sparse representer theorem to the sequence space $\ell_1(\mathbb{N})$. By comparing this space with the sequence spaces $\ell_p(\mathbb{N})$ for $1 < p < +\infty$, we show that $\ell_1(\mathbb{N})$ can promote sparsity of learning solutions but spaces $\ell_p(\mathbb{N})$ have no such a feature. Section 6 concerns a specific RKBS constructed by the measure space. We establish the sparse representer theorems for the solutions of the MNI problem and the regularization problem in the space by verifying that the RKBS satisfies the imposed assumptions. We include in Appendix A a complete proof of the explicit representer theorem of the MNI problem in a general Banach space established in Section 3. In Appendix B, we characterize the dual problem of the MNI problem in order to acquire the dual element emerging in the representer theorem established in this paper.

2. Learning in RKBSs

Motivated by developing sparse learning algorithms, RKBSs have been proposed as appropriate hypothesis spaces for learning an objective function from its values. Since the introduction of the notion of RKBSs, theory and applications of these function spaces have attracted much research interest (Bartolucci et al. (2023); Fasshauer et al. (2015); Lin et al. (2021, 2022); Song et al. (2013); Spek et al. (2023); Xu and Ye (2019); Zhang et al. (2009); Zhang and Zhang (2012)). In this section, we recall the notion of RKBSs. We reveal the important role of the family of the point evaluation functionals in the dual space of an RKBS and identify a unique reproducing kernel for the RKBS as closed-form function representations for the point evaluation functionals. We also describe the MNI problem and the regularization problem in an RKBS and introduce the notion of sparse kernel representations for solutions of such learning problems.

We start with recalling the notion of RKBSs. A Banach space $\mathcal{B}$ is called a space of functions on a prescribed set X if $\mathcal{B}$ is composed of functions defined on X and for each $f \in \mathcal{B}$, $\|f\|_{\mathcal{B}} = 0$ implies that $f(x) = 0$ for all $x \in X$. The notion of RKBSs was originally introduced in Zhang et al. (2009), to ensure the stability of sampling function values from functions in the hypothesis space.

Definition 1 *A Banach space $\mathcal{B}$ of functions on a prescribed set X is called an RKBS if all the point evaluation functionals δ_x, $x \in X$, are continuous on $\mathcal{B}$, that is, for each $x \in X$, there exists a constant $C_x > 0$ such that*

$$|\delta_x(f)| \leq C_x \|f\|_{\mathcal{B}}, \quad \textit{for all } f \in \mathcal{B}.$$

This definition appeared originally in Zhang et al. (2009) in a somewhat restricted version and its current form is adopted from Xu (2023). We let

$$\Delta := \operatorname{span}\{\delta_x : x \in X\}. \tag{1}$$

It is essential to understand the closures of Δ under various types of topology.

We first review necessary notions in Banach spaces. For a Banach space $\mathcal{B}$ with a norm $\|\cdot\|_{\mathcal{B}}$, we denote by $\mathcal{B}^*$ the dual space of $\mathcal{B}$, which is composed of all bounded linear functionals on $\mathcal{B}$ endowed with the norm

$$\|\nu\|_{\mathcal{B}^*} := \sup_{\|f\|_{\mathcal{B}} \leq 1} |\nu(f)|, \text{ for all } \nu \in \mathcal{B}^*.$$

The dual bilinear form $\langle \cdot, \cdot \rangle_{\mathcal{B}}$ on $\mathcal{B}^* \times \mathcal{B}$ is defined by $\langle \nu, f \rangle_{\mathcal{B}} := \nu(f)$ for all $\nu \in \mathcal{B}^*$ and all $f \in \mathcal{B}$. The weak* topology of the dual space $\mathcal{B}^*$ is the smallest topology for $\mathcal{B}^*$ such that, for each $f \in \mathcal{B}$, the linear functional $\nu \to \langle \nu, f \rangle_{\mathcal{B}}$ on $\mathcal{B}^*$ is continuous with respect to the topology. A topological property that holds with respect to the weak* topology is said to hold weakly*. For example, a sequence $\nu_n, n \in \mathbb{N}$, in $\mathcal{B}^*$ is said to converge weakly* to $\nu \in \mathcal{B}^*$ if $\lim_{n \to +\infty} \langle \nu_n, f \rangle_{\mathcal{B}} = \langle \nu, f \rangle_{\mathcal{B}}$, for all $f \in \mathcal{B}$. Let $\mathcal{M}$ and $\mathcal{M}'$ be subsets of $\mathcal{B}$ and $\mathcal{B}^*$, respectively. The annihilator, in $\mathcal{B}^*$, of $\mathcal{M}$ is defined by

$$\mathcal{M}^\perp := \{\nu \in \mathcal{B}^* : \langle \nu, f \rangle_{\mathcal{B}} = 0, \text{ for all } f \in \mathcal{M}\},$$

and the annihilator, in $\mathcal{B}$, of $\mathcal{M}'$ is defined by

$${}^\perp\mathcal{M}' := \{f \in \mathcal{B} : \langle \nu, f \rangle_{\mathcal{B}} = 0, \text{ for all } \nu \in \mathcal{M}'\}.$$

We denote by $\overline{\mathcal{M}'}$ and $\overline{\mathcal{M}'}^{w^*}$ the closure of $\mathcal{M}'$ in the norm topology and the weak* topology of $\mathcal{B}^*$, respectively.

We now turn to characterizing the weak* density of the linear span Δ in $\mathcal{B}^*$. We observe from Definition 1 that if $\mathcal{B}$ is an RKBS on X, then there holds $\delta_x \in \mathcal{B}^*$ for all $x \in X$. That is, $\Delta \subseteq \mathcal{B}^*$. It is known that if $\mathcal{B}$ is an RKHS, then there holds $\overline{\Delta} = \mathcal{B}^*$ due to the Riesz representation theorem of the Hilbert space, where the closure is in the sense of the norm $\|\cdot\|_{\mathcal{B}}$ of the Hilbert space. This result cannot be extended to a general RKBS $\mathcal{B}$, because of the lack of the representation theorem in a general Banach space. However, we can show that the linear span Δ is weakly* dense in $\mathcal{B}^*$. We present this result in the next proposition.

Proposition 2 *If $\mathcal{B}$ is an RKBS on X, then $\overline{\Delta}^{w^*} = \mathcal{B}^*$.*

Proof By Definition 1, we clearly see that $\Delta \subseteq \mathcal{B}^*$. Note that $f \in {}^\perp\Delta$ if and only if $\delta_x(f) = 0$, for all $x \in X$. That is, $f(x) = 0$, for all $x \in X$, or $f = 0$. Therefore, we obtain that ${}^\perp\Delta = \{0\}$, which further leads to $({}^\perp\Delta)^\perp = \mathcal{B}^*$. According to Proposition 2.6.6 of Megginson (1998), there holds that $({}^\perp\Delta)^\perp = \overline{\Delta}^{w^*}$. Combining the above two equations, we conclude that $\overline{\Delta}^{w^*} = \mathcal{B}^*$. ■

When an RKBS $\mathcal{B}$ is assumed to have a pre-dual space $\mathcal{B}_*$ satisfying $\Delta \subseteq \mathcal{B}_*$, we can identify $\overline{\Delta}$ with the pre-dual $\mathcal{B}_*$, where the closure $\overline{\Delta}$ is taken in the norm of $\mathcal{B}^*$. We say that the Banach space $\mathcal{B}$ has a pre-dual space if there exists a Banach space $\mathcal{B}_*$ such that $(\mathcal{B}_*)^* = \mathcal{B}$ and we call the space $\mathcal{B}_*$ a pre-dual space of $\mathcal{B}$. The existence of a pre-dual space $\mathcal{B}_*$ makes it valid for $\mathcal{B}$ to be equipped with weak* topology. Since the pre-dual space $\mathcal{B}_*$ can be isometrically embedded into $\mathcal{B}^*$, any element in $\mathcal{B}_*$ can be viewed as a bounded linear functional on $\mathcal{B}$. Namely, $\mathcal{B}_* \subset \mathcal{B}^*$ and there holds

$$\langle \nu, f\rangle_{\mathcal{B}} = \langle f, \nu\rangle_{\mathcal{B}_*}, \text{ for all } f \in \mathcal{B} \text{ and all } \nu \in \mathcal{B}_*. \tag{2}$$

Proposition 3 *Suppose that an RKBS $\mathcal{B}$ on X has a pre-dual space $\mathcal{B}_*$. If $\Delta \subseteq \mathcal{B}_*$, then $\overline{\Delta} = \mathcal{B}_*$.*

Proof Suppose that $f \in \mathcal{B}$ satisfies $\langle f, \delta_x\rangle_{\mathcal{B}_*} = 0$ for all $x \in X$. By equation (2), we have that $\langle \delta_x, f\rangle_{\mathcal{B}} = 0$ for all $x \in X$. This leads to $f(x) = 0$ for all $x \in X$ and thus $f = 0$. Due to the arbitrariness of $f \in \mathcal{B}$, we get the desired density result. ■

A reflexive RKBS $\mathcal{B}$ always takes the dual space $\mathcal{B}^*$ as a pre-dual space $\mathcal{B}_*$. As a direct consequence of Proposition 3, we obtain that $\overline{\Delta} = \mathcal{B}^*$ for a reflexive RKBS $\mathcal{B}$.

It is known that each RKHS enjoys a unique reproducing kernel, which provides closed-form function representations for all point evaluation functionals on the RKHS. The existence of the reproducing kernel lies in the well-known Riesz representation theorem, which states that the dual space of a Hilbert space is isometrically isomorphic to itself. However, in general, the dual space of a Banach space is not isometrically isomorphic to itself. To ensure the existence of the reproducing kernel for an RKBS, we need to impose additional assumptions. Various hypotheses have been imposed on RKBSs (Lin et al. (2022); Xu (2023); Xu and Ye (2019); Zhang et al. (2009)) in the literature. A hypothesis, which better captures the essence of reproducing kernels, was described in Xu (2023). Following Xu (2023), we call $\overline{\Delta}$ the δ-dual space of $\mathcal{B}$ and denote it by $\mathcal{B}'$. Note that $\mathcal{B}'$ is the smallest Banach space that contains all the point evaluation functionals on $\mathcal{B}$. We assume that the δ-dual space $\mathcal{B}'$ is isometrically isomorphic to a Banach space $\mathcal{F}$ of functions on a set X'. In the rest of this paper, we will not distinguish $\mathcal{B}'$ from $\mathcal{F}$. We now identify a unique reproducing kernel with each RKBS satisfying the hypothesis.

Proposition 4 *Suppose that $\mathcal{B}$ is an RKBS on X and its δ-dual space $\mathcal{B}'$ is isometrically isomorphic to a Banach space of functions on X'. Then there exists a unique function $K : X \times X' \to \mathbb{R}$ such that the following statements hold.*

(1) For each $x \in X$, $K(x, \cdot) \in \mathcal{B}'$ and

$$f(x) = \langle K(x, \cdot), f\rangle_{\mathcal{B}}, \text{ for all } f \in \mathcal{B}. \tag{3}$$

(2) The linear span $K(X) := \operatorname{span}\{K(x, \cdot) : x \in X\}$ is dense in $\mathcal{B}'$.

Proof We first prove statement (1). For each $x \in X$, since δ_x is a continuous linear functional on $\mathcal{B}$, there exists $k_x \in \mathcal{B}'$ such that $f(x) = \langle k_x, f\rangle_{\mathcal{B}}$, for all $f \in \mathcal{B}$. By defining a function $K : X \times X' \to \mathbb{R}$ as $K(x, x') := k_x(x')$, $x \in X$, $x' \in X'$, we obtain that $K(x, \cdot) \in \mathcal{B}'$ for all $x \in X$. Thus, we observe that equation (3) holds.

It suffices to verify that the function K on $X \times X'$ satisfying the above properties is unique. Assume that there exists another $\widetilde{K} : X \times X' \to \mathbb{R}$ such that $\widetilde{K}(x, \cdot) \in \mathcal{B}'$ for all $x \in X$, and $f(x) = \langle \widetilde{K}(x, \cdot), f \rangle_{\mathcal{B}}$, for all $f \in \mathcal{B}$ and all $x \in X$. It follows from the above equation and equation (3) that

$$\langle K(x, \cdot) - \widetilde{K}(x, \cdot), f \rangle_{\mathcal{B}} = 0, \quad \text{for all} \quad f \in \mathcal{B} \quad \text{and all} \quad x \in X.$$

That is, for all $x \in X$, $K(x, \cdot) - \widetilde{K}(x, \cdot) = 0$. Noting that $\mathcal{B}'$ is isometrically isomorphic to a Banach space of functions on X', we conclude that $K(x, x') - \widetilde{K}(x, x') = 0$ for all $x \in X$, $x' \in X'$, which is equivalent to $K = \widetilde{K}$.

We next show the density stated in (2). It follows from statement (1) that the linear span $K(X)$ is isometrically isomorphic to Δ defined by (1). Hence, the closure $\overline{K(X)}$ in the norm topology is isometrically isomorphic to the closure $\overline{\Delta}$ in the norm topology, which together with $\overline{\Delta} = \mathcal{B}'$ leads to $\overline{K(X)} = \mathcal{B}'$. ■

We call the function $K : X \times X' \to \mathbb{R}$, satisfying $K(x, \cdot) \in \mathcal{B}'$ for all $x \in X$, and equation (3), the reproducing kernel for the RKBS $\mathcal{B}$. Moreover, equation (3) is called the reproducing property.

Motivated by representing the solutions of learning problems in an RKBS via its reproducing kernel, we introduce the notion of the adjoint RKBS. Specifically, if in addition $\mathcal{B}'$ is an RKBS on X', $K(\cdot, x') \in \mathcal{B}$ for all $x' \in X'$, and

$$g(x') = \langle g, K(\cdot, x') \rangle_{\mathcal{B}}, \text{ for all } g \in \mathcal{B}' \text{ and all } x' \in X', \tag{4}$$

we call $\mathcal{B}'$ an adjoint RKBS of $\mathcal{B}$ and call $\mathcal{B}$, $\mathcal{B}'$ a pair of RKBSs. We introduce the linear span of the point evaluation functions on $\mathcal{B}'$ by

$$\Delta' := \text{span}\{\delta_{x'} : x' \in X'\}. \tag{5}$$

Observing from equation (4), the linear span $K(X') := \text{span}\{K(\cdot, x') : x' \in X'\}$, as a subset of $\mathcal{B}$, is isometrically isomorphic to Δ' defined by (5). Moreover, the next result shows that if $\mathcal{B}'$ is a pre-dual space of $\mathcal{B}$, then the function space $K(X')$ is large enough to fill in $\mathcal{B}$ in the sense that any function f in $\mathcal{B}$ could be approximated arbitrarily well by elements in $K(X')$ with respect to weak* topology.

Proposition 5 *Suppose that $\mathcal{B}$ is an RKBS on X, the δ-dual space $\mathcal{B}'$ is an adjoint RKBS on X' of $\mathcal{B}$ and K is the reproducing kernel. If $\mathcal{B}'$ is a pre-dual space of $\mathcal{B}$, then there holds $\overline{K(X')}^{w^*} = \mathcal{B}$.*

Proof Since $\mathcal{B}'$ is an RKBS on X', Proposition 2 with $\mathcal{B}$ being replaced by $\mathcal{B}'$ ensures that $\overline{\Delta'}^{w^*} = (\mathcal{B}')^*$. This together with the assumption that $(\mathcal{B}')^* = \mathcal{B}$ leads to $\overline{\Delta'}^{w^*} = \mathcal{B}$. Note that the linear span $K(X')$ is isometrically isomorphic to Δ'. Hence, we get the desired weak* density of $K(X')$ in $\mathcal{B}$.

■

We also characterize the RKBS in terms of the feature representation.

Proposition 6 *A Banach space $\mathcal{B}$ of functions on X is an RKBS if and only if there exist a Banach space W and a map $\Phi : X \to W^*$ satisfying*

$$\overline{\operatorname{span}\{\Phi(x) : x \in X\}}^{w^*} = W^*, \tag{6}$$

such that

$$\mathcal{B} = \{\langle \Phi(\cdot), u\rangle_W : u \in W\}, \tag{7}$$

equipped with

$$\|\langle \Phi(\cdot), u\rangle_W\|_{\mathcal{B}} = \|u\|_W, \quad \textit{for all} \quad u \in W. \tag{8}$$

Proof Suppose that $\mathcal{B}$ is an RKBS on X. We choose the Banach space $W := \mathcal{B}$ and the map $\Phi : X \to \mathcal{B}^*$ defined by $\Phi(x) := \delta_x$, $x \in X$. Hence, we have that $\operatorname{span}\{\Phi(x) : x \in X\} = \Delta$. It follows from Proposition 2 that the density condition (6) holds true. In addition, we can trivially represent any $f \in \mathcal{B}$ as $f = \langle \Phi(\cdot), u\rangle_W$ with $u := f$ and hence $\|f\|_{\mathcal{B}} = \|u\|_W$.

Conversely, suppose that there exist W and $\Phi : X \to W^*$ such that equation (6) holds true. We then define a vector space $\mathcal{B}$ by (7) and a map $\|\cdot\|_{\mathcal{B}} : \mathcal{B} \to \mathbb{R}$ by (8). We point out that under the assumption that $\operatorname{span}\{\Phi(x) : x \in X\}$ is weakly* dense in W^*, any function in $\mathcal{B}$ has a unique representation. Indeed, if $\langle \Phi(x), u\rangle_W = \langle \Phi(x), v\rangle_W$, for all $x \in X$, then $\langle \Phi(x), u - v\rangle_W = 0$, for all $x \in X$. This together with equation (6) leads to $u = v$. As a result, the map defined by (8) is well-defined. It is easy to verify that $\|\cdot\|_{\mathcal{B}}$ is a norm on $\mathcal{B}$. Moreover, $\mathcal{B}$ is isometrically isomorphic to W and then is a Banach space. It suffices to show that point evaluation functionals are continuous on $\mathcal{B}$. For all $x \in X$ and all $f = \langle \Phi(\cdot), u\rangle_W \in \mathcal{B}$ with $u \in W$, it holds that

$$|f(x)| \leq \|\Phi(x)\|_{W^*}\|u\|_W = \|\Phi(x)\|_{W^*}\|f\|_{\mathcal{B}}.$$

This ensures that the point evaluation functionals are all continuous on $\mathcal{B}$. According to Definition 1, $\mathcal{B}$ is an RKBS. ■

To close this section, we describe two learning problems in RKBSs to be considered in this paper. Learning a function from a finite number of sampled data is often formulated as a MNI problem or a regularization problem. For each $n \in \mathbb{N}$, let $\mathbb{N}_n := \{1, 2, \ldots, n\}$. Suppose that $\mathcal{B}$ is an RKBS having a pre-dual space $\mathcal{B}_*$ and $\nu_j \in \mathcal{B}_*$, $j \in \mathbb{N}_n$, are linearly independent. Associated with these functionals, we set

$$\mathcal{V} := \operatorname{span}\{\nu_j : j \in \mathbb{N}_n\}, \tag{9}$$

and define an operator $\mathcal{L} : \mathcal{B} \to \mathbb{R}^n$ by

$$\mathcal{L}(f) := \left[\langle \nu_j, f\rangle_{\mathcal{B}} : j \in \mathbb{N}_n\right], \text{ for all } f \in \mathcal{B}. \tag{10}$$

Let $\mathbf{y} := [y_j : j \in \mathbb{N}_n] \in \mathbb{R}^n$ be a given vector. Learning a target function in $\mathcal{B}$ from the given sampled data $\{(\nu_j, y_j) : j \in \mathbb{N}_n\}$ consists of solving the first kind operator equation

$$\mathcal{L}(f) = \mathbf{y}, \tag{11}$$

for $f \in \mathcal{B}$. The MNI aims at finding an element in $\mathcal{B}$, having the smallest norm and satisfying equation (11). By introducing a subset of $\mathcal{B}$ as

$$\mathcal{M}_{\mathbf{y}} := \{f \in \mathcal{B} : \mathcal{L}(f) = \mathbf{y}\}, \tag{12}$$

we formulate the MNI problem with the given data $\{(\nu_j, y_j) : j \in \mathbb{N}_n\}$ as

$$\inf \{\|f\|_{\mathcal{B}} : f \in \mathcal{M}_{\mathbf{y}}\}. \tag{13}$$

To address the ill-posedness of equation (11), the regularization approach adds a regularization term to a data fidelity term constructed from equation (11) such that the resulting optimization problem is much less sensitive to disturbances. Specifically, the regularization problem has the form

$$\inf \{\mathcal{Q}_{\mathbf{y}}(\mathcal{L}(f)) + \lambda\varphi(\|f\|_{\mathcal{B}}) : f \in \mathcal{B}\}, \tag{14}$$

where $\mathcal{Q}_{\mathbf{y}} : \mathbb{R}^n \to \mathbb{R}_+ := [0, +\infty)$ is a loss function, $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ is a regularizer and λ is a positive regularization parameter. We always assume $\mathcal{Q}_{\mathbf{y}}$ and φ to be both lower semi-continuous. A function $\mathcal{T}$ mapping from a topological space $\mathcal{X}$ to $\mathbb{R}$ is said to be lower semi-continuous if $\mathcal{T}(f) \leq \liminf_\alpha \mathcal{T}(f_\alpha)$ whenever f_α, $\alpha \in I$, for some index set I is a net in $\mathcal{X}$ converging to some element $f \in \mathcal{X}$. We also assume that φ is increasing and coercive, that is, $\lim_{t\to+\infty} \varphi(t) = +\infty$. Throughout this paper, we denote by $\mathrm{S}(\mathbf{y})$ and $\mathrm{R}(\mathbf{y})$ the solution sets of the MNI problem (13) and the regularization problem (14) with $\mathbf{y} \in \mathbb{R}^n$, respectively.

Proposition 5 motivates us to consider representing the solutions of the MNI problem and the regularization problems in $\mathcal{B}$ by the kernel sessions $K(\cdot, x')$, $x' \in X'$. In this paper, a solution of a learning problem in an RKBS from n data points is said to have a sparse kernel representation if there exist a nonnegative integer m less than n and $x'_j \in X'$, $\alpha_j \in \mathbb{R}\backslash\{0\}$, $j \in \mathbb{N}_m$, such that

$$f(\cdot) = \sum_{j\in\mathbb{N}_m} \alpha_j K(\cdot, x'_j).$$

We call the smallest nonnegative integer m such that the above equation holds the sparsity level of f under the kernel representation.

3. Sparse Representer Theorem for MNI

The goal of this section is to establish a sparse representer theorem for solutions of the MNI problem in an RKBS. For this purpose, we first provide a representer theorem for solutions of the MNI problem in a general Banach space. We then formulate two assumptions on the RKBS and the functionals used to produce the sampled data, which together with the representer theorem transfer the MNI problem in the RKBS of infinite dimension to a finite dimensional one. We finally establish the sparse representer theorem for solutions of the MNI problem in the RKBS using the sparse representation for the solutions of the finite dimensional MNI problem.

We begin with the MNI problem (13) in an RKBS $\mathcal{B}$. To obtain sparse kernel representations for the solutions, we first need an explicit representer theorem. A representer theorem for the solutions of the MNI problem (13) in a general Banach space having a pre-dual space was established in Wang and Xu (2021). To describe this result, we recall the notion of the subdifferential of a convex function on a Banach space $\mathcal{B}$. A convex function $\phi : \mathcal{B} \to \mathbb{R} \cup \{+\infty\}$ is said to be subdifferentiable at $f \in \mathcal{B}$ if there exists $\nu \in \mathcal{B}^*$ such that

$$\phi(g) - \phi(f) \geq \langle \nu, g - f \rangle_{\mathcal{B}}, \text{ for all } g \in \mathcal{B}.$$

The set of all functionals in $\mathcal{B}^*$ satisfying the above inequalities is called the *subdifferential* of ϕ at f and denoted by $\partial\phi(f)$. A subset A of a vector space $\mathcal{X}$ is called a convex set if $tx + (1-t)y \in A$ for all $x, y \in A$ and all $t \in [0,1]$. It can be directly verified by definition that for any $f \in \mathcal{B}$, $\partial\phi(f)$ is a convex and weakly* closed subset of $\mathcal{B}^*$. In particular, the subdifferential of the norm function $\|\cdot\|_{\mathcal{B}}$ at each $f \in \mathcal{B}\backslash\{0\}$ has the following essential property (Cioranescu (1990))

$$\partial\|\cdot\|_{\mathcal{B}}(f) = \{\nu \in \mathcal{B}^* : \|\nu\|_{\mathcal{B}^*} = 1, \langle \nu, f \rangle_{\mathcal{B}} = \|f\|_{\mathcal{B}}\}. \tag{15}$$

The elements in $\partial\|\cdot\|_{\mathcal{B}}(f)$ are also called norming functionals of f as applying such functional on f turns out exactly the norm of f. Suppose that $\mathcal{B}$ is a Banach space having a pre-dual space $\mathcal{B}_*$ and $\nu_j \in \mathcal{B}_*$, $j \in \mathbb{N}_n$, are linearly independent. Theorem 12 in Wang and Xu (2021) states that $\hat{f} \in \mathcal{B}$ is a solution of the MNI problem (13) with $\mathbf{y} \in \mathbb{R}^n$ if and only if $\hat{f} \in \mathcal{M}_{\mathbf{y}}$ and there exists $\hat{\nu} \in \mathcal{V}$, such that

$$\hat{f} \in \|\hat{\nu}\|_{\mathcal{B}_*} \partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu}). \tag{16}$$

This representer theorem provides a characterization for the solutions of the MNI problem (13) by using an inclusion relation. We will develop an explicit representer theorem for the solution $\hat{f}$ of the MNI problem (13) based on the aforementioned result.

We recall the notion of extreme points of a closed convex subset. Let A be a nonempty closed convex subset of a Hausdorff topological vector space $\mathcal{X}$. An element $z \in A$ is said to be an extreme point of A if $x, y \in A$ and $tx + (1-t)y = z$ for some $t \in (0,1)$ implies that $x = y = z$. By $\text{ext}(A)$ we denote the set of extreme points of A. It is known (Megginson (1998)) that $z \in \text{ext}(A)$ if and only if whenever $x, y \in A$ and $z = (x+y)/2$, it follows that $x = y = z$. We also need the notion of the convex hull of a subset. The convex hull of a subset A of a vector space $\mathcal{X}$, denoted by $\text{co}(A)$, is the smallest convex set that contains A. It is easy to show that

$$\text{co}(A) = \left\{ \sum_{j \in \mathbb{N}_n} t_j x_j : x_j \in A, t_j \in [0, +\infty), \sum_{j \in \mathbb{N}_n} t_j = 1, j \in \mathbb{N}_n, n \in \mathbb{N} \right\}.$$

If $\mathcal{X}$ has a topology, then the closed convex hull of A, denoted by $\overline{\text{co}}(A)$, is the smallest closed convex set that contains A. The celebrated Krein-Milman theorem (Megginson (1998)) states that if A is a nonempty compact convex subset of a Hausdorff locally convex topological vector space $\mathcal{X}$, then A is the closed convex hull of its set of extreme points, that is, $A = \overline{\text{co}}(\text{ext}(A))$. As a direct consequence of this result, the set $\text{ext}(A)$ must be nonempty. Notice that if the Banach space $\mathcal{B}$ has a pre-dual space, then it equipped with the weak* topology is a Hausdorff locally convex topological vector space. The solution set $\text{S}(\mathbf{y})$, guaranteed by Lemma 31 in Appendix A, is a nonempty, convex and weakly* compact subset of $\mathcal{B}$. Then the Krein-Milman Theorem enables us to express $\mathrm{S}(\mathbf{y})$ by its extreme points as

$$\mathrm{S}(\mathbf{y}) = \overline{\mathrm{co}}\left(\mathrm{ext}\left(\mathrm{S}(\mathbf{y})\right)\right), \tag{17}$$

where the closed convex hull is taken under the weak* topology. Observing from equation (17), we will only provide closed-form representations for the extreme points of $\mathrm{S}(\mathbf{y})$.

With the help of the dual element $\hat{\nu}$ appearing in inclusion (16), we establish in the following proposition an explicit representer theorem for the extreme points of the solution set $\mathrm{S}(\mathbf{y})$ of problem (13). In this result, we consider a general Banach space which has a pre-dual space. Its complete proof is included in Appendix A.

Proposition 7 *Suppose that $\mathcal{B}$ is a Banach space having a pre-dual space $\mathcal{B}_*$. Let $\nu_j \in \mathcal{B}_*$, $j \in \mathbb{N}_n$, be linearly independent and $\mathbf{y} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$. Suppose that $\mathcal{V}$ and $\mathcal{M}_{\mathbf{y}}$ are defined by* (9) *and* (12)*, respectively, and $\hat{\nu} \in \mathcal{V}$ satisfies*

$$(\|\hat{\nu}\|_{\mathcal{B}_*} \partial \|\cdot\|_{\mathcal{B}_*}(\hat{\nu})) \cap \mathcal{M}_{\mathbf{y}} \neq \emptyset. \tag{18}$$

Then for any $\hat{f} \in \mathrm{ext}(\mathrm{S}(\mathbf{y}))$, there exist $\gamma_j \in \mathbb{R}$, $j \in \mathbb{N}_n$, with $\sum_{j\in\mathbb{N}_n} \gamma_j = \|\hat{\nu}\|_{\mathcal{B}_*}$ and $u_j \in \mathrm{ext}\left(\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})\right)$, $j \in \mathbb{N}_n$, such that

$$\hat{f} = \sum_{j\in\mathbb{N}_n} \gamma_j u_j. \tag{19}$$

Two remarks about this explicit representer theorem are in order. First, a representer theorem was proposed in Boyer et al. (2019); Bredies and Carioni (2020) based on which additional results were obtained in Bartolucci et al. (2023); Lin et al. (2020); Unser and Aziznejad (2022). While specializing the result in Boyer et al. (2019) to the MNI problem (13) in a Banach space which has a pre-dual space, any $\hat{f} \in \mathrm{ext}(\mathrm{S}(\mathbf{y}))$ can be represented as in (19) with $\gamma_j \in \mathbb{R}$, $j \in \mathbb{N}_n$ and $u_j \in \mathrm{ext}(B_0)$, $j \in \mathbb{N}_n$, where B_0 denotes the closed unit ball of $\mathcal{B}$ with center in the origin. This representation does not depend on the given data that define the MNI problem (13). Proposition 7 strengthens the representer theorem of Boyer et al. (2019) by specifying the elements u_j, $j \in \mathbb{N}_n$, used to represent the extreme points of $\mathrm{S}(\mathbf{y})$, to belong to the data-dependent set $\mathrm{ext}(\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu}))$, where $\hat{\nu}$ is some linear combination of the given functionals ν_j, $j \in \mathbb{N}_n$. Moreover, as shown in Proposition 35 of Appendix A, the set $\mathrm{ext}(\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu}))$ is indeed smaller than the set $\mathrm{ext}(B_0)$. In Section 5, we will show that in the special case that $\mathcal{B} := \ell_1(\mathbb{N})$, the set $\mathrm{ext}(B_0)$ includes infinitely many elements while $\mathrm{ext}(\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu}))$ has only finite elements. In a word, Proposition 7 provides a more precise characterization for u_j, $j \in \mathbb{N}_n$ by using the given data. Second, observing from Proposition 7, the element $\hat{\nu} \in \mathcal{V}$ satisfying (18) plays an important role in the representation (19) of the extreme points of $\mathrm{S}(\mathbf{y})$. To characterize such an element, we establish in Appendix B a dual problem for the MNI problem (13). Proposition 37 in Appendix B demonstrates that the element $\hat{\nu}$ can be obtained by solving the associated dual problem.

The explicit representer theorem enables us to establish sparse kernel representations for solutions of the MNI problem (13) in certain RKBSs. In the remaining part of this section, we always assume that $\mathcal{B}$ and $\mathcal{B}'$ are a pair of RKBSs and $K : X \times X' \to \mathbb{R}$ is the reproducing kernel for $\mathcal{B}$. In addition, we suppose that $\mathcal{B}_*$ is a pre-dual space of $\mathcal{B}$

and ν_j, $j \in \mathbb{N}_n$, are linearly independent elements in $\mathcal{B}_*$. Proposition 7 ensures that any extreme point of the solution set $\mathrm{S}(\mathbf{y})$ of problem (13) with $\mathbf{y} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ can be expressed by the linear combination of elements in the set $\mathrm{ext}(\partial \|\cdot\|_{\mathcal{B}_*}(\hat{\nu}))$ for some $\hat{\nu} \in \mathcal{V}$. In order to promote sparse kernel representations for the solutions of the MNI problem, it is intuitive to require that the elements in $\mathrm{ext}(\partial \|\cdot\|_{\mathcal{B}_*}(\nu))$, the building blocks of the solution set, are as sparse as possible. For this purpose, we require that the RKBS $\mathcal{B}$ and the functionals $\nu_j \in \mathcal{B}_*$, $j \in \mathbb{N}_n$, satisfy the following assumption.

(A1) For any nonzero $\nu \in \mathcal{V}$, there exists a finite subset X'_ν of X' such that

$$\mathrm{ext}(\partial \|\cdot\|_{\mathcal{B}_*}(\nu)) \subset \left\{-K(\cdot, x'), K(\cdot, x') : x' \in X'_\nu\right\}.$$

Under Assumption (A1), we can express any extreme point of the solution set $\mathrm{S}(\mathbf{y})$ as a linear combination of the kernel sessions. We impose an additional assumption on the RKBS $\mathcal{B}$ by relating its norm with the well-known sparsity-promoting norm $\|\cdot\|_1$.

(A2) There exists a positive constant C such that for any $m \in \mathbb{N}$, distinct points $x'_j \in X'$, $j \in \mathbb{N}_m$, and $\boldsymbol{\alpha} = [\alpha_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$, there holds

$$\left\|\sum_{j \in \mathbb{N}_m} \alpha_j K(\cdot, x'_j)\right\|_{\mathcal{B}} = C\|\boldsymbol{\alpha}\|_1.$$

We now turn to establishing sparse kernel representations for the solutions of the MNI problem (13) under the above assumptions. To this end, we introduce a finite dimensional MNI problem. Suppose that Assumption (A1) holds and $\hat{\nu} \in \mathcal{V}$ satisfies (18) with $\mathbf{y} \in \mathbb{R}^n$. We denote by $n(\hat{\nu})$ the cardinality of the set $X'_{\hat{\nu}}$ and suppose that

$$X'_{\hat{\nu}} := \left\{x'_j : j \in \mathbb{N}_{n(\hat{\nu})}\right\}. \tag{20}$$

By defining a matrix

$$\mathbf{L}_{\hat{\nu}} := [\langle \nu_i, K(\cdot, x'_j)\rangle_{\mathcal{B}} : \ i \in \mathbb{N}_n,\ j \in \mathbb{N}_{n(\hat{\nu})}] \in \mathbb{R}^{n \times n(\hat{\nu})}, \tag{21}$$

we introduce the finite dimensional MNI problem by

$$\inf\left\{\|\boldsymbol{\alpha}\|_1 : \mathbf{L}_{\hat{\nu}}\boldsymbol{\alpha} = \mathbf{y},\ \boldsymbol{\alpha} \in \mathbb{R}^{n(\hat{\nu})}\right\}, \tag{22}$$

and denote by $\mathrm{S}_{\hat{\nu}}(\mathbf{y})$ the solution set of problem (22) with $\mathbf{y} \in \mathbb{R}^n$. The next result concerns the sparsity of the elements in $\mathrm{S}_{\hat{\nu}}(\mathbf{y})$.

Proposition 8 *Suppose that* $\mathbf{y} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$, $\mathcal{V}$ *and* $\mathcal{M}_{\mathbf{y}}$ *be defined by* (9) *and* (12), *respectively, and* $\hat{\nu} \in \mathcal{V}$ *satisfy* (18). *If Assumption (A1) holds and* $\mathbf{L}_{\hat{\nu}}$ *is defined by* (21), *then* $\hat{\boldsymbol{\alpha}} \in \mathrm{ext}\,(\mathrm{S}_{\hat{\nu}}(\mathbf{y}))$ *has at most* $\mathrm{rank}(\mathbf{L}_{\hat{\nu}})$ *nonzero components.*

Proof We prove this result by employing Proposition 7. In this case, the Banach space $\widetilde{\mathcal{B}} := \mathbb{R}^{n(\hat{\nu})}$ endowed with the ℓ_1 norm has the pre-dual space $\widetilde{\mathcal{B}}_* := \mathbb{R}^{n(\hat{\nu})}$ endowed with the ℓ_∞ norm. In addition, for each $j \in \mathbb{N}_n$, the element $\widetilde{\nu}_j \in \widetilde{\mathcal{B}}_*$ is chosen as the j-th row of $\mathbf{L}_{\hat{\nu}}$. For $\mathbf{y} := [y_j : j \in \mathbb{N}_n]$, we set $\widetilde{\mathcal{M}}_{\mathbf{y}} := \{\boldsymbol{\alpha} \in \widetilde{\mathcal{B}} : \langle \widetilde{\nu}_j, \boldsymbol{\alpha}\rangle_{\widetilde{\mathcal{B}}} = y_j, j \in \mathbb{N}_n\}$. If $\widetilde{\nu}_j \in \widetilde{\mathcal{B}}_*$, $j \in \mathbb{N}_n$, are linearly dependent, we may select a maximal linearly independent subset $\{\widetilde{\nu}_{n_j} : j \in \mathbb{N}_{\dim(\widetilde{\mathcal{V}})}\}$ with $\widetilde{\mathcal{V}} := \operatorname{span}\{\widetilde{\nu}_j : j \in \mathbb{N}_n\}$. For $\mathbf{y}' := [y_{n_j} : j \in \mathbb{N}_{\dim(\widetilde{\mathcal{V}})}] \in \mathbb{R}^{\dim(\widetilde{\mathcal{V}})}$, we define

$$\widetilde{\mathcal{M}}_{\mathbf{y}'} := \left\{\boldsymbol{\alpha} \in \widetilde{\mathcal{B}} : \langle \widetilde{\nu}_{n_j}, \boldsymbol{\alpha} \rangle_{\widetilde{\mathcal{B}}} = y_{n_j}, j \in \mathbb{N}_{\dim(\widetilde{\mathcal{V}})}\right\}.$$

It is obvious that if $\widetilde{\mathcal{M}}_{\mathbf{y}}$ is nonempty, then $\widetilde{\mathcal{M}}_{\mathbf{y}}$ and $\widetilde{\mathcal{M}}_{\mathbf{y}'}$ are exactly the same. This allows us to consider an equivalent MNI problem $\inf\{\|\boldsymbol{\alpha}\|_{\widetilde{\mathcal{B}}} : \boldsymbol{\alpha} \in \widetilde{\mathcal{M}}_{\mathbf{y}'}\}$, in which the given functionals are linearly independent.

By choosing $\widetilde{\nu} \in \widetilde{\mathcal{V}}$ satisfying $\|\widetilde{\nu}\|_{\widetilde{\mathcal{B}}_*} \partial\|\cdot\|_{\widetilde{\mathcal{B}}_*}(\widetilde{\nu}) \cap \widetilde{\mathcal{M}}_{\mathbf{y}} \neq \emptyset$, Proposition 7 ensures that any extreme point $\hat{\boldsymbol{\alpha}}$ of the solution set $\mathrm{S}_{\hat{\nu}}(\mathbf{y})$ can be represented as a linear combination of at most $\dim(\widetilde{\mathcal{V}})$ extreme points of $\partial\|\cdot\|_{\widetilde{\mathcal{B}}_*}(\widetilde{\nu})$. Then the desired result follows from the fact that $\dim(\widetilde{\mathcal{V}}) = \operatorname{rank}(\mathbf{L}_{\hat{\nu}})$ and each extreme points of $\partial\|\cdot\|_{\widetilde{\mathcal{B}}_*}(\widetilde{\nu})$ has just one nonzero component. ■

Empirical results show that the MNI problem with the ℓ_1 norm can promote sparsity of a solution. Proposition 8 provides a theoretical characterization of the sparsity of the solutions of problem (22). In fact, the sparsity can be further characterized by the positional relationship between the hyperplanes constructed by the given data and the unit ball of $\mathbb{R}^{n(\hat{\nu})}$ under the ℓ_1 norm. Such relationship can be multifarious making the comprehensive analysis rather complicated.

Below, we reveal the relation between the solutions of problems (13) and (22).

Lemma 9 *Suppose that* $\mathbf{y} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$, $\mathcal{V}$ *and* $\mathcal{M}_{\mathbf{y}}$ *be defined by* (9) *and* (12), *respectively, and* $\hat{\nu} \in \mathcal{V}$ *satisfy* (18). *If Assumptions (A1) and (A2) hold,* $X'_{\hat{\nu}}$ *and* $\mathbf{L}_{\hat{\nu}}$ *are defined by* (20) *and* (21), *respectively, then the following statements hold.*

1. $\hat{f} := \sum_{j \in \mathbb{N}_{n(\hat{\nu})}} \hat{\alpha}_j K(\cdot, x'_j) \in \mathrm{S}(\mathbf{y})$ *if and only if* $\hat{\boldsymbol{\alpha}} := [\hat{\alpha}_j : j \in \mathbb{N}_{n(\hat{\nu})}] \in \mathrm{S}_{\hat{\nu}}(\mathbf{y})$.
2. *If* $\hat{f} := \sum_{j \in \mathbb{N}_{n(\hat{\nu})}} \hat{\alpha}_j K(\cdot, x'_j) \in \operatorname{ext}(\mathrm{S}(\mathbf{y}))$, *then* $\hat{\boldsymbol{\alpha}} := [\hat{\alpha}_j : j \in \mathbb{N}_{n(\hat{\nu})}] \in \operatorname{ext}(\mathrm{S}_{\hat{\nu}}(\mathbf{y}))$.

Proof For any $\boldsymbol{\alpha} := [\alpha_j : j \in \mathbb{N}_{n(\hat{\nu})}] \in \mathbb{R}^{n(\hat{\nu})}$, we set $f := \sum_{j \in \mathbb{N}_{n(\hat{\nu})}} \alpha_j K(\cdot, x'_j)$. It follows from Assumption (A2) that

$$\|f\|_{\mathcal{B}} = C\|\boldsymbol{\alpha}\|_1. \tag{23}$$

By definition (21) of the matrix $\mathbf{L}_{\hat{\nu}}$, we have that

$$\mathbf{L}_{\hat{\nu}}\boldsymbol{\alpha} = \left[\sum_{j \in \mathbb{N}_{n(\hat{\nu})}} \alpha_j \langle \nu_i, K(\cdot, x'_j) \rangle_{\mathcal{B}} : i \in \mathbb{N}_n\right].$$

This together with definition (10) of the operator $\mathcal{L}$ leads to

$$\mathbf{L}_{\hat{\nu}}\boldsymbol{\alpha} = \mathcal{L}(f). \tag{24}$$

We first prove statement 1. Suppose that $\hat{f} := \sum_{j \in \mathbb{N}_{n(\hat{\nu})}} \hat{\alpha}_j K(\cdot, x'_j) \in \mathrm{S}(\mathbf{y})$ and set $\hat{\boldsymbol{\alpha}} := [\hat{\alpha}_j : j \in \mathbb{N}_{n(\hat{\nu})}]$. It follows that $\mathcal{L}(\hat{f}) = \mathbf{y}$. This together with equation (24) with $\boldsymbol{\alpha} := \hat{\boldsymbol{\alpha}}$

and $f := \hat{f}$ leads to $\mathbf{L}_{\hat{\nu}}\hat{\boldsymbol{\alpha}} = \mathbf{y}$. To verify $\hat{\boldsymbol{\alpha}} \in \mathrm{S}_{\hat{\nu}}(\mathbf{y})$, it suffices to show that $\|\hat{\boldsymbol{\alpha}}\|_1 \leq \|\boldsymbol{\alpha}\|_1$ for any $\boldsymbol{\alpha} \in \mathbb{R}^{n(\hat{\nu})}$ satisfying $\mathbf{L}_{\hat{\nu}}\boldsymbol{\alpha} = \mathbf{y}$. Suppose that $\boldsymbol{\alpha} := [\alpha_j : j \in \mathbb{N}_{n(\hat{\nu})}] \in \mathbb{R}^{n(\hat{\nu})}$ satisfying $\mathbf{L}_{\hat{\nu}}\boldsymbol{\alpha} = \mathbf{y}$. By equation (24), the function $f := \sum_{j\in\mathbb{N}_{n(\hat{\nu})}} \alpha_j K(\cdot, x'_j)$ satisfies $\mathcal{L}(f) = \mathbf{y}$. That is, $f \in \mathcal{M}_{\mathbf{y}}$. This combined with $\hat{f} \in \mathrm{S}(\mathbf{y})$ leads to $\|\hat{f}\|_{\mathcal{B}} \leq \|f\|_{\mathcal{B}}$. Substituting equation (23) with the pair $f, \boldsymbol{\alpha}$ and the same equation with the pair $\hat{f}, \hat{\boldsymbol{\alpha}}$ into the above inequality, we conclude that $\|\hat{\boldsymbol{\alpha}}\|_1 \leq \|\boldsymbol{\alpha}\|_1$. Conversely, suppose that $\hat{\boldsymbol{\alpha}} := [\hat{\alpha}_j : j \in \mathbb{N}_{n(\hat{\nu})}] \in \mathrm{S}_{\hat{\nu}}(\mathbf{y})$ and set $\hat{f} := \sum_{j\in\mathbb{N}_{n(\hat{\nu})}} \hat{\alpha}_j K(\cdot, x'_j)$. It follows from $\mathbf{L}_{\hat{\nu}}\hat{\boldsymbol{\alpha}} = \mathbf{y}$ and equation (24) with $\boldsymbol{\alpha} := \hat{\boldsymbol{\alpha}}$, $f := \hat{f}$ that $\mathcal{L}(\hat{f}) = \mathbf{y}$. That is $\hat{f} \in \mathcal{M}_{\mathbf{y}}$. We choose $f \in \mathrm{ext}(\mathrm{S}(\mathbf{y}))$ and proceed to show that $\|\hat{f}\|_{\mathcal{B}} \leq \|f\|_{\mathcal{B}}$. Combining Proposition 7 with Assumption (A1), we represent f as $f = \sum_{j\in\mathbb{N}_{n(\hat{\nu})}} \alpha_j K(\cdot, x'_j)$ for some $\boldsymbol{\alpha} := [\alpha_j : j \in \mathbb{N}_{n(\hat{\nu})}] \in \mathbb{R}^{n(\hat{\nu})}$. Equation (24) ensures that $\mathbf{L}_{\hat{\nu}}\boldsymbol{\alpha} = \mathbf{y}$, which together with $\hat{\boldsymbol{\alpha}} \in \mathrm{S}_{\hat{\nu}}(\mathbf{y})$ leads to $\|\hat{\boldsymbol{\alpha}}\|_1 \leq \|\boldsymbol{\alpha}\|_1$. Again substituting equation (23) with the pair $f, \boldsymbol{\alpha}$ and the same equation with the pair $\hat{f}, \hat{\boldsymbol{\alpha}}$ into the above inequality, we obtain that $\|\hat{f}\|_{\mathcal{B}} \leq \|f\|_{\mathcal{B}}$. By noting that $f \in \mathrm{S}(\mathbf{y})$, we get the conclusion that $\hat{f} \in \mathrm{S}(\mathbf{y})$.

We next verify statement 2. Suppose that $\hat{f} := \sum_{j\in\mathbb{N}_{n(\hat{\nu})}} \hat{\alpha}_j K(\cdot, x'_j) \in \mathrm{ext}(\mathrm{S}(\mathbf{y}))$. Statement 1 of this lemma ensures that $\hat{\boldsymbol{\alpha}} := [\hat{\alpha}_j : j \in \mathbb{N}_{n(\hat{\nu})}] \in \mathrm{S}_{\hat{\nu}}(\mathbf{y})$. According to the definition of extreme points, it suffices to prove that whenever $\hat{\boldsymbol{\beta}} := [\hat{\beta}_j : j \in \mathbb{N}_{n(\hat{\nu})}] \in \mathrm{S}_{\hat{\nu}}(\mathbf{y})$ and $\hat{\boldsymbol{\gamma}} := [\hat{\gamma}_j : j \in \mathbb{N}_{n(\hat{\nu})}] \in \mathrm{S}_{\hat{\nu}}(\mathbf{y})$ satisfying $\hat{\boldsymbol{\alpha}} = (\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\gamma}})/2$, we have $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\gamma}}$. Set $\hat{g} := \sum_{j\in\mathbb{N}_{n(\hat{\nu})}} \hat{\beta}_j K(\cdot, x'_j)$ and $\hat{h} := \sum_{j\in\mathbb{N}_{n(\hat{\nu})}} \hat{\gamma}_j K(\cdot, x'_j)$. Clearly, $\hat{f} = (\hat{g} + \hat{h})/2$. Again using statement 1 of this lemma, we obtain that $\hat{g}, \hat{h} \in \mathrm{S}(\mathbf{y})$. Combining $f \in \mathrm{ext}(\mathrm{S}(\mathbf{y}))$ with the definition of extreme points, we get that $\hat{g} = \hat{h}$. It follows from Assumption (A2) that $\|\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\gamma}}\|_1 = \|\hat{g} - \hat{h}\|_{\mathcal{B}}/C = 0$. Thus, $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\gamma}}$, which completes the proof. ■

Combining the relation between the solutions of problems (13) and (22) and the sparsity characterization of the latter, we are ready to provide sparse kernel representations for the solutions of the MNI problem (13).

Theorem 10 *Suppose that* $\mathbf{y} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$, $\mathcal{V}$ *and* $\mathcal{M}_{\mathbf{y}}$ *are defined by* (9) *and* (12), *respectively, and* $\hat{\nu} \in \mathcal{V}$ *satisfy* (18). *If Assumptions (A1) and (A2) hold with a positive constant* C, $X'_{\hat{\nu}}$ *and* $\mathbf{L}_{\hat{\nu}}$ *are defined by* (20) *and* (21), *respectively, then for any* $\hat{f} \in \mathrm{ext}\,(\mathrm{S}(\mathbf{y}))$, *there exist* $\hat{\alpha}_j \neq 0$, $j \in \mathbb{N}_M$, *with* $\sum_{j\in\mathbb{N}_M} |\hat{\alpha}_j| = \|\hat{\nu}\|_{\mathcal{B}_*}/C$ *and* $x'_j \in X'_{\hat{\nu}}$, $j \in \mathbb{N}_M$, *such that*

$$\hat{f} = \sum_{j\in\mathbb{N}_M} \hat{\alpha}_j K(\cdot, x'_j), \tag{25}$$

for some positive integer $M \leq \mathrm{rank}(\mathbf{L}_{\hat{\nu}})$.

Proof Suppose that $\hat{f} \in \mathrm{ext}\,(\mathrm{S}(\mathbf{y}))$. By Proposition 7 and noting that Assumption (A1) holds, we represent f as

$$\hat{f} = \sum_{j\in\mathbb{N}_{n(\hat{\nu})}} \hat{\alpha}_j K(\cdot, x'_j), \tag{26}$$

for some $\hat{\alpha}_j \in \mathbb{R}$, $x'_j \in X'_{\hat{\nu}}$, $j \in \mathbb{N}_{n(\hat{\nu})}$. Since $\hat{f} \in \mathrm{ext}(\mathrm{S}(\mathbf{y}))$, we get by statement 2 of Lemma 9 that $\hat{\boldsymbol{\alpha}} := [\hat{\alpha}_j : j \in \mathbb{N}_{n(\hat{\nu})}] \in \mathrm{ext}(\mathrm{S}_{\hat{\nu}}(\mathbf{y}))$. Proposition 8 ensures that $\hat{\boldsymbol{\alpha}}$ has at most $\mathrm{rank}(\mathbf{L}_{\hat{\nu}})$

nonzero entries, which allows us to rewrite equation (26) as (25) with $\hat{\alpha}_j \neq 0$, $j \in \mathbb{N}_M$, for some positive integer $M \leq \operatorname{rank}(\mathbf{L}_{\hat{\nu}})$. It remains to show that $\sum_{j\in\mathbb{N}_M} |\hat{\alpha}_j| = \|\hat{\nu}\|_{\mathcal{B}_*}/C$. It follows from Assumption (A2) that

$$\sum_{j\in\mathbb{N}_M} |\hat{\alpha}_j| = \|\hat{f}\|_{\mathcal{B}}/C. \tag{27}$$

Theorem 12 in Wang and Xu (2021) guarantees that if $\hat{\nu} \in \mathcal{V}$ satisfies (18), then any $f \in (\|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})) \cap \mathcal{M}_{\mathbf{y}}$ is a solution of the MNI problem (13). According to property (15), any $f \in (\|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})) \cap \mathcal{M}_{\mathbf{y}}$ satisfies that $\|f\|_{\mathcal{B}} = \|\hat{\nu}\|_{\mathcal{B}_*}$. That is, the infimum of the MNI problem (13) is $\|\hat{\nu}\|_{\mathcal{B}_*}$. By noting that $\hat{f} \in \mathrm{S}(\mathbf{y})$, we obtain that $\|\hat{f}\|_{\mathcal{B}} = \|\hat{\nu}\|_{\mathcal{B}_*}$. Substituting the above equation into equation (27), we get that $\sum_{j\in\mathbb{N}_M} |\hat{\alpha}_j| = \|\hat{\nu}\|_{\mathcal{B}_*}/C$, which completes the proof. ■

Theorem 10 provides kernel representations, for the solutions of the MNI problem (13), with the number of the kernel sessions, which appear in the resulting representations, being no more than the number n of the observed data. In Section 5, we will show by specific examples that $\operatorname{rank}(\mathbf{L}_{\hat{\nu}})$ is usually less than the number of the data. Hence, Theorem 10 may be taken as a sparse representer theorem for the solutions of the MNI problem. Such a sparse kernel representation profits from the fact that Assumptions (A1) and (A2) allow us to transform the original MNI problem to an equivalent finite dimensional MNI problem with the ℓ_1 norm. As a result, a further characterization of the sparsity of the solutions of problem (22) may lead to a more precise sparsity of the solutions of problem (13).

4. Sparse Representer Theorem for Regularization Problems

In this section, we establish a sparse representer theorem for regularization problems in the RKBS. This is done by translating the sparse representer theorem established in the last section for the MNI problem to regularization problems via the connection between the solutions of these problems. Unlike the MNI problem, the regularization problem involves a regularization parameter which allows us to further promote the sparsity level of the solution. Specifically, we convert the regularization problem in the infinite dimensional RKBS to a finite dimensional one by using the sparse representer theorem. We then obtain choices of the regularization parameter for sparse solutions of the regularization problem in the RKBS by using the existing results for the finite dimensional regularization problem.

Throughout this section, we suppose that $\mathcal{B}$ and $\mathcal{B}'$ are a pair of RKBSs and $K : X \times X' \to \mathbb{R}$ is the reproducing kernel for $\mathcal{B}$. Let $\mathcal{B}_*$ be a pre-dual space of $\mathcal{B}$ and ν_j, $j \in \mathbb{N}_n$, be linearly independent elements in $\mathcal{B}_*$. We also assume that for any given $\mathbf{y} \in \mathbb{R}^n$, both $\mathcal{Q}_{\mathbf{y}} : \mathbb{R}^n \to \mathbb{R}_+$ and $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ are lower semi-continuous and moreover, φ is increasing and coercive. It is known (Unser (2021)) that the solution set $\mathrm{R}(\mathbf{y})$ of the regularization problem (14) with $\mathbf{y} \in \mathbb{R}^n$ and $\lambda > 0$ is nonempty and weakly* compact, and moreover, if both $\mathcal{Q}_{\mathbf{y}}$ and φ are convex, then $\mathrm{R}(\mathbf{y})$ is also convex.

We begin with recalling the relation between a solution of the MNI problem (13) and that of the regularization problem (14) which was put forward in Wang and Xu (2021).

Recalling $\mathcal{L}:\mathcal{B}\to\mathbb{R}^n$ defined by (10), we introduce a subset $\mathcal{D}_{\lambda,\mathbf{y}}$ of $\mathbb{R}^n$ by

$$\mathcal{D}_{\lambda,\mathbf{y}} := \mathcal{L}(\mathrm{R}(\mathbf{y})). \tag{28}$$

In this notation, Proposition 41 in Wang and Xu (2021) shows that

$$\bigcup_{\mathbf{z}\in\mathcal{D}_{\lambda,\mathbf{y}}} \mathrm{S}(\mathbf{z}) \subset \mathrm{R}(\mathbf{y}), \tag{29}$$

and if φ is further assumed to be strictly increasing, then

$$\bigcup_{\mathbf{z}\in\mathcal{D}_{\lambda,\mathbf{y}}} \mathrm{S}(\mathbf{z}) = \mathrm{R}(\mathbf{y}). \tag{30}$$

The next lemma concerns a relation between the extreme points of the solution sets of problems (13) and (14).

Lemma 11 *Suppose that $\mathbf{y}_0\in\mathbb{R}^n$ and $\lambda>0$. Let $\mathcal{D}_{\lambda,\mathbf{y}_0}$ be defined by (28) with $\mathbf{y}:=\mathbf{y}_0$. If both $\mathcal{Q}_{\mathbf{y}_0},\varphi$ are convex and moreover, φ is strictly increasing, then*

$$\mathrm{ext}\left(\mathrm{R}(\mathbf{y}_0)\right) \subset \bigcup_{\mathbf{z}\in\mathcal{D}_{\lambda,\mathbf{y}_0}} \mathrm{ext}\left(\mathrm{S}(\mathbf{z})\right). \tag{31}$$

Proof We first note that the Krein-Milman theorem ensures that the extreme point sets appearing in inclusion (31) are all nonempty.

We next prove that inclusion (31) holds true. We assume that

$$\hat{f}\in\mathrm{ext}(\mathrm{R}(\mathbf{y}_0)), \tag{32}$$

and proceed to show that $\hat{f}$ belongs to the set on the right-hand-side of (31). To this end, we choose $\hat{\mathbf{z}}:=\mathcal{L}(\hat{f})$ and clearly, $\hat{\mathbf{z}}\in\mathcal{D}_{\lambda,\mathbf{y}_0}$. It suffices to show that $\hat{f}\in\mathrm{ext}(\mathrm{S}(\hat{\mathbf{z}}))$. It follows from (30) with $\mathbf{y}:=\mathbf{y}_0$ that

$$\mathrm{S}(\hat{\mathbf{z}})\subset\mathrm{R}(\mathbf{y}_0), \tag{33}$$

and $\hat{f}\in\mathrm{S}(\hat{\mathbf{z}})$. For any $f_1,f_2\in\mathrm{S}(\hat{\mathbf{z}})$ satisfying $\hat{f}=(f_1+f_2)/2$, according to (33), it follows that $f_1,f_2\in\mathrm{R}(\mathbf{y}_0)$. This combined with (32) and the definition of extreme points leads to $f_1=f_2=\hat{f}$. Again using the definition of extreme points, we obtain that $\hat{f}\in\mathrm{ext}(\mathrm{S}(\hat{\mathbf{z}}))$. ■

Through the connection between the solutions of these two problems, we translate the sparse representer theorem 10 for the MNI problem (13) to the regularization problem (14).

Theorem 12 *Suppose that $\mathbf{y}_0\in\mathbb{R}^n$ and $\lambda>0$. Let $\mathcal{V}$ be defined by (9) and $\mathcal{D}_{\lambda,\mathbf{y}_0}$ be defined by (28) with $\mathbf{y}:=\mathbf{y}_0$. If Assumptions (A1) and (A2) hold with a positive constant C, then the following statements hold.*

1. *If $\mathcal{D}_{\lambda,\mathbf{y}_0}\neq\{\mathbf{0}\}$, then there exists $\hat{f}\in\mathrm{R}(\mathbf{y}_0)$ such that*

$$\hat{f}=\sum_{j\in\mathbb{N}_M}\hat{\alpha}_j K(\cdot,x'_j), \tag{34}$$

for some $\hat{\nu}\in\mathcal{V}$, positive integer $M\le\mathrm{rank}(\mathbf{L}_{\hat{\nu}})$, $x'_j\in X'_{\hat{\nu}}$, $j\in\mathbb{N}_M$, and $\hat{\alpha}_j\neq0$, $j\in\mathbb{N}_M$, with $\sum_{j\in\mathbb{N}_M}|\hat{\alpha}_j|=\|\hat{\nu}\|_{\mathcal{B}_*}/C$.

2. *If both $\mathcal{Q}_{\mathbf{y}_0}$ and φ are convex and φ is strictly increasing, then every nonzero extreme point $\hat{f}$ of* $\mathrm{R}(\mathbf{y}_0)$ *satisfies* (34) *for some $\hat{\nu} \in \mathcal{V}$, positive integer $M \leq \mathrm{rank}(\mathbf{L}_{\hat{\nu}})$, $x'_j \in X'_{\hat{\nu}}$, $j \in \mathbb{N}_M$, and $\hat{\alpha}_j \neq 0$, $j \in \mathbb{N}_M$, with $\sum_{j\in\mathbb{N}_M} |\hat{\alpha}_j| = \|\hat{\nu}\|_{\mathcal{B}_*}/C$.*

Proof We first prove Statement 1. Since $\mathrm{R}(\mathbf{y}_0)$ is nonempty and $\mathcal{D}_{\lambda,\mathbf{y}_0} \neq \{\mathbf{0}\}$, there exists $\hat{g} \in \mathrm{R}(\mathbf{y}_0)$ such that $\hat{\mathbf{z}} := \mathcal{L}(\hat{g}) \neq \mathbf{0}$. It follows from (29) that $\mathrm{S}(\hat{\mathbf{z}}) \subset \mathrm{R}(\mathbf{y}_0)$. As a result, $\mathrm{ext}\,(\mathrm{S}(\hat{\mathbf{z}})) \subset \mathrm{R}(\mathbf{y}_0)$. Noting that $\mathrm{ext}\,(\mathrm{S}(\hat{\mathbf{z}}))$ is nonempty, we choose $\hat{f} \in \mathrm{ext}\,(\mathrm{S}(\hat{\mathbf{z}}))$ and clearly, $\hat{f} \in \mathrm{R}(\mathbf{y}_0)$. It suffices to represent $\hat{f}$ as in (34). According to Proposition 37, we select $\hat{\nu} \in \mathcal{V}$ satisfying (18) with $\mathbf{y} := \hat{\mathbf{z}}$. Then Theorem 10 guarantees that $\hat{f}$, as an element of $\mathrm{ext}\,(\mathrm{S}(\hat{\mathbf{z}}))$, can be represented as in (34) for some positive integer $M \leq \mathrm{rank}(\mathbf{L}_{\hat{\nu}})$, $x'_j \in X'_{\hat{\nu}}$, $j \in \mathbb{N}_M$, and $\hat{\alpha}_j \neq 0$, $j \in \mathbb{N}_M$, with $\sum_{j\in\mathbb{N}_M} |\hat{\alpha}_j| = \|\hat{\nu}\|_{\mathcal{B}_*}/C$.

We next verify Statement 2. Suppose that $\hat{f}$ is a nonzero extreme point of $\mathrm{R}(\mathbf{y}_0)$ and set $\hat{\mathbf{z}} := \mathcal{L}(\hat{f})$. We will show that $\hat{\mathbf{z}} \neq \mathbf{0}$. It follows from Lemma 11 that $\hat{f} \in \mathrm{ext}\,(\mathrm{S}(\hat{\mathbf{z}}))$. If $\hat{\mathbf{z}} = \mathbf{0}$, we obtain that $\mathrm{S}(\hat{\mathbf{z}}) = \{0\}$ and thus $\hat{f} = 0$. This is a contradiction. Again, we choose $\hat{\nu} \in \mathcal{V}$ satisfying (18) with $\mathbf{y} := \hat{\mathbf{z}}$ by Proposition 37. Theorem 10 enables us to represent $\hat{f} \in \mathrm{ext}\,(\mathrm{S}(\hat{\mathbf{z}}))$ as in (34) for some positive integer $M \leq \mathrm{rank}(\mathbf{L}_{\hat{\nu}})$, $x'_j \in X'_{\hat{\nu}}$, $j \in \mathbb{N}_M$, and $\hat{\alpha}_j \neq 0$, $j \in \mathbb{N}_M$, with $\sum_{j\in\mathbb{N}_M} |\hat{\alpha}_j| = \|\hat{\nu}\|_{\mathcal{B}_*}/C$. ∎

The regularization parameters play an important role in promoting the sparsity of the regularized solutions. Based upon the sparse representer theorem 12, we reveal in the following how the regularization parameter can further promote the sparsity level of the solution. For this purpose, we convert the regularization problem (14) to a finite dimensional regularization problem with the ℓ_1 norm. For given $\mathbf{y} \in \mathbb{R}^n$ and $\lambda > 0$, we choose $\hat{\mathbf{z}} \in \mathcal{D}_{\lambda,\mathbf{y}} \backslash \{\mathbf{0}\}$ and $\hat{\nu} \in \mathcal{V}$ satisfying (18) with $\mathbf{y} := \hat{\mathbf{z}}$. We then introduce the regularization problem in $\mathbb{R}^{n(\hat{\nu})}$ by

$$\inf\left\{\mathcal{Q}_{\mathbf{y}}(\mathbf{L}_{\hat{\nu}}\boldsymbol{\alpha}) + \lambda\varphi\left(C\|\boldsymbol{\alpha}\|_1\right) : \boldsymbol{\alpha} \in \mathbb{R}^{n(\hat{\nu})}\right\}. \tag{35}$$

Denote by $\mathrm{R}_{\hat{\nu}}(\mathbf{y})$ the solution set of problem (35). We show in the following lemma the relation between the solutions of the regularization problems (14) and (35).

Lemma 13 *Suppose that $\mathbf{y}_0 \in \mathbb{R}^n$ and $\lambda > 0$. Let $\mathcal{V}$ be defined by* (9)*, $\mathcal{D}_{\lambda,\mathbf{y}_0}$ be defined by* (28) *with $\mathbf{y} := \mathbf{y}_0$, and let $\hat{\mathbf{z}} \in \mathcal{D}_{\lambda,\mathbf{y}_0}\backslash\{\mathbf{0}\}$, $\hat{\nu} \in \mathcal{V}$ satisfy* (18) *with $\mathbf{y} := \hat{\mathbf{z}}$. If Assumptions (A1) and (A2) hold, $X'_{\hat{\nu}}$ and $\mathbf{L}_{\hat{\nu}}$ are defined by* (20) *and* (21)*, respectively, then $\hat{f} := \sum_{j\in\mathbb{N}_{n(\hat{\nu})}} \hat{\alpha}_j K(\cdot, x'_j) \in \mathrm{R}(\mathbf{y}_0)$ if and only if $\hat{\boldsymbol{\alpha}} := [\hat{\alpha}_j : j \in \mathbb{N}_{n(\hat{\nu})}] \in \mathrm{R}_{\hat{\nu}}(\mathbf{y}_0)$.*

Proof We suppose that $\hat{f} := \sum_{j\in\mathbb{N}_{n(\hat{\nu})}} \hat{\alpha}_j K(\cdot, x'_j) \in \mathrm{R}(\mathbf{y}_0)$ and proceed to prove that $\hat{\boldsymbol{\alpha}} := [\hat{\alpha}_j : j \in \mathbb{N}_{n(\hat{\nu})}] \in \mathrm{R}_{\hat{\nu}}(\mathbf{y}_0)$. It suffices to show that

$$\mathcal{Q}_{\mathbf{y}_0}(\mathbf{L}_{\hat{\nu}}\hat{\boldsymbol{\alpha}}) + \lambda\varphi(C\|\hat{\boldsymbol{\alpha}}\|_1) \leq \mathcal{Q}_{\mathbf{y}_0}(\mathbf{L}_{\hat{\nu}}\boldsymbol{\alpha}) + \lambda\varphi\left(C\|\boldsymbol{\alpha}\|_1\right), \tag{36}$$

for any $\boldsymbol{\alpha} \in \mathbb{R}^{n(\hat{\nu})}$. Let $\boldsymbol{\alpha} := [\alpha_j : j \in \mathbb{N}_{n(\hat{\nu})}] \in \mathbb{R}^{n(\hat{\nu})}$ and set $f := \sum_{j\in\mathbb{N}_{n(\hat{\nu})}} \alpha_j K(\cdot, x'_j)$. Since $\hat{f} \in \mathrm{R}(\mathbf{y}_0)$, we get that

$$\mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(\hat{f})) + \lambda\varphi(\|\hat{f}\|_{\mathcal{B}}) \leq \mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(f)) + \lambda\varphi\left(\|f\|_{\mathcal{B}}\right). \tag{37}$$

Substituting equations (23) and (24) with the pair $f, \boldsymbol{\alpha}$ and the same equations with the pair $\hat{f}, \hat{\boldsymbol{\alpha}}$ into inequality (37), we get the desired inequality (36). Conversely, suppose that $\hat{\boldsymbol{\alpha}} := [\hat{\alpha}_j : j \in \mathbb{N}_{n(\hat{\nu})}] \in \mathrm{R}_{\hat{\nu}}(\mathbf{y}_0)$ and set $\hat{f} := \sum_{j\in\mathbb{N}_{n(\hat{\nu})}} \hat{\alpha}_j K(\cdot, x'_j)$. Noting that $\mathcal{D}_{\lambda,\mathbf{y}_0} \neq \{\mathbf{0}\}$, statement 1 of Theorem 12 ensures that there exists $\widetilde{\boldsymbol{\alpha}} := [\widetilde{\alpha}_j : j \in \mathbb{N}_{n(\hat{\nu})}] \in \mathbb{R}^{n(\hat{\nu})}$ such that $\widetilde{f} := \sum_{j\in\mathbb{N}_{n(\hat{\nu})}} \widetilde{\alpha}_j K(\cdot, x'_j) \in \mathrm{R}(\mathbf{y}_0)$. Since $\hat{\boldsymbol{\alpha}} \in \mathrm{R}_{\hat{\nu}}(\mathbf{y}_0)$, inequality (36) holds with $\boldsymbol{\alpha} := \widetilde{\boldsymbol{\alpha}}$. Again substituting equations (23) and (24) with the pair $\widetilde{f}, \widetilde{\boldsymbol{\alpha}}$ and the same equations with the pair $\hat{f}, \hat{\boldsymbol{\alpha}}$ into inequality (36) with $\boldsymbol{\alpha} := \widetilde{\boldsymbol{\alpha}}$, we obtain inequality (37) with $f := \widetilde{f}$. This together with $\widetilde{f} \in \mathrm{R}(\mathbf{y}_0)$ leads to $\hat{f} \in \mathrm{R}(\mathbf{y}_0)$. ■

The role of the regularization parameter on the sparsity of the solutions of the finite dimensional regularization problem with the ℓ_1 norm has been studied in Liu et al. (2023). By similar arguments in Liu et al. (2023), we present a sparsity characterization of the solutions of problem (35) as follows. For each $j \in \mathbb{N}_{n(\hat{\nu})}$, we denote by $\mathbf{e}_j$ the unit vector with 1 for the jth component and 0 otherwise. Using these vectors, we define $n(\hat{\nu})$ numbers of subsets of $\mathbb{R}^{n(\hat{\nu})}$ by

$$\Omega_l := \left\{ \sum_{j\in\mathbb{N}_l} \alpha_{k_j}\mathbf{e}_{k_j} : \alpha_{k_j} \in \mathbb{R}\setminus\{0\},\ \text{for } 1 \le k_1 < k_2 < \cdots < k_l \le n(\hat{\nu}) \right\},\ \text{for all } l \in \mathbb{N}_{n(\hat{\nu})}.$$

Proposition 14 *Suppose that $\mathbf{y}_0 \in \mathbb{R}^n$, $\lambda > 0$, both $\mathcal{Q}_{\mathbf{y}_0} : \mathbb{R}^n \to \mathbb{R}_+$ and $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ are convex, and moreover, φ is differentiable and strictly increasing. Let $\mathcal{V}$ be defined by* (9)*, $\mathcal{D}_{\lambda,\mathbf{y}_0}$ be defined by* (28) *with $\mathbf{y} := \mathbf{y}_0$, and let $\hat{\mathbf{z}} \in \mathcal{D}_{\lambda,\mathbf{y}_0}\setminus\{\mathbf{0}\}$, $\hat{\nu} \in \mathcal{V}$ satisfy* (18) *with $\mathbf{y} := \hat{\mathbf{z}}$. If Assumptions (A1) and (A2) hold with a positive constant C and $\mathbf{L}_{\hat{\nu}}$ be defined by* (21)*, then problem* (35) *with $\mathbf{y} := \mathbf{y}_0$ has a solution $\hat{\boldsymbol{\alpha}} = \sum_{i\in\mathbb{N}_l} \hat{\alpha}_{k_i}\mathbf{e}_{k_i} \in \Omega_l$ for some $l \in \mathbb{N}_{n(\hat{\nu})}$ if and only if there exists $\mathbf{a} \in \partial\mathcal{Q}_{\mathbf{y}_0}(\mathbf{L}_{\hat{\nu}}\hat{\boldsymbol{\alpha}})$ such that*

$$\lambda = -(\mathbf{L}_{\hat{\nu}}^\top\mathbf{a})_{k_i}\mathrm{sign}(\hat{\alpha}_{k_i})/(C\varphi'(C\|\hat{\boldsymbol{\alpha}}\|_1)),\quad i \in \mathbb{N}_l, \tag{38}$$

$$\lambda \ge |(\mathbf{L}_{\hat{\nu}}^\top\mathbf{a})_j|/(C\varphi'(C\|\hat{\boldsymbol{\alpha}}\|_1)),\ j \in \mathbb{N}_{n(\hat{\nu})}\setminus\{k_i : i \in \mathbb{N}_l\}. \tag{39}$$

Proof Due to the convexity of $\mathcal{Q}_{\mathbf{y}_0}$ and the linearity of $\mathbf{L}_{\hat{\nu}}$, the fidelity term $\mathcal{Q}_{\mathbf{y}_0} \circ \mathbf{L}_{\hat{\nu}}$ is convex. Moreover, since φ is increasing, convex and the norm function $\|\cdot\|_1$ is convex, we claim that the regularization term $\varphi(C\|\cdot\|_1)$ is also convex. By using the Fermat rule (Zălinescu (2002)) and the continuity of $\varphi(C\|\cdot\|_1)$, we conclude that $\hat{\boldsymbol{\alpha}}$ is a solution of problem (35) with $\mathbf{y} := \mathbf{y}_0$ if and only if

$$\mathbf{0} \in \partial(\mathcal{Q}_{\mathbf{y}_0} \circ \mathbf{L}_{\hat{\nu}})(\hat{\boldsymbol{\alpha}}) + \lambda\partial(\varphi(C\|\cdot\|_1))(\hat{\boldsymbol{\alpha}}). \tag{40}$$

According to the chain rule of the subdifferential and the differentiability of φ, inclusion (40) can be rewritten as

$$\mathbf{0} \in \mathbf{L}_{\hat{\nu}}^\top\partial\mathcal{Q}_{\mathbf{y}_0}(\mathbf{L}_{\hat{\nu}}\hat{\boldsymbol{\alpha}}) + \lambda C\varphi'(C\|\hat{\boldsymbol{\alpha}}\|_1)\partial\|\cdot\|_1(\hat{\boldsymbol{\alpha}}).$$

Equivalently, there exists $\mathbf{a} \in \partial\mathcal{Q}_{\mathbf{y}_0}(\mathbf{L}_{\hat{\nu}}\hat{\boldsymbol{\alpha}})$ such that

$$-\mathbf{L}_{\hat{\nu}}^\top\mathbf{a} \in \lambda C\varphi'(C\|\hat{\boldsymbol{\alpha}}\|_1)\partial\|\cdot\|_1(\hat{\boldsymbol{\alpha}}). \tag{41}$$

Noting that $\hat{\boldsymbol{\alpha}} = \sum_{i\in\mathbb{N}_l} \hat{\alpha}_{k_i}\mathbf{e}_{k_i} \in \Omega_l$ with $\hat{\alpha}_{k_i} \in \mathbb{R}\setminus\{0\}$, $i\in\mathbb{N}_l$, we obtain that

$$\partial\|\cdot\|_1(\hat{\boldsymbol{\alpha}}) = \left\{\mathbf{z}\in\mathbb{R}^{n(\hat{\nu})} : z_{k_i} = \operatorname{sign}(\hat{\alpha}_{k_i}), i\in\mathbb{N}_l \text{ and } |z_j|\leq 1, j\in\mathbb{N}_{n(\hat{\nu})}\setminus\{k_i : i\in\mathbb{N}_l\}\right\}.$$

By using the above equation and noting that $\varphi'(t) > 0$ for all $t\in(0,+\infty)$, we rewrite inclusion (41) as (38) and (39). This completes the proof of this proposition. ■

Combining Lemma 13 with Proposition 14, we are ready to obtain choices of the regularization parameter for sparse solutions of the regularization problem (14).

Theorem 15 *Suppose that $\mathbf{y}_0\in\mathbb{R}^n$, $\lambda>0$, both $\mathcal{Q}_{\mathbf{y}_0}:\mathbb{R}^n\to\mathbb{R}_+$ and $\varphi:\mathbb{R}_+\to\mathbb{R}_+$ are convex, and moreover, φ is differentiable and strictly increasing. Let $\mathcal{V}$ be defined by* (9)*, $\mathcal{D}_{\lambda,\mathbf{y}_0}$ be defined by* (28) *with $\mathbf{y}:=\mathbf{y}_0$, and let $\hat{\mathbf{z}}\in\mathcal{D}_{\lambda,\mathbf{y}_0}\backslash\{\mathbf{0}\}$, $\hat{\nu}\in\mathcal{V}$ satisfy* (18) *with $\mathbf{y}:=\hat{\mathbf{z}}$. If Assumptions (A1) and (A2) hold, $X'_{\hat{\nu}}$ and $\mathbf{L}_{\hat{\nu}}$ are defined by* (20) *and* (21)*, respectively, then problem* (14) *with $\mathbf{y}:=\mathbf{y}_0$ has a solution $\hat{f}=\sum_{i\in\mathbb{N}_l}\hat{\alpha}_{k_i}K(\cdot,x'_{k_i})$ with $\hat{\alpha}_{k_i}\in\mathbb{R}\backslash\{0\}$, $x'_{k_i}\in X'_{\hat{\nu}}$, $i\in\mathbb{N}_l$ for some $l\in\mathbb{N}_{n(\hat{\nu})}$ if and only if there exists $\mathbf{a}\in\partial\mathcal{Q}_{\mathbf{y}_0}(\mathbf{L}_{\hat{\nu}}\hat{\boldsymbol{\alpha}})$ with $\hat{\boldsymbol{\alpha}}:=\sum_{i\in\mathbb{N}_l}\hat{\alpha}_{k_i}\mathbf{e}_{k_i}$ such that* (38) *and* (39) *hold.*

Proof It follows from Lemma 13 that $\hat{f}:=\sum_{i\in\mathbb{N}_l}\hat{\alpha}_{k_i}K(\cdot,x'_{k_i})$ with $\hat{\alpha}_{k_i}\in\mathbb{R}\backslash\{0\}$, $x'_{k_i}\in X'_{\hat{\nu}}$, $i\in\mathbb{N}_l$, is a solution of problem (14) with $\mathbf{y}:=\mathbf{y}_0$ if and only if $\hat{\boldsymbol{\alpha}}:=\sum_{i\in\mathbb{N}_l}\hat{\alpha}_{k_i}\mathbf{e}_{k_i}\in\Omega_l$ is a solution of problem (35) with $\mathbf{y}:=\mathbf{y}_0$. Proposition 14 ensures that the latter is equivalent to that there exists $\mathbf{a}\in\partial\mathcal{Q}_{\mathbf{y}_0}(\mathbf{L}_{\hat{\nu}}\hat{\boldsymbol{\alpha}})$ such that (38) and (39) hold. This completes the proof of this theorem. ■

Observing from Theorem 15, the choice of the regularization parameter can influence the sparsity of the solution. Specifically, the equalities in (38) and the inequalities in (39) correspond to the nonzero components and the zero components of the solution, respectively. As the number of the inequalities increases, the solution becomes more sparse. Such a characterization of the sparsity of the solutions also benefits from Assumptions (A1) and (A2) satisfied by the RKBS.

5. Sparse Learning in $\ell_1(\mathbb{N})$

In this section, we consider the MNI problem and the regularization problem in the sequence space $\ell_1(\mathbb{N})$ which is a typical RKBS. We first verify that $\ell_1(\mathbb{N})$ satisfies Assumptions (A1) and (A2). We then specialize Theorems 10 and 12 to the RKBS $\ell_1(\mathbb{N})$ and establish sparse representer theorems for the solutions of the MNI problem and the regularization problem in this space. For the regularization problem in $\ell_1(\mathbb{N})$, we further study the influence of the regularization parameter on the sparsity of the solutions. Finally, we show that unlike $\ell_1(\mathbb{N})$, the RKBSs $\ell_p(\mathbb{N})$, for all $1<p<+\infty$, cannot promote sparsity of a learning solution in them.

We first recall the sequence space $\ell_1(\mathbb{N})$. The Banach space $\ell_1(\mathbb{N})$ consists of all real sequences $\mathbf{x}:=[x_j : j\in\mathbb{N}]$ such that $\|\mathbf{x}\|_1:=\sum_{j\in\mathbb{N}}|x_j|<+\infty$. It is known that $\ell_1(\mathbb{N})$ has $c_0(\mathbb{N})$ as its pre-dual space, where $c_0(\mathbb{N})$ denotes the space of all real sequences $\mathbf{v}:=$

$[v_j : j \in \mathbb{N}]$ converging to 0 as $j \to \infty$, endowed with $\|\mathbf{v}\|_\infty := \sup\{|v_j| : j \in \mathbb{N}\} < +\infty$. The dual bilinear form $\langle \cdot, \cdot \rangle_{\ell_1(\mathbb{N})}$ on $c_0(\mathbb{N}) \times \ell_1(\mathbb{N})$ is defined by $\langle \mathbf{v}, \mathbf{x} \rangle_{\ell_1(\mathbb{N})} := \sum_{j \in \mathbb{N}} v_j x_j$, for all $\mathbf{v} := [v_j : j \in \mathbb{N}] \in c_0(\mathbb{N})$ and all $\mathbf{x} := [x_j : j \in \mathbb{N}] \in \ell_1(\mathbb{N})$. It has been established in Xu (2023) that $\ell_1(\mathbb{N})$ is an RKBS composed of functions defined on $\mathbb{N}$ and its δ-dual is isometrically isomorphic to $c_0(\mathbb{N})$ which is also the adjoint RKBS of $\ell_1(\mathbb{N})$. That is, $\ell_1(\mathbb{N})$ and $c_0(\mathbb{N})$ are a pair of RKBSs. Moreover, the function $K : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ defined by

$$K(i,j) := \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \quad \text{for any } (i,j) \in \mathbb{N} \times \mathbb{N}, \tag{42}$$

is the reproducing kernel for $\ell_1(\mathbb{N})$.

We next describe the MNI problem and the regularization problem in $\ell_1(\mathbb{N})$. We suppose that $\mathbf{v}_i := [v_{i,j} : j \in \mathbb{N}]$, $i \in \mathbb{N}_n$, are a finite number of linearly independent elements in $c_0(\mathbb{N})$ and set

$$\mathcal{V} := \operatorname{span}\{\mathbf{v}_j : j \in \mathbb{N}_n\}. \tag{43}$$

The operator $\mathcal{L} : \ell_1(\mathbb{N}) \to \mathbb{R}^n$ with the form (10) may be taken as a semi-infinite matrix

$$\mathbf{V} := [v_{i,j} : i \in \mathbb{N}_n, j \in \mathbb{N}]. \tag{44}$$

For a given vector $\mathbf{y} \in \mathbb{R}^n$, the hyperplane $\mathcal{M}_\mathbf{y}$, defined by (12), has the form

$$\mathcal{M}_\mathbf{y} := \{\mathbf{x} \in \ell_1(\mathbb{N}) : \mathbf{V}\mathbf{x} = \mathbf{y}\}. \tag{45}$$

With the notation above, the MNI problem in $\ell_1(\mathbb{N})$ is formulated as

$$\inf\left\{\|\mathbf{x}\|_1 : \mathbf{x} \in \mathcal{M}_\mathbf{y}\right\}. \tag{46}$$

By introducing a loss function $\mathcal{Q}_\mathbf{y} : \mathbb{R}^n \to \mathbb{R}_+$ and a regularization parameter $\lambda > 0$, the regularization problem in $\ell_1(\mathbb{N})$ has the form

$$\inf\left\{\mathcal{Q}_\mathbf{y}(\mathbf{V}\mathbf{x}) + \lambda\|\mathbf{x}\|_1 : \mathbf{x} \in \ell_1(\mathbb{N})\right\}. \tag{47}$$

In this section, we still denote by $\mathrm{S}(\mathbf{y})$ and $\mathrm{R}(\mathbf{y})$ the solution sets of the MNI problem (46) and the regularization problem (47), respectively.

We now turn to establishing the sparse representer theorem for the solutions of problems (46) and (47). We begin with verifying that the RKBS $\ell_1(\mathbb{N})$ satisfies Assumptions (A1) and (A2). To this end, we need a result about the subdifferential of the ℓ_∞ norm at any $\mathbf{v} \in c_0(\mathbb{N})$. For each $\mathbf{v} := [v_j : j \in \mathbb{N}] \in c_0(\mathbb{N})$, by $\mathbb{N}(\mathbf{v})$ we denote the index set where the sequence $\mathbf{v}$ achieves its supremum norm $\|\mathbf{v}\|_\infty$, namely,

$$\mathbb{N}(\mathbf{v}) := \left\{j \in \mathbb{N} : |v_j| = \|\mathbf{v}\|_\infty\right\}. \tag{48}$$

Note that the sequence $\mathbf{v} \in c_0(\mathbb{N})$ goes to 0 while j approaches to infinity and hence the index set $\mathbb{N}(\mathbf{v})$ is of finite cardinality. Let $n(\mathbf{v})$ denote the cardinality of $\mathbb{N}(\mathbf{v})$. We also introduce for each $\mathbf{v} := [v_j : j \in \mathbb{N}] \in c_0(\mathbb{N})$ a subset of $\ell_1(\mathbb{N})$ as

$$\Omega(\mathbf{v}) := \left\{\operatorname{sign}(v_j) K(\cdot, j) : j \in \mathbb{N}(\mathbf{v})\right\}. \tag{49}$$

As has been shown in Cheng and Xu (2021), it holds for any $\mathbf{v} \in c_0(\mathbb{N})\backslash\{\mathbf{0}\}$ that

$$\partial\|\cdot\|_\infty(\mathbf{v}) = \mathrm{co}(\Omega(\mathbf{v})). \tag{50}$$

This together with noting that $\Omega(\mathbf{v})$ is a finite set further leads to

$$\mathrm{ext}\,(\partial\|\cdot\|_\infty(\mathbf{v})) = \Omega(\mathbf{v}). \tag{51}$$

The next lemma shows that the RKBS $\ell_1(\mathbb{N})$ satisfies Assumptions (A1) and (A2).

Lemma 16 *If $\mathbf{v}_j, j \in \mathbb{N}_n$, are linearly independent elements in $c_0(\mathbb{N})$ and $\mathcal{V}$ is defined by* (43)*, then the space $\ell_1(\mathbb{N})$ satisfies Assumption (A1) with $X'_{\mathbf{v}} := \mathbb{N}(\mathbf{v})$ for any $\mathbf{v} \in \mathcal{V}$ and Assumption (A2) with $C := 1$.*

Proof We first show that Assumption (A1) holds. It follows from definition (49) and equation (51) that for any nonzero $\mathbf{v} \in \mathcal{V}$, there holds

$$\mathrm{ext}\,(\partial\|\cdot\|_\infty(\mathbf{v})) = \{\mathrm{sign}(v_j)K(\cdot, j) : j \in \mathbb{N}(\mathbf{v})\}.$$

That is, for any nonzero $\mathbf{v} \in \mathcal{V}$, there exists a finite subset $X'_{\mathbf{v}} := \mathbb{N}(\mathbf{v})$ of $X' := \mathbb{N}$ such that

$$\mathrm{ext}\,(\partial\|\cdot\|_\infty(\mathbf{v})) \subset \{-K(\cdot, j), K(\cdot, j) : j \in X'_{\mathbf{v}}\}.$$

Thus, the RKBS $\ell_1(\mathbb{N})$ satisfies Assumption (A1) with $X'_{\mathbf{v}} := \mathbb{N}(\mathbf{v})$ for any $\mathbf{v} \in \mathcal{V}$.

We next verify that Assumption (A2) holds. Note that for each $j \in \mathbb{N}$, the kernel session $K(\cdot, j)$ coincides with the vector $\mathbf{e}_j$. As a result, for any $m \in \mathbb{N}$, distinct points $l_j \in \mathbb{N}$, $j \in \mathbb{N}_m$, and $\boldsymbol{\alpha} = [\alpha_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$, there holds

$$\left\|\sum_{j\in\mathbb{N}_m} \alpha_j K(\cdot, l_j)\right\|_1 = \|\boldsymbol{\alpha}\|_1.$$

Clearly, the RKBS $\ell_1(\mathbb{N})$ satisfies Assumption (A2) with $C := 1$. ■

We are ready to specialize the sparse representer theorem 10 to the MNI problem (46). For each $\mathbf{v} := [v_j : j \in \mathbb{N}] \in c_0(\mathbb{N})$, we define a truncation matrix of $\mathbf{V}$ as follows. Suppose that $\mathbb{N}(\mathbf{v}) = \{k_j \in \mathbb{N} : j \in \mathbb{N}_{n(\mathbf{v})}\}$. We truncate the semi-infinite matrix $\mathbf{V}$ with the form (44) by throwing away the columns with index not appearing in $\mathbb{N}(\mathbf{v})$. Specifically, we define the truncation matrix

$$\mathbf{V}_{\mathbf{v}} := [v_{i,k_j} : i \in \mathbb{N}_n, j \in \mathbb{N}_{n(\mathbf{v})}] \in \mathbb{R}^{n\times n(\mathbf{v})}. \tag{52}$$

Theorem 17 *Suppose that $\mathbf{v}_j$, $j \in \mathbb{N}_n$, are linearly independent elements in $c_0(\mathbb{N})$ and $\mathbf{y} \in \mathbb{R}^n\backslash\{\mathbf{0}\}$. Let $\mathcal{V}$ and $\mathcal{M}_{\mathbf{y}}$ be defined by* (43) *and* (45)*, respectively. If $\hat{\mathbf{v}} \in \mathcal{V}$ satisfies*

$$(\|\hat{\mathbf{v}}\|_\infty \mathrm{co}(\Omega(\hat{\mathbf{v}}))) \cap \mathcal{M}_{\mathbf{y}} \neq \emptyset, \tag{53}$$

$\mathbb{N}(\hat{\mathbf{v}})$ and $\mathbf{V}_{\hat{\mathbf{v}}}$ are defined by (48) *and* (52) *with $\mathbf{v} := \hat{\mathbf{v}}$, respectively, then for any $\hat{\mathbf{x}} \in \mathrm{ext}\,(\mathrm{S}(\mathbf{y}))$, there exist $\hat{\alpha}_j \neq 0$, $j \in \mathbb{N}_M$, with $\sum_{j\in\mathbb{N}_M} |\hat{\alpha}_j| = \|\hat{\mathbf{v}}\|_\infty$ and $k_j \in \mathbb{N}(\hat{\mathbf{v}})$, $j \in \mathbb{N}_M$, such that*

$$\hat{\mathbf{x}} = \sum_{j\in\mathbb{N}_M} \hat{\alpha}_j K(\cdot, k_j). \tag{54}$$

for some positive integer $M \leq \mathrm{rank}(\mathbf{V}_{\hat{\mathbf{v}}})$.

Proof We prove this result by specializing Theorem 10 to the MNI problem in $\ell_1(\mathbb{N})$. We first note that $\ell_1(\mathbb{N})$ and $c_0(\mathbb{N})$ are a pair of RKBSs and K defined by (42) is the reproducing kernel for $\ell_1(\mathbb{N})$. Moreover, $\ell_1(\mathbb{N})$ has $c_0(\mathbb{N})$ as its pre-dual space. We next show that $\hat{\mathbf{v}}$ satisfies (18). Substituting equation (50) with $\mathbf{v}$ being replaced by $\hat{\mathbf{v}}$ into assumption (53) leads directly to $(\|\hat{\mathbf{v}}\|_\infty \partial\|\cdot\|_\infty(\hat{\mathbf{v}})) \cap \mathcal{M}_{\mathbf{y}} \neq \emptyset$. That is, $\hat{\mathbf{v}}$ satisfies (18). Finally, Lemma 16 guarantees that both Assumption (A1) holds with $X'_{\mathbf{v}} := \mathbb{N}(\mathbf{v})$ for any $\mathbf{v} \in \mathcal{V}$ and Assumption (A2) holds with $C := 1$. Consequently, the hypotheses of Theorem 10 are satisfied. By Theorem 10 with noting that $X'_{\hat{\mathbf{v}}} := \mathbb{N}(\hat{\mathbf{v}})$, $C := 1$ and the matrix $\mathbf{L}_{\hat{\mathbf{v}}}$, defined by (21), coincides exactly with the truncation matrix $\mathbf{V}_{\hat{\mathbf{v}}}$, any extreme point $\hat{\mathbf{x}}$ of $\mathrm{S}(\mathbf{y})$ can be expressed as in equation (54). ■

Since the kernel session $K(\cdot, k_j)$ appearing in equation (54) has merely one nonzero entry at k_j for each $j \in \mathbb{N}_M$, Theorem 17 shows that any extreme point of the solution set $\mathrm{S}(\mathbf{y})$ of the MNI problem (46) has at most $\mathrm{rank}(\mathbf{V}_{\hat{\mathbf{v}}})$ nonzero components. Obviously, $\mathrm{rank}(\mathbf{V}_{\hat{\mathbf{v}}}) \leq n$. According to Proposition 37 in Appendix B, the sequence $\hat{\mathbf{v}} \in \mathcal{V}$ satisfying (53) can be obtained by solving the dual problem (113) with $\mathcal{B}_* := c_0(\mathbb{N})$ and $\nu_j := \mathbf{v}_j$, $j \in \mathbb{N}_n$. It has been proved in Cheng and Xu (2021) that the resulting dual problem, as a finite dimensional optimization problem, may be solved by linear programming. Admittedly, the solutions of the dual problem may not be unique and the quantity $\mathrm{rank}(\mathbf{V}_{\hat{\mathbf{v}}})$ hinges on the choice of $\hat{\mathbf{v}} \in \mathcal{V}$, or the choice of the solution of the corresponding dual problem. We further remark that a data-independent representer theorem for MNI problem (46), established in Unser et al. (2016), expresses the extreme point $\hat{\mathbf{x}}$ of the solution set in terms of a linear combination of n extreme points of the unit ball in $\ell_1(\mathbb{N})$, which is data-independent. Theorem 17 differs from the data-independent representer theorem of Unser et al. (2016) in the kernel representation (54), where k_j, $j \in \mathbb{N}_M$, depend on the element $\hat{\mathbf{v}}$, which is a linear combination of the given data $\mathbf{v}_j$, $j \in \mathbb{N}_n$. Moreover, it follows from definition (48) that the set $\mathbb{N}(\hat{\mathbf{v}})$ has a finite cardinality while there are infinitely many extreme points of the unit ball in $\ell_1(\mathbb{N})$.

Below, we present a specific example to demonstrate that the quantity $\mathrm{rank}(\mathbf{V}_{\hat{\mathbf{v}}})$ could be equal or strictly less than n depending on the distinct solution of the dual problem. We choose the functionals $\mathbf{v}_1, \mathbf{v}_2 \in c_0(\mathbb{N})$ as

$$\mathbf{v}_1 := \left[\frac{1}{n} : n \in \mathbb{N}\right], \ \mathbf{v}_2 := \left[\frac{1}{(-2)^{n-1}} : n \in \mathbb{N}\right],$$

and consider the MNI problem (46) with $\mathbf{y} := [1, 1]^\top$. In this case, the solution set of the dual problem (113) can be characterized as

$$\left\{\hat{\mathbf{c}} := [\hat{c}_1, \hat{c}_2]^\top \in \mathbb{R}^2 : \hat{c}_1 + \hat{c}_2 = 1, -\frac{1}{2} \leq \hat{c}_1 \leq \frac{3}{2}\right\},$$

and the optimal value is $m_0 = 1$. We choose the vector $\hat{\mathbf{v}}$ satisfying (53) according to two distinct elements in the above solution set. We first select $\hat{\mathbf{c}}_1 := [-\frac{1}{2}, \frac{3}{2}]^\top$ as a solution of the dual problem. Proposition 37 in Appendix B ensures that $\hat{\mathbf{v}}_1 = -\frac{1}{2}\mathbf{v}_1 + \frac{3}{2}\mathbf{v}_2$ satisfies (53). Clearly, $\mathbb{N}(\hat{\mathbf{v}}_1) = \{1, 2\}$ and $n(\hat{\mathbf{v}}_1) = 2$. It follows that the matrix $\mathbf{V}_{\hat{\mathbf{v}}_1}$ has the form

$$\mathbf{V}_{\hat{\mathbf{v}}_1} := \begin{bmatrix} 1 & \frac{1}{2} \\ 1 & -\frac{1}{2} \end{bmatrix},$$

and $\mathrm{rank}(\mathbf{V}_{\hat{\mathbf{v}}_1}) = 2$. We next select an alternative solution of the dual problem, that is, $\hat{\mathbf{c}}_2 := [0,1]^\top$. Accordingly, $\hat{\mathbf{v}}_2 = \mathbf{v}_2$ satisfies (53). It follows from $\mathbb{N}(\hat{\mathbf{v}}_2) = \{1\}$, $n(\hat{\mathbf{v}}_2) = 1$ that $\mathbf{V}_{\hat{\mathbf{v}}_2} = [1,1]^\top$ and $\mathrm{rank}(\mathbf{V}_{\hat{\mathbf{v}}_2}) = 1$. Notice that $\mathrm{rank}(\mathbf{V}_{\hat{\mathbf{v}}_1})$ is equal to the number of the given data points while $\mathrm{rank}(\mathbf{V}_{\hat{\mathbf{v}}_2})$ is strictly less than the number. In this example, the MNI problem (46) with $\mathbf{y} := [1,1]^\top$ has a unique solution $\hat{\mathbf{x}} = [1,0,0,\ldots] \in \ell_1(\mathbb{N})$ with sparsity level $l = 1$. Clearly, the quantity $\mathrm{rank}(\mathbf{V}_{\hat{\mathbf{v}}_1})$ provides a precise characterization for the sparsity of the solution.

By applying Theorem 12 to the regularization problem (47), we get the following sparse representer theorem for the solutions.

Theorem 18 *Suppose that $\mathbf{v}_j, j \in \mathbb{N}_n$, are linearly independent elements in $c_0(\mathbb{N})$, $\mathbf{y}_0 \in \mathbb{R}^n$ and $\lambda > 0$. Let $\mathcal{V}$ be defined by (43). If $\mathcal{Q}_{\mathbf{y}_0}$ is lower semi-continuous and convex, then every nonzero $\hat{\mathbf{x}} \in \mathrm{ext}\,(\mathrm{R}(\mathbf{y}_0))$ has the form (54) for some $\hat{\mathbf{v}} \in \mathcal{V}$, positive integer $M \leq \mathrm{rank}(\mathbf{V}_{\hat{\mathbf{v}}})$, $\alpha_j \neq 0$, $j \in \mathbb{N}_M$, with $\sum_{j\in\mathbb{N}_M} |\alpha_j| = \|\hat{\mathbf{v}}\|_\infty$ and $k_j \in \mathbb{N}(\hat{\mathbf{v}})$, $j \in \mathbb{N}_M$.*

Proof Again, we point out that $\ell_1(\mathbb{N})$ and $c_0(\mathbb{N})$ are a pair of RKBSs, K defined by (42) is the reproducing kernel for $\ell_1(\mathbb{N})$ and in addition, $\ell_1(\mathbb{N})$ has $c_0(\mathbb{N})$ as its pre-dual space. The space $\ell_1(\mathbb{N})$, guaranteed by Lemma 16, satisfies Assumption (A1) with $X'_{\mathbf{v}} := \mathbb{N}(\mathbf{v})$ for any $\mathbf{v} \in \mathcal{V}$ and Assumption (A2) with $C := 1$. In this case, the function $\varphi(t) := t$, $t \in \mathbb{R}_+$, is continuous, convex and strictly increasing. That is, the hypotheses of Theorem 12 are all satisfied. Hence, by Theorem 12 with $X'_{\hat{\mathbf{v}}} := \mathbb{N}(\hat{\mathbf{v}})$, $C := 1$ and $\mathbf{L}_{\hat{\mathbf{v}}} := \mathbf{V}_{\hat{\mathbf{v}}}$, any nonzero extreme point $\hat{\mathbf{x}}$ of $\mathrm{R}(\mathbf{y}_0)$ has the form (54) for some $\hat{\mathbf{v}} \in \mathcal{V}$, positive integer $M \leq \mathrm{rank}(\mathbf{V}_{\hat{\mathbf{v}}})$, $\alpha_j \neq 0$, $j \in \mathbb{N}_M$, with $\sum_{j\in\mathbb{N}_M} |\alpha_j| = \|\hat{\mathbf{v}}\|_\infty$ and $k_j \in \mathbb{N}(\hat{\mathbf{v}})$, $j \in \mathbb{N}_M$. ■

The regularization parameter λ involved in the regularization problem (47) can be used to further promote the sparsity level of the solution. As a consequence of Theorem 15, we get the following choices of the regularization parameter for sparse solutions of problem (47). For each $\mathbf{y} \in \mathbb{R}^n$ and each $\lambda > 0$, the subset $\mathcal{D}_{\lambda,\mathbf{y}}$, defined by (28), takes the form

$$\mathcal{D}_{\lambda,\mathbf{y}} := \{\mathbf{V}\mathbf{x} : \mathbf{x} \in \mathrm{R}(\mathbf{y})\}. \tag{55}$$

Theorem 19 *Suppose that $\mathbf{v}_j, j \in \mathbb{N}_n$, are linearly independent elements in $c_0(\mathbb{N})$, $\mathbf{y}_0 \in \mathbb{R}^n$, $\lambda > 0$ and that $\mathcal{Q}_{\mathbf{y}_0}$ is lower semi-continuous and convex. Let $\mathcal{D}_{\lambda,\mathbf{y}_0}$ be defined by (55) with $\mathbf{y} := \mathbf{y}_0$, $\hat{\mathbf{z}} \in \mathcal{D}_{\lambda,\mathbf{y}_0}\backslash\{\mathbf{0}\}$ and let $\mathcal{V}$ be defined by (43), $\hat{\mathbf{v}} \in \mathcal{V}$ satisfy (53) with $\mathbf{y} := \hat{\mathbf{z}}$ and $\mathbb{N}(\hat{\mathbf{v}})$, $\mathbf{V}_{\hat{\mathbf{v}}}$ be defined by (48) and (52) with $\mathbf{v} := \hat{\mathbf{v}}$, respectively. Then problem (47) with $\mathbf{y} := \mathbf{y}_0$ has a solution $\hat{\mathbf{x}} = \sum_{i\in\mathbb{N}_l} \hat{\alpha}_{k_i} K(\cdot, k_i)$ with $\hat{\alpha}_{k_i} \in \mathbb{R}\setminus\{0\}$, $k_i \in \mathbb{N}(\hat{\mathbf{v}})$, $i \in \mathbb{N}_l$ for some $l \in \mathbb{N}_{n(\hat{\mathbf{v}})}$ if and only if there exists $\mathbf{a} \in \partial\mathcal{Q}_{\mathbf{y}_0}(\mathbf{V}_{\hat{\mathbf{v}}}\hat{\boldsymbol{\alpha}})$ with $\hat{\boldsymbol{\alpha}} := \sum_{i\in\mathbb{N}_l} \hat{\alpha}_{k_i}\mathbf{e}_{k_i}$ such that*

$$\lambda = -(\mathbf{V}_{\hat{\mathbf{v}}}^\top \mathbf{a})_{k_i}\mathrm{sign}(\hat{\alpha}_{k_i}), \quad i \in \mathbb{N}_l, \ \ \text{and } \lambda \geq |(\mathbf{V}_{\hat{\mathbf{v}}}^\top \mathbf{a})_j|, \quad j \in \mathbb{N}_{n(\hat{\mathbf{v}})}\backslash\{k_i : i \in \mathbb{N}_l\}. \tag{56}$$

Proof Lemma 16 ensures that the RKBS $\ell_1(\mathbb{N})$ satisfies Assumption (A1) with $X'_{\mathbf{v}} := \mathbb{N}(\mathbf{v})$ for any $\mathbf{v} \in \mathcal{V}$ and Assumption (A2) with $C := 1$. As has been shown in the proof of Theorem 17, assumption (53) yields that $\hat{\mathbf{v}}$ satisfies (18) with $\mathbf{y} := \hat{\mathbf{z}}$. Note that $\varphi(t) := t$, $t \in \mathbb{R}_+$, is continuous, convex and strictly increasing. The hypotheses of Theorem 15 are satisfied. Thus, by Theorem 15 with noting that $\mathbf{L}_{\hat{\mathbf{v}}} := \mathbf{V}_{\hat{\mathbf{v}}}$ and $X'_{\hat{\mathbf{v}}} := \mathbb{N}(\hat{\mathbf{v}})$, problem (47) with

$\mathbf{y} := \mathbf{y}_0$ has a solution $\hat{\mathbf{x}} = \sum_{i\in\mathbb{N}_l} \hat{\alpha}_{k_i} K(\cdot, k_i)$ with $\hat{\alpha}_{k_i} \in \mathbb{R}\setminus\{0\}$, $k_i \in \mathbb{N}(\hat{\mathbf{v}})$, $i \in \mathbb{N}_l$ for some $l \in \mathbb{N}_{n(\hat{\mathbf{v}})}$ if and only if there exists $\mathbf{a} \in \partial\mathcal{Q}_{\mathbf{y}_0}(\mathbf{V}_{\hat{\mathbf{v}}}\hat{\boldsymbol{\alpha}})$ with $\hat{\boldsymbol{\alpha}} := \sum_{i\in\mathbb{N}_l} \hat{\alpha}_{k_i}\mathbf{e}_{k_i}$ such that

$$\lambda = -(\mathbf{V}_{\hat{\mathbf{v}}}^\top \mathbf{a})_{k_i}\operatorname{sign}(\hat{\alpha}_{k_i})/(\varphi'(C\|\hat{\boldsymbol{\alpha}}\|_1)), \quad i \in \mathbb{N}_l, \tag{57}$$

$$\lambda \geq |(\mathbf{V}_{\hat{\mathbf{v}}}^\top \mathbf{a})_j|/(C\varphi'(C\|\hat{\boldsymbol{\alpha}}\|_1)), \quad j \in \mathbb{N}_{n(\hat{\mathbf{v}})}\setminus\{k_i : i \in \mathbb{N}_l\}. \tag{58}$$

Substituting $\varphi'(t) = 1$, $t \in \mathbb{R}_+$ and $C = 1$ into (38) and (39), we get the desired characterization (56). ■

Theorems 17 and 18 provide sparse representations for the solutions of the MNI problem and the regularization problem in $\ell_1(\mathbb{N})$, respectively. The sparsity of the solutions benefits from the capacity of $\ell_1(\mathbb{N})$ in promoting sparsity, that is, $\ell_1(\mathbb{N})$ satisfies Assumptions (A1) and (A2). To close this section, we show that in general the sequence spaces $\ell_p(\mathbb{N})$, for $1 < p < +\infty$, will not promote sparsity. Recall that $\ell_p(\mathbb{N})$ with $1 < p < +\infty$ is the Banach space of all real sequences $\mathbf{x} := [x_j : j \in \mathbb{N}]$ such that $\|\mathbf{x}\|_p := \left(\sum_{j\in\mathbb{N}} |x_j|^p\right)^{1/p} < +\infty$. It is known that $\ell_p(\mathbb{N})$ is reflexive and then the dual space $\ell_q(\mathbb{N})$ is its pre-dual space, where $1/p + 1/q = 1$. The dual bilinear form $\langle\cdot,\cdot\rangle_{\ell_p(\mathbb{N})}$ on $\ell_q(\mathbb{N}) \times \ell_p(\mathbb{N})$ is defined by $\langle \mathbf{v}, \mathbf{x}\rangle_{\ell_p(\mathbb{N})} := \sum_{j\in\mathbb{N}} v_j x_j$, for all $\mathbf{v} := [v_j : j \in \mathbb{N}] \in \ell_q(\mathbb{N})$ and all $\mathbf{x} := [x_j : j \in \mathbb{N}] \in \ell_p(\mathbb{N})$. Clearly, $\ell_p(\mathbb{N})$, $\ell_q(\mathbb{N})$ are a pair of RKBSs and the function K defined by (42) is also the reproducing kernel of $\ell_p(\mathbb{N})$. We claim that $\ell_p(\mathbb{N})$ with $1 < p < +\infty$ does not satisfy Assumptions (A1) and (A2). Indeed, the subdifferential of the norm $\|\cdot\|_q$ at any $\mathbf{v} := [v_k : k \in \mathbb{N}] \in \ell_q(\mathbb{N})$ is a singleton, that is,

$$\partial\|\cdot\|_q(\mathbf{v}) := \left\{\mathbf{u} := \left[v_k\,|v_k|^{q-2}/\|\mathbf{v}\|_q^{q-1} : k \in \mathbb{N}\right]\right\}. \tag{59}$$

The vector $\mathbf{u}$ usually does not have the form $\alpha K(\cdot, j)$ for some $\alpha \in \{-1, 1\}$ and $j \in \mathbb{N}$ unless $\mathbf{v}$ is in such a form. That is, Assumption (A1) does not hold. For any $m \in \mathbb{N}$, distinct points $l_j \in \mathbb{N}$, $j \in \mathbb{N}_m$, and $\boldsymbol{\alpha} = [\alpha_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$, there holds

$$\left\|\sum_{j\in\mathbb{N}_m} \alpha_j K(\cdot, l_j)\right\|_p = \|\boldsymbol{\alpha}\|_p,$$

which yields that Assumption (A2) is not satisfied either for $\ell_p(\mathbb{N})$ with $1 < p < +\infty$.

We remark that Assumptions (A1) and (A2) are sufficient conditions for an RKBS to enjoy the ability of promoting the sparsity of the learning solutions. Even though the spaces $\ell_p(\mathbb{N})$ for $1 < p < +\infty$ do not satisfy these assumptions, we still need to understand why these RKBSs cannot promote sparsity. To this end, we consider the MNI problem in $\ell_p(\mathbb{N})$ with $1 < p < +\infty$. Suppose that $\mathbf{v}_j := [v_{j,k} : k \in \mathbb{N}]$, $j \in \mathbb{N}_n$, are a finite number of linearly independent elements in $\ell_q(\mathbb{N})$. The operator $\mathcal{L} : \ell_p(\mathbb{N}) \to \mathbb{R}^n$, defined by (10), can also be taken as the semi-infinite matrix $\mathbf{V}$ with the form (44). For a given vector $\mathbf{y} \in \mathbb{R}^n$, the subset $\mathcal{M}_{\mathbf{y}}$ of $\ell_p(\mathbb{N})$, defined by (12), has the form

$$\mathcal{M}_{\mathbf{y}} := \{\mathbf{x} \in \ell_p(\mathbb{N}) : \mathbf{V}\mathbf{x} = \mathbf{y}\}. \tag{60}$$

The MNI problem with $\mathbf{y}$ in $\ell_p(\mathbb{N})$ with $1 < p < +\infty$ is formulated as

$$\inf\left\{\|\mathbf{x}\|_p : \mathbf{x} \in \mathcal{M}_{\mathbf{y}}\right\}, \tag{61}$$

and its dual problem has the form

$$\sup\left\{\sum_{j\in\mathbb{N}_n} c_j y_j : \left\|\sum_{j=1}^{n} c_j \mathbf{v}_j\right\|_q = 1\right\}. \tag{62}$$

Note that problem (61) has a unique solution. By employing Propositions 7 and 37, we represent the unique solution as follows.

Proposition 20 *Let $p, q \in (1, +\infty)$ be such that $1/p + 1/q = 1$. Suppose that $\mathbf{v}_j$, $j \in \mathbb{N}_n$, are linearly independent elements in $\ell_q(\mathbb{N})$ and $\mathbf{y} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$. If $\hat{\mathbf{c}} := [\hat{c}_j : j \in \mathbb{N}_n] \in \mathbb{R}^n$ is the solution of the dual problem (62) and $\hat{\mathbf{v}}_p := (\mathbf{y}^\top \hat{\mathbf{c}}) \sum_{j\in\mathbb{N}_n} \hat{c}_j \mathbf{v}_j$, then the unique solution of problem (61) with $\mathbf{y}$ has the form $\hat{\mathbf{x}}_p := [\hat{x}_k : k \in \mathbb{N}]$ with $\hat{x}_k := (\hat{\mathbf{v}}_p)_k \left|(\hat{\mathbf{v}}_p)_k\right|^{q-2} / \|\hat{\mathbf{v}}_p\|_q^{q-2}$, $k \in \mathbb{N}$.*

Proof Since $\hat{\mathbf{c}} := [\hat{c}_j : j \in \mathbb{N}_n] \in \mathbb{R}^n$ is the solution of the dual problem (62), Proposition 37 ensures that $\hat{\mathbf{v}}_p := (\mathbf{y}^\top \hat{\mathbf{c}}) \sum_{j\in\mathbb{N}_n} \hat{c}_j \mathbf{v}_j$ satisfies (18) with $\mathcal{B}_* = \ell_q(\mathbb{N})$. That is, the hypothesis of Proposition 7 is satisfied. By Proposition 7, the unique solution $\hat{\mathbf{x}}_p := [\hat{x}_k : k \in \mathbb{N}]$ of problem (61) can be represented as

$$\hat{\mathbf{x}}_p = \sum_{j\in\mathbb{N}_n} \gamma_j \mathbf{u}_j, \tag{63}$$

for some $\gamma_j \in \mathbb{R}$, $j \in \mathbb{N}_n$, with $\sum_{j\in\mathbb{N}_n} \gamma_j = \|\hat{\mathbf{v}}_p\|_q$ and $\mathbf{u}_j \in \mathrm{ext}\left(\partial\|\cdot\|_q(\hat{\mathbf{v}}_p)\right)$, $j \in \mathbb{N}_n$. Noting that the set $\partial\|\cdot\|_q(\hat{\mathbf{v}}_p)$ is a singleton, representation (63) reduces to

$$\hat{\mathbf{x}}_p = \|\hat{\mathbf{v}}_p\|_q \mathbf{u}, \tag{64}$$

with $\mathbf{u} := [u_k : k \in \mathbb{N}] \in \partial\|\cdot\|_q(\hat{\mathbf{v}}_p)$. It follows from equation (59) with $\mathbf{v} := \hat{\mathbf{v}}_p$ that $u_k = (\hat{\mathbf{v}}_p)_k \left|(\hat{\mathbf{v}}_p)_k\right|^{q-2} / \|\hat{\mathbf{v}}_p\|_q^{q-1}$, $k \in \mathbb{N}$. Hence, we get for each $k \in \mathbb{N}$ that $\hat{x}_k = (\hat{\mathbf{v}}_p)_k \left|(\hat{\mathbf{v}}_p)_k\right|^{q-2} / \|\hat{\mathbf{v}}_p\|_q^{q-2}$, which completes the proof. ■

Although there is only one term involved in it, representation (64) cannot lead to sparsity of the solution $\hat{\mathbf{x}}_p$, since $\mathbf{u}$ may not be sparse under the kernel representation. In what follows, we give a necessary condition on the elements $\mathbf{v}_j$, $j \in \mathbb{N}_n$, such that the solution $\hat{\mathbf{x}}_p$ has finite nonzero components. For each $N \in \mathbb{N}$, we define a truncation operator $\mathcal{T}_N : \ell_p(\mathbb{N}) \to \ell_p(\mathbb{N})$ by $\mathcal{T}_N(\mathbf{x}) := [x_{N+k} : k \in \mathbb{N}]$ for all $\mathbf{x} = [x_k : k \in \mathbb{N}] \in \ell_p(\mathbb{N})$. For each $\mathbf{x} := [x_j : j \in \mathbb{N}]$, the support of $\mathbf{x}$, denoted by $\mathrm{supp}(\mathbf{x})$, is defined to be the index set on which $\mathbf{x}$ is nonzero, that is, $\mathrm{supp}(\mathbf{x}) := \{j \in \mathbb{N} : x_j \neq 0\}$.

Proposition 21 *Let $p, q \in (1, +\infty)$ be such that $1/p + 1/q = 1$. Suppose that $\mathbf{v}_j, j \in \mathbb{N}_n$, are linearly independent elements in $\ell_q(\mathbb{N})$, $\mathbf{y} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ and $\hat{\mathbf{x}}_p$ is the unique solution of the MNI problem (61) with $\mathbf{y}$ in $\ell_p(\mathbb{N})$. If there exists $N \in \mathbb{N}$ such that $\mathrm{supp}(\hat{\mathbf{x}}_p) \subset \mathbb{N}_N$, then $\mathcal{T}_N(\mathbf{v}_j)$, $j \in \mathbb{N}_n$, are linearly dependent.*

Proof Let $\hat{\mathbf{c}} := [\hat{c}_j : j \in \mathbb{N}_n] \in \mathbb{R}^n$ be the solution of the dual problem (62) and set $\hat{\mathbf{v}}_p := (\mathbf{y}^\top \hat{\mathbf{c}}) \sum_{j\in\mathbb{N}_n} \hat{c}_j \mathbf{v}_j$. Proposition 20 guarantees that $\hat{\mathbf{x}}_p = [\hat{x}_k : k \in \mathbb{N}]$ with $\hat{x}_k := (\hat{\mathbf{v}}_p)_k \, |(\hat{\mathbf{v}}_p)_k|^{q-2} / \|\hat{\mathbf{v}}_p\|_q^{q-2}$, $k \in \mathbb{N}$. This together with the assumption that $\mathrm{supp}(\hat{\mathbf{x}}_p) \subset \mathbb{N}_N$ leads to $\mathrm{supp}(\hat{\mathbf{v}}_p) \subset \mathbb{N}_N$. This implies that $\mathcal{T}_N(\hat{\mathbf{v}}_p) = 0$. Substituting the definition of $\hat{\mathbf{v}}_p$ into this equation, we get that

$$(\mathbf{y}^\top \hat{\mathbf{c}}) \sum_{j\in\mathbb{N}_n} \hat{c}_j \mathcal{T}_N(\mathbf{v}_j) = 0. \tag{65}$$

Since $\mathbf{y} \neq \mathbf{0}$, the infimum of the MNI problem (61) is nonzero. Then by Proposition 36, the quantity $\mathbf{y}^\top \hat{\mathbf{c}}$, as the supremum of the dual problem (62), is also nonzero. As a result, we have that $\hat{\mathbf{c}} \neq \mathbf{0}$. Hence, it follows from (65) that $\sum_{j\in\mathbb{N}_n} \hat{c}_j \mathcal{T}_N(\mathbf{v}_j) = 0$ for $\hat{\mathbf{c}} \neq \mathbf{0}$. This ensures that $\mathcal{T}_N(\mathbf{v}_j)$, $j \in \mathbb{N}_n$, are linearly dependent. ■

Proposition 21 indicates that in general the solution $\hat{\mathbf{x}}_p$ of the MNI problem in the sequence spaces $\ell_p(\mathbb{N})$, for $1 < p < +\infty$, will not have a finite number of nonzero components, unless strict conditions are imposed to the elements $\mathbf{v}_j$, $j \in \mathbb{N}_n$. This has been demonstrated in an example presented in Cheng and Xu (2021), where the solution of the MNI problem in $\ell_2(\mathbb{N})$ has infinite nonzero components, that is, the solution can only be expressed by infinitely many kernel sessions. Indeed, for the elements $\mathbf{v}_1$ and $\mathbf{v}_2$ chosen in the example, $\mathcal{T}_N(\mathbf{v}_1)$ and $\mathcal{T}_N(\mathbf{v}_2)$ will never be linearly dependent for any choice of $N \in \mathbb{N}$.

6. Sparse Learning in the Measure Space

We study in this section the MNI problem and the regularization problem in an RKBS constructed by the measure space. This RKBS is proved to have the space of continuous functions as both the adjoint RKBS and the pre-dual space. By verifying Assumptions (A1) and (A2) for the RKBS, we specialize Theorems 10 and 12 to this space and obtain the sparse representer theorems for the solutions of the MNI problem and the regularization problem in this space.

We begin with introducing an RKBS, which has been considered in Bartolucci et al. (2023); Lin and Xu (2022); Lin et al. (2022); Spek et al. (2023). Let X be a prescribed set and X' be a locally compact Hausdorff space. Denote by $C_0(X')$ the space of all continuous functions $f : X' \to \mathbb{R}$ such that for any $\epsilon > 0$, the set $\{x' \in X' : |f(x')| \geq \epsilon\}$ is compact. We equip the maximum norm on $C_0(X')$, namely $\|f\|_\infty := \sup_{x'\in X'} |f(x')|$ for all $f \in C_0(X')$. The Riesz-Markov representation theorem (Conway (1990)) states that the dual space of $C_0(X')$ is isometrically isomorphic to the space $\mathfrak{M}(X')$ of real-valued regular Borel measures on X' endowed with the total variation norm $\|\cdot\|_{\mathrm{TV}}$. Suppose that $K : X \times X' \to \mathbb{R}$ satisfies $K(x, \cdot) \in C_0(X')$ for all $x \in X$ and the density condition

$$\overline{\mathrm{span}}\{K(x, \cdot) : x \in X\} = C_0(X'). \tag{66}$$

Associated with the function K, we introduce a space of functions on X by

$$\mathcal{B}_K := \left\{ f_\mu := \int_{X'} K(\cdot, x') d\mu(x') : \mu \in \mathfrak{M}(X') \right\}, \tag{67}$$

equipped with

$$\|f_\mu\|_{\mathcal{B}_K} := \|\mu\|_{\mathrm{TV}}. \tag{68}$$

In passing, we provide an example of a kernel K that satisfies the density condition (66). We choose $X = X' := \mathbb{R}^d$ and consider the Gaussian kernel defined by

$$K(x, y) := e^{-\frac{\|x-y\|^2}{2\sigma^2}}, \quad x, y \in \mathbb{R}^d, \tag{69}$$

with $\sigma > 0$. To show that the Gaussian kernel K satisfies the density condition (66), we establish that for $\nu \in \mathfrak{M}(\mathbb{R}^d)$ satisfying

$$\int_{\mathbb{R}^d} K(x, y) d\nu(y) = 0, \quad \text{for all } x \in \mathbb{R}^d, \tag{70}$$

we must have that $\nu = 0$. For this purpose, we re-express the Gaussian kernel K as

$$K(x, y) = \int_{\mathbb{R}^d} p(t) e^{i\langle t, x-y\rangle} dt, \quad x, y \in \mathbb{R}^d, \tag{71}$$

where

$$p(t) := \left(\frac{\sigma}{\sqrt{2\pi}}\right)^d e^{-\frac{\sigma^2\|t\|^2}{2}}, \quad t \in \mathbb{R}^d.$$

Substituting representation (71) into equation (70) yields

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} p(t) e^{i\langle t, x-y\rangle} dt d\nu(y) = 0, \quad x \in \mathbb{R}^d,$$

which further leads to

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} p(t) e^{i\langle t, x-y\rangle} dt d\nu(y) d\nu(x) = 0.$$

By the Fubini theorem, we get that

$$\int_{\mathbb{R}^d} p(t) \left|\hat{\nu}(t)\right|^2 dt = 0, \quad \text{where } \ \hat{\nu}(t) := \int_{\mathbb{R}^d} e^{i\langle t, x\rangle} d\nu(x), \quad t \in \mathbb{R}^d.$$

Since $p(t) > 0$ for all $t \in \mathbb{R}^d$, the above equation yields that $\hat{\nu}(t) = 0$ for all $t \in \mathbb{R}^d$ and hence $\nu = 0$.

We now return to the investigation of the space $\mathcal{B}_K$ defined by (67) and (68). We first show that it is an RKBS on X which has $C_0(X')$ as a pre-dual space.

Proposition 22 *Let X be a prescribed set and X' be a locally compact Hausdorff space. If the function $K : X \times X' \to \mathbb{R}$ satisfies $K(x, \cdot) \in C_0(X')$ for any $x \in X$ and the density condition* (66)*, then the space $\mathcal{B}_K$ defined by* (67) *endowed with the norm* (68) *is an RKBS on X and $C_0(X')$ is a pre-dual space of $\mathcal{B}_K$.*

Proof We first establish that $\mathcal{B}_K$ is an RKBS of functions on X. It follows from the density condition (66) that the mapping $\Phi : \mathfrak{M}(X') \to \mathcal{B}_K$, defined by

$$\Phi(\mu) := \int_{X'} K(\cdot, x')d\mu(x'), \quad \text{for all} \quad \mu \in \mathfrak{M}(X'), \tag{72}$$

is an isometric isomorphism from $\mathfrak{M}(X')$ to $\mathcal{B}_K$. Thus, $\mathcal{B}_K$ is a Banach space of functions on X. By Definition 1, it suffices to verify that δ_x, $x \in X$, are all continuous on $\mathcal{B}_K$. For all $x \in X$ and all $f \in \mathcal{B}_K$, by (67), we have that

$$f(x) = \int_{X'} K(x, x')d\mu(x'), \quad \text{for some} \quad \mu \in \mathfrak{M}(X').$$

It follows from the definition of the total variation norm and (68) that

$$\begin{aligned}|\delta_x(f)| &= \left|\int_{X'} K(x, x')d\mu(x')\right| \\ &\leq \|K(x,\cdot)\|_\infty \|\mu\|_{\mathrm{TV}} \\ &= \|K(x,\cdot)\|_\infty \|f\|_{\mathcal{B}_K}.\end{aligned}$$

That is, δ_x is continuous on $\mathcal{B}_K$.

It remains to show that

$$(C_0(X'))^* = \mathcal{B}_K. \tag{73}$$

It is known from the Riesz-Markov representation theorem (Conway (1990)) that

$$(C_0(X'))^* = \mathfrak{M}(X'). \tag{74}$$

Moreover, the mapping $\Phi : \mathfrak{M}(X') \to \mathcal{B}_K$ defined by (72) is an isometric isomorphism. Thus, the measure space $\mathfrak{M}(X')$ is isometrically isomorphic to $\mathcal{B}_K$. This together with (74) leads to (73). ■

We next reveal that the δ-dual space $\mathcal{B}'_K$ of $\mathcal{B}_K$ is isometrically isomorphic to $C_0(X')$, a Banach space of functions, and the function K coincides with the reproducing kernel of $\mathcal{B}_K$.

Proposition 23 *Let X be a prescribed set and X' be a locally compact Hausdorff space. Suppose that the function $K : X \times X' \to \mathbb{R}$ satisfies $K(x,\cdot) \in C_0(X')$ for any $x \in X$ and the density condition* (66)*. If $\mathcal{B}_K$ is an RKBS on X defined by* (67) *endowed with the norm* (68)*, then the δ-dual space $\mathcal{B}'_K$ of $\mathcal{B}_K$ is isometrically isomorphic to $C_0(X')$ and the function K is the reproducing kernel of $\mathcal{B}_K$.*

Proof We first prove that $\mathcal{B}'_K$ is isometrically isomorphic to $C_0(X')$. According to the definition of $\mathcal{B}'_K$ and the density condition (66), it suffices to verify that $\Delta := \mathrm{span}\{\delta_x : x \in X\}$ is isometrically isomorphic to the linear span $K_X := \mathrm{span}\{K(x,\cdot) : x \in X\}$. We introduce a mapping $\Psi : \Delta \to K_X$ by

$$\Psi\left(\sum_{j\in\mathbb{N}_m} \alpha_j \delta_{x_j}\right) := \sum_{j\in\mathbb{N}_m} \alpha_j K(x_j,\cdot), \quad \text{for all } m \in \mathbb{N}, \alpha_j \in \mathbb{R}, x_j \in X, j \in \mathbb{N}_m, \tag{75}$$

and proceed to prove that Ψ is an isometric isomorphism between Δ and K_X. Clearly, Ψ is linear and surjective. It remains to prove that Ψ is isometric, that is, $\|\ell\|_{\mathcal{B}_K^*} = \|\Psi(\ell)\|_\infty$, for all $\ell \in \Delta$. For any $\ell := \sum_{j\in\mathbb{N}_m} \alpha_j \delta_{x_j} \in \Delta$ with $m \in \mathbb{N}$, $\alpha_j \in \mathbb{R}$, $x_j \in X$, $j \in \mathbb{N}_m$ and any $f_\mu := \int_{X'} K(\cdot, x')d\mu(x') \in \mathcal{B}_K$ with $\mu \in \mathfrak{M}(X')$, it holds that

$$\ell(f_\mu) = \int_{X'} \left(\sum_{j\in\mathbb{N}_m} \alpha_j K(x_j, x') \right) d\mu(x'),$$

which together with definition (75) further leads to

$$\ell(f_\mu) = \int_{X'} (\Psi(\ell))(x')d\mu(x'). \tag{76}$$

It follows from the definition of the total variation norm that $|\ell(f_\mu)| \le \|\Psi(\ell)\|_\infty \|\mu\|_{\mathrm{TV}}$. By definition (68), we rewrite the above inequality as $|\ell(f_\mu)| \le \|\Psi(\ell)\|_\infty \|f_\mu\|_{\mathcal{B}_K}$ for all $f_\mu \in \mathcal{B}_K$ which implies that $\|\ell\|_{\mathcal{B}_K^*} \le \|\Psi(\ell)\|_\infty$. To prove $\|\ell\|_{\mathcal{B}_K^*} = \|\Psi(\ell)\|_\infty$, it suffices to identify a specific function f_μ in $\mathcal{B}_K$ such that $|\ell(f_\mu)| = \|\Psi(\ell)\|_\infty \|f_\mu\|_{\mathcal{B}_K}$. Since $\Psi(\ell) \in C_0(X')$, there exists $z \in X'$ at which the function $\Psi(\ell)$ attains its norm, that is, $|\Psi(\ell)(z)| = \|\Psi(\ell)\|_\infty$. By noting that $\delta_z \in \mathfrak{M}(X')$ satisfies $\|\delta_z\|_{\mathrm{TV}} = 1$, we obtain that $f_{\delta_z} \in \mathcal{B}_K$ and $\|f_{\delta_z}\|_{\mathcal{B}_K} = 1$. It follows from equation (76) that

$$\ell(f_{\delta_z}) = \int_{X'} (\Psi(\ell))(x')d\delta_z(x') = \Psi(\ell)(z).$$

Noting that $|\Psi(\ell)(z)| = \|\Psi(\ell)\|_\infty$ and $\|f_{\delta_z}\|_{\mathcal{B}_K} = 1$, we get from above equation that $|\ell(f_{\delta_z})| = \|\Psi(\ell)\|_\infty \|f_{\delta_z}\|_{\mathcal{B}_K}$. Consequently, we conclude that $\|\ell\|_{\mathcal{B}_K^*} = \|\Psi(\ell)\|_\infty$ for all $\ell \in \Delta$ and hence, Ψ is an isometric isomorphism between Δ and K_X.

We next identify the reproducing kernel of $\mathcal{B}_K$. Since the δ-dual space $\mathcal{B}'_K$ is isometrically isomorphic to a Banach space of functions on X', Proposition 4 ensures that there exists a unique reproducing kernel. To prove that K is the reproducing kernel, we need to verify the reproducing property. According to the isometric isomorphism Ψ defined by (75), the bilinear form on $C_0(X') \times \mathcal{B}_K$ can be defined by

$$\langle f, f_\mu \rangle_{C_0(X')\times\mathcal{B}_K} := \left\langle \Psi^{-1}(f), f_\mu \right\rangle_{\mathcal{B}_K}, \tag{77}$$

for all $f \in C_0(X')$ and $f_\mu \in \mathcal{B}_K$. By equation (77) with noting that $\Psi^{-1}(K(x,\cdot)) = \delta_x$ for all $x \in X$, we obtain for any $x \in X$ and any $f_\mu \in \mathcal{B}_K$ that

$$f_\mu(x) = \langle \delta_x, f_\mu \rangle_{\mathcal{B}_K} = \langle K(x,\cdot), f_\mu \rangle_{C_0(X')\times\mathcal{B}_K}.$$

That is, the reproducing property holds. Thus, we get the conclusion that K is the reproducing kernel of $\mathcal{B}_K$. ■

In the next result, we show that $C_0(X')$ is the adjoint RKBS of $\mathcal{B}_K$.

Proposition 24 *Let X be a prescribed set and X' be a locally compact Hausdorff space. Suppose that the function $K : X \times X' \to \mathbb{R}$ satisfies $K(x,\cdot) \in C_0(X')$ for any $x \in X$ and the density condition* (66)*. If $\mathcal{B}_K$ is an RKBS on X defined by* (67) *endowed with the norm* (68)*, then $C_0(X')$ is the adjoint RKBS of $\mathcal{B}_K$.*

Proof We first establish that $C_0(X')$ is an RKBS. It is clear that $C_0(X')$ is a space of functions. For any $g \in C_0(X')$ and $x' \in X'$, there holds $|\delta_{x'}(g)| = |g(x')| \leq \|g\|_\infty$ which implies that the point evaluation functionals are continuous and hence $C_0(X')$ is an RKBS. We then show that $K(\cdot, x') \in \mathcal{B}_K$ for all $x \in X'$. It follows for each $x \in X$, $x' \in X'$ that

$$K(x, x') = \int_{X'} K(x, y') d\delta_{x'}(y'), \tag{78}$$

which together with definition (67) of $\mathcal{B}_K$ yields that $K(\cdot, x') \in \mathcal{B}_K$. It remains to verify the reproducing property (4) in $C_0(X')$. As pointed out in the proof of Proposition 22, the mapping $\Phi : \mathfrak{M}(X') \to \mathcal{B}_K$ defined by (72) is an isometric isomorphism from $\mathfrak{M}(X')$ to $\mathcal{B}_K$. Moreover, it follows from equation (78) that $\Phi(\delta_{x'}) = K(\cdot, x')$ for all $x' \in X'$. Then we have for any $g \in C_0(X')$ and $x' \in X'$ that

$$g(x') = \langle g, \delta_{x'} \rangle_{\mathfrak{M}(X')} = \langle g, K(\cdot, x') \rangle_{\mathcal{B}_K},$$

which proves (4) with $\mathcal{B}' := C_0(X')$. Consequently, the desired result follows from the definition of the adjoint RKBS. ∎

We now turn to describing the MNI problem and the regularization problem in $\mathcal{B}_K$. In this section, we consider learning a target function in $\mathcal{B}_K$ from the point-evaluation functional data. Specifically, suppose that $x_j \in X$, $j \in \mathbb{N}_n$, are n distinct points in X and $K(x_j, \cdot)$, $j \in \mathbb{N}_n$, are linearly independent elements in $C_0(X')$. Associated with these point-evaluation functionals, we set

$$\mathcal{V} := \operatorname{span} \{K(x_j, \cdot) : j \in \mathbb{N}_n\}. \tag{79}$$

The operator $\mathcal{L} : \mathcal{B}_K \to \mathbb{R}^n$, defined by (10), is specialized as

$$\mathcal{L}(f_\mu) := [f_\mu(x_j) : j \in \mathbb{N}_n], \text{ for all } f_\mu \in \mathcal{B}_K. \tag{80}$$

For a given vector $\mathbf{y} \in \mathbb{R}^n$, the subset $\mathcal{M}_\mathbf{y}$ of $\mathcal{B}_K$, defined by (12), has the form

$$\mathcal{M}_\mathbf{y} := \{f_\mu \in \mathcal{B}_K : \mathcal{L}(f_\mu) = \mathbf{y}\}. \tag{81}$$

With the notation above, we formulate the MNI problem in $\mathcal{B}_K$ as

$$\inf \left\{ \|f_\mu\|_{\mathcal{B}_K} : f_\mu \in \mathcal{M}_\mathbf{y} \right\}, \tag{82}$$

and the regularization problem in $\mathcal{B}_K$ as

$$\inf \{\mathcal{Q}_\mathbf{y}(\mathcal{L}(f_\mu)) + \lambda \|f_\mu\|_{\mathcal{B}_K} : f_\mu \in \mathcal{B}_K\}, \tag{83}$$

where $\mathcal{Q}_\mathbf{y} : \mathbb{R}^n \to \mathbb{R}_+$ is a loss function and λ is a positive regularization parameter.

Before establishing the sparse representer theorems for the solutions of problems (82) and (83), we verify Assumptions (A1) and (A2) for the RKBS $\mathcal{B}_K$. To this end, we characterize the extreme points of the subdifferential set of the maximum norm at any $g \in C_0(X')$. We start from the result concerning the extreme points of the unit ball in $\mathcal{B}_K$.

Lemma 25 *Let X be a prescribed set and X' be a locally compact Hausdorff space. Suppose that the function $K : X \times X' \to \mathbb{R}$ satisfies $K(x, \cdot) \in C_0(X')$ for any $x \in X$ and the density condition* (66)*. If $\mathcal{B}_K$ is an RKBS on X defined by* (67) *endowed with the norm* (68) *and B_K^0 is the closed unit ball of $\mathcal{B}_K$ with center at the origin, then*

$$\operatorname{ext}\left(B_K^0\right) = \left\{-K(\cdot, x'), K(\cdot, x') : x' \in X'\right\}. \tag{84}$$

Proof Recall that the mapping $\Phi : \mathfrak{M}(X') \to \mathcal{B}_K$ defined by (72) is an isometric isomorphism from $\mathfrak{M}(X')$ to $\mathcal{B}_K$. Let $\mathfrak{M}^0(X')$ denote the closed unit ball of $\mathfrak{M}(X')$ with center at the origin. Clearly, Ψ is also an isometric isomorphism from $\mathfrak{M}^0(X')$ to B_K^0. Note that isometric isomorphisms from one normed space onto another preserve extreme points (Megginson (1998)). Hence, we have that

$$\operatorname{ext}(B_K^0) = \Phi(\operatorname{ext}(\mathfrak{M}^0(X'))). \tag{85}$$

It is known (Bredies and Carioni (2020)) that $\operatorname{ext}\left(\mathfrak{M}^0(X')\right) = \{-\delta_{x'}, \delta_{x'} : x' \in X'\}$, which together with equation (85) leads to

$$\operatorname{ext}(B_K^0) = \{-\Phi(\delta_{x'}), \Phi(\delta_{x'}) : x' \in X'\}. \tag{86}$$

Substituting $\Phi(\delta_{x'}) = K(\cdot, x')$ for all $x' \in X'$ into (86), we get the desired result (84) of this lemma. ■

Next, we characterize the extreme points of the subdifferential set $\partial\|\cdot\|_\infty(g)$ for a nonzero $g \in C_0(X')$. For each $g \in C_0(X')$, let $\mathcal{N}(g)$ denote the subset of X' where the function g attains its supremum norm $\|g\|_\infty$, namely

$$\mathcal{N}(g) := \left\{x' \in X' : |g(x')| = \|g\|_\infty\right\}. \tag{87}$$

For each $g \in C_0(X')$, we introduce a subset of $\mathcal{B}_K$ by

$$\Omega(g) := \left\{\operatorname{sign}(g(x'))K(\cdot, x') : x' \in \mathcal{N}(g)\right\}. \tag{88}$$

We show in the next lemma that for each nonzero $g \in C_0(X')$, the set of the extreme points of the subdifferential set $\partial\|\cdot\|_\infty(g)$ coincides with $\Omega(g)$.

Lemma 26 *Let X' be a locally compact Hausdorff space. If $g \in C_0(X')\backslash\{0\}$ and $\Omega(g)$ is defined by* (88)*, then*

$$\operatorname{ext}\left(\partial\|\cdot\|_\infty(g)\right) = \Omega(g). \tag{89}$$

Proof We first prove that $\Omega(g) \subset \operatorname{ext}\left(\partial\|\cdot\|_\infty(g)\right)$. We assume that $f \in \Omega(g)$ and proceed to show $f \in \operatorname{ext}\left(\partial\|\cdot\|_\infty(g)\right)$. It follows from definition (88) of $\Omega(g)$ that there exists $x' \in \mathcal{N}(g)$ such that

$$f = \operatorname{sign}(g(x'))K(\cdot, x'). \tag{90}$$

We start with proving $f \in \partial\|\cdot\|_\infty(g)$. Recall Φ defined by (72) provides an isometric isomorphism from $\mathfrak{M}(X')$ to $\mathcal{B}_K$ and $\Phi(\delta_{x'}) = K(\cdot, x')$. Noticing that $\delta_{x'} \in \mathfrak{M}(X')$ satisfies $\|\delta_{x'}\|_{\mathrm{TV}} = 1$, we obtain that $\|K(\cdot, x')\|_{\mathcal{B}_K} = 1$. This together equation (90) leads directly to $\|f\|_{\mathcal{B}_K} = 1$. According to the reproducing property (4), we have that $\langle g, f\rangle_{\mathcal{B}_K} = \mathrm{sign}(g(x'))g(x')$. Noting that $x' \in \mathcal{N}(g)$, the above equation implies $\langle g, f\rangle_{\mathcal{B}_K} = \|g\|_\infty$. Hence, we conclude by equation (15) that $f \in \partial\|\cdot\|_\infty(g)$. It suffices to show that for any $f_1, f_2 \in \partial\|\cdot\|_\infty(g)$ satisfying $f = (f_1 + f_2)/2$, there holds $f = f_1 = f_2$. Since $g \neq 0$ and $x' \in \mathcal{N}(g)$, we get that $g(x') \neq 0$. Lemma 25 ensures that function f with the form (90) satisfies $f \in \mathrm{ext}(B_K^0)$. Moreover, it follows from $\partial\|\cdot\|_\infty(g) \subset B_K^0$ that $f_1, f_2 \in B_K^0$. This combined with $f \in \mathrm{ext}(B_K^0)$ and the definition of extreme points leads to $f = f_1 = f_2$. Again, using the definition of extreme points, we obtain that $f \in \mathrm{ext}\,(\partial\|\cdot\|_\infty(g))$.

It remains to show that $\mathrm{ext}\,(\partial\|\cdot\|_\infty(g)) \subset \Omega(g)$. Assume that $f \in \mathrm{ext}\,(\partial\|\cdot\|_\infty(g))$. It follows from Proposition 35 that $\mathrm{ext}\,(\partial\|\cdot\|_\infty(g)) \subset \mathrm{ext}\left(B_K^0\right)$. Hence, $f \in \mathrm{ext}\left(B_K^0\right)$. Lemma 25 ensures that there exist $\alpha \in \{-1, 1\}$ and $x' \in X'$ such that

$$f = \alpha K(\cdot, x'). \tag{91}$$

Since $f \in \mathrm{ext}\,(\partial\|\cdot\|_\infty(g))$, we get by equation (15) that $\langle g, f\rangle_{\mathcal{B}_K} = \|g\|_\infty$. Substituting representation (91) into the above equation, we get that $\|g\|_\infty = \alpha\langle g, K(\cdot, x')\rangle_{\mathcal{B}_K}$, which together with the reproducing property (4) leads to $\|g\|_\infty = \alpha g(x')$. Clearly, $x' \in \mathcal{N}(g)$ and $\alpha = \mathrm{sign}(g(x'))$. Due to the definition (88) of $\Omega(g)$, we conclude that $f \in \Omega(g)$. ■

Note that for any $g \in C_0(X')$, the set $\mathcal{N}(g)$ is bounded in X' but its cardinality may not be finite. To ensure that the RKBS $\mathcal{B}_K$ satisfies Assumption (A1), we need to impose the following assumption on the reproducing kernel K.

(A3) For any nonzero $g \in \mathcal{V}$, with $\mathcal{V}$ being defined by (79), the cardinality of $\mathcal{N}(g)$ is finite.

Below, we present an example of kernel K which satisfies Assumption (A3). Let $X = X' = \mathbb{R}$ and K be the Gaussian kernel defined by (69) with $d = 1$. Suppose that x_j, $j \in \mathbb{N}_n$, are n distinct points in $\mathbb{R}$. We will show that for any nonzero $g \in C_0(\mathbb{R})$ having the form $g := \sum_{j\in\mathbb{N}_n} \alpha_j K(x_j, \cdot)$ with $\alpha_j \in \mathbb{R}$, $j \in \mathbb{N}_n$, the subset $\mathcal{N}(g)$, defined by (87), has a finite cardinality. By the definition (69) of K, g is differentiable. We denote by $\mathcal{Z}(g)$ the set of the zeros of the derivative g', that is, $\mathcal{Z}(g) := \{x \in \mathbb{R} : g'(x) = 0\}$. Clearly, $\mathcal{N}(g) \subset \mathcal{Z}(g)$. We next show that $\mathcal{N}(g)$ has a finite cardinality. Assume to the contrary that $\mathcal{N}(g)$ is infinite. We let x_j, $j \in \mathbb{N}$, be a sequence of distinct points in $\mathcal{N}(g)$. It follows from $\mathcal{N}(g) \subset \mathcal{Z}(g)$ that $x_j \in \mathcal{Z}(g)$, $j \in \mathbb{N}$. Since $\mathcal{N}(g)$ is bounded, the sequence x_j, $j \in \mathbb{N}$, is also bounded. Hence, Bolzano-Weierstrass theorem ensures that there is a subsequence x_{j_k}, $k \in \mathbb{N}$, which converges to some $x \in \mathbb{R}$. According to the continuity of g', we obtain that $\lim_{k\to\infty} g'(x_{j_k}) = g'(x)$, which together with $x_{j_k} \in \mathcal{Z}(g)$, $k \in \mathbb{N}$, implies that $x \in \mathcal{Z}(g)$. Consequently, we conclude that x is a cumulative point of $\mathcal{Z}(g)$. However, the set $\mathcal{Z}(g)$, as the set of the zeros of the analytic function g', has no cumulative points, a contradiction. Therefore, the set $\mathcal{N}(g)$ must have a finite cardinality. That is, the Gaussian kernel with the form (69) satisfies Assumption (A3).

Below, we validate that the RKBS $\mathcal{B}_K$ satisfies Assumptions (A1) and (A2) when the kernel K satisfies Assumption (A3).

Lemma 27 *Let X be a prescribed set and X' be a locally compact Hausdorff space. Suppose that the function $K : X \times X' \to \mathbb{R}$ satisfies $K(x, \cdot) \in C_0(X')$ for any $x \in X$ and the density* *condition* (66)*. Let* $x_j \in X$*,* $j \in \mathbb{N}_n$*, be* n *distinct points in* X *and* $K(x_j,\cdot)$*,* $j \in \mathbb{N}_n$*, be linearly independent elements in* $C_0(X')$*, and let* $\mathcal{V}$ *be defined by* (79)*. If* K *satisfies Assumption (A3), then the RKBS* $\mathcal{B}_K$ *defined by* (67) *endowed with the norm* (68) *satisfies Assumption (A1) with* $X'_g := \mathcal{N}(g)$ *for any* $g \in C_0(X')$ *and Assumption (A2) with* $C := 1$*.*

Proof We first show that Assumption (A1) holds. For a nonzero $g \in \mathcal{V}$, Lemma 26 ensures that $\mathrm{ext}\,(\partial\|\cdot\|_\infty(g)) = \Omega(g)$, which together with definition (88) of $\Omega(g)$ leads to

$$\mathrm{ext}(\partial\|\cdot\|_\infty(g)) = \left\{\mathrm{sign}(g(x'))K(\cdot,x') : x' \in \mathcal{N}(g)\right\}.$$

Note that for any $x' \in \mathcal{N}(g)$, we have that $g(x') \neq 0$ as g is nonzero. Hence, we get from the above equation that

$$\mathrm{ext}(\partial\|\cdot\|_\infty(g)) \subset \left\{-K(\cdot,x'_j), K(\cdot,x'_j) : x'_j \in \mathcal{N}(g)\right\}.$$

Assumption (A3) guarantees the finite cardinality of the set $\mathcal{N}(g)$. Thus, Assumption (A1) holds with $X'_g := \mathcal{N}(g)$ for any $g \in \mathcal{V}$.

We next prove that Assumption (A2) holds. For any $m \in \mathbb{N}$, distinct points $x'_j \in X'$, $j \in \mathbb{N}_m$, and $\boldsymbol{\alpha} = [\alpha_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$, it follows from definition (68) of the norm of $\mathcal{B}_K$ that

$$\left\|\sum_{j\in\mathbb{N}_m} \alpha_j K(\cdot,x'_j)\right\|_{\mathcal{B}_K} = \left\|\sum_{j\in\mathbb{N}_m} \alpha_j \delta_{x'_j}\right\|_{\mathrm{TV}}.$$

It is clear that $\|\sum_{j\in\mathbb{N}_m} \alpha_j\delta_{x'_j}\|_{\mathrm{TV}} = \|\boldsymbol{\alpha}\|_1$, which together with the above equation leads to $\|\sum_{j\in\mathbb{N}_m} \alpha_j K(\cdot,x'_j)\|_{\mathcal{B}_K} = \|\boldsymbol{\alpha}\|_1$. Therefore, Assumption (A2) holds with $C := 1$. ■

We are ready to establish the sparse representation theorems for the solutions of the MNI problem (82) and the regularization problem (83). Suppose that kernel K satisfies Assumption (A3). For each $g \in \mathcal{V}$, we denote by $n(g)$ the cardinality of $\mathcal{N}(g)$ and suppose that $\mathcal{N}(g) = \{x'_j \in X' : j \in \mathbb{N}_{n(g)}\}$. For each $g \in \mathcal{V}$, we introduce a kernel matrix by

$$\mathbf{V}_g := [K(x_i,x'_j) : i \in \mathbb{N}_n, j \in \mathbb{N}_{n(g)}] \in \mathbb{R}^{n\times n(g)}. \tag{92}$$

In the following theorems, we always let X be a prescribed set and X' be a locally compact Hausdorff space. Suppose that the function $K : X \times X' \to \mathbb{R}$ satisfies $K(x,\cdot) \in C_0(X')$ for any $x \in X$ and the density condition (66) and that $\mathcal{B}_K$ is the RKBS on X defined by (67) endowed with the norm (68). In addition, let $x_j \in X$, $j \in \mathbb{N}_n$, be n distinct points in X and $K(x_j,\cdot)$, $j \in \mathbb{N}_n$, be linearly independent elements in $C_0(X')$.

For the MNI problem (82), we get the following sparse representer theorem by employing Theorem 10.

Theorem 28 *Let* $\mathbf{y} \in \mathbb{R}^n\backslash\{\mathbf{0}\}$*,* $\mathcal{V}$*, and* $\mathcal{M}_{\mathbf{y}}$ *be defined by* (79) *and* (81)*, respectively. Suppose that* $\hat{g} \in \mathcal{V}$ *satisfy*

$$(\|\hat{g}\|_\infty \mathrm{co}(\Omega(\hat{g}))) \cap \mathcal{M}_{\mathbf{y}} \neq \emptyset. \tag{93}$$

If Assumption (A3) holds, $\mathcal{N}(\hat{g})$ *and* $\mathbf{V}_{\hat{g}}$ *are defined by* (87) *and* (92) *with* $g := \hat{g}$*, respectively, then for any* $\hat{f} \in \mathrm{ext}\,(\mathrm{S}(\mathbf{y}))$*, there exist* $\hat{\alpha}_j \neq 0$*,* $j \in \mathbb{N}_M$*, with* $\sum_{j\in\mathbb{N}_M} |\hat{\alpha}_j| = \|\hat{g}\|_\infty$ *and* $x'_j \in \mathcal{N}(\hat{g})$*,* $j \in \mathbb{N}_M$*, such that*

$$\hat{f} = \sum_{j\in\mathbb{N}_M} \hat{\alpha}_j K(\cdot, x'_j), \tag{94}$$

for some positive integer $M \leq \mathrm{rank}(\mathbf{V}_{\hat{g}})$*,*

Proof We prove this result by employing Theorem 10. We first point out that the hypotheses about the RKBS in Theorem 10 are satisfied. The RKBS $\mathcal{B}_K$ guaranteed by Proposition 23 has $C_0(X')$ as its pre-dual space and K as its reproducing kernel. In addition, according to Proposition 24, $\mathcal{B}_K$ and $C_0(X')$ are a pair of RKBSs. We next verify that $\hat{g}$ satisfies (18). Lemma 26 ensures that $\mathrm{ext}\,(\partial\|\cdot\|_\infty(\hat{g})) = \Omega(\hat{g})$, which together with Krein-Milman theorem leads to $\partial\|\cdot\|_\infty(\hat{g}) = \overline{\mathrm{co}}(\Omega(\hat{g}))$. It follows from Assumption (A3) that $\Omega(\hat{g})$ is of finite cardinality. Thus, the above equation reduces to $\partial\|\cdot\|_\infty(\hat{g}) = \mathrm{co}(\Omega(\hat{g}))$. Substituting this equation into relation (93) leads to $(\|\hat{g}\|_\infty\partial\|\cdot\|_\infty(\hat{g}))\cap\mathcal{M}_{\mathbf{y}} \neq \emptyset$. That is, $\hat{g}$ satisfies (18). Finally, according to Lemma 27, we have that Assumption (A1) holds with $X'_g := \mathcal{N}(g)$ for any $g \in \mathcal{V}$ and Assumption (A2) holds with $C := 1$. Thus, the hypotheses in Theorem 10 are all satisfied. By the reproducing property (4), there holds for each $i \in \mathbb{N}_n$, $j \in \mathbb{N}_{n(\hat{g})}$, that $\langle K(x_i, \cdot), K(\cdot, x'_j)\rangle_{\mathcal{B}_K} = K(x_i, x'_j)$, which shows that the matrix $\mathbf{L}_{\hat{g}}$ defined by (21) coincides with the matrix $\mathbf{V}_{\hat{g}}$. Hence, Theorem 10 combined with $X'_{\hat{g}} := \mathcal{N}(\hat{g})$ and $\mathbf{L}_{\hat{g}} := \mathbf{V}_{\hat{g}}$ guarantees that any extreme point $\hat{f}$ of $\mathrm{S}(\mathbf{y})$ can be expressed in the form of (94). ■

Theorem 28 is again data-dependent, similar to Theorem 17. It improves the data-independent representer theorem, presented in Bartolucci et al. (2023), which expresses the extreme point of the solution set of the MNI problem (82) in terms of a linear combination of n elements from the set $\mathrm{ext}(B^0_K)$, where B^0_K is the closed unit ball in $\mathcal{B}_K$, independent of the given data. While Theorem 28 expresses an extreme point of the solution set of the MNI problem (82) as a linear combination of at most M elements of $\mathrm{ext}(\partial\|\cdot\|_\infty(\hat{g}))$, which is a much smaller data-dependent subset of $\mathrm{ext}(B^0_K)$.

By specializing Theorem 12 to the RKBS $\mathcal{B}_K$, we establish below the sparse representer theorem for the solutions of the regularization problem (83).

Theorem 29 *Suppose that* $\mathbf{y}_0 \in \mathbb{R}^n$*,* $\lambda > 0$ *and that* $\mathcal{Q}_{\mathbf{y}_0}$ *is lower semi-continuous and convex. Let* $\mathcal{V}$ *be defined by* (79)*. If Assumption (A3) holds, then every nonzero* $\hat{f} \in \mathrm{ext}\,(\mathrm{R}(\mathbf{y}_0))$ *has the representation* (94) *for some* $\hat{g} \in \mathcal{V}$*, positive integer* $M \leq \mathrm{rank}(\mathbf{V}_{\hat{g}})$*,* $\alpha_j \neq 0$*,* $j \in \mathbb{N}_M$*, with* $\sum_{j\in\mathbb{N}_M} |\alpha_j| = \|\hat{g}\|_\infty$ *and* $x'_j \in \mathcal{N}(\hat{g})$*,* $j \in \mathbb{N}_M$*.*

Proof As has been shown in Propositions 23 and 24, $\mathcal{B}_K$ and $C_0(X')$ are a pair of RKBSs with K being the reproducing kernel of $\mathcal{B}_K$ and moreover, $\mathcal{B}_K$ takes $C_0(X')$ as its pre-dual. Since Assumption (A3) holds, Lemma 27 ensures that Assumption (A1) is satisfied with $X'_g := \mathcal{N}(g)$ for any $g \in \mathcal{V}$ and Assumption (A2) is satisfied with $C := 1$. We notice that the regularizer having the form $\varphi(t) := t$, $t \in \mathbb{R}_+$, is continuous, convex and strictly increasing. That is, the hypotheses of Theorem 12 are satisfied. We then obtain the desired result by using Theorem 12 and replacing $X'_{\hat{g}}$ and $\mathbf{L}_{\hat{g}}$ by $\mathcal{N}(\hat{g})$ and $\mathbf{V}_{\hat{g}}$, respectively. ■

Remarks on the relation of Theorems 28 and 29 with the result of Song et al. (2013) are in order. The RKBS with the ℓ_1 norm considered in Song et al. (2013) is a subspace of the RKBS defined by (67) and (68). Representer theorems for the MNI and the regularization problems were also obtained in Song et al. (2013) under an additional admissibility condition. Our results differ from those of Song et al. (2013) in several aspects. First of all, the sparse representer theorems presented in Theorems 28 and 29 do not require the admissibility condition on the kernel function as Song et al. (2013) requires. Second, Song et al. (2013) represented solutions as in (94) with x'_j, $j \in \mathbb{N}_n$, being given training samples, while we represent the extreme points of the solution sets in the form (94), with the points x'_j, $j \in \mathbb{N}_M$, not necessarily training samples. We also provide a precise characterization for the points x'_j, $j \in \mathbb{N}_M$, which belong to the data-dependent set $\mathcal{N}(\hat{g})$. Finally, the number n of the kernel sections in the solution representation of Song et al. (2013) coincides with the number of the training samples, while the number M of the kernel sections in representation (94) may be smaller than the number of the training samples.

The regularization parameter λ in (83) allows to further promote the sparsity level of the regularized solutions. We show in the next theorem how the choice of the parameter λ can accomplish it by employing Theorem 15.

Theorem 30 *Suppose that $\mathbf{y}_0 \in \mathbb{R}^n$, $\lambda > 0$ and that $\mathcal{Q}_{\mathbf{y}_0}$ is lower semi-continuous and convex. Let $\mathcal{D}_{\lambda,\mathbf{y}_0}$ be defined by (28) with $\mathbf{y} := \mathbf{y}_0$, $\hat{\mathbf{z}} \in \mathcal{D}_{\lambda,\mathbf{y}_0}\backslash\{\mathbf{0}\}$ and let $\mathcal{V}$ be defined by (79), $\hat{g} \in \mathcal{V}$ satisfy (93) with $\mathbf{y} := \hat{\mathbf{z}}$. If Assumption (A3) holds, $\mathcal{N}(\hat{g})$ and $\mathbf{V}_{\hat{g}}$ be defined by (87) and (92) with $g := \hat{g}$, respectively, then problem (83) with $\mathbf{y} := \mathbf{y}_0$ has a solution $\hat{f} = \sum_{i\in\mathbb{N}_l} \hat{\alpha}_{k_i} K(\cdot, x_{k_i})$ with $\hat{\alpha}_{k_i} \in \mathbb{R} \setminus \{0\}$, $x_{k_i} \in \mathcal{N}(\hat{g})$, $i \in \mathbb{N}_l$ for some $l \in \mathbb{N}_{n(\hat{g})}$ if and only if there exists $\mathbf{a} \in \partial\mathcal{Q}_{\mathbf{y}}(\mathbf{V}_{\hat{g}}\hat{\boldsymbol{\alpha}})$ for $\hat{\boldsymbol{\alpha}} := \sum_{i\in\mathbb{N}_l} \hat{\alpha}_{k_i}\mathbf{e}_{k_i}$ such that*

$$\lambda = -(\mathbf{V}_{\hat{g}}^\top \mathbf{a})_{k_i}\mathrm{sign}(\hat{\alpha}_{k_i}), \quad i \in \mathbb{N}_l, \ \text{ and } \lambda \geq |(\mathbf{V}_{\hat{g}}^\top \mathbf{a})_j|, \quad j \in \mathbb{N}_{n(\hat{g})}\backslash\{k_i : i \in \mathbb{N}_l\}. \tag{95}$$

Proof We prove this theorem by specializing Theorem 15 to problem (83). Once again, as has been shown in Propositions 23 and 24, $\mathcal{B}_K$ and $C_0(X')$ are a pair of RKBSs with K being the reproducing kernel of $\mathcal{B}_K$ and moreover, $\mathcal{B}_K$ takes $C_0(X')$ as its pre-dual. Moreover, Assumption (A3) ensures that Assumption (A1) holds with $X'_g := \mathcal{N}(g)$ for any $g \in \mathcal{V}$ and Assumption (A2) holds with $C := 1$. As in the proof of Theorem 28, one can see that assumption (93) ensures that $\hat{g}$ satisfies (18) with $\mathbf{y} := \hat{\mathbf{z}}$. We also notice that $\varphi(t) := t$, $t \in \mathbb{R}_+$ in the regularization problem (83), is continuous, convex and strictly increasing. The hypotheses of Theorem 15 are all fulfilled. Therefore, by Theorem 15 with $\mathbf{L}_{\hat{g}} := \mathbf{V}_{\hat{g}}$ and $X'_{\hat{g}} := \mathcal{N}(\hat{g})$, problem (83) with $\mathbf{y} := \mathbf{y}_0$ has a solution $\hat{f} = \sum_{i\in\mathbb{N}_l} \hat{\alpha}_{k_i} K(\cdot, x_{k_i})$ with $\hat{\alpha}_{k_i} \in \mathbb{R}\backslash\{0\}$, $x_{k_i} \in \mathcal{N}(\hat{g})$, $i \in \mathbb{N}_l$ for some $l \in \mathbb{N}_{n(\hat{g})}$ if and only if there exists $\mathbf{a} \in \partial\mathcal{Q}_{\mathbf{y}}(\mathbf{V}_{\hat{g}}\hat{\boldsymbol{\alpha}})$ with $\hat{\boldsymbol{\alpha}} := \sum_{i\in\mathbb{N}_l} \hat{\alpha}_{k_i}\mathbf{e}_{k_i} \in \Omega_l$ such that

$$\lambda = -(\mathbf{V}_{\hat{g}}^\top \mathbf{a})_{k_i}\mathrm{sign}(\hat{\alpha}_{k_i})/(\varphi'(C\|\hat{\boldsymbol{\alpha}}\|_1)), \quad i \in \mathbb{N}_l, \tag{96}$$

$$\lambda \geq |(\mathbf{V}_{\hat{g}}^\top \mathbf{a})_j|/(C\varphi'(C\|\hat{\boldsymbol{\alpha}}\|_1)), \quad j \in \mathbb{N}_{n(\hat{g})}\backslash\{k_i : i \in \mathbb{N}_l\}. \tag{97}$$

Substituting $\varphi'(t) = 1$, $t \in \mathbb{R}_+$ and $C = 1$ into (96) and (97) yields the desired result (95). ■

7. Conclusion

We have studied attributes of RKBSs that can promote sparsity of learning solutions in the spaces. We have proposed sufficient conditions on RKBSs which give rise to explicit and data-dependent sparse representer theorems for solutions of the MNI problem and the regularization problem in the spaces. Following the established general sparse representer theorems, we have shown that two specific RKBSs, the sequence space $\ell_1(\mathbb{N})$ and the measure space, have sparse representations for solutions of the MNI problem and the regularization problem in these spaces.

Acknowledgments

R. Wang is supported in part by the Natural Science Foundation of China under grant 12171202 and by the National Key Research and Development Program of China (grants no. 2020YFA0714101 and 2020YFA0713600); Y. Xu is supported in part by the US National Science Foundation under grants DMS-1912958 and DMS-2208386, and by the US National Institutes of Health under grant R21CA263876. Send all correspondence to Y. Xu.

Appendix A. Explicit Representer Theorems for MNI in Banach Spaces

An explicit representer theorem for the solutions of the MNI problem in a general Banach space having a pre-dual space is stated in Proposition 7 of Section 3. In this appendix, we give a complete proof for the proposition.

We first present several useful properties of the solution set $\mathrm{S}(\mathbf{y})$ of problem (13) with $\mathbf{y} \in \mathbb{R}^n$. For this purpose, we recall two well-known results. The generalized Weierstrass Theorem (Kurdila and Zabarankin (2005)) shows that if $\mathcal{X}$ is a compact topological space and a functional $\mathcal{T} : \mathcal{X} \to \mathbb{R}$ is lower semi-continuous, then there exists $f_0 \in \mathcal{X}$ such that

$$\mathcal{T} f_0 = \inf\{\mathcal{T}(f) : f \in \mathcal{X}\}.$$

A consequence of the Banach-Alaoglu Theorem (Megginson (1998)) ensures that any bounded and weakly* closed subset of $\mathcal{B}^*$ is weakly* compact.

Lemma 31 *If the Banach space $\mathcal{B}$ has a pre-dual space $\mathcal{B}_*$ and $\nu_j \in \mathcal{B}_*$, $j \in \mathbb{N}_n$, are linearly independent, then for any $\mathbf{y} \in \mathbb{R}^n$, the solution set $\mathrm{S}(\mathbf{y})$ of problem (13) is nonempty, convex and weakly* compact.*

Proof We first show that $\mathrm{S}(\mathbf{y})$ is nonempty. Note that the linear independence of ν_j, $j \in \mathbb{N}_n$, guarantees that $\mathcal{M}_{\mathbf{y}}$ is nonempty. Pick $g \in \mathcal{M}_{\mathbf{y}}$ and set $r := \|g\|_{\mathcal{B}} + 1$. It follows that $B_r \cap \mathcal{M}_{\mathbf{y}} \neq \emptyset$, where B_r denotes the closed ball centered at the origin with radius r under the norm $\|\cdot\|_{\mathcal{B}}$. Accordingly, we rewrite the MNI problem (13) as

$$\inf\left\{\|f\|_{\mathcal{B}} : f \in B_r \cap \mathcal{M}_{\mathbf{y}}\right\}. \tag{98}$$

We prove that problem (98) has at least one solution by using the generalized Weierstrass Theorem. Note that the norm $\|\cdot\|_{\mathcal{B}}$ is weakly* lower semi-continuous on $\mathcal{B}$. It suffices to verify that $B_r \cap \mathcal{M}_{\mathbf{y}}$ is weakly* compact. We first claim that $B_r \cap \mathcal{M}_{\mathbf{y}}$ is weakly* closed. Indeed, suppose that the sequence f_m, $m \in \mathbb{N}$, in $B_r \cap \mathcal{M}_{\mathbf{y}}$ converges weakly* to f. We then get that $\lim_{m\to+\infty} \langle f_m, \nu_j\rangle_{\mathcal{B}_*} = \langle f, \nu_j\rangle_{\mathcal{B}_*}$, for all $j \in \mathbb{N}$, which together with equation (2) leads to $\lim_{m\to+\infty} \langle \nu_j, f_m\rangle_{\mathcal{B}} = \langle \nu_j, f\rangle_{\mathcal{B}}$, for all $j \in \mathbb{N}$. Noting that $\langle \nu_j, f_m\rangle_{\mathcal{B}} = y_j$, for all $m \in \mathbb{N}$ and all $j \in \mathbb{N}_n$, the above equation yields that $\langle \nu_j, f\rangle_{\mathcal{B}} = y_j$, for all $j \in \mathbb{N}_n$. That is to say, $f \in \mathcal{M}_{\mathbf{y}}$. According to the weakly* lower semi-continuity of the norm $\|\cdot\|_{\mathcal{B}}$ on $\mathcal{B}$, we have that

$$\|f\|_{\mathcal{B}} \leq \liminf_{m} \|f_m\|_{\mathcal{B}} \leq r. \tag{99}$$

We get the conclusion that $f \in B_r \cap \mathcal{M}_{\mathbf{y}}$ and hence $B_r \cap \mathcal{M}_{\mathbf{y}}$ is weakly* closed. Obviously, $B_r \cap \mathcal{M}_{\mathbf{y}}$ is bounded in $\mathcal{B}$. Consequently, the Banach-Alaoglu Theorem guarantees that $B_r \cap \mathcal{M}_{\mathbf{y}}$ is weakly* compact. By virtue of the generalized Weierstrass theorem, there exists an $f_0 \in B_r \cap \mathcal{M}_{\mathbf{y}}$ such that $\|f_0\|_{\mathcal{B}} = \inf\{\|f\|_{\mathcal{B}} : f \in B_r \cap \mathcal{M}_{\mathbf{y}}\}$. That is, $f_0 \in \mathrm{S}(\mathbf{y})$.

We next verify that $\mathrm{S}(\mathbf{y})$ is convex. It is clear that for any $f_1, f_2 \in \mathrm{S}(\mathbf{y})$ and any $\theta \in [0,1]$, there holds $\theta f_1 + (1-\theta) f_2 \in \mathcal{M}_{\mathbf{y}}$. Moreover, we also have that

$$\|\theta f_1 + (1-\theta) f_2\|_{\mathcal{B}} \leq \theta \|f_1\|_{\mathcal{B}} + (1-\theta)\|f_2\|_{\mathcal{B}},$$

which together with f_1 and f_2 both attaining the infimum in (13) yields that $\theta f_1 + (1-\theta) f_2$ is also a solution of (13). That is, $\theta f_1 + (1-\theta) f_2 \in \mathrm{S}(\mathbf{y})$. Thus, $\mathrm{S}(\mathbf{y})$ is a convex set.

We finally prove that $\mathrm{S}(\mathbf{y})$ is weakly* compact. Again by the Banach-Alaoglu theorem, it suffices to prove that $\mathrm{S}(\mathbf{y})$ is bounded and weakly* closed. Since all the elements in $\mathrm{S}(\mathbf{y})$ have the infimum in (13) as their norms, the set $\mathrm{S}(\mathbf{y})$ is obviously bounded. Suppose that the sequence f_m, $m \in \mathbb{N}$, in $\mathrm{S}(\mathbf{y})$ converges weakly* to f. As pointed out earlier, we have that $f \in \mathcal{M}_{\mathbf{y}}$. It follows from the weakly* lower semi-continuity of $\|\cdot\|_{\mathcal{B}}$ that $\|f\|_{\mathcal{B}} \leq \liminf_m \|f_m\|_{\mathcal{B}} = \|f_0\|_{\mathcal{B}}$. Note that $\|f\|_{\mathcal{B}} = \|f_0\|_{\mathcal{B}}$ otherwise it will contradict to the fact that f_0 is a solution. Consequently, we conclude that $f \in \mathrm{S}(\mathbf{y})$ and thus $\mathrm{S}(\mathbf{y})$ is weakly* closed. ■

We next express the solution set $\mathrm{S}(\mathbf{y})$ by using the elements in the set $\mathcal{V}$ defined by (9). For any $f \in \mathcal{B}$, we introduce a subset of $\mathcal{V}$ by

$$\mathcal{J}(f) := \{\nu \in \mathcal{V} : f \in \|\nu\|_{\mathcal{B}_*} \partial \|\cdot\|_{\mathcal{B}_*}(\nu)\}. \tag{100}$$

Theorem 12 in Wang and Xu (2021) guarantees that for any solution $\hat{f}$ of problem (13), the set $\mathcal{J}(\hat{f})$ is nonempty. The next lemma shows that the sets defined by (100) associated with different solutions of problem (13) are exactly the same.

Lemma 32 *If $\mathcal{B}$ is a Banach space having a pre-dual space $\mathcal{B}_*$ and $\nu_j \in \mathcal{B}_*$, $j \in \mathbb{N}_n$, are linearly independent, then for any two solutions $\hat{f}, \hat{g}$ of the MNI problem* (13) *with $\mathbf{y} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$, there holds $\mathcal{J}(\hat{f}) = \mathcal{J}(\hat{g})$.*

Proof It suffices to show that $\mathcal{J}(\hat{f}) \subseteq \mathcal{J}(\hat{g})$. Suppose that $\mu \in \mathcal{J}(\hat{f})$. That is, $\mu \in \mathcal{V}$ and $\hat{f} \in \|\mu\|_{\mathcal{B}_*} \partial \|\cdot\|_{\mathcal{B}_*}(\mu)$. It is clear that $\mu \neq 0$ since $\hat{f} \neq 0$. By equation (15) with $\mathcal{B}$ being replaced by $\mathcal{B}_*$ and equation (2), we have that

$$\|\hat{f}\|_{\mathcal{B}} = \|\mu\|_{\mathcal{B}_*} \quad \text{and} \quad \langle \mu, \hat{f}\rangle_{\mathcal{B}} = \|\hat{f}\|_{\mathcal{B}} \|\mu\|_{\mathcal{B}_*}. \tag{101}$$

Since both $\hat{f}$ and $\hat{g}$ are solutions of the MNI problem (13) with $\mathbf{y}$, it holds that $\|\hat{f}\|_{\mathcal{B}} = \|\hat{g}\|_{\mathcal{B}}$ and $\langle \nu_j, \hat{f}\rangle_{\mathcal{B}} = \langle \nu_j, \hat{g}\rangle_{\mathcal{B}}$, for all $j \in \mathbb{N}_n$, which together with $\mu \in \mathcal{V}$ lead to

$$\|\hat{f}\|_{\mathcal{B}} = \|\hat{g}\|_{\mathcal{B}} \quad \text{and} \quad \langle \mu, \hat{f}\rangle_{\mathcal{B}} = \langle \mu, \hat{g}\rangle_{\mathcal{B}}. \tag{102}$$

Combining equations (101) with (102), we obtain that $\|\hat{g}\|_{\mathcal{B}} = \|\mu\|_{\mathcal{B}_*}$ and $\langle \mu, \hat{g}\rangle_{\mathcal{B}} = \|\hat{g}\|_{\mathcal{B}}\|\mu\|_{\mathcal{B}_*}$. This implies $\mu \in \mathcal{J}(\hat{g})$. Consequently, we conclude that $\mathcal{J}(\hat{f}) \subseteq \mathcal{J}(\hat{g})$. ∎

As we have seen in Lemma 32, the set $\mathcal{J}(\hat{f})$ is independent of the choice of the solution $\hat{f}$. This result allows us to express the solution set $\mathrm{S}(\mathbf{y})$ by using any fixed $\hat{\nu} \in \mathcal{J}(\hat{f})$.

Proposition 33 *Suppose that $\mathcal{B}$ is a Banach space having a pre-dual space $\mathcal{B}_*$, $\nu_j \in \mathcal{B}_*$, $j \in \mathbb{N}_n$, are linearly independent and $\mathbf{y} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$. If $\mathcal{V}$ and $\mathcal{M}_{\mathbf{y}}$ are defined by* (9) *and* (12)*, respectively, and $\hat{\nu} \in \mathcal{V}$ satisfies* (18)*, then*

$$\mathrm{S}(\mathbf{y}) = (\|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})) \cap \mathcal{M}_{\mathbf{y}}. \tag{103}$$

Proof We first assume that $\hat{f} \in (\|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})) \cap \mathcal{M}_{\mathbf{y}}$. That is, $\hat{f} \in \mathcal{M}_{\mathbf{y}}$ and there exists $\hat{\nu} \in \mathcal{V}$ satisfying the inclusion relation (16). Theorem 12 in Wang and Xu (2021) ensures that $\hat{f} \in \mathrm{S}(\mathbf{y})$. Hence, we have that $(\|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})) \cap \mathcal{M}_{\mathbf{y}} \subseteq \mathrm{S}(\mathbf{y})$. Conversely, assume that $\hat{f} \in \mathrm{S}(\mathbf{y})$. We choose $\hat{g} \in (\|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})) \cap \mathcal{M}_{\mathbf{y}}$. Similar arguments show that $\hat{g}$ is also a solution of problem (13) with $\mathbf{y}$ and $\hat{\nu} \in \mathcal{J}(\hat{g})$. With the help of Lemma 32, we obtain that $\hat{\nu} \in \mathcal{J}(\hat{f})$. That is, $\hat{f} \in \|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})$, which together with $\hat{f} \in \mathcal{M}_{\mathbf{y}}$ leads further to $\hat{f} \in (\|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})) \cap \mathcal{M}_{\mathbf{y}}$. We thus get that $\mathrm{S}(\mathbf{y}) \subseteq (\|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})) \cap \mathcal{M}_{\mathbf{y}}$, which completes the proof of the desired result. ∎

We will characterize the extreme points of $\mathrm{S}(\mathbf{y})$ by employing a well-known result about the extreme points. Dubins (1962) concerns a characterization of the extreme points of a subset, which is the intersection of a bounded, closed and convex subset with a finite number of hyperplanes. Specifically, let $\mathcal{X}$ be a topological vector space over the field of real numbers. Suppose that A is a bounded, closed and convex subset of $\mathcal{X}$. Dubins (1962) proved that every extreme point of the intersection of A with n hyperplanes is a convex combination of at most $n+1$ extreme points of A. Below, we give a technical lemma.

Lemma 34 *Suppose that $\mathcal{B}$ is a Banach space having a pre-dual space $\mathcal{B}_*$ and $\nu_j \in \mathcal{B}_*$, $j \in \mathbb{N}_n$, are linearly independent. Let $\mathcal{L}$ and $\mathcal{V}$ be defined by* (10) *and* (12)*, respectively. If $\nu \in \mathcal{V}\backslash\{0\}$ and $w_j \in \partial\|\cdot\|_{\mathcal{B}_*}(\nu)$, $j \in \mathbb{N}_{n+1}$, then there exist $a_j \in \mathbb{R}$, $j \in \mathbb{N}_{n+1}$, not all zero, such that $\sum_{j\in\mathbb{N}_{n+1}} a_j\mathcal{L}(w_j) = \mathbf{0}$ and $\sum_{j\in\mathbb{N}_{n+1}} a_j = 0$.*

Proof Assume that $\nu := \sum_{i\in\mathbb{N}_n} c_i\nu_i$ and set $\mathbf{c} := [c_i : i \in \mathbb{N}_n] \in \mathbb{R}^n$. By definition (10) of the operator $\mathcal{L}$, we have for each $j \in \mathbb{N}_{n+1}$ that

$$\langle \mathcal{L}(w_j), \mathbf{c}\rangle_{\mathbb{R}^n} = \left\langle \sum_{i\in\mathbb{N}_n} c_i v_i, w_j \right\rangle_{\mathcal{B}} = \langle \nu, w_j\rangle_{\mathcal{B}},$$

which together with $w_j \in \partial\|\cdot\|_{\mathcal{B}_*}(\nu)$ further leads to $\langle \mathcal{L}(w_j), \mathbf{c}\rangle_{\mathbb{R}^n} = \|\nu\|_{\mathcal{B}_*}$. It is clear that $\mathcal{L}(w_j)$, $j \in \mathbb{N}_{n+1}$, as $n+1$ elements in $\mathbb{R}^n$, are linearly dependent. Hence, there exist $a_j \in \mathbb{R}$, $j \in \mathbb{N}_{n+1}$, not all zero, such that $\sum_{j\in\mathbb{N}_{n+1}} a_j\mathcal{L}(w_j) = \mathbf{0}$. By equation $\langle \mathcal{L}(w_j), \mathbf{c}\rangle_{\mathbb{R}^n} = \|\nu\|_{\mathcal{B}_*}$, we get that

$$0 = \left\langle \sum_{j\in\mathbb{N}_{n+1}} a_j\mathcal{L}(w_j), \mathbf{c} \right\rangle_{\mathbb{R}^n} = \|\nu\|_{\mathcal{B}_*} \sum_{j\in\mathbb{N}_{n+1}} a_j,$$

which together with $\nu \neq 0$ results that $\sum_{j\in\mathbb{N}_{n+1}} a_j = 0$. ∎

We are ready to provide a complete proof for Proposition 7.

Proof [of Proposition 7] With the help of Proposition 33, we represent the solution set $\mathrm{S}(\mathbf{y})$ as in equation (103). It follows from equation (15) and $\hat{\nu} \neq 0$ that the subset $\|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})$ is bounded and thus weakly* bounded. By the definition of subdifferential, we get that $\|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})$ is also convex and weakly* closed. In addition, it is clear that $\mathcal{M}_{\mathbf{y}}$ is the intersection of n hyperplanes. Consequently, the solution set $\mathrm{S}(\mathbf{y})$ may be seen as the intersection between a subset of $\mathcal{B}$, which is weakly* bounded, weakly* closed and convex, and n hyperplanes. Then Dubins (1962) ensures that any extreme point $\hat{f}$ of $\mathrm{S}(\mathbf{y})$ has the form $\hat{f} = \sum_{j\in\mathbb{N}_{n+1}} \gamma_j^{'} u_j^{'}$, with $\gamma_j^{'} \geq 0$ satisfying $\sum_{j\in\mathbb{N}_{n+1}} \gamma_j^{'} = 1$ and $u_j^{'} \in \mathrm{ext}\left(\|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})\right)$. By setting $\gamma_j^{''} := \|\hat{\nu}\|_{\mathcal{B}_*}\gamma_j^{'}$, $u_j^{''} := u_j^{'}/\|\hat{\nu}\|_{\mathcal{B}_*}$, $j \in \mathbb{N}_{n+1}$, and noting that $\mathrm{ext}\left(\|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})\right) = \|\hat{\nu}\|_{\mathcal{B}_*}\mathrm{ext}\left(\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})\right)$, we rewrite $\hat{f}$ as in

$$\hat{f} = \sum_{j\in\mathbb{N}_{n+1}} \gamma_j^{''} u_j^{''}, \tag{104}$$

with $\gamma_j^{''} \geq 0$ satisfying $\sum_{j\in\mathbb{N}_{n+1}} \gamma_j^{''} = \|\hat{\nu}\|_{\mathcal{B}_*}$ and $u_j^{''} \in \mathrm{ext}\left(\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})\right)$.

We will represent $\hat{f}$ with the form (104) as in (19). Note that if there exists $j_0 \in \mathbb{N}_{n+1}$ such that $\gamma_{j_0}^{''} = 0$, then representation (104) reduces to the desired result (19). Hence, it remains to consider the case that $\gamma_j^{''} > 0$ for all $j \in \mathbb{N}_{n+1}$. We will show that $u_j^{''}$, $j \in \mathbb{N}_{n+1}$ are linearly dependent. By Lemma 34 and noting that $\hat{\nu} \in \mathcal{V}\backslash\{0\}$ and $u_j^{''} \in \partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})$, $j \in \mathbb{N}_{n+1}$, there exist $a_j \in \mathbb{R}$, $j \in \mathbb{N}_{n+1}$, not all zero, such that

$$\sum_{j\in\mathbb{N}_{n+1}} a_j\mathcal{L}(u_j^{''}) = \mathbf{0} \ \text{ and } \ \sum_{j\in\mathbb{N}_{n+1}} a_j = 0. \tag{105}$$

By setting $\alpha := \frac{\min_{j\in\mathbb{N}_{n+1}} \gamma_j^{''}}{\max_{j\in\mathbb{N}_{n+1}} |a_j|}$, we introduce two elements f_1, f_2 in $\mathcal{B}$ by

$$f_1 := \hat{f} + \frac{\alpha}{2} \sum_{j\in\mathbb{N}_{n+1}} a_j u_j^{''}, \quad f_2 := \hat{f} - \frac{\alpha}{2} \sum_{j\in\mathbb{N}_{n+1}} a_j u_j^{''}.$$

It follows from the first equation in (105) and $\mathcal{L}(\hat{f}) = \mathbf{y}$ that $\mathcal{L}(f_1) = \mathbf{y}$. That is, $f_1 \in \mathcal{M}_{\mathbf{y}}$. Substituting equation (104) into the representation of f_1, we obtain that

$$f_1 = \sum_{j\in\mathbb{N}_{n+1}} \left(\gamma_j^{''} + \frac{\alpha a_j}{2}\right) u_j^{''}. \tag{106}$$

According to the definition of α, the coefficient of each u_j'' appearing in (106) is positive. Moreover, combining $\sum_{j\in\mathbb{N}_{n+1}}\gamma_j''=\|\hat{\nu}\|_{\mathcal{B}_*}$ with the second equation in (105), we get that

$$\sum_{j\in\mathbb{N}_{n+1}}\left(\gamma_j''+\frac{\alpha a_j}{2}\right)=\sum_{j\in\mathbb{N}_{n+1}}\gamma_j''+\frac{\alpha}{2}\sum_{j\in\mathbb{N}_{n+1}}a_j=\|\hat{\nu}\|_{\mathcal{B}_*}.$$

We thus conclude that $f_1/\|\hat{\nu}\|_{\mathcal{B}_*}$ is a convex combination of u_j'', $j\in\mathbb{N}_{n+1}$. Recall that $u_j''\in\mathrm{ext}\,(\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu}))$ and hence $f_1/\|\hat{\nu}\|_{\mathcal{B}_*}\in\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})$, that is, $f_1\in\|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})$. Above all, we obtain that $f_1\in(\|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu}))\cap\mathcal{M}_{\mathbf{y}}$, which guaranteed by Proposition 33 is equivalent to $f_1\in\mathrm{S}(\mathbf{y})$. By similar arguments we also get that $f_2\in\mathrm{S}(\mathbf{y})$. It is clear that $\hat{f}=(f_1+f_2)/2$, which together with the assumption that $\hat{f}\in\mathrm{ext}\,(\mathrm{S}(\mathbf{y}))$ further leads to $\hat{f}=f_1=f_2$. Substituting the above relation into the representations of f_1 and f_2 with noting that $\alpha>0$, we have that

$$\sum_{j\in\mathbb{N}_{n+1}}a_ju_j''=0. \tag{107}$$

Since $a_j\in\mathbb{R}$, $j\in\mathbb{N}_{n+1}$ are not all zero, (107) implies that u_j'', $j\in\mathbb{N}_{n+1}$, are linearly dependent.

Without loss of generality, we assume that $a_{n+1}\neq 0$. It follows from equation (107) that

$$u_{n+1}''=-\sum_{j\in\mathbb{N}_n}\frac{a_j}{a_{n+1}}u_j''. \tag{108}$$

By substituting relation (108) into (104), we obtain that

$$\hat{f}=\sum_{j\in\mathbb{N}_n}\left(\gamma_j''-\frac{\gamma_{n+1}''}{a_{n+1}}a_j\right)u_j'' \tag{109}$$

For each $j\in\mathbb{N}_n$, by letting $\gamma_j:=\gamma_j''-\frac{\gamma_{n+1}''}{a_{n+1}}a_j$ and $u_j:=u_j''$, we can rewrite (109) as (19). In addition, due to $\sum_{j\in\mathbb{N}_{n+1}}\gamma_j''=\|\hat{\nu}\|_{\mathcal{B}_*}$ and the second equation of (105), we obtain that $\sum_{j\in\mathbb{N}_n}\gamma_j=\|\hat{\nu}\|_{\mathcal{B}_*}$. This completes the proof. ■

The next proposition shows that for any $\nu\in\mathcal{B}_*\backslash\{0\}$, the set $\mathrm{ext}\,(\partial\|\cdot\|_{\mathcal{B}_*}(\nu))$ is smaller than the set of extreme points of the closed unit ball B_0. That is to say, Proposition 7 provides an even more precise characterization for u_j, $j\in\mathbb{N}_n$, appearing in representation (19).

Proposition 35 *If Banach space $\mathcal{B}$ has a pre-dual space $\mathcal{B}_*$, then for any $\nu\in\mathcal{B}_*\backslash\{0\}$, there holds* $\mathrm{ext}\,(\partial\|\cdot\|_{\mathcal{B}_*}(\nu))\subset\mathrm{ext}\,(B_0)$.

Proof We assume that $\nu\in\mathcal{B}_*\backslash\{0\}$ and $f\in\mathrm{ext}\,(\partial\|\cdot\|_{\mathcal{B}_*}(\nu))$. It is sufficient to present $f\in\mathrm{ext}(B_0)$. By equation (15), we get that $\|f\|_{\mathcal{B}}=1$ and thus $f\in B_0$. For any $f_1,f_2\in B_0$ satisfying $f=(f_1+f_2)/2$, we shall prove that $f_1=f_2=f$. We first show that $\|f_1\|_{\mathcal{B}}=$

$\|f_2\|_{\mathcal{B}} = 1$. Otherwise, without loss of generality, we suppose that $\|f_1\|_{\mathcal{B}} < 1$. Then there holds

$$1 = \|f\|_{\mathcal{B}} = \left\|\frac{f_1+f_2}{2}\right\|_{\mathcal{B}} \leq \frac{\|f_1\|_{\mathcal{B}} + \|f_2\|_{\mathcal{B}}}{2} < 1,$$

which leads to a contradiction. Hence, $\|f_1\|_{\mathcal{B}} = \|f_2\|_{\mathcal{B}} = 1$. We next prove that $\langle f_i, \nu\rangle_{\mathcal{B}_*} = \|\nu\|_{\mathcal{B}_*}$, for $i = 1, 2$. If the claim is not true, without loss of generality, we suppose that $\langle f_1, \nu\rangle_{\mathcal{B}_*} < \|\nu\|_{\mathcal{B}_*}$. It follows that

$$\|\nu\|_{\mathcal{B}_*} = \langle f, \nu\rangle_{\mathcal{B}_*} = \frac{\langle f_1, \nu\rangle_{\mathcal{B}_*} + \langle f_2, \nu\rangle_{\mathcal{B}_*}}{2} < \|\nu\|_{\mathcal{B}_*},$$

which is a contradiction as well. Thus, $\langle f_i, \nu\rangle_{\mathcal{B}_*} = \|\nu\|_{\mathcal{B}_*}$, for $i = 1, 2$. We then conclude by equation (15) that $f_1, f_2 \in \partial\|\cdot\|_{\mathcal{B}_*}(\nu)$. This combined with $f \in \mathrm{ext}\,(\partial\|\cdot\|_{\mathcal{B}_*}(\nu))$ and the definition of extreme points leads to $f_1 = f_2 = f$. Again using the definition of extreme point, we obtain that $f \in \mathrm{ext}(B_0)$, which completes the proof. ■

Appendix B. A Dual Problem of MNI in Banach Spaces

In this appendix, we formulate a dual problem, a finite dimensional optimization problem, of the MNI problem (13) in a general Banach space having a pre-dual space. We then show that the element $\hat{\nu} \in \mathcal{V}$ appearing in Proposition 7 can be obtained by solving the resulting dual problem. Throughout this appendix, we suppose that $\mathcal{B}$ is a Banach space having the pre-dual space $\mathcal{B}_*$ and $\nu_j \in \mathcal{B}_*$, $j \in \mathbb{N}_n$, are linearly independent.

We establish the dual problem of problem (13) via a functional analysis approach. For this purpose, we recall the notion of the quotient space. If $\mathcal{M}$ is a closed subspace of a Banach space $\mathcal{B}$, then the quotient space $\mathcal{B}/\mathcal{M}$ is defined to be the collection of cosets $f + \mathcal{M}$, for all $f \in \mathcal{B}$. The quotient space is a Banach space when endowed with the norm

$$\|f + \mathcal{M}\|_{\mathcal{B}/\mathcal{M}} := \inf\left\{\|f + g\|_{\mathcal{B}} : g \in \mathcal{M}\right\}.$$

It is known (Conway (1990)) that $\mathcal{M}^*$ is isometrically isomorphic to $\mathcal{B}^*/\mathcal{M}^{\perp}$.

We now transform the MNI problem (13) into an equivalent dual problem.

Proposition 36 *For $\mathbf{y} := [y_j : j \in \mathbb{N}_n] \in \mathbb{R}^n\backslash\{\mathbf{0}\}$, let $\mathcal{M}_{\mathbf{y}}$ be defined by (12). Then*

$$\inf\left\{\|f\|_{\mathcal{B}} : f \in \mathcal{M}_{\mathbf{y}}\right\} = \sup\left\{\sum_{j\in\mathbb{N}_n} c_j y_j : \left\|\sum_{j\in\mathbb{N}_n} c_j\nu_j\right\|_{\mathcal{B}_*} = 1\right\}. \tag{110}$$

Proof By setting $\mathcal{M}_0$ to be $\mathcal{M}_{\mathbf{y}}$ with $\mathbf{y} = \mathbf{0}$, we represent $\mathcal{M}_{\mathbf{y}}$ as a translation of $\mathcal{M}_0$, that is, $\mathcal{M}_{\mathbf{y}} := \mathcal{M}_0 + f_0$ for some $f_0 \in \mathcal{M}_{\mathbf{y}}$. Then the MNI problem (13) may be rewritten as

$$\inf\left\{\|f\|_{\mathcal{B}} : f \in \mathcal{M}_{\mathbf{y}}\right\} = \inf\left\{\|f_0 + g\|_{\mathcal{B}} : g \in \mathcal{M}_0\right\},$$

which further leads to

$$\inf\left\{\|f\|_{\mathcal{B}} : f \in \mathcal{M}_{\mathbf{y}}\right\} = \|f_0 + \mathcal{M}_0\|_{\mathcal{B}/\mathcal{M}_0}. \tag{111}$$

By the isometric isomorphism between $(\mathcal{B}_*)^*/\mathcal{V}^\perp$ and $\mathcal{V}^*$, with noting that $\mathcal{B} = (\mathcal{B}_*)^*$ and $\mathcal{M}_0 = \mathcal{V}^\perp$, we have that

$$\|f_0 + \mathcal{M}_0\|_{\mathcal{B}/\mathcal{M}_0} = \sup\left\{\left\langle \sum_{j\in\mathbb{N}_n} c_j\nu_j, f_0\right\rangle_{\mathcal{B}} : \left\|\sum_{j\in\mathbb{N}_n} c_j\nu_j\right\|_{\mathcal{B}_*} = 1\right\}. \tag{112}$$

Substituting $\langle \nu_j, f_0\rangle_{\mathcal{B}} = y_j$, $j \in \mathbb{N}_n$, into the right-hand-side of equation (112), we get that

$$\|f_0 + \mathcal{M}_0\|_{\mathcal{B}/\mathcal{M}_0} = \sup\left\{\sum_{j\in\mathbb{N}_n} c_j y_j : \left\|\sum_{j\in\mathbb{N}_n} c_j\nu_j\right\|_{\mathcal{B}_*} = 1\right\}.$$

Again substituting the above equation into (111), we get the desired equation (110). ■

As a finite dimensional optimization problem, the dual problem

$$\sup\left\{\sum_{j\in\mathbb{N}_n} c_j y_j : \left\|\sum_{j\in\mathbb{N}_n} c_j\nu_j\right\|_{\mathcal{B}_*} = 1\right\}, \tag{113}$$

shares the same optimal value with the MNI problem (13). We remark that such a dual problem was considered in Cheng and Xu (2021) for the MNI problem with $\mathcal{B} := \ell_1(\mathbb{N})$ and in Cheng et al. (2024) for a class of regularization problems.

The next result concerns how to obtain an element $\hat{\nu} \in \mathcal{V}$ satisfying (18) once we have a solution of the dual problem (113) at hand.

Proposition 37 *For* $\mathbf{y} := [y_j : j \in \mathbb{N}_n] \in \mathbb{R}^n\backslash\{\mathbf{0}\}$, *let* $\mathcal{M}_{\mathbf{y}}$ *be defined by* (12). *If* m_0 *is the infimum of the MNI problem* (13) *with* $\mathbf{y}$ *and* $\hat{\mathbf{c}} := [\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n] \in \mathbb{R}^n$ *is a solution of the dual problem* (113) *with* $\mathbf{y}$, *then* $\hat{\nu} := m_0 \sum_{j\in\mathbb{N}_n} \hat{c}_j\nu_j$ *satisfies* (18).

Proof Noting that $\mathrm{S}(\mathbf{y})$ is nonempty, we choose $\hat{f} \in \mathrm{S}(\mathbf{y})$ and proceed to prove that $\hat{f} \in \|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu}) \cap \mathcal{M}_{\mathbf{y}}$. Clearly, $\hat{f} \in \mathcal{M}_{\mathbf{y}}$. It suffices to show that $\hat{f} \in \|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})$. Since $\hat{\mathbf{c}}$ is a solution of problem (113), we get that $\|\sum_{j\in\mathbb{N}_n} \hat{c}_j\nu_j\|_{\mathcal{B}_*} = 1$. This together with the definition of $\hat{\nu}$ leads to $\|\hat{\nu}\|_{\mathcal{B}_*} = m_0$. By noting that $\|\hat{f}\|_{\mathcal{B}} = m_0$, we obtain that $\left\|\frac{\hat{f}}{\|\hat{\nu}\|_{\mathcal{B}_*}}\right\|_{\mathcal{B}} = 1$. It follows that

$$\left\langle \frac{\hat{f}}{\|\hat{\nu}\|_{\mathcal{B}_*}}, \hat{\nu}\right\rangle_{\mathcal{B}_*} = \sum_{j\in\mathbb{N}_n} \hat{c}_j\langle \hat{f}, \nu_j\rangle_{\mathcal{B}_*}.$$

Substituting $\langle \hat{f}, \nu_j\rangle_{\mathcal{B}_*} = y_j$, $j \in \mathbb{N}_n$, into the above equation, we have that

$$\left\langle \frac{\hat{f}}{\|\hat{\nu}\|_{\mathcal{B}_*}}, \hat{\nu}\right\rangle_{\mathcal{B}_*} = \sum_{j\in\mathbb{N}_n} \hat{c}_j y_j.$$

Noting that $\hat{\mathbf{c}}$ is a solution of problem (113), the above equation, guaranteed by Proposition 36, yields that $\langle \hat{f}/\|\hat{\nu}\|_{\mathcal{B}_*}, \hat{\nu}\rangle_{\mathcal{B}_*} = m_0$. This together with $\|\hat{\nu}\|_{\mathcal{B}_*} = m_0$ leads directly to $\langle \hat{f}/\|\hat{\nu}\|_{\mathcal{B}_*}, \hat{\nu}\rangle_{\mathcal{B}_*} = \|\hat{\nu}\|_{\mathcal{B}_*}$. Consequently, we conclude that $\hat{f}/\|\hat{\nu}\|_{\mathcal{B}_*} \in \partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})$, that is, $\hat{f} \in \|\hat{\nu}\|_{\mathcal{B}_*}\partial\|\cdot\|_{\mathcal{B}_*}(\hat{\nu})$. This completes the proof of this proposition. ■

Proposition 37 ensures that the element $\hat{\nu} \in \mathcal{V}$ appearing in Proposition 7 can be obtained by solving the dual problem (113).

References

A. Argyriou, C. A. Micchelli, and M. Pontil. When is there a representer theorem? Vector versus matrix regularizers. *Journal of Machine Learning Research*, 10:2507–2529, 2009.

F. Bach. Breaking the curse of dimensionality with convex neural networks. *Journal of Machine Learning Research*, 18:629–681, 2017.

F. Bartolucci, E. D. Vito, L. Rosasco, and S. Vigogna. Understanding neural networks with reproducing kernel Banach spaces. *Applied and Computational Harmonic Analysis*, 62: 194–236, 2023.

C. Boyer, A. Chambolle, Y. D. Castro, V. Duval, F. D. Gournay, and P. Weiss. On representer theorems and convex regularization. *SIAM Journal on Optimization*, 29:1260–1281, 2019.

K. Bredies and M. Carioni. Sparsity of solutions for variational inverse problems with finite-dimensional data. *Calculus of Variations and Partial Differential Equations*, 59: 1–26, 2020.

R. Cheng and Y. Xu. Minimum norm interpolation in the $\ell_1(\mathbb{N})$ space. *Analysis and Applications*, 19:21–42, 2021.

R. Cheng, R. Wang, and Y. Xu. A duality approach to regularized learning problems in Banach spaces. *Journal of Complexity*, 81:101818, 2024.

I. Cioranescu. *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems.* Mathematics and its Applications, volume 62. Kluwer Academic Publishers Group, Dordrecht, 1990.

J. B. Conway. *A Course in Functional Analysis.* 2nd Edition. Springer-Verlag, New York, 1990.

D. D. Cox and F. O'Sullivan. Asymptotic analysis of penalized likelihood and related estimators. *The Annals of Statistics*, 18:1676–1695, 1990.

C. deBoor and R. E. Lynch. On splines and their minimum properties. *Journal of Mathematics and Mechanics*, 15:953–969, 1966.

L. E. Dubins. On extreme points of convex sets. *Journal of Mathematical Analysis and Applications*, 5:237–244, 1962.

G. E. Fasshauer, F. J. Hickernell, and Q. Ye. Solving support vector machines in reproducing kernel Banach spaces with positive definite functions. *Applied and Computational Harmonic Analysis*, 38:115–139, 2015.

T. Hoefler, D. A. Alistarh, N. Dryden, and T. Ben-Nun. The future of deep learning will be sparse. *SIAM News, May*, 3, 2021.

L. Huang, C. Liu, L. Tan, and Q. Ye. Generalized representer theorems in Banach spaces. *Analysis and Applications*, 19:125–146, 2021.

G. Kimeldorf and G. Wahba. A correspondence between bayesian estimation on stochastic processes and smoothing by splines. *Annals of Mathematical Statistics*, 41:495–502, 1970.

A. J. Kurdila and M. Zabarankin. *Convex Functional Analysis*. Systems & Control: Foundations & Applications. Birkhäuser Verlag, Basel, 2005.

R. Lin and Y. Xu. Reproducing kernel Banach spaces of bounded continuous functions. *Preparation*, 2022.

R. Lin, G. Song, and H. Zhang. Multi-task learning in vector-valued reproducing kernel Banach spaces with the ℓ_1 norm. *Journal of Complexity*, 63:101514, 2021.

R. Lin, H. Zhang, and J. Zhang. On reproducing kernel Banach spaces: generic definitions and unified framework of constructions. *Acta Mathematica Sinica, English Series*, 38: 1459–1483, 2022.

Y. Lin, R. Lin, and Q. Ye. Sparse regularized learning in the reproducing kernel Banach spaces with the ℓ_1 norm. *Mathematical Foundations of Computing*, 3:205–218, 2020.

Q. Liu, R. Wang, Y. Xu, and M. Yan. Parameter choices for sparse regularization with the ℓ_1 norm. *Inverse Problems*, 39:025004, 2023.

R. E. Megginson. *An Introduction to Banach Space Theory*. Graduate Texts in Mathematics, volume 183. Springer-Verlag, New York, 1998.

C. A. Micchelli and M. Pontil. A function representation for learning in Banach spaces. In *International Conference on Computational Learning Theory*, pages 255–269. Springer, 2004.

R. Parhi and R. D. Nowak. Banach space representer theorems for neural networks and ridge splines. *Journal of Machine Learning Research*, 22:1960–1999, 2021.

S. Rosset, G. Swirszcz, N. Srebro, and J. Zhu. ℓ_1 regularization in infinite dimensional feature spaces. In *International Conference on Computational Learning Theory*, pages 544–558. Springer, 2007.

B. Schölkopf, R. Herbrich, and 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*, pages 416–426. Springer-Verlag, London, 2001.

J. Shenouda, R. Parhi, K. Lee, and R. D. Nowak. Vector-valued variation spaces and width bounds for DNNs: insights on weight decay regularization. *arXiv preprint arXiv:2305.16534*, 2023.

G. Song, H. Zhang, and F. J. Hickernell. Reproducing kernel Banach spaces with the ℓ_1 norm. *Applied and Computational Harmonic Analysis*, 34:96–116, 2013.

L. Spek, T. J. Heeringa, F. Schwenninger, and C. Brune. Duality for neural networks through reproducing kernel Banach spaces. *arXiv preprint arXiv:2211.05020*, 2023.

M. Unser. A unifying representer theorem for inverse problems and machine learning. *Foundations of Computational Mathematics*, 21:941–960, 2021.

M. Unser and S. Aziznejad. Convex optimization in sums of Banach spaces. *Applied and Computational Harmonic Analysis*, 56:1–25, 2022.

M. Unser, J. Fageot, and H. Gupta. Representer theorems for sparsity-promoting ℓ_1 regularization. *IEEE Transactions on Information Theory*, 62:5167–5180, 2016.

M. Unser, J. Fageot, and J. P. Ward. Splines are universal solutions of linear inverse problems with generalized TV regularization. *SIAM Review*, 59:769–793, 2017.

R. Wang and Y. Xu. Representer theorems in Banach spaces: minimum norm interpolation, regularized learning and semi-discrete inverse problems. *Journal of Machine Learning Research*, 22:1–65, 2021.

Y. Xu. Sparse machine learning in Banach spaces. *Applied Numerical Mathematics*, 187: 138–157, 2023.

Y. Xu and Q. Ye. Generalized Mercer kernels and reproducing kernel Banach spaces. *Memoirs of the American Mathematical Society*, 258(1243), 2019.

H. Zhang and J. Zhang. Regularized learning in Banach spaces as an optimization problem: representer theorems. *Journal of Global Optimization*, 54:235–250, 2012.

H. Zhang, Y. Xu, and J. Zhang. Reproducing kernel Banach spaces for machine learning. *Journal of Machine Learning Research*, 10:2741–2775, 2009.

C. Zălinescu. *Convex Analysis in General Vector Spaces*. World Scientific, New Jersey, 2002.